

A historical study of structures for communication of organic chemistry information prior to 1950†

Helen Cooke*‡

Chemistry Department, UMIST, P.O. Box 88, Sackville Street, Manchester, UK M60 1QD.
E-mail: hcooke@comcast.net

Received 1st July 2004, Accepted 7th September 2004
First published as an Advance Article on the web 11th October 2004

The evolution of the graphical structure diagram as a means of communication of chemical structure information is traced from its origins through to the mid-20th century. The impact of developments in structural theory on the representation of structures is discussed. A study of how structures were represented in the 19th and early 20th centuries in the *Journal of the Chemical Society* and its predecessors, the *Journal of the American Chemical Society* and United States patents was made, making use of electronic journal and patent archives. The problems associated with representing structures graphically are discussed.

Introduction

A study of structures as representations of compounds and their use as a language has been made by Hoffmann and Laszlo.¹ These authors suggest that structural formulae are a form

† This is one of a number of contributions on the theme of molecular informatics, published to coincide with the RSC Symposium "New Horizons in Molecular Informatics", December 7th 2004, Cambridge UK.

‡ Current address: 268 Bella Lane, King of Prussia, PA 19406, USA.

Upon graduating in chemistry from the University of Manchester, I embarked upon my career as a chemical information scientist. My first position was with Derwent Publications as a chemical coder, my second with the International Union of Crystallography as an editorial assistant. After completing a post-graduate diploma in librarianship, I worked at UMIST (University of Manchester Institute of Science and Technology) for many years, initially in the Chemistry Department Library. From the mid-1990s I became increasingly involved with chemical information teaching and research. I worked partly for MIMAS (Manchester Information and Associated Services) training chemists in universities in the UK, Ireland and Scandinavia to use CrossFire Beilstein and Gmelin, and most recently was the course leader of UMIST's MSc in Cheminformatics. I completed my PhD, "Perspectives on the history of the research chemical literature and chemical information retrieval (1850–1960)", in 2003.



Helen Cooke

of language as they combine symbolic and representational (iconic) values, and hence are somewhere between being symbols and models:

"Most, if not all, scientists make use of visual imagery for problem-solving, in order to sort out and organise information, to find analogies, to think. But chemists are unusual among scientists...in having an iconic vernacular, that of formulas.

A chemical formula is like a word. It purports to identify, to single out the chemical species it stands for."

Structure diagrams convey more direct meaning than most normal textual words, whose representation in themselves do not communicate any image directly apart from in the mind of the informed reader. Unlike normal language, where words can often have several meanings, structures (particularly three-dimensional) are usually representations of unique molecules.

This paper will explore the origins of structure diagrams, their representation in the chemistry literature and the resulting effect on chemical information retrieval. Some work has already been reported on the representation of formulae and equations in textbooks,² but no previous systematic study of journals and patents has been identified, which will be the focus of this paper. The emphasis will be on organic compounds, since structure diagrams are of greatest importance as a means of representation and communication to organic chemists. The use of conventions, representation of three-dimensional structures, different types of compounds and reactions will be discussed.

Chemical nomenclature will not be discussed in any detail. This has been well documented,^{3,4} although it will on occasion be necessary to refer to chemical names and nomenclature conventions.

Early use of symbols

The Egyptians, Greeks and alchemists used symbols widely, to represent metals, other materials and the elements of earth, air, fire and water. The alchemists borrowed symbols from the Greeks, who in turn had sometimes borrowed from the Egyptians;⁵ the alchemists also used astrological and planetary symbols to represent the metals gold, silver, iron, copper, lead, tin and mercury.⁶

An account of the development of the symbols is given by Crosland.⁷ Early examples of the use of chemical symbols are present in the *Basilica Chymia* of Oswald Croll (1609) and in Robert Hooke's diary of the years 1672–1680; more extensive use was made by Lemery in his *Cours de Chimie* in the 1675 and later editions.⁸ Lemery's symbols were also quoted in John Harris's *Lexicon Technicum*, the first technical dictionary, published in London in 1704.

Use of symbols to represent elements and graphical structures to represent compounds is closely linked to the development of chemical structural theory. By the 18th century, symbols were becoming more sophisticated and meaningful, as more indication of the composition of substances was represented. Geoffroy's table of affinities (1718) was an early attempt at representing the concept of chemical change or reaction through indication of the attractions between substances.⁹ Bergman, in 1715, also compiled a table of "elective attractions", which

employed symbols to represent acids and other substances listed beneath them in the table in order of their decreasing affinity for the acid.¹⁰ Alchemical symbols continued to be used until the end of the 18th century. It is noteworthy, however, that the key experimenters of this time, Priestley, Cavendish and Scheele, did not use symbols, but instead described their findings using normal language.¹¹ However, those that were attempting to systematise chemistry at the end of the 18th century did feel the need for symbols, and Lavoisier used them to represent the elements and compounds known to him. Lavoisier also used chemical equations to represent reactions.¹² In 1787, de Morveau, Lavoisier, Berthollet and de Fourcroy published *Méthode de nomenclature chimique* (which was published in English in 1788). This was supplemented by *A new system of chymical characters adapted to the nomenclature* by Hassenfratz and Adet; these workers recognised the confused state of chemical representation, and attempted to produce a set of standard chemical characters and symbols commensurate with Guyton de Morveu *et al.*'s new nomenclature.¹³ But because this work preceded atomic theory, it was not destined to stand the test of time.

Another factor contributing generally to the lack of adoption of chemical symbols in the late 18th and early 19th centuries is concerned with typographical limitations. Dalton, the originator of chemical structure theory, first used symbols in 1803,¹⁴ and these were first published by Thomas Thomson in his *System of chemistry* in 1807. Dalton's own work, *A new system of chemical philosophy*, was published in 1808. Dalton chose a notation which was not based on initial letter abbreviations, as he wanted his symbols not only to represent the types of atoms present in a compound, but also their spatial arrangement. Although Dalton's symbols were used as structure representations, they suffered, albeit to a lesser degree, from the same problems as those of Hassenfratz and Adet in that they could not readily be incorporated into a printed text.

It was the work of Berzelius which led to the development of modern symbols for chemical representation. In 1819 Berzelius published his proposals in *Théorie des proportions chimiques, et table synoptique des poids*. Unlike his predecessors, Berzelius's symbols deliberately used ordinary letters, which overcame the typographical problems associated with earlier symbols; thus they were more readily deployable in printed publications, and this contributed to their wide adoption.

The use of symbols was debated widely in the British chemistry community in the 1830s, in the halls of the Royal Institution and the pages of the journal *Philosophical Magazine*.¹⁵ The most common justification for use of pictorial symbols such as those of Dalton was the fact that they conveyed the spatial configuration of compounds, but Berzelius's symbols were ultimately to prevail.

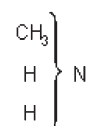
Representation of organic compounds

In the early 19th century, the representation of all compounds was based on that used for inorganic compounds, since the development of inorganic chemistry preceded that of organic. Even though many organic compounds were known, particularly organic acids, it was not possible using the symbolism of the time to represent them such that any significant structural meaning was conveyed, if indeed any clear idea of the structure was known.

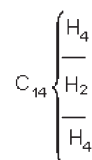
Considerable advances were being made in the early 19th century concerning the organisation of organic chemistry. For example, in 1827 William Prout made the distinction between fats, carbohydrates and proteins. In 1843 Gerhardt introduced the term "homologous series", a key development in the classification of organic compounds, which also allowed the prediction of properties of compounds in the series not yet isolated.

During the first half of the 19th century, attempts were made to convey the compositions of chemical compounds

by formulae, and Berzelius introduced the terms "empirical formula" and "rational formula" in 1833.¹⁶ Empirical formulae gave the atoms in the structures in the correct ratio according to their relative proportions. Rational, or "type", formulae indicated the "rational composition" showing which particular groupings replaced hydrogen in the parent compound.¹⁷ A means of representing rational formulae was through the use of brackets, or braces, *e.g.* methylamine (a compound of the ammonia type):



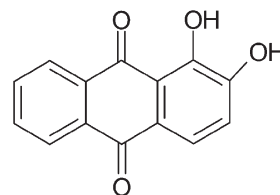
These were similar to constitutional formulae which were used by Hofmann and others from around 1850. Such formulae were used for aromatic compounds after 1865;¹⁸ for example, structures containing fused rings could be represented by use of horizontal lines drawn between symbols for the numbers of hydrogen atoms on each ring. The constitutional formula of anthracene would thus be represented:



Constitutional formulae were also used to represent more complex molecules. For example, Graebe and Liebermann proposed two possible structures for alizarin in 1870:



The actual structure of alizarin is:

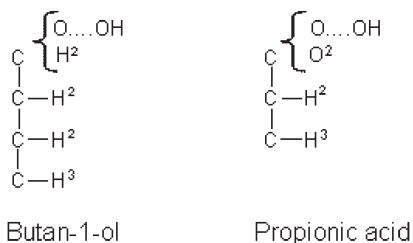


Methods of representing compounds which made use of braces continued until around the end of the 19th century.

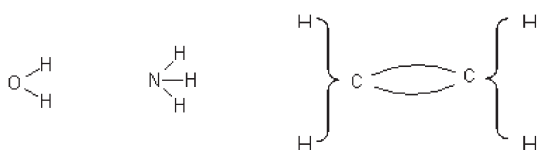
The concept of valence was developed in the middle of the 19th century. Kekulé proposed the marsh-gas structure type in 1854 (*e.g.* CH₄, CHCl₃, CCl₃NO₂ were considered to belong to this type).¹⁹ Kekulé observed that carbon atoms always combined with four atoms of a monatomic element and two of a diatomic element and concluded that carbon is tetravalent. Other chemists who contributed to the concept of valence were Kolbe and Frankland, but the correct valences of elements could not be finalised until the distinction between equivalent, atomic and molecular weights had been established through the work of Gerhardt and Laurent, and methods of atomic weight determination by Cannizzaro.²⁰ This was an essential step on the route to the structure diagram as a means of representing compounds.

In 1854 Kekulé stated that the binding between atoms was by means of bonds, and indeed later represented bonds graphically, but it was Archibald Scott Couper of Kirkintilloch rather than Kekulé who first drew bonds and schematic structures

similar to those deployed today.^{21,22} The first of Couper's graphical structures appeared in 1858,^{23,24} in *Comptes Rendus*, *Philosophical Magazine* and *Annales de Chimie et de Physique*. Examples of some of Couper's structures are given below:²⁵



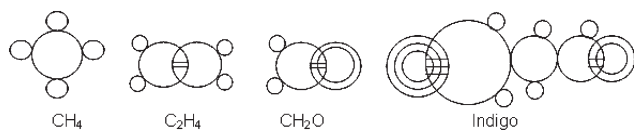
In these representations, a single oxygen is written as O–O, and oxygen was assumed to have an atomic weight of 8. Couper's formulae became well known, and were referred to by Lothar Meyer in his book *Modernen theorien der chemie* published in 1864.²⁶ Lothar Meyer drew both single and double bonds:



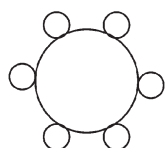
Lothar Meyer also proposed ring structures for sulfur trioxide and ozone:



A publication by Loschmidt²⁷ in 1861 contains 368 graphic formulae, 121 of which were aromatic compounds. Loschmidt's drawings are based on Dalton's symbols. These structures indicate double and triple bonds,²⁸ e.g.



Using the same graphical representation, Loschmidt differentiated between the isomers 1-propenol and 3-propenol. He also speculated as to the structure of benzene, settling on:

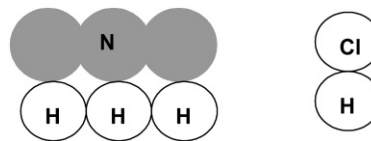


Caven and Cranston²⁹ argue that Loschmidt therefore proposed a ring structure for benzene four years before Kekulé independently did so.

Loschmidt's representations suffered from the same problem as those of Dalton, as they did not readily lend themselves to larger structures: having to draw circles to represent atoms was cumbersome, and led to typography problems.

Loschmidt also published the first line formula descriptions, such as C₂H₅.O.C₂H₅ for diethyl ether.³⁰ By 1866, a number of chemists were using line formulae in the form CH₃.CO.OH or H₃C–CO–OH (with no standardisation in the use of commas, full stops or hyphens); these convey more structural information than empirical formulae which merely give the relative quantities of the elements present. Wiswesser, founder of the famous line notation in the 1950s, states that after 1861 no new practices in the use of line notations to represent structures occurred for 79 years.³¹

In 1859 Kekulé started to construct graphical formulae, which bore some resemblance to those of John Dalton. Kekulé used pictures of atoms of different sizes effectively representing different valences:³² e.g.



These graphical formulae appeared in footnotes to the text of his publication *Lehrbuch der organischen Chemie*. By 1865 Kekulé had started to use bonds in structures. In the same year, Hofmann delivered a famous lecture on homologous series to the Royal Institution in which he used croquet balls of different colours to represent atoms provided with arms to represent units of attraction, which could be regarded as an early form of molecular model.³³ Hofmann's models were two-dimensional – the era of three-dimensional structures and stereochemistry had not yet arrived.

It was Kekulé's hexagonal representation of benzene which was a turning point in the representation of structures containing aromatic rings.³⁴ Kekulé's early representations were famously inspired by a dream, through which he envisioned chains of carbon atoms like snakes chasing their own tails, forming rings.³⁵ Kekulé's work explaining the structure of aromatic rings was published in 1865.³⁶ Other workers proposed a variety of different structures for benzene, and it was one of his pupils, Körner, who introduced the terms *ortho*-, *meta*- and *para*-.

In 1861 Crum Brown developed an atomic symbol notation which used circles containing the initial letters of atom names, linked by dotted lines, which were later replaced by continuous lines.³⁷ Crum Brown's representations also included double and triple bonds. They first appeared in his publications in 1864. Frankland, who introduced the term "bond" (which became known as Frankland's notation), represented structures in a similar fashion to Crum Brown.³⁸ Continuous lines to represent bonds were not employed universally, and bonds were represented as two lines separated by a dot in *Berichte* from 1870 to 1886. In 1877, Claus used a combination of this representation with a dotted line to indicate the double bond in a phenyl ring:



Further workers to devise graphical representations were Wilbrand (Germany) and Foster (Scotland).

Overall, there was considerable confusion as to the best way to represent structures during the second half of the 19th century.

Stereochemistry

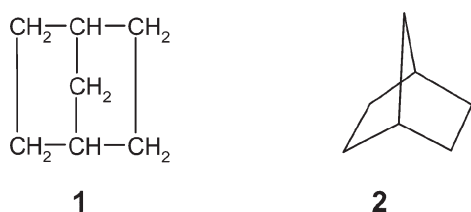
The three-dimensional nature of chemical compounds was reported by Louis Pasteur in 1860, when he gave an account of the relation between the *d*- and *l*-configurations of tartaric acid to the Société Chimique de France, and Kekulé also attempted a three-dimensional representation of benzene in 1867. But it was Le Bel in France and van't Hoff in Belgium who in 1874 simultaneously proposed the tetrahedral arrangement of atoms around carbon, and hence also the possibility of asymmetric carbon atoms. They also related this to the optical properties of asymmetric compounds.³⁹

Van't Hoff also showed that there was another type of isomerism, around double bonds. The terms *cis*- and *trans*- were introduced by Baeyer in 1888. Le Bel and van't Hoff represented the stereochemistry around double bonds as two tetrahedra united along a common edge.

Discovery of the three-dimensional nature of structures was the major achievement in organic chemistry between 1880 and

1900, and also led to the theory of steric hindrance. With this knowledge, Fischer was able to determine the structures of isomeric sugars, and this formed the foundation of carbohydrate chemistry. Fischer devised his "Fischer projection" for the representation of three-dimensional structures.

At the end of the 19th century, when structure representations were being developed, engraving was the main means of type-setting drawings. This was an expensive process, and even more so for lines drawn at an angle, hence three-dimensional representations tended to be drawn in two dimensions.¹ This situation continued until the 1950s after engraving had been replaced by photogravure. Three-dimensional structures were therefore represented in a two-dimensional fashion, *e.g.* norbornadiene was represented as structure **1** rather than structure **2**, the latter providing a clearer representation of the three-dimensional nature of this compound.

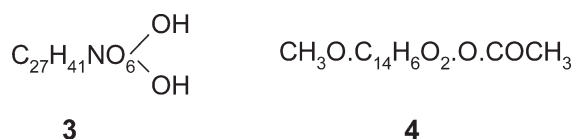


The use of wedge bonds to represent the three-dimensional nature of compounds did not start until the 1950s.

Generic structures

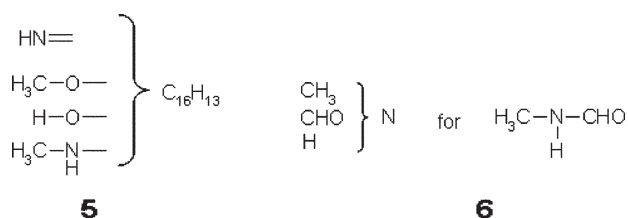
The earliest means by which generic groups or atoms (*i.e.* symbols representing the possibility of structural variation in compounds) were denoted was by means of inclusion of symbols in formulae other than those used to represent specific atom symbols, such as R to represent an alkyl group, X a halogen or M a metal. These symbols have been in use since around 1880,⁴⁰ and are still used today. (Although X and M are fairly standard, nowadays R is frequently defined locally, *i.e.* within the specific document or even formula.)

One type of generic structure which could represent compounds with certain specific groups present while maintaining uncertainty about other parts of the structure, although restricting the overall composition, is of the type **3** and **4** below:



These were used in publications around the turn of the 20th century.

If more than two connections were needed, then braces were often used, as in structure **5**, which arose out of representations of the type **6** used to represent specific structures. Brace structures have variable connectivity, but not variable composition.



Structure **5**, representing the alkaloid coclaurin, has been reported as being the last occurrence of the use of braces in this way to represent structures.⁴¹

Markush structures

Markush structures are a type of generic chemical structure which occur mainly in patents. In Markush structures, a single compound incorporating one or more groups represents a variety of more specific compounds through the use of defined generic groups. The generic groups may be clearly defined structure fragments, such as methyl groups, keto groups, *etc.*, or less well defined textual descriptions, such as "substituted phenol", "aryl", "carbocyclic", *etc.*

Structures of this type were permitted after a ruling by the United States Patent Office, following the publication of a patent by Eugene Markush,⁴² which claimed a process for the manufacture of dyes that comprised coupling with a halogen-substituted pyrazolone, a diazotized unsulfonated material selected from the group consisting of aniline, homologues of aniline, and halogen substitution products of aniline. However, Markush's name was not associated with the Markush structure because his patent was the first to include structures of this type. There had already been patents with claims concerned with compounds defined from a selection of variable groups, but a patent examiner stubbornly rejected Markush's claims. Markush appealed, and the commissioner issued a published decision overruling the examiner and approving the claim format. Future applicants argued for patentability on the basis of this precedent, and Markush's name became thereafter attached to the format.

Lynch in 1984 suggested that the first patent containing the concept of a Markush structure was to Perkin in 1856 for the first aniline dye (mauve),⁴³ his specification included aniline, toluidine, xylydine and cumidine and is shown in Fig. 1. Another more recent example reported in the literature which predates Markush's patent was granted to Eli Lilly in 1915.⁴⁴

The nature of my Invention consists in producing a new coloring matter for dyeing with a lilac or purple color stuffs of silk, cotton, wool, and other materials in the manner following :—

I take a cold solution of sulphate of aniline, or a cold solution of sulphate of toluidine, or a cold solution of sulphate of xylydine, or a cold solution of sulphate of cumidine, or a mixture of any one of such solutions with any others or other of them, and as much of a cold solution of a soluble bichromate as contains base enough to convert the sulphuric acid in any of the above-mentioned solutions into a neutral sulphate. I then mix the solutions, and allow them to stand for ten or twelve hours, when the mixture will consist of a black powder and a solution of a neutral sulphate. I then throw this mixture upon a fine filter, and wash it with water till free from the neutral sulphate. I then dry the substance thus obtained at a temperature of 100° centigrade, or 212° Fahrenheit, and digest it repeatedly with coal tar naphtha, until it is free from a brown substance which is extracted by the naphtha. Any other substance than coal tar naphtha may be used in which the brown substance is soluble and the coloring matter is not soluble. I then free the residue from the naphtha by evaporation, and digest it with methylated spirit, or any other liquid in which the coloring matter is soluble, which dissolves out the new coloring matter. I then separate the methylated spirit from the coloring matter by distillation, at a temperature of 100° centigrade, or 212° Fahrenheit.

Fig. 1 Excerpt from Perkin's mauve patent from 1856.

Originally, Markush structures were only permitted when nomenclature could not adequately describe an applicant's invention, but gradually they became more widely accepted, and are now used when claims are made for a number of similar compounds, not all of which have necessarily been synthesised.

Occurrence of structure diagrams in journals and patents

In order to determine how and when structure representation conventions were deployed in publications, a study was made of journals and patents. pdf files, which reproduce documents in their original formats, can often be obtained by use of electronic versions of publications. These can be browsed fairly rapidly, and this facilitates a search for structure representations.

Table 1 Dates that structural features would be expected to appear in the literature (on the basis of information in history of chemistry literature) and first dates observed in this study (NB: these were as found in this study, and are not necessarily the first occurrences)

Earliest date expected	Date observed	Structural representation	Publication type where observed (J = journal, P = patent)
	1843	Use of braces, little structural information	J
	1843	Reaction schemes	J
1850s	1860s	Constitutional formulae	J
	1862	Use of primes providing structural information	J
	1865	Use of braces, little structural information	P
1870	1870	Graphic structures	J
	1876	Phenyl rings with alternating bonds	J
1880	1877	R group in a brace structure	J
1830	1879	Brace structures conveying structural information	J
1870	1879	<i>Ortho-, meta-, para-</i> used for substitution of benzene	J
	1882	Reaction schemes	P
1866	1882	Dots and commas separating fragments in linear formulae	P
1830	1884	Brace structures conveying structural information	P
1888	1895	Use of <i>cis-</i> and <i>trans-</i>	J
	1899	Markush structures	P
1900		End of use of braces (brackets) to convey structural information	
1900		Use of braces (brackets) to represent generic structures	
	1900	Representation of asymmetric carbon atoms	J
1891	1905	Fischer projections in carbohydrates	J
	1910	Use of <i>d-</i> and <i>l-</i> terminology	J
	1920s	Bridged compound with no 3D perspective	J
1924		Increase in Markush structures	
	1930s	Carbohydrate rings showing 3D perspective	J
	1930s	Steroids with dashed bonds to show 3D information	J
	1940s	Steroids with dashed bonds to show 3D information	P

On the basis of the development of structure representations discussed earlier, we might expect to see structural features appearing as shown in Table 1, which also summarises the results of this study. As would be expected, the features did not usually start to appear in publications until several years after the time the representations were first proposed; in addition, not all types of representation were observed. This is probably because in this work only a sample of publications was studied.

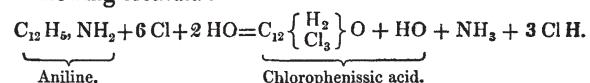
Study of the journal literature

The archives of many journals are becoming available in electronic form. For example, all RSC (and formerly Chemical Society) journals have recently been made available, back to the first issue of the *Memoirs and Proceedings of the Chemical Society (MPCS)* in 1841. The *Journal of the American Chemical Society (JACS)* is also available in electronic form in its entirety, back to 1879.⁴⁵ Unfortunately, the absence of electronic versions of early German journals (such as *Berichte der Deutschen Chemischen Gesellschaft* which started publication in 1868) precludes ready study of structures in the German literature, and for this reason German journals have not been included in this work.

Issues of *MPCS*, *Quarterly Journal of the Chemical Society of London (QJCS)*, *Journal of the Chemical Society (JCS)* and *JACS* were examined by browsing the electronic tables of contents available through the Societies' web pages. Issues were examined at intervals of approximately five years. For the early years of each publication, several years' worth of journals were examined as the total number of papers published each year was small. When the title of a paper suggested that the subject matter was likely to be concerned with organic chemistry, the full article was explored. After the start of its publication in 1936, the *Journal of Organic Chemistry (JOC)* was also examined.

No representations of compounds containing structural information were found in the early volumes of *MPCS*, though a number of papers reported the percentages of the constituent elements present in compounds. As would be expected from Table 1, examples of the use of braces (brackets) were found after 1830, the first example discovered being in a paper by Hofmann published in 1843 (see Fig. 2).⁴⁶ There is little structural information conveyed, and it is not possible to determine the structures Hofmann had in mind, if any, from

aniline. Besides a neutral compound, which was not produced in sufficient quantity for analysis, there was always formed as the chief product chlorindoptinic acid (chlorophenissic acid of Laurent), whose relation to aniline is shown by the following formula:



The action of bromine on aniline, investigated by Fritzsche†, is much less energetic. A simple substitution of bromine for hydrogen takes place, giving rise to the compound named by him bromaniloid,

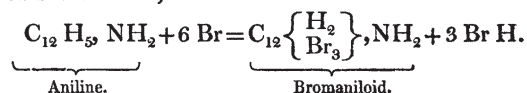


Fig. 2 Excerpt from an *MPCS* paper from 1843.

these representations. Examples of this type of representation were observed into the 1850s. Note also the use of equals signs rather than arrows in the reaction schemes.

In 1849, the *QJCS* superseded the *MPCS*, and by 1860 examples of attempts to convey more structural information were found in this journal, although in the example shown in Fig. 3⁴⁷ the arrangement of atoms and bonds is not clear.

The following experiments were undertaken, at the suggestion of Professor Kolbe, with the view of testing the correctness of the supposition, that the compound recently obtained by Wurtz from lactic acid, and described by him as chloride of laetyl $\text{C}_6\text{H}_4\text{O}_2\text{Cl}_2$ is the chloride of chloropropoxyyl $\left(\text{C}_4\left\{\begin{array}{c} \text{H}_4 \\ \text{Cl} \end{array}\right\}\right)\text{C}_2\text{O}_2\text{Cl}$. If this view be correct, the supposed chloride of chloropropoxyyl may be expected to decompose with water into chloropropionic and hydrochloric acids, and the product of the action of alcohol on that compound, described by Wurtz as chlorolactic ether, will exhibit the characters of chloropropionate of ethyl:

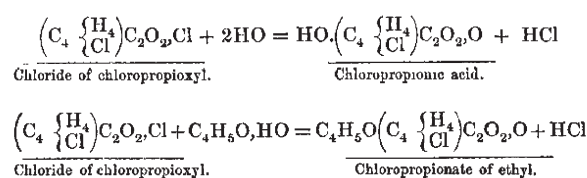


Fig. 3 Excerpt from a *QJCS* paper from 1860.

Note that parentheses are present in the structure shown in addition to braces.

The *Journal of the Chemical Society* was founded in 1862. An early paper in the journal shows an interesting convention in the form of the use of primes (Fig. 4).⁴⁸ Superscript primes were introduced in the 1850s, almost certainly in 1853 by William Odling, to represent “replaceable value” as compared with hydrogen.⁴⁹ They were not valences, but could be considered equivalent to what were later known as valences. Valence came into use around 1865, so from then primes must have represented valences, as in the case of the extract from a paper by Perkin from 1870 (Fig. 5).⁵⁰ In this example, double primes indicate oxygens in the C=O, or carbonyl, bond as found in quinone and alizarin. This paper makes for fascinating reading, and also includes swatches of fabrics dyed with compounds synthesised by Perkin, reproduced in colour in the RSC’s journal archive. Several conventions for structure representation in use at the time are present in this paper. In some compounds, lines are used to represent bonds in organic compounds. Perkin also uses braces, and his structures are thus between constitutional and structural formulae.

The reactions just described lend, I think, some support to Frankland and Kolbe’s view of the constitution of these acids, namely, that they are composed of two equivalents of carbonic acid, in which two equivalents of oxygen are replaced by a diatomic radical. However this may be, it is convenient, at all events, to formulate these bodies according to the carbonic acid type :—

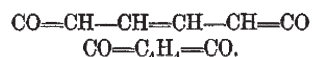


I propose to continue my researches in this direction, and to extend them to the cyanides of the triatomic radicals.

Fig. 4 Excerpt from a *JCS* paper from 1862.

This substance was discovered as early as 1838 by Woskresensky as a product of the oxidation of quinic acid, and although investigated by many chemists, as Wöhler, Laurent, Hofmann, Stenhouse, and others, no clue to its chemical structure had been obtained.

Kekulé, however, expressed the opinion that it might be a compound containing carbonyl, and represented thus :—



Graebe, after working with great energy upon quinone compounds, came to the conclusion that it is not constituted according to the above formula, but is a substitution-product of benzol in which two atoms of hydrogen are replaced by the group (O—O)”, in which half the combining values of oxygen saturate each other,* thus :—



Fig. 5 Excerpt from a *JCS* paper from 1870.

An early occurrence of benzene rings drawn with alternating double and single bonds was found in 1876 (Fig. 6).⁵¹ Generally, however, in *JCS* benzene rings were drawn without double bonds shown.

The properties and composition of the acid prove it to be terephthalic acid; the results therefore show that tolyl-phenyl gives, as a first product of oxidation (1 : 4), diphenylcarbonic acid, and as a final product terephthalic acid (1 : 4), from which it follows, as well as from the mode of its formation, viz., from bromobenzene and (1 : 4) bromotoluene, that its constitution is—

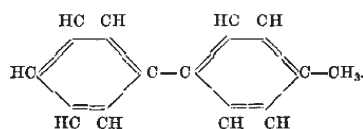
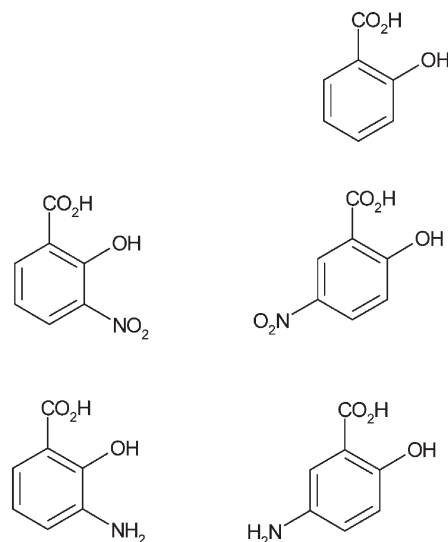


Fig. 6 Excerpt from a *JCS* paper from 1876.

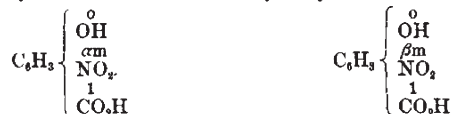
During this study, the use of an R symbol (in a brace structure) was observed in another *JCS* paper by Perkin, though there is no definition of R given and this cannot be ascertained with certainty from the context.⁵²

By the time of the first volume of *JACS* in 1879, structural chemistry was becoming better understood. An early example of a structure in *JACS* is in a report of an article published in *Justus Liebig’s Annalen der Chemie*. Examples of structural representations in this paper are given in Fig. 7.⁵³ As had already been observed in *JCS*, brackets or braces were used to convey structural information. The structures in this paper are easier to interpret than those published in earlier years and the compounds they represent are shown below (the structures are arranged in the same configuration on the page as those in the extract from *JACS* in Fig 7).



Orthohydroxybenzoic acid or salicylic acid, $\text{C}_6\text{H}_4 \left\{ \begin{array}{l} \text{OH} \\ \text{CO}_2\text{H} \end{array} \right.$, by

nitration, yields two metanitroorthohydroxybenzoic acids:



α metanitroorthohydroxybenzoic acid ; β metanitroorthohydroxybenzoic acid. The diethyl ethers of these acids on treatment with alcoholic ammonia, yield:



α orthoamidometanitrobenzoic acid ; β orthoamidometanitrobenzoic acid, which by azotation yield:

Fig. 7 Excerpt from a *JACS* paper from 1879. (Reprinted with permission from *Journal of the American Chemical Society*. © 1879 American Chemical Society.)

In the structures, in each case the carboxyl group is indicated at position 1, and the OH group at the *ortho* (*o*) position, *i.e.* position 2 in the modern convention. The positions of the nitro groups and the amino groups are indicated as being *meta* and β *meta* (representing the present-day 3- and 5-positions respectively). In accordance with Table 1, these representations employ the features which might be expected: *ortho*, *meta* and *para* were established terminology indicating substitution positions on phenyl rings; braces were still in common usage.

Another example of the type of organic compound representation from around that time is in a paper published in *JACS* in 1880 (Fig. 8).⁵⁴ In this paper the phenyl rings are represented by the short cut C_6H_n , where *n* is the number of hydrogen atoms on the ring, and the other substitution positions

Metazosulphoxybenzolephenol, $C_6H_4(HSO_3)-\overset{3}{N}=\overset{1}{N}-C_6H_4(\overset{4}{OH})$. This is prepared by the action of metadiazobenzosulpho acid on an alkaline solution of phenol. It is easily soluble in water and alcohol, but insoluble in ether.

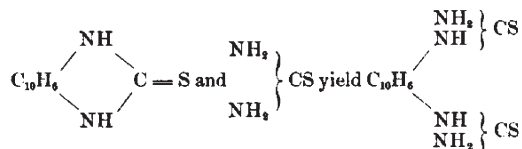
Azobenzosulphoxyphenol, $C_6H_5-N=\overset{1}{N}-C_6H_3(HSO_3)(\overset{4}{OH})$, prepared by the action of diazobenzole nitrate on an alkaline solution of orthophenolsulpho acid. The free acid is easily soluble in water and alcohol, but nearly insoluble in ether.

Fig. 8 Excerpt from a *JACS* paper from 1880. (Reprinted with permission from *Journal of the American Chemical Society*. © 1880 American Chemical Society.)

are shown as numbers above the substituent atoms. This is an example of a type of constitutional formula.

A further method of structure representation from the first volume of *JACS* is given in Fig. 9.⁵⁵ It seems that wherever possible the drawing of benzene rings is avoided – no rings are drawn for the naphthalene component of the compounds in Fig. 9. Instead there is a combination of brace formulae along with bonds and it is assumed that the reader will be able to interpret the structure from the name and formula. Note there are no arrows or equals signs in the reaction scheme. (Although equals signs were observed in papers as far back as the 1840s, arrows were not observed until the early 1900s.)

Now I think that, as the reaction progresses, No. 2 reacts on No. 1, and forms a compound containing 2CS, in place of 1, viz.:



This, then, may be looked upon as a disulpho-ureate of naphthalene.

Fig. 9 Excerpt from a *JACS* paper from 1879. (Reprinted with permission from *Journal of the American Chemical Society*. © 1879 American Chemical Society.)

In 1880 in *JACS* we see the use of equals signs in a reaction scheme⁵⁶ and in the same paper benzene rings (Fig. 10). Note the use of double bonds in the phenanthrene component which are omitted from the phenyl rings on the reactant side of the equation, and the absence of bonds between the rings and the CH_2Br groups. This difference in the non-use and use of double bonds in isolated and fused rings respectively was commonplace for many decades. Note that the phenanthrene ring system is described as anthracene. Also there is an extra hydrogen at one of the ring fusion carbon atoms, presumably a typographical error. A representation of phenanthrene in *JCS* in the same year is correctly named as such (Fig. 11); there is a direct acknowledgement of the work of Kelulé in this paper.⁵⁷

Generally, the structure diagrams from the late 1870s until the early 20th century were of better quality in *JCS* than *JACS*. Drawings were more consistent within and between papers,

The formation of anthracene may be represented as follows, whereby it will be seen that two mols. of orthobrombenzyl bromide unite with one another, giving off HBr , viz.:

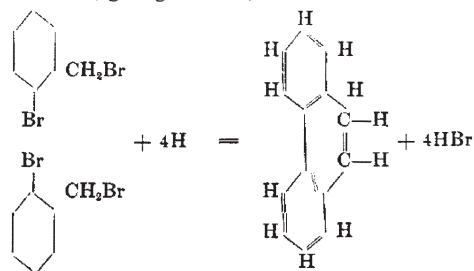


Fig. 10 Excerpt from a *JACS* paper from 1880. (Reprinted with permission from *Journal of the American Chemical Society*. © 1880 American Chemical Society.)

In the synthesis of phenanthrene from symmetrical diphenyl-ethylene we have, therefore, constructed a benzene nucleus from the ethylene residue $-\text{CH}=\text{CH}-$ on the one hand (which, as long as it existed in diphenyl-ethylene, showed by its entire behaviour that it did not belong to a benzene nucleus) and, on the other, two pairs of carbon atoms, each of which pairs was contained in the ortho-position by a separate already existing nucleus:—

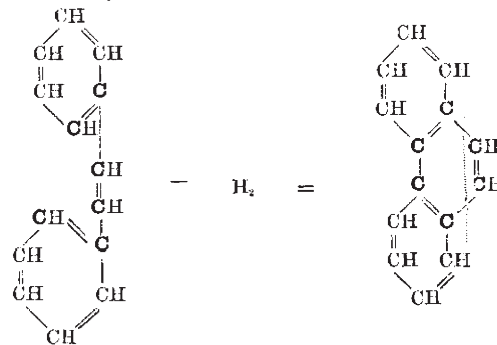


Fig. 11 Excerpt from a *JCS* paper from 1880.

there were fewer typographical errors and it seems that editorial control was superior.

In the 1890s, linear empirical formulae were still in common use, and often more structural information was conveyed through the trivial names of compounds than the formulae themselves, with reliance placed on the reader's ability to translate such names into structures.

In a paper by Fischer from 1890 in *JACS*⁵⁸ we see a representation of a sugar molecule and the use of parentheses to indicate the substitution of the OH groups on the chain (Fig. 12), although this structure conveys no information about the three-dimensional structure of the sugar, as would be expected from the dates in Table 1.

On the other hand the structure of the three other important sugars was essentially known before my own researches began.

The formulae now in use for glucose and galactose:

$\text{CH}_2(\text{OH})$. $\text{CH}(\text{OH})$. $\text{CH}(\text{OH})$. $\text{CH}(\text{OH})$. $\text{CH}(\text{OH})$. COH ; and for laevulose:

$\text{CH}_2(\text{OH})$. $\text{CH}(\text{OH})$. $\text{CH}(\text{OH})$. $\text{CH}(\text{OH})$. CO . $\text{CH}_2(\text{OH})$ are derived from the following facts:

Fig. 12 Excerpt from a *JACS* paper from 1890. (Reprinted with permission from *Journal of the American Chemical Society*. © 1890 American Chemical Society.)

Referring to Table 1, the terms *cis*- and *trans*- would be expected to appear in the literature from around 1888. An occurrence of their usage was found in a *JCS* paper from 1895 (Fig. 13).⁵⁹ Note that this paper also specifically mentions the concept of stereoisomers.

In 1890, an example of the use of X to represent a generic group in a *JCS* paper was found.⁶⁰ From the context, it appears that X can be any group, and is not specifically a halogen as might be expected in accordance with common practice today.

A careful examination of the formula of hexahydro-orthotoluic acid indicates at once that this acid, like hexahydrophthalic acid, may exist in two stereoisomeric modifications

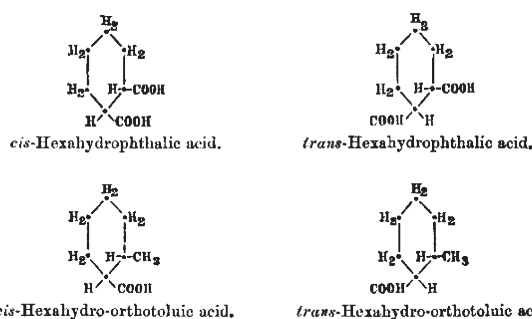
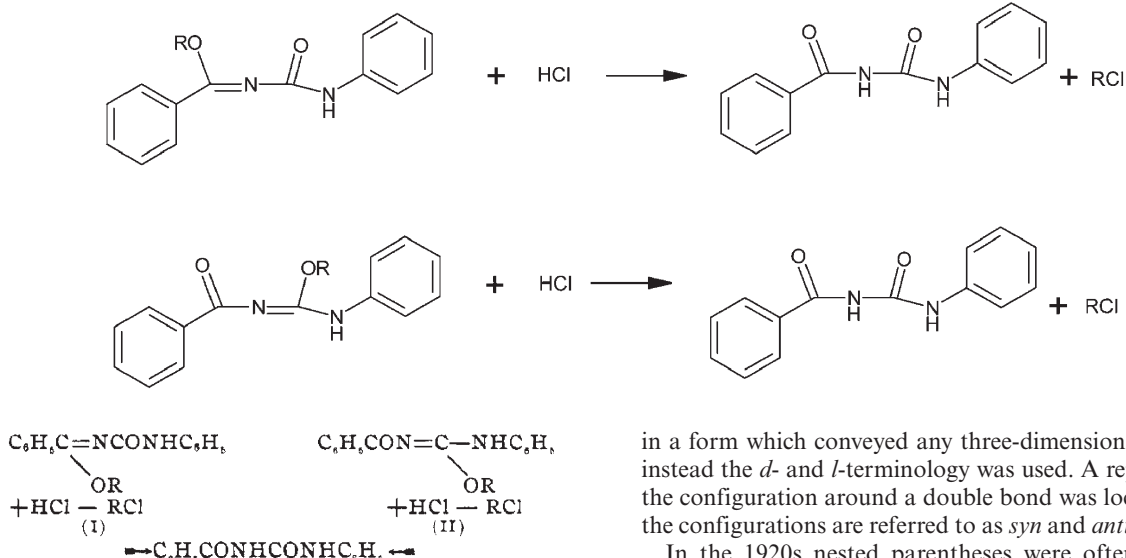


Fig. 13 Excerpt from a *JCS* paper from 1895.

In a later paper by the same authors in the same year of *JCS*, an X was used specifically to represent a halogen, however.⁶¹

A paper from *JACS* from 1900 (Fig. 14) shows an interesting representation of a reaction:⁶² both reactants, which are isomers, react with HCl to give the same product, the dash between the HCl and the RCl being a minus sign, presumably indicating that an alkyl chloride is a product of the reaction. It appears that the R group represents a generic alkyl group, as no definition is given in the paper. A modern representation of these reactions is:

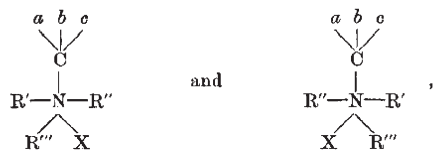


It will be noticed that the imido ester grouping in the different positions in these isomers are affected alike, inasmuch as under similar conditions both compounds give with hydrogen chloride the same derivative, benzoylphenylurea. The urea-imido esters

Fig. 14 Excerpt from a *JACS* paper from 1900. (Reprinted with permission from *Journal of the American Chemical Society*. © 1900 American Chemical Society.)

By the early 1900s a number of papers were appearing which discussed stereochemistry and asymmetric carbon atoms. An example from *JCS* is shown in Fig. 15;⁶³ note that there is no use of wedges, dashed bonds or other notations to indicate the stereochemistry. The same author described a kind of molecular models in a later paper in the same year, along with a detailed explanation of their operation.⁶⁴ In 1910 in *JCS* stereoisomeric compounds are clearly represented by use of a Fischer projection (see Fig. 16).⁶⁵

The two compounds which could be produced from an active amine might be conventionally represented by plane projections of the space formulæ, thus:



and would bear to one another a relation similar to that of the α - and β -glucoheptonic acids or the two $\alpha\beta$ -dihydroxybutyric acids.

Fig. 15 Excerpt from a *JCS* paper from 1905.

of two isomeric levorotatory glycols, each one of which contains two asymmetric carbon atoms, is theoretically possible:

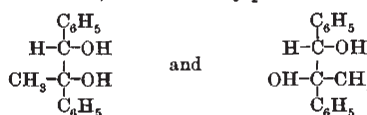


Fig. 16 Excerpt from a *JCS* paper from 1905.

In 1910, there are further examples of linear formulae of the type used in 1879 previously shown, with numbers above the atom symbols to indicate substitution positions. A further notation spotted in 1910 was use of a colon to represent a double bond in a linear formula, e.g. $\text{C}_6\text{H}_5\text{C}(:\text{N})\text{OCH}_3$. Generally, structure diagrams are becoming more common by 1910.

Representation of three-dimensional structures around 1910 is inconsistent. For example, in a paper discussing the separation of optical isomers (which refers back to the work of van't Hoff and Le Bel),⁶⁶ no attempt was made to represent these graphically

in a form which conveyed any three-dimensional information; instead the *d*- and *l*-terminology was used. A representation of the configuration around a double bond was located, although the configurations are referred to as *syn* and *anti*.⁶⁷

In the 1920s nested parentheses were often employed to represent side chains. Local definitions of R groups were becoming more prevalent, as opposed to the assumption that R always represented an alkyl group. Equals signs seem to have been almost entirely replaced by arrows in reactions, and examples of the modern equilibrium sign were found. Some structures in *JACS* were hand drawn, as in the example in Fig. 17, which shows a representation of a three-dimensional structure.⁶⁸ This paper explains the mechanism of reactions, and electrons are shown in the structure, although not the "curly arrows" used more recently to show mechanisms. In the same paper, there are also more usual structures, with the element symbols type-set.

these hydrogens. In the presence of an alkyl group, (R), a third type of elimination (C) becomes possible, and the intermediate structure thus formed may rearrange, by trans-migrations² of R and the hydroxyl-hydrogen, into a ketone:

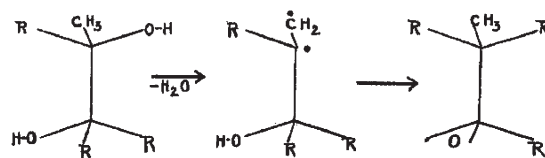


Fig. 17 Excerpt from a *JACS* paper from 1920. (Reprinted with permission from *Journal of the American Chemical Society*. © 1920 American Chemical Society.)

An example of a three-dimensional structure in a bridge compound is given in Fig. 18.⁶⁹ Structures of this type were observed during this study in both *JCS* and *JACS*.

This substance they have prepared, and propose to term *homocamphor* (III). It was originally hoped to make the research a

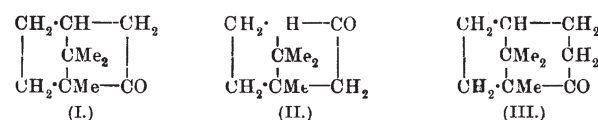


Fig. 18 Excerpt from a *JCS* paper from 1920.

In 1925 a further type of diagram was found to represent sugars in *JCS*, as shown in Fig. 19.⁷⁰

It is of great interest that galactose and arabinose exhibit this marked tendency to react in both 1:5- and 1:4-forms—the former being the normal form in both cases—under conditions which, so far as our own observations and those of other workers go, lead to the production of only one form in the cases of glucose

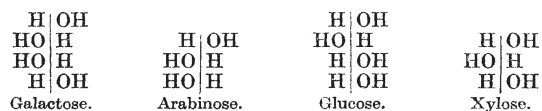


Fig. 19 Excerpt from a *JCS* paper from 1925.

In 1930 benzene rings continued to be drawn without the double bonds shown. Sometimes bonds were drawn between rings and their substituents, sometimes not. Occasionally an equals sign was found in a reaction, but mostly these had been replaced by arrows. In *JACS*, structures are generally better drawn than previously, but still with some inconsistencies and some representations which could be confusing for the modern reader (Fig. 20).⁷¹

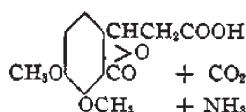
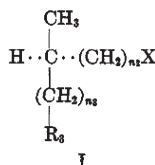


Fig. 20 Excerpt from a *JACS* paper from 1930. (Reprinted with permission from *Journal of the American Chemical Society*. © 1930 American Chemical Society.)

A further type of three-dimensional representation (Fig. 21) was found in the first volume of *JOC* (note also the local definition of X and R).⁷² In *JCS* in 1930 there were examples of carbohydrate rings drawn to show the three-dimensional configuration (Fig. 22).⁷³

Given a substance of known configuration of the general type of formula I,



(where X = a functional group; R₃ = an alkyl or an aryl radical; n₂ and n₃ = 0 or an integer), the configuration of the derivative obtained by substitution of X by Y cannot be predicted when n₂ = 0, whereas when

Fig. 21 Excerpt from a *JOC* paper from 1936. (Reprinted with permission from *Journal of Organic Chemistry*. © 1936 American Chemical Society.)

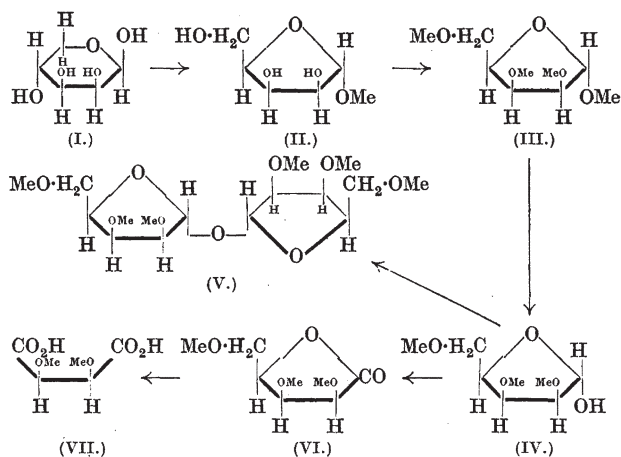
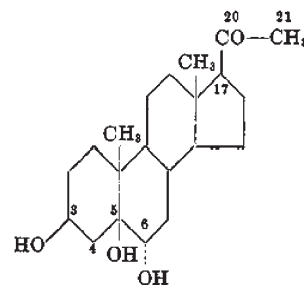


Fig. 22 Excerpt from a *JCS* paper from 1930.

Steroid structures appear in the literature in the late 1930s, and indications of the stereochemistry are given. In Fig. 23 the OH group at the 6-position is indicated as being into the plane of the page, other bonds at asymmetric carbon atoms being out of the page and represented by solid, but not wedge, bonds.⁷⁴ This type of representation for steroids and their stereochemistry continued until the end of the time period investigated in this study.



II. Pregnone-20-one-3(β)-5,6(trans)-triol.

Fig. 23 Excerpt from a *JOC* paper from 1940. (Reprinted with permission from *Journal of Organic Chemistry*. © 1940 American Chemical Society.)

Unlike most of the general structure representations in two dimensions, for which there are no scientific reasons for the structures to be drawn differently (e.g. the presence or absence of a bond from a benzene ring carbon to a substituent), the various types of three-dimensional representation can to some degree be justified as they are often more suitable for a particular type of compound. Even with modern computerised structure drawing packages, there are still differences today in the representation of different types of compound in three dimensions.

A further convention encountered in an early volume of *JOC* was the use of φ to represent a phenyl ring, and benzene rings with alternating single and double bonds were becoming more commonplace by around 1940. There were few other changes of significance in the 1940s and 1950s.

Study of United States patents

A similar study to that above was made for patents, to examine how structures were represented, how structure drawings have varied with time and how the structures in patents compare with those in journals. It would be expected that Markush structures would occur in patents more frequently than journals, but otherwise there is no evidence to suggest that structures in patents would be different to those in journals.

Patents were identified by searching the USPTO website⁷⁵ for patents at organic chemistry classification numbers 534 (azo dyes), 536 (carbohydrates) and 540 (heterocycles) in order to identify candidate patents likely to contain organic structural information for study. These classifications were selected because there were likely to be sufficient patents in these subject areas during the period under investigation. United States patents were chosen as all US patents back to 1836 are available in .pdf or .tiff image format, enabling them to be reproduced. Patents were chosen at random, approximately one out of every ten patents being selected, although many were quickly dismissed as the titles of the patents were not available prior to display of the full document so no preliminary filtering was possible to determine the likelihood of structures being present.

Little in the way of structural information through diagrams is presented in patents from before 1860, which is as would be expected on the basis of the information in Table 1. In 1865, an example of the use of braces was found (Fig. 24).⁷⁶ The brace structure in this patent is similar to that in the 1843 paper by Hofmann already cited⁴⁶ and conveys little structural information, only the NO₂ part suggesting a structure fragment, though it is not clear what the structure of this would be.

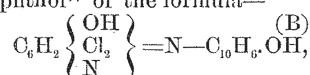
ters into the composition, forming an explosive compound of the chemical formula $C_{12}H_7N_3O_{22}$, which may also be expressed $C_{12} \left\{ \begin{matrix} H_7 \\ 3NO_4 \end{matrix} \right\} O_{10}$.

One hundred parts of the said explosive compound contains—
 $C = 24.24$
 $H = 2.37$
 $N = 14.14$
 $O = 59.25$ } as expressed by weight and corresponding to the chemical formula above given.

Fig. 24 Excerpt from a US patent from 1865.

A more conventional use of braces was found in 1884 (Fig. 25), representing structures in a similar fashion to those found in *JACS* and *JCS*.⁷⁷ In the first structure shown the position of the OH group on the naphthalene ring system is shown by the symbol (B), which represents the β position. The positions of the substituents on the phenyl ring on the left of the structures are not indicated in the formulae. In the patent for a dyestuff in Fig. 26 from 1882, there are examples of explicit reaction schemes.⁷⁸ Here double bonds are clearly represented and although no rings are drawn, the information conveyed is enhanced by the names of the compounds. An equals sign is present between the reactants and products. Elsewhere in the patent there appears to be a typographical error, with $NaNO_2$ written as $NaNO_2$. Typographical errors are commonplace in formulae and structures in patents at the time and to a greater degree than was found in *JCS* and *JACS*.

beta-naphthol" of the formula—



hereinafter using this name to designate it.

To prepare this compound I take the body known as "ortho-amido-dichlorphenol," which may have been prepared in various ways—as, for instance, by reducing the ortho-nitro-dichlorphenol of F. Fischer, (*Zeitschrift für Chemie*, 1868, p. 386.) The amido-dichlorphenol is diazotized in the well-known manner, and the diazo-dichlorphenol of Schmidt and Glutz, (*Berichte der Deutschen Chem., Ges.*, II, p. 52,)—a yellowish-brown body of

the formula $C_6H_5 \left\{ \begin{matrix} Cl_2 \\ O \\ N \end{matrix} \right\} = N$ —is obtained. The

Fig. 25 Excerpt from a US patent from 1884.

nia of ten per cent. strength. My new coloring-matter is the result of the reaction of the substances in solution. The following equation exhibits the changes more specifically:

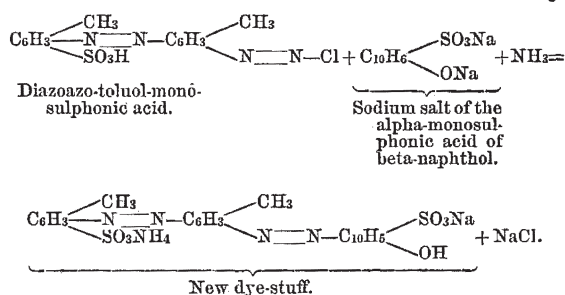


Fig. 26 Excerpt from a US patent from 1882.

Also in 1882 an example of the use of dots and commas separating structural fragments was found, as might be expected from Table 1. Use of R groups to represent alkyl groups in structures in patents was also observed in 1882, as shown in Fig. 27.⁷⁹

The problem of drawing diagonal lines for bonds is illustrated in the structure in Fig. 28, although this structure does have the β symbol shown correctly indicating the substitution position of the NH_2 group on the naphthalene.⁸⁰

these substances I proceed in the following way: Any primary amine is combined with resorcine (in aqueous alkaline solution) in the way sufficiently known.

The general reaction may be shown by the following equations:

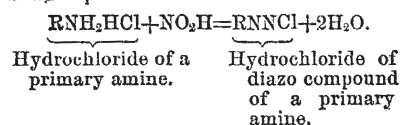


Fig. 27 Excerpt from a US patent from 1882.

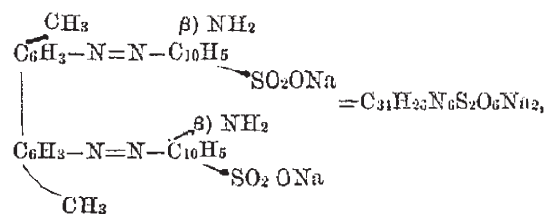


Fig. 28 Excerpt from a US patent from 1885.

As with journals at the time, there is little standardisation of the representation of structures. Particularly for patents, sometimes a considerable amount of educated guesswork is required in order to attempt to determine the structures being discussed.

An example of a structure which is difficult, if not impossible, to interpret is given in Fig. 29.⁸¹ By 1895 there are examples of clearly written structures (see Fig. 30),⁸² though with braces still present, as would be expected from the dates in Table 1. Even though the rings are not drawn in the structure in Fig. 31,⁸³ this is one of the more explicit structure representations discovered in a patent so far, but there is still uncertainty about the substitution positions on the phenyl rings (although there are some clues in the text of the patent and an expert in dye chemistry may be able to interpret this).

disulpho-acid with mineral acids. If the acid H is combined with one equivalent of a diazo compound in a solution containing a free mineral acid, products are obtained which show the property of reacting with a second equivalent of diazo compounds in alkaline solution, forming bodies of the general formula:

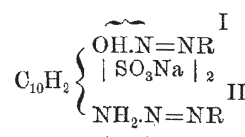


Fig. 29 Excerpt from a US patent from 1892.

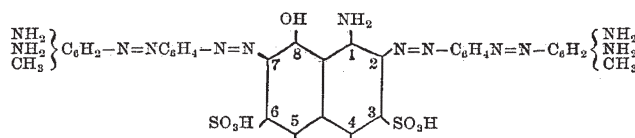


Fig. 30 Excerpt from a US patent from 1895.

As already mentioned, Markush structures assumed that name in 1924. An early example of a Markush structure with a structure diagram, from 1899, was identified in the course of this work; the Markush structure is reproduced in Fig. 32, and shows variation in both the nature of the substituents and also the position of substitution.⁸⁴ A number of additional examples of Markush structures were found around 1900.⁸⁵

The representations of structures in patents generally became clearer and less ambiguous after 1900, and the numbers of compounds which have structure drawings present increases as time progresses. Benzene rings in patents were usually drawn without double bonds until the middle of the 20th century,

stuff having in case the tetrazo derivative
tolidin is employed, the formula

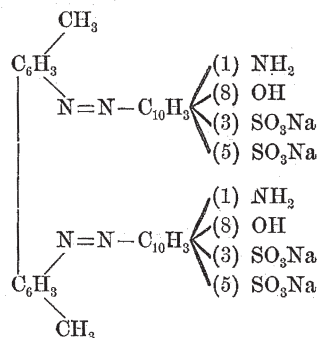
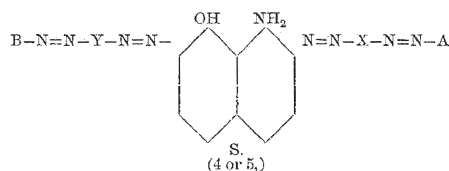


Fig. 31 Excerpt from a US patent from 1898.

My new coloring-matters therefore possess
the following general formula:



in which $-N=N-X-N=N-A$ is a tetrazo
residue containing one or more basic groups
and $-N=N.Y-N=N-B$ is a tetrazo residue
containing one or more hydroxyl groups.

Fig. 32 Excerpt from a US patent from 1899.

examples with double bonds being found in the 1940s,⁸⁶ although
pyridine and other non-benzene rings often have double bonds
shown much earlier; this is in line with the findings from *JCS*
and *JACS*.

Concerning stereochemistry, examples of steroid structures
without the stereo-configuration indicated were found in the
1940s, a steroid with dotted bonds to represent the orientation
of the bond not being found until the 1950s, though as the
study involved selecting patents at random this does not
necessarily mean this was the first such structure in a patent. An
example of a sugar with its stereochemistry defined is shown in
Fig. 33⁸⁷ and an even clearer representation was located in 1949
(see Fig 34).⁸⁸

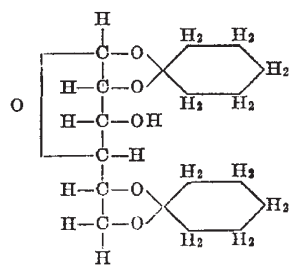


Fig. 33 Excerpt from a US patent from 1933.

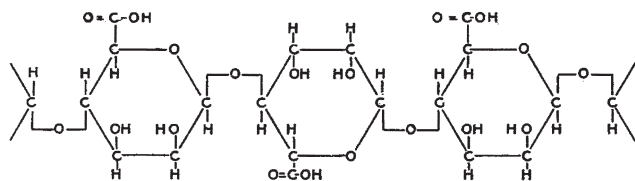


Fig. 34 Excerpt from a US patent from 1949.

Many of the patents discussed in the late 19th century were
for dyestuffs, which is not surprising as this was a major growth
area in organic chemistry at the time and led to production of
the materials most suitable for commercial application, and thus
were likely to be patented. The patent classification for dyestuffs
was deliberately chosen for this study, along with two others

likely to produce patents containing structures, but during the
19th century almost all patents which contained structures were
from the dyestuffs classification.

Searching and interpreting chemical structures

It is probably safe to say that no searching for chemical
information by means of structure drawings occurred until the
1940s, when publications such as the American Chemical
Society's *Ring Index* started to be published.⁸⁹ The *Ring Index*
provided ring systems arranged systematically along with
their bibliographic references. Ever since chemists started to
represent chemical substances by means of names, formulae and
symbols, there was the need to find information about specific
compounds, and this was primarily through the use of indexes
where names would have been searched for in the same way as
other topics. This was not ideal, as the innovative creators of
chemical information sources, such as Beilstein and Gmelin,
realised, and they sought to organise chemical information
with the focus much more on chemical structures. As there
were no direct means of searching by structures, the methods
they used were by sophisticated classification schemes, linear
formula and chemical name indexes. Prior to the development
of classification systems, the chemist searcher would have a
structure diagram in his/her head or written down which would
have needed to be translated into the appropriate text format for
the particular index or search tool to be used.

The chemical structure was becoming the natural language
of chemists by the start of the 20th century, as indicated by the
numbers of structure diagrams which were appearing in papers
and patents by that time. This meant that the chemist needed
to learn chemical nomenclature systems in order to search the
chemical literature for chemical compounds.

Once relevant publications had been identified, the chemist
then had the challenge of interpreting the structural information
in the documents. Much structural information in papers in the
late 1800s and early 1900s was buried in chemical names, very
often trivial names rather than systematic names – a chemist had
to be able to translate the non-systematic name into a structural
representation either mentally or on to paper. Sometimes
structures were represented as formulae, some parts of the
formulae not conveying structural information but assuming
that the chemist could translate the formula because of the
context in which it occurred. Much knowledge of the language
of chemistry of the time was assumed, and this language was
still confused and had many dialects until well into the 1900s.
The situation must have been particularly difficult for patents,
in which, as already mentioned, the structural information was
often poorly presented and even potentially misleading.

In the 20th century, the structure diagram increasingly
replaced chemical nomenclature in publications, reducing the
need for the chemist reader to know the structure implied
by a name or linear formula. This has in part been related to
improvements in printing technology and also the evolution of
conventions for the representation of structures.

Discussion

In the late 19th century the chemistry community was not
accustomed to seeing structures in print, being more familiar
with chemical names, so the absence of structural diagrams,
which we would regard today as a major barrier to the
communication of chemical information, probably was not seen
as such at the time.

At several points in this paper, it has been mentioned that a
major barrier to the representation of chemical structures has
been the limitations of printing technology. Some of the early
representations proposed for atoms used complex symbols, easy
to sketch by hand, but not simply transferred to print format.
Even once modern symbols for atoms had been adopted, new
structure theories and consequent means of representing bonds

and rings meant that there were on-going technical problems associated with transferring structure diagrams to print.

Possibly for reasons of conservation of space as well as limitations of printing, there appears to have been reluctance by publishers of journals or patents to print structure diagrams. There are many occasions when a unique structure diagram would have simplified the communication of chemical information in *JCS*, *JACS* and patents, especially as the complexity of structures increased.

However, as time progressed, publishers strove to print structure diagrams, or at least to produce formulae which conveyed structural meaning. Probably in part due to the numerous representations of structures that existed, particularly in the 19th century, there was little standardisation in their presentation. It might be expected that publications such as *JCS* and *JACS*, which have a house style, would have had definitive conventions for representing structures, but this does not appear to have been the case in the first few decades of their existence. However, in *JCS* and *JACS* after around 1870 a diagram for a compound under discussion can usually be drawn from the information provided, either from the structure fragments or names assigned, or a combination of the two.

Many of the 19th century patents contained ambiguities in their structure representations. This raises questions as to whether these were deliberate, or simply a consequence of the inadequate printing technology available. There was also sometimes genuine uncertainty about the precise structures of the compounds under investigation, for example the location of substituents in benzene and naphthalene derivatives, the structures often only being determined explicitly later in the 20th century. In general, it was the process that was generally more important in patents rather than the precise chemical structure.

Structural formulae were more common after around 1900, by which time many structure determination problems had been resolved. Before that time the lack of structural knowledge was sometimes used by patentees in disputes to protect their inventions, sometimes as a means of revealing the minimum of information. Certainly there were claims that the Germans in some cases deliberately resorted to disinformation, for example by omitting precise conditions for critical steps in reactions.⁹⁰ William Henry Perkin Junior, in the Presidential Address to the Chemical Society in 1915, highlighted some of the problems associated with using German patents, and the fact the British Government seemed to think that it was a simple matter to synthesise the dyes described by following instructions in the patent documents:⁹¹

“It seems to be imagined in many quarters that, in order to manufacture a dye which had previously been made in Germany all that is necessary is to follow the directions given in the patent dealing with that particular dye. No greater mistake could possibly be made. It is common knowledge that German manufacturers have for many years devoted large sums to the establishment of an efficient staff of patent experts, whose business it is to so word a patent that, whilst it satisfies the requirements of the patent laws of the various countries in which it is taken out, it only gives such information as is absolutely necessary, and contains no indication of the process which is used in the actual manufacture. In many cases patents are devised which are of no practical value, and are merely intended to mislead and throw the competitors on the wrong scent.”

If carried through to US patents, which is quite likely bearing in mind the number of German companies which took out patents in the USA and were identified in the present work, then the confusing structures may well have on occasions been deliberately misleading.

It is also interesting to note that there were numerous references to the *Beilstein Handbuch* in patents studied in this work, confirming its importance to industrial chemists. During the earlier part of the time period under consideration, *Beilstein* was the only authoritative information tool available, and its

reliability by virtue of only including evaluated information must have been a benefit, the fact that it was published in German not appearing to be a deterrent to its use.

It would be expected that there would be less standardisation in patents than in respected journals such as *JCS* and *JACS*, and this is indeed the case. It almost seems that no two patents, unless produced by the same assignee, used the same conventions in the late 19th century. This combined with the ambiguities, due to lack of knowledge of precise structures, makes patents from that era difficult to interpret.

Standardisation of structure diagrams was not organised in the same way as chemical nomenclature. The IUPAC⁹² and its predecessors have been involved with standardisation of chemical nomenclature, as have organisations such as the Chemical Society, which first published nomenclature and notation guidelines in 1882, and the Chemical Abstracts Service. Whereas publishers might have insisted on a name being supplied with a manuscript in accordance with a particular naming convention, this does not seem to have been the case for structure diagrams. As printing techniques improved and structures became easier to reproduce, over time the problems of representation of structures to some extent solved themselves, and unwritten rules ensured that structure diagrams became more standardised. Indeed, only now is the IUPAC addressing the issue of standardisation of structure representations.

Many factors have led us to the position where we are today, and these have resulted from synergies between structural theory, communication of chemistry by symbols, printing techniques, conventions and, nowadays, computer technology.

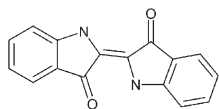
Acknowledgements

Sincere thanks to Dr Tony Travis, Sidney Edelstein Center for the History and Philosophy of Science, Technology and Medicine, Hebrew University of Jerusalem, for his help with interpretation of the patenting of dyes in the 19th century, his advice and support.

References

- 1 R. Hoffmann and P. Laszlo, *Angew. Chem., Int. Ed. Engl.*, 1991, **30**, 1–16.
- 2 K. Kolli, in *Pioneering Ideas for the Physical and Chemical Sciences: Josef Loschmidt's Contributions and Modern Developments in Structural Organic Chemistry, Atomistics and Statistical Mechanics* (Proceedings of the Josef Loschmidt Symposium, held 25–27 June 1995, Vienna), W. Fleischhacker and T. Schönfeld, eds., Plenum Press, New York, 1997, pp. 297–300.
- 3 M. P. Crosland, *Historical Studies in the Language of Chemistry*, Heinemann, London, 1962.
- 4 P. E. Verkade, *A History of the Nomenclature of Organic Chemistry*, D. Reidel, Dordrecht, Boston, Lancaster, 1985.
- 5 F. S. Taylor, *Ambix*, 1937, **1**, 64–67.
- 6 J. R. Partington, *Ambix*, 1937, **1**, 61–64.
- 7 M. P. Crosland, *Historical Studies in the Language of Chemistry*, Heinemann, London, 1962, p. 245–255.
- 8 D. McKie, *Ambix*, 1937, **1**, 75–77.
- 9 J. R. Partington, *A History of Chemistry*, MacMillan, London, 1962, vol. 3, p. 54.
- 10 J. R. Partington, *A History of Chemistry*, MacMillan, London, 1962, vol. 3, p. 192.
- 11 R. M. Caven and J. A. Cranston, *Symbols and Formulae in Chemistry: an Historical Study*, Blackie & Son, London, Glasgow, 1928, p. 18.
- 12 Crosland states that one of the first chemical equations was sketched by Jean Beguin in 1615, which showed that he understood the compositions of several substances (M. Crosland, *Science in Context*, 1996, **9**, 225–240).
- 13 J. R. Partington, *A History of Chemistry*, MacMillan, London, 1962, vol. 3, p. 483.
- 14 R. M. Caven and J. A. Cranston, *Symbols and Formulae in Chemistry: an Historical Study*, Blackie & Son, London, Glasgow, 1928, p. 25.
- 15 T. L. Alborn, *Ann. Sci.*, 1989, **46**, 437–460.
- 16 J. R. Partington, *A History of Chemistry*, MacMillan, London, 1964, vol. 4, p. 259.

- 17 A. J. Idhe, *The Development of Modern Chemistry*, Harper Row, New York, 1964, pp. 191–198; A. J. Idhe, *The Development of Modern Chemistry*, Harper Row, New York, 1964, pp. 208–216.
- 18 A. S. Travis, *The Rainbow Makers: the Origins of the Synthetic Dyestuffs Industry in Western Europe*, Associated University Press, London, Toronto, 1993, p. 76; A. S. Travis, *The Rainbow Makers: the Origins of the Synthetic Dyestuffs Industry in Western Europe*, Associated University Press, London, Toronto, 1993, Appendix pp. 245–254.
- 19 A. J. Idhe, *The Development of Modern Chemistry*, Harper Row, New York, 1964, pp. 224–225.
- 20 R. M. Caven and J. A. Cranston, *Symbols and Formulae in Chemistry: an Historical Study*, Blackie & Son, London, Glasgow, 1928, p. 66.
- 21 J. Walker, *J. Chem. Soc.*, 1923, 940–946.
- 22 R. M. Caven, J. A. Cranston, *Symbols and Formulae in Chemistry: an Historical Study*, Blackie & Son, London, Glasgow, 1928, p. 68.
- 23 M. P. Crosland, *Historical Studies in the Language of Chemistry*, Heinemann, London, 1962, p. 331.
- 24 Priesner (see: C. Priesner, *Chem. Int.*, 1989, **11**, 216–224; C. Priesner, *Chem. Int.*, 1989, **11**, 237–238) states that Couper's structures were in fact representations of forces between molecules rather than actual structure diagrams, and that the concept of a molecular structure was introduced by the Russian chemist Butlerov. Farrar and Farrar (see ref. 25) suggest that the claims of Butlerov depend more on repetition than documentary evidence.
- 25 W. V. Farrar and K. R. Farrar, *Proc. Chem. Soc.*, 1959, 285–290.
- 26 M. P. Crosland, *Historical Studies in the Language of Chemistry*, Heinemann, London, 1962, p. 332.
- 27 Originally reported by J. Loschmidt, *Chemische Studien*, Vienna, 1861, and described by R. Anschütz, *Ber.*, 1912, 539–553. Also reported by R. M. Caven and J. A. Cranston (see ref. 29).
- 28 Note that Loschmidt's structure for indigo is not a correct representation of the structure, which was determined in 1883 by Bayer:



- 29 R. M. Caven and J. A. Cranston, *Symbols and Formulae in Chemistry: an Historical Study*, Blackie & Son, London, Glasgow, 1928, p. 75.
- 30 W. J. Wiswesser, *J. Chem. Inf. Comput. Sci.*, 1985, **26**, 258–263.
- 31 W. J. Wiswesser, *J. Chem. Doc.*, 1968, **3**, 146–150.
- 32 J. R. Partington, *A History of Chemistry*, MacMillan, London, 1964, vol. 4, p. 540.
- 33 J. Hudson, *The History of Chemistry*, 1992, Chapman & Hall, New York, pp. 138–139.
- 34 J. R. Partington, *A History of Chemistry*, MacMillan, London, 1964, vol. 4, pp. 555–558.
- 35 There have been a number of discussions about the explanation of the origins of Kekulé's representations. See, for example, *The Kekulé Riddle: a Challenge for Chemists and Psychiatrists*, J. H. Wotiz, ed., Cache River Press, Clearwater, Florida, 1993.
- 36 K. A. Kekulé, *Bull. Soc. Chim. Paris*, 1865, **3**, 98–111.
- 37 R. M. Caven and J. A. Cranston, *Symbols and Formulae in Chemistry: an Historical Study*, Blackie & Son, London, Glasgow, 1928, p. 77.
- 38 C. A. Russell, *Chem. Br.*, 1977, **13**(1), 4–7; C. A. Russell, *Chem. Br.*, 1977, **13**(1), 20.
- 39 J. R. Partington, *A History of Chemistry*, MacMillan, London, 1964, vol. 4, pp. 749–759.
- 40 Examples were present in *Justus Liebigs Annalen der Chemie*, as reported by J. E. Gordon and J. C. Brockwell (see ref. 41).
- 41 H. Kondo and T. Kondo, *J. Prakt. Chem.*, 1930, **126**, 24–52, reported in J. E. Gordon and J. C. Brockwell, *J. Chem. Inf. Comput. Sci.*, 1983, **23**, 117–134.
- 42 E. A. Markush, Pharma-Chemical Corporation, New York, 'Pyrazolone Dye and Process of Making the Salt', *US Pat.*, 1,506,316, 1924.
- 43 W. H. Perkin, 'Dyeing Fabrics', *GB Pat.*, 1984, 1856; M. F. Lynch, private communication of observations made as part of a conference address in 1984.
- 44 *US Pat.*, 1150580, 1915, described by J. F. Sibley, *J. Chem. Inf. Comput. Sci.*, 1991, **31**, 5–9.
- 45 For an account of the history of chemistry journals, see: H. Cooke, *Learned Publishing*, 2004, **17**, 125–134.
- 46 A. W. Hofmann, *Mem. Proc. Chem. Soc.*, 1843, **2**, 265–300.
- 47 C. Ulrich, *Q. J. Chem. Soc.*, 1860, **12**, 23–26.
- 48 M. Simpson, *J. Chem. Soc.*, 1862, **15**, 134–141.
- 49 A. S. Travis, private communication, 2004.
- 50 W. H. Perkin, *J. Chem. Soc.*, 1870, **23**, 133–143.
- 51 T. Carnelley, *J. Chem. Soc.*, 1876, **29**, 13–23.
- 52 W. H. Perkin, *J. Chem. Soc.*, 1877, **32**, 90–103.
- 53 G. A. Prochazka (abstractor), *J. Am. Chem. Soc.*, 1879, **1**, 233–260.
- 54 J. H. Stebbins, *J. Am. Chem. Soc.*, 1880, **2**, 236–246.
- 55 J. H. Stebbins, *J. Am. Chem. Soc.*, 1879, **1**, 415.
- 56 J. H. Stebbins, *J. Am. Chem. Soc.*, 1880, **2**, 199–203.
- 57 F. R. Japp, *J. Chem. Soc.*, 1880, **37**, 83–90.
- 58 E. Fischer, *J. Am. Chem. Soc.*, 1890, **12**, 340–348.
- 59 W. Goodwin and W. H. Perkin, jun., *J. Chem. Soc.*, 1895, **67**, 119–128.
- 60 R. Meldola, *J. Chem. Soc.*, 1890, **57**, 328–331.
- 61 R. Meldola and F. Hughes, *J. Chem. Soc.*, 1890, **57**, 393–404.
- 62 H. L. Wheeler and W. M. Sanders, *J. Am. Chem. Soc.*, 1900, **22**, 365–378.
- 63 H. O. Jones, *J. Chem. Soc.*, 1905, **87**, 135–144.
- 64 H. O. Jones, *J. Chem. Soc.*, 1905, **87**, 1721–1735.
- 65 A. McKenzie and H. Wren, *J. Chem. Soc.*, 1910, **97**, 473–486.
- 66 W. A. Noyes and L. P. Kyriakides, *J. Am. Chem. Soc.*, 1910, **32**, 1057–1061.
- 67 C. V. Gortner and R. A. Gortner, *J. Am. Chem. Soc.*, 1910, **32**, 1294–1296.
- 68 A. Michael, *J. Am. Chem. Soc.*, 1920, **42**, 787–821.
- 69 A. Lapworth and F. A. Royle, *J. Chem. Soc.*, 1920, **117**, 743–750.
- 70 J. Pryde, E. L. Hirst and R. W. Humphreys, *J. Chem. Soc.*, 1925, **127**, 348–357.
- 71 W. M. Rodionow and A. M. Federova, *J. Am. Chem. Soc.*, 1930, **52**, 368–371.
- 72 P. A. Levene and A. Rothen, *J. Org. Chem.*, 1936, **1**, 76–133.
- 73 H. G. Bott, E. L. Hirst and J. A. B. Smith, *J. Chem. Soc.*, 1930, 658–669.
- 74 M. Ehrenstein and T. O. Stevens, *J. Org. Chem.*, 1940, **5**, 318–328.
- 75 United States Patent and Trademark Office, Patents, www.uspto.gov/main/patents.htm, 2003.
- 76 J. J. Révy, 'Improvement in the Manufacture of Gun-Cotton', *US Pat.*, 50082, 1865.
- 77 A. Spiegel, 'Azo Coloring-Matter', *US Pat.*, 302790, 1884.
- 78 C. Rumpff, 'Dye-Stuff or Coloring-Matter', *US Pat.*, 256375, 1882.
- 79 O. Wallach, 'Manufacture of Coloring-Matter', *US Pat.*, 269359, 1882.
- 80 C. Duisberg, 'Coloring-Matter from Tetraazo-Ditoly', *US Pat.*, 329633, 1885.
- 81 M. Hofmann, 'Diazo Dye', *US Pat.*, 480326, 1892.
- 82 W. Herzberg and O. Weber, 'Black Dye', *US Pat.*, 541750, 1895.
- 83 F. Bender, 'Blue Dye', *US Pat.*, 606436, 1898.
- 84 G. Koerner, 'Black Azo Dye', *US Pat.*, 638127, 1899.
- 85 See, for example: A. Israel and R. Kothe, 'Blue Trisazo Dye', *US Pat.*, 658897, 1900; M. Ulrich and K. Heidenreich, 'Orange Disazodye and Method of Making Same', *US Pat.*, 678323, 1901.
- 86 See, for example: B. Gaspar and P. Daniel, 'Acid Azo Dyes', *US Pat.*, 2612496, 1948.
- 87 G. Kränzlein, 'Acetals of Cyclic Ketones with Polyhydric Alcohols and Process of Preparing the Same', *US Pat.*, 1902866, 1933.
- 88 A. B. Steiner, 'Substituted Alkylene Glycol', *US Pat.*, 2463824, 1949.
- 89 A. M. Patterson and L. T. Capell, *The Ring Index*, American Chemical Society, Washington, 1940.
- 90 A. S. Travis, private communication, 2003. For information about ambiguities in German patents, see: K. Steen, *Isis*, 2001, **92**, 91–124.
- 91 W. H. Perkin, *J. Chem. Soc.*, 1915, **107**, 557–578.
- 92 International Union of Pure and Applied Chemistry (IUPAC), www.iupac.org, 2004.