Computational data analysis shows that key developments towards the periodic system occurred in the 1840s

Wilmer Leal[1,2], Eugenio J. Llanos[1,2,3], Andrés Bernal[2,4],
Peter F. Stadler[1,2,5,6,7,8], Jürgen Jost[2,8] & Guillermo Restrepo[2,5]

October 11, 2021

[1] Bioinformatics Group, Department of Computer Science, Universität Leipzig, Härtelstraße 16-18, 04107 Leipzig, Germany
[2] Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, 04103 Leipzig, Germany
[3] Corporación SCIO, Calle 57b 50-50 bloque d22 of. 412, 111321 Bogotá, Colombia
[4] Departamento de Ciencias Básicas, Universidad Jorge Tadeo Lozano, Carrera 4 # 22-61, 110311 Bogotá, Colombia
[5] Interdisciplinary Center for Bioinformatics, Universität Leipzig, Härtelstraße 16-18, 04107 Leipzig, Germany
[6] Institute for Theoretical Chemistry, University of Vienna, Währingerstraße 17, 1090 Vienna, Austria
[7] Facultad de Ciencias, Universidad Nacional de Colombia, Sede Bogotá, Colombia
[8] The Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, New Mexico 87501

Abstract

The periodic system arose from knowledge about substances, which constitute the chemical space. Despite the importance of this interplay, little is known about how the expanding space affected the system. Here we show, by analysing the space between 1800 and 1869, how the periodic system evolved until its formulation. We found that after an unstable period culminating around 1826, the system began to converge to a backbone structure, unveiled in the 1860s, which was clearly evident in the 1840s. Hence, contrary to the belief that the "ripe moment" to formulate the system was in the 1860s, it was in the 1840s. The evolution of the system is marked by the rise of organic chemistry in the first quarter of the nineteenth-century, which prompted the recognition of relationships among main group elements and obscured some of transition metals, which explains why the formulators of the periodic system struggled accommodating them. We also introduced an algorithm to adjust the chemical space according to different sets of atomic weights, which allowed for estimating the resulting periodic systems of chemists using one or the other nineteenth-century atomic weights. These weights produce orderings of the elements very similar to that of 1869, while providing different similarity relationships among the elements, therefore producing different periodic systems. By analysing these systems, from Dalton up to Mendeleev, we found that Gmelin's atomic weights of 1843 produce systems remarkably similar to that of 1869, a similarity that was reinforced by the atomic weights on the years to come.

Introduction

Meyer and Mendeleev's periodic systems culminated a series of attempts at building Systems of Chemical Elements (SCEs) [1], which resulted from ordering and classifying elements [2] through the knowledge of their compounds [3, 4]. Thus, by finding the smallest common combining weight of a large set of compounds containing a reference element, atomic weights were determined and used to order elements [5]. Likewise, compounds played a major role in assessing chemical resemblance among elements, which was mainly determined on the basis of similarities of empirical and molecular formulae [3, 4]. Thus, SCEs refer not just to chemical elements, but to substances in general.

Every discovered substance enlarges the set of known chemicals, which we call the *chemical space* [6]. Given the central role of this space for the formulation of the SCE, every discovery of new elements and compounds, may affect the SCE by introducing or perturbing similarities among chemical elements or by affecting the ordering of their atomic weights. Up to now there is no account of how the evolution of the chemical space affected the SCE. Nevertheless, historians have concluded that the ripe moment for formulating the system came in the 1860s [7, 1], thanks largely to the normalisation of molecular formulae through the standardised set of atomic weights resulting from the 1860 Karlsruhe conference.

Here we analyse how the expansion of the chemical space between 1800 and 1869 affected the SCE by using two approaches: presentist and retrodictive. The first one assesses the effect of the amount of chemical data according to current standards and solves the question of when the size and diversity of the chemical space

allowed the emergence of the SCE. The second approach focuses on the evolution of the atomic theory and solves the question on when the theory led to the emergence of the SCE. Taken together, these approaches help us to determine whether specific changes in the 1860s chemical space actually led to the SCE; or whether the patterns of the SCE were already present earlier in history, which leads to ponder whether the SCE could have been formulated earlier.
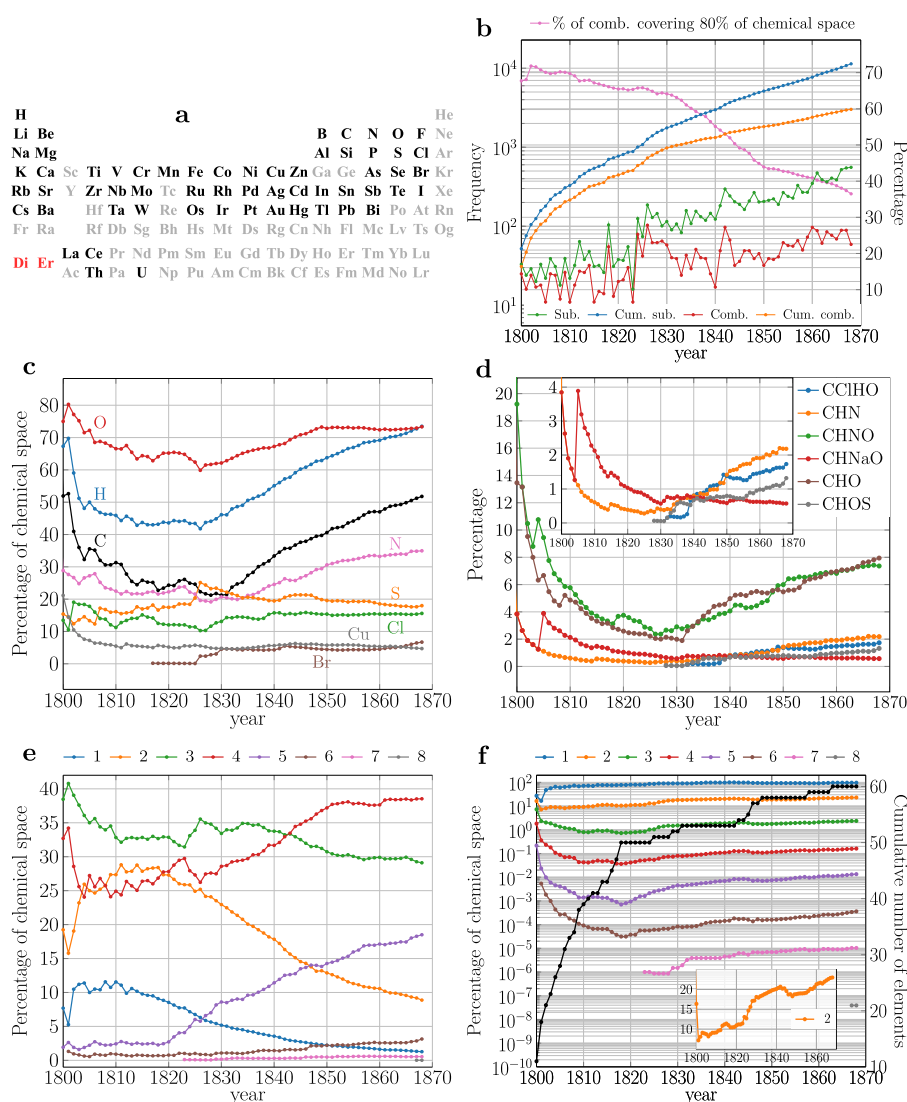

Figure 1: Chemical elements, growth and diversity of the chemical space up to 1869. a) Current system of chemical elements (SCE) depicting elements known by 1869 (black), undiscovered elements (grey) and mixtures that were thought to be elements (red) (Supporting Information). Elements in black were considered in this study. b) Absolute (left axis [l.a.]) and cumulative values (right axis [r.a.]) of new substances and combinations. c) Percentage of chemical space spanned by some elements. These percentages are non-additive because a single substance adds to each one of its elements, e.g. H_2O contributes to both H and O counts. d) Percentage of chemical space spanned by different combinations. e) Percentage of chemical space spanned by substances made of n elements. After 1811 the number of uncombined forms (unary substances) in which elements appeared exceeded the number of known elements as a consequence of the allotropic forms and polymorphs of elements. For instance, by 1868 sulfur had nine uncombined forms. f) Cumulative number of elements (r.a.) and percentage of theoretical combinations of different sizes actually observed (l.a.) (Materials and Methods)

Evolution of the chemical space (1800-1868)

Gmelin and Beilstein's Handbooks, initiated in the nineteenth-century, gather records of extractions, synthesis and properties of substances [6, 8]. Nowadays Reaxys©, a large electronic database of chemical information, which merges these two handbooks plus several other sources of chemical information, constitutes a suitable corpus for historical studies of chemistry [6, 8].

We collected records from Reaxys[1] (January 2017) from 1771 up to 1868, that is two months before the publication of the first Mendeleev' SCE [9]. These amounted to 11,356 substances involved in 21,521 single-step reactions (Materials and Methods), mainly reported in Gmelin's Handbook and gathered from leading

nineteenth-century journals [10]. These substances span a growing number of elements over time, from nine in 1800 up to 60 in 1868 (Figure 1a and Interactive Information). The most complete system was formulated by Mendeleev [11], including the 60 elements of Figure 1a plus Er, Yt and Di [12]. Nonetheless, we excluded these three elements because of their unreliable information by 1869: Yt (currently Y [13]) first reported reaction dates back to 1872. Er and Di were later found to be mixtures of other elements [14] (Supporting Information).

Chemists expanded the chemical space at an exponential rate, from a handful of new substances in 1800 up to the 11 thousand of 1868 [6] (Figure 1b). From each substance, we extracted its element combination, that is HOS for H_2SO_4. We used the number of combinations as a measure of diversity. Figure 1b shows that although combinations also grew exponentially [6], unlike substances their growth was reduced after 1830, indicating decreasing diversity. After this year, substance discovery focused on fewer combinations, so that by the end of the period 36% of the combinations covered 80% of the chemical space (Figure 1b). Figures 1c-d provide further details about the 1830 turning point: at the dawn of the century the chemical space was mainly populated by compounds of C, H, O and N; then, during the first quarter of the century, chemists found new combinations, which reduced the percentage of chemical space spanned by each combination (Figure 1d, Figure S1). This is a period where the number of new substances and of combinations grew hand in hand (Figure 1b). A minimum was reached for CHO and CHNO compounds around 1830. Afterwards, there was again a clear emphasis on CHO and CHNO (Figure 1d), which resulted in a less rapid production of new combinations (Figure 1b). More CHO and CHNO substances distributed over a slow growing number of combinations increased the space spanned by these combinations (Figures 1d and S2). This is clearly a consequence of the organic revolution [5, 15]: before 1830 most new combinations were metallic, while afterwards most were organic (Figure 1d, Figures S1 and S3, Table S1). The importance of organic chemistry after 1830 is observed in Figures S4 and S5, where substances containing typical organic chemistry molecular fragments skyrocketed, in contrast with those containing inorganic ones.

Another attribute of a combination is its size, that is the number of elements present in it. The theoretical number of combinations depends on the available elements. Thus, by 1800, with 11 elements, there were 2,036 possible combinations (Materials and Methods), which grew up to 1.15×10^{18} by 1868 with 60 elements. We found that despite the growth of new combinations (Figure 1b), chemists reported compounds of no more than eight elements (Figure 1e). During the first quarter of the nineteenth-century the chemical space was mainly populated by compounds of size 2-3, presumably due to the prevalence of dualism in chemical theories [15]. Afterwards there was a surge in the number of larger combinations involving 4-5 elements, mostly organic.

By analysing how close were chemists to actually realising the theoretical combinations of different sizes, we found that during the first years of the nineteenth-century, where a rapid discovery of elements took place (Figure 1f), the amount of theoretical combinations skyrocketed causing a rapid drop of the proportion of realised combinations. Once the discovery of new elements slowed around 1820, more combinations were actually observed increasing the proportion of theoretical combinations realised (Figure 1f). In the mid 1840s came another batch of new elements, reducing again the proportion of realised combinations, which coincides with a strong drop in the number of new substances, from 300 in 1842 to 163 in 1846 (Figure 1b). After a decade, chemists were again discovering more combinations and increasing this proportion. As expected, given its relatively small number, binary combinations were always closer to their theoretical possibilities than combinations of more elements. By 1825, after the stabilisation of the number of elements, about 13% of the theoretical number of binary compounds was reported (Figure 1f, inset), a growing percentage not even affected by the emphasis on compounds of three and four elements (Figure 1e). In fact, by 1868 about 23% of the possible binary compounds (made with combinations of 60 elements) were already known (Figure 1f, inset).

Evolution of the system of chemical elements (1800-1868)

We analysed the interplay between the chemical space and the SCE from two perspectives, one contemporary or presentist and another historical or retrodictive. The *presentist approach* "sees" the chemical space of the nineteenth-century through the eyes of twenty first-century chemistry. Here nineteenth-century formulae, for example Dalton's OH for water, are replaced by their contemporary versions. Reaxys data suit this approach. The *retrodictive approach* considers the evolution of the chemical space as historically witnessed. It acknowledges the historical construction of consent on atomic weights and its associated formulae. Therefore, when analysing the chemical space, for instance of 1810, it attempts to use the formulae by the leading chemists of that time, for example Dalton. This approach allows for studying possible SCEs according to several nineteenth-century chemists.

Presentist approach to the evolution of the system of chemical elements

Figure 2 explains our approach to quantifying similarity among chemical elements (Supporting Information). We associate similarity with the possibility of substituting one element by another in an empirical formula. This follows from Mendeleev's idea that "the elements, which are most chemically analogous, are characterised by the fact of their giving compounds of similar form RX_n" [16].

As SCEs intend to display only the most remarkable similarities among elements, we display only maximum similarities for each element. This choice is justified because SCEs are customarily presented as tables in which similarities between neighbouring elements are the largest. If this is the case, non-maximal but important similarities can be recovered from sequences of maximum similarity relationships, for instance Li being most
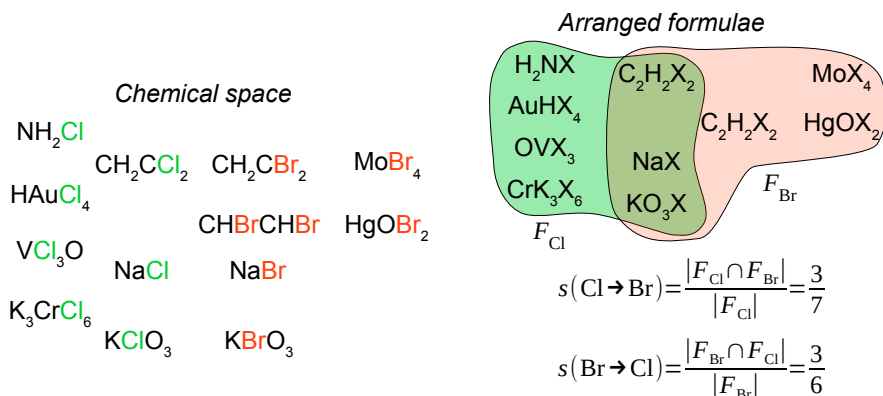
Figure 2: Similarity among chemical elements. Toy chemical space of 13 substances. Each compound provides an arranged formula for an element in the given formula when Cl or Br is replaced by X and the elements are lexicographically ordered. Arranged formulae of element X are gathered in F_X, which is a multiset as elements may appear more than once, e.g. $C_2H_2X_2$ appears twice in F_{Br} (Supporting Information). The similarity of element x to element y is given by $s(x \to y)$, which is the probability of x having a common arranged formula with y. In chemical terms, it is a measure of substitutability. This similarity is an asymmetric relation [17] e.g. $s(Br \to Cl) > s(Cl \to Br)$. For instance, by 1869 we have $s(Br \to Cl)=344/659=0.52$, while $s(Cl \to Br)=349/1556=0.22$ (Interactive Information). This means that Br could be substituted by Cl to obtain a known compound in roughly half of Br combinations, whereas Cl could be substituted by Br in about one fourth of those of Cl. This similarity measure generalises that presented in [18]

similar to Na and Na to K means that likely Li is quite similar to K as well (Figure 3). Therefore, elements related by maximum similarities correspond to the notion of families (groups) of elements on periodic tables.

Having determined the key similarities among chemical elements, all that remains is to arrange them according to their atomic weights to retrieve the SCE of each year between 1800 and 1868. We depict these systems as similarity networks. Figure 3a-c present three of them. All 69 networks can be found in the Interactive Information.

Despite the increase in the number of elements (Figure 1f), the amount of "most similar" relationships decreased over time, dropping from a maximum of 166 in 1818 down to 69 in 1862 (Figure S6). This indicates that as chemists expanded the chemical space, they were converging towards a core set of similarity relationships. To better assess this convergence, we calculated the similarity between SCEs of different years (Figure 4a, Materials and Methods).

The reddish region around the diagonal in Figure 4a indicates continuity in the evolution of the SCE, as the most similar periodic system of any year is always one of an adjacent year. Nevertheless, it also shows qualitative shifts, the most visible of which appears in 1826, that suggest convergence to a stable SCE (Interactive Information). The dark blue regions around the early years indicate that the SCEs of those years did not stand the test of the time: they transformed completely within the span of few years and are essentially different from the SCEs at the end of the period (Interactive Information). Similarities in this early quarter of the century were mainly related to substitutions in chlorides, oxides, hydroxides, sulfates and other typical inorganic compounds (Interactive Information). But then, in 1826, there was a sharp stabilisation of the SCE, as revealed by the light blue-yellow square in Figure 4a, which indicates that more than 40% of the similarities found by 1826 remained in the SCE all the way to the end of the period. Some of these early known similarities were Ag $\to$ K and Pt $\to$ Pd, caused mainly by their inorganic compounds (Interactive Information). Other similarities showing up a decade later were K $\to$ Na, Hg $\to$ Cu, Si $\to$ Ti, Fe $\to$ Co and Ni $\to$ Co, which were also mainly caused by inorganic compounds (Interactive Information). Interestingly, famous nineteenth-century SCEs as those published by Meyer and Mendeleev in the 1860s, depict Cu and Ag as similar elements (Figure S7), mainly due to the similar low oxygen content of their oxides [19]. However, our results show that, ever since 40 years earlier, Cu had been more related to Zn group, mainly for its dominating +2 valence, and Ag to K, for its dominating +1 valence (Figure 3, Interactive Information). Also, by 1826, halogens and Fe, Co, Ni became SCE families (Figure 3a-c).

The reddish region between columns 1826-1860 and rows 1835-1845 (Figure 4a) shows that about 80% of the similarities of the SCE observed between 1835 and 1845 were present since 1826 and lasted until 1860. During the 1860s, this resemblance dropped down to about 60%. The period after 1845 shows that the similarities observed after this year lasted, but that there were also some transient similarities, for instance those of Nb, Ta, Rb and Cs (Interactive Information).

The mechanisms behind this convergence were of substance discovery and tie-breaking. In the early years of the century not enough compounds were known to determine the similarities that unveil the patterns of the SCE. For example, up to 1825 no combinations of Ce were known, which caused this element to be similar to almost all the rest (Figure 3a). In 1826 Ce_2S_3, the first Ce compound, was reported, which caused its similarity to collapse on elements that presented the same kind of sulfides. The discrete shift on 1826 is due largely to this particularity of Ce. The same mechanism operated all across the expanding chemical space, eventually leading

Figure 3: Evolution of the system of chemical elements (SCE). a-c) SCEs of three different years. Arrows $x \to y$ indicate that x is most similar to y. Node (element) size is proportional to the number of substances composed by the element. Similarities of Ce are coloured in light grey for the sake of readability. In Figure a, all Ce similarities are collapsed for the sake of simplicity. Some of the formulae shared by elements with the same colour (X) are shown in c. Readers can also select in the Interactive Information any set of elements to retrieve the formulae making similar the elements in any particular year. d) Backbone of the SCE depicting the pairs of most similar chemical elements appearing in more than 60% of the SCEs between 1800-1869 (Materials and Methods)

to the pattern observed in the period 1837-1845 that already reproduced 80% of the similarities observed up to the 1860s. This suggests that a fairly accurate SCE could have been proposed as early as the 1840s. However, the problem of uncertainty on atomic weights still needs to be addressed, and we shall do it in the next section.

A system of elements containing the most frequent pairs of most similar elements is depicted in Figure 3d, which we regard as the backbone of the periodic system from 1800 up to 1868. It shows salient families of elements including alkali metals, halogens, chalcogens, pnictogens (without N) and {Fe, Co, Ni}, plus well-known families of transition metals such as {Pd, Pt, Ir} and {Mo, W, Ta}.

Regarding the magnitude of similarities among chemical elements, that is their actual similarity values, Figure S8 shows that they were very weak. In fact, all over the history here analysed more than 80% of the similarities had values lower than 0.1. That is, over the history less than 10% of the formulae of any element have been shared with its most similar element(s). The lowest similarity values ever recorded corresponded to those of organogenic elements (Interactive Information). This fits Mendeleev's concept of "typical elements" [12, 20], today called singularity principle or uniqueness of second period elements [21], which indicates that these elements possess weak similarities with elements of their families [22].

If the similarities were so small, how could they become so noticeable to chemists? This is even more surprising if we consider that each nineteenth-century chemist may not have had the complete knowledge provided to us by the database. We believe it has to do with ubiquity: these similarities extend over the whole spread of the chemical space, so that they are equally visible in any reasonably-sized portion of the chemical space. To test this hypothesis, we took random samples of different sizes of the space, for every year, and analysed how often the "most similar" relationships among elements were present in the samples (Materials and Methods). We found that most of the similarities observed in the first quarter of the nineteenth-
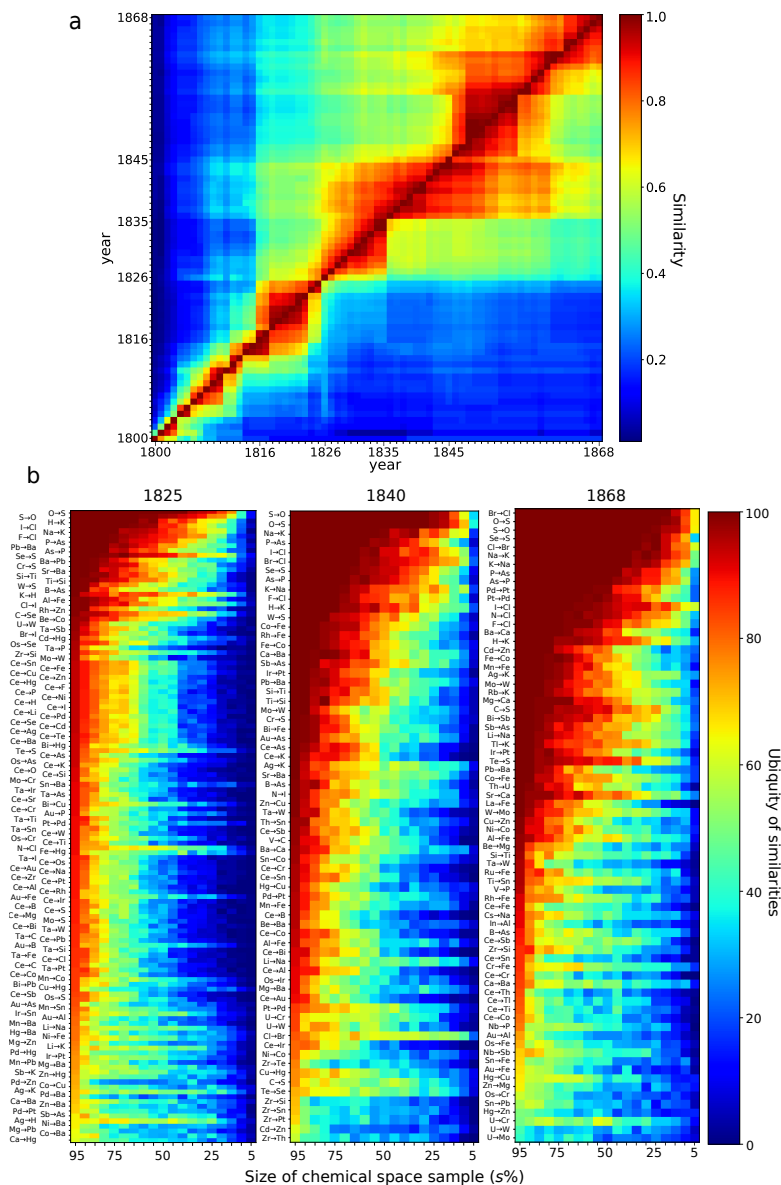
Figure 4: Similarity among systems of chemical elements (SCEs) and ubiquity of element resemblances. a) Resemblance between SCEs. The heatmap depicts similarity from the SCE of the column to the SCE of the row (Materials and Methods). Any row y indicates how similar the SCEs are, year after year, to the SCE of year y. Any column x shows which fraction of the SCE, year after year, is similar to the SCE of year x. b) Ubiquity of the similarities of the SCEs of 1825, 1840 and 1868. The ubiquity of each similarity $x \rightarrow y$ corresponds to the percentage of appearance of such similarity in the sampled space of size $s\%$ (Materials and Methods). Plots for all years 1800-1868 are found in Figure S6

century required more than 50% of the chemical space to be detected, indicating that in this period similarities of different elements were focused on different regions of the chemical space, hindering the discovery of the patterns of the SCE (Figures 4b and S9). As time went by, especially after 1830, similarities became more ubiquitous and easier to detect. This effect is particularly intense for similarities among organogenic elements, as they were spread among the increasingly large number of organic compounds, which promptly became the majority of the chemical space. For instance, the similarity $S \rightarrow O$, detected as early as 1800, required at least 65% of the 1800 space to be observed, while by 1840 this fraction plummeted to 10% and dropped to 5% by 1868 (Figures 4b and S9). In contrast, $Ba \rightarrow Ca$ required in 1800 three quarters of the space, by 1840 60% and by 1868 a quarter, that is five times more chemical space than $S \rightarrow O$ to be detected (Figures 4b and S9). Therefore, the redundancy of the space on organogenic compounds facilitated the detection of similarities among organogenic elements, which contrasts with those of the transition metals, whose detection demanded examination of a larger fraction of the chemical space.

This explains why nineteenth-century chemists, such as Meyer and Mendeleev, struggled with similarities among transition metals [23, 24, 25, 26, 27, 28, 29, 12, 20] (Figure S7, Table S2). Mendeleev also faced problems with the similarities of In and the rare earths he included in his system (Table S2) [30]. Remarkably, detecting $In \rightarrow Al$ by 1869, as Meyer did, required more than 75% of the chemical space (Figure 4d). Examples of other similarities requiring large amounts of chemical space to be detected were $Zn \rightarrow Mg$, $Nb \rightarrow P$ and $Nb \rightarrow Sb$. The first of these is explicit in Mendeleev and Meyer's 1869/70 systems (Figure S7), and the other two are explicit in Meyer's system and discussed as similarities by Mendeleev [20] (Figure S7, Table S2). Overall, we found

that about 53% of the similarities among chemical elements arising from the chemical space were recovered by Meyer in his 1864 and 1868 systems (true positives, Supporting Information, Table S3). Almost a quarter of non similarities of the 1864 chemical space were observed as similarities by Meyer (false positives, Table S3). This fraction plummeted in 1868 to about 7%. At any rate, the best agreement between Meyer's systems and the system allowed by the chemical space was achieved in 1869/70, when 62% of the similarities of the space were gauged by his system, while there were only 6% non similarities observed as similarities. Mendeleev, in turn, attained 58% and 10% of true and false positives, respectively (Table S3). Note that the (dis)agreements here discussed are based on the similarities reported by the two chemists in their systems, which were abundant and detailed in Mendeleev's case and very seldom discussed by Meyer, in which case similarities needed to be interpreted from his periodic tables. Also, the greater detail of Mendeleev's discussions on similarity is expected to yield a higher rate of false positives, due to our methodology being based on maximum similarities.

Retrodictive approach to the evolution of the system of chemical elements

The presentist approach to the evolution of the SCE takes for granted a stable set of atomic weights and of empirical and molecular formulae corresponding to current standards. Nevertheless, analysing the historical evolution of the chemical space and its influence upon the SCE requires considering the history of the atomic theory. That is, it requires considering the various nineteenth-century competing sets of atomic weights associated to different theoretical and experimental settings [31, 5], which led to a chaos of formulae [32]. Hence, different atomic weights produce different orderings of the elements and different formulae, so that different chemists working with different sets of atomic weights could find widely different SCEs, even if they worked with the same experimental data. Here we analyse the possible SCEs resulting from different perspectives of the nineteenth-century chemical space spawned by several distinct sets of atomic weights proposed over the period.

In the nineteenth-century, empirical data on composition came in the form of mass percentages for each element. For instance, Dalton knew that water was made of 88% and 12% by weight of oxygen and hydrogen, respectively. From Dalton on, chemists assumed formulae for key compounds such as water, ammonia and oxides. Thus, chemists selected an element and assigned a reference atomic weight to it, and recorded atomic weights of other elements relative to that one. The initial assumptions thus propagate through all the calculations, therefore creating a different chemical space for each chemist (Figure 5a). For example, Dalton's reference was an atomic weight of 1 for hydrogen. He assumed HO as the formula of water, therefore yielding an atomic weight of 7 for oxygen. This leads to molecular formulae of oxides whose coefficients are around half of those we know today. The determinations were made even more difficult by the varying quality of the experimental data [31, 5].

We gathered 13 sets of atomic weights (Table S4), corresponding to data published by Dalton (1810) [33], Thomson (1813) [34], Berzelius (1819 [35] and 1826 [36, 37, 38]), Gmelin (1843) [23], Lenßen (1857) [24], Meyer (1864 [25], 1868 [39] and 1869/70 [27]), Odling (1864) [28], Hinrichs (1867) [29] and Mendeleev (1869) [12], plus the currently accepted atomic weights. Starting with Gmelin, these sets of atomic weights were proposed by authors who actually devised SCEs [1]. Although neither Dalton, nor Thomson nor Berzelius aimed at devising SCEs, they were some of the key figures in the development of the atomic theory [31, 5], which is why we also explored the effects of their atomic weights upon the SCEs that could have been obtained from their respective chemical spaces. Figure S10 shows the elements comprised by each system of atomic weights, which range from 30 for Dalton to 60 for Mendeleev. Information on the selection of these elements is found in Table S5.

As any SCE is based on ordering and similarity of its chemical elements [2], we analysed the different orderings of elements associated to each set of weights. In all cases they agreed in more than 80%, even with the current atomic weights (Table S6). This indicates that the ordering relationships among elements were rather stable since the beginning of the nineteenth-century. To determine element similarities it is necessary to reconstruct the formulae spanned by each system of atomic weights (Figure 5a). As there is no systematic record of the chemical formulae corresponding to the assumptions of each chemist, we devised an algorithm to obtain approximate formulae meeting the assumptions of the chemists here analysed (Materials and Methods). This entails, for instance, approximating the current Fe_2O_3 to FeO_3 according to Dalton (Figure 5a). Our procedure takes all Reaxys formulae known by the time of publication of each chemist's atomic weights and rescales the modern formulae to fit chemist's atomic weights within 20 different levels of *tolerance* (Materials and Methods, Supporting Information). Often, the higher the tolerance, the lower the perturbation of Reaxys formulae.

For each chemist' set of atomic weights and level of tolerance we obtained a corresponding chemical space, which led to an associated SCE holding a set of similarities among chemical elements. In order to quantify how close was a chemist' set of atomic weights of gauging the similarities allowed by the actual chemical space of chemist's time (calculated with our contemporary atomic weights), we computed the fraction depicted in the red plot of Figure 5b (see also Figure S11). This corresponds to the true positives rate, indicating to which extent the old atomic weights sharpened our ancestors' capabilities of discovering the SCE of 1868. As a chemist' space could lead to several transient similarities not remaining until 1868, we also quantified chemist's fraction of similarities of this sort (Figures 5b, blue and S12). They correspond to the false positives rate (Figure 5b, blue).

By inspecting Figure 5b, we observe how, as the century progressed, SCEs resulting from fitting chemists' sets of atomic weights contain more and more 1868 similarities and how transient similarities were further reduced. There is a remarkable leap with Gmelin, who becomes a turning point in the trends, separating SCEs with
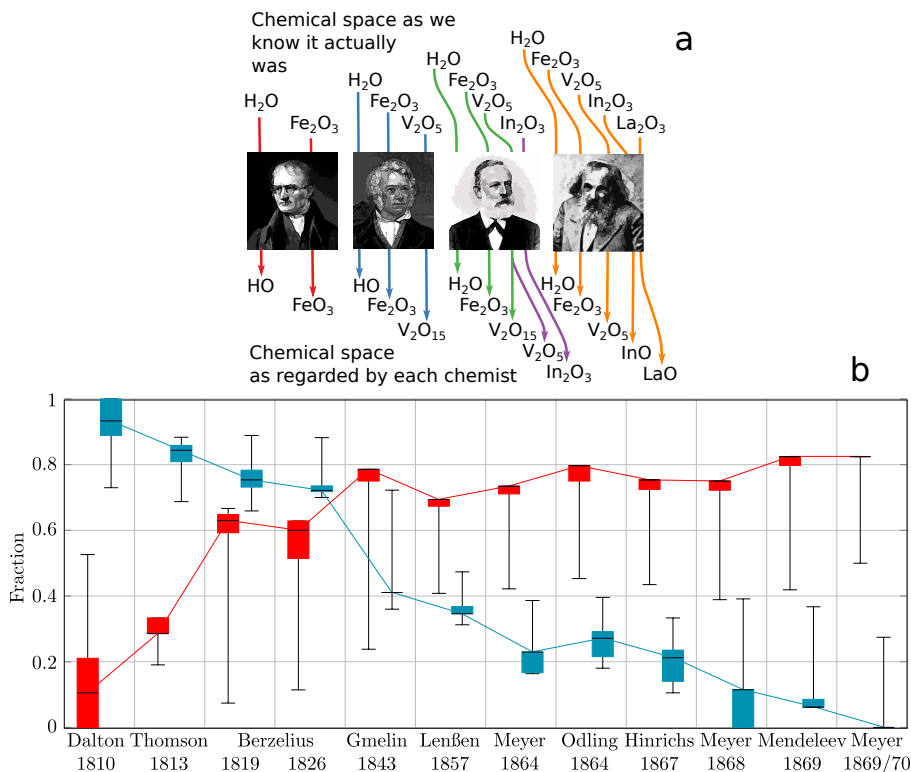
Figure 5: Contrast between systems of chemical elements (SCEs) calculated with nineteenth-century and with modern atomic weights. a) Examples of modified formulae according to the atomic weights of (left to right) Dalton (1810), Gmelin (1843), Meyer (1864, green; 1869/70, purple) and Mendeleev (1869). For every chemist publishing a set of atomic weights in year y, known Reaxys substances (S_{y-1}) up to year $y - 1$ (inclusive) were retrieved and the corresponding SCE P_{y-1} was obtained. Afterwards, formulae of substances S_{y-1} were transformed to fit chemist's atomic weights within 20 different tolerance values (τ), each τ yielding a SCE with similarities gathered in P_{y-1}^{τ} (Materials and Methods, Supporting Information). b) Red: efficacy of chemist's atomic weights in approaching P_{1868}, measured as $|P_{y-1}^{\tau} \cap P_{1868}|/|P_{y-1} \cap P_{1868}|$; a value of 1 means that they are just as effective as our modern atomic weights. Blue: fraction of transient similarities in P_{y-1}^{τ}, that is similarities not observed by 1868. These were calculated as $|P_y^{\tau} \setminus P_{1868}|/|P_y^{\tau}|$; a small value means that most of chemist's SCEs obtained from his atomic weights were observed in 1868. Boxplots depict median (black horizontal line) and min/max values as whiskers

many transient similarities and few standing the test of the time (on Gmelin's left in Figure 5b) from SCEs rich in 1868 similarities and with very few transient similarities (on Gmelin's right in Figure 5b). Gmelin's atomic weights lead to SCEs containing about 78% of the 1868 similarities and about 40% of the similarities of those SCEs are transient. This is actually an improvement, when contrasted with the SCEs obtained with the atomic weights of Gemlin's predecessors. For instance, the SCEs obtained with Dalton's weights have about 10% of 1868 similarities and 93% of transient ones, and those of Berzelius (1826) 60% and 73%, respectively. The lack of accuracy of pre-Gmelin SCEs is caused by the many changes the chemical space underwent. Nevertheless, in those times the SCEs obtained using Berzelius' 1819 weights stand out. In spite of their 75% of transient similarities, Berzelius' atomic weights lead to SCEs with 63% of 1868 similarities.

The remarkable separation of the two plots after Gmelin (Figure 5b) shows the strong relationship between the theoretical and experimental advances the atomic theory brought about and the raise of the backbone of the periodic system. Interestingly, this is particularly evident in the 1840s and not in the 1860s as traditionally accepted, which agrees with the results of our presentist approach.

By analysing the SCEs obtained from Meyer and Mendeleev's atomic weights, we found that each new version of Meyer's weights achieves more 1868 similarities and reduces the amount of transient similarities. His last set of atomic weights leads to SCEs with no transient similarities matching 82% of the 1868 similarities. In turn, Mendeleev's atomic weights produce SCEs with 83% of 1868 similarities and 6% of transient similarities. These results coincide with the different stances the two chemists had regarding the SCE. Meyer favoured accurate atomic weights and experimental information and Mendeleev completeness [11, 40], as noted in the several elements left aside by Meyer that were included by Mendeleev.

Conclusion

Contrary to common opinion, we found that different systems of atomic weights in vogue in the nineteenth-century did not affect to a large extent the ordering of elements in a system of chemical elements (SCE). In contrast, similarity played a major role in shaping the SCE.

By analysing the size and diversity of the chemical space between 1800 and 1869 we found that before

1830 the expansion of the space involved discovering new instances of a wide variety of combinations, which allowed observing, as early as 1826, several similarities standing the test of the time. The rise of organic chemistry after 1830 caused a strong emphasis on the expansion of the chemical space towards compounds of organogenic elements, mainly O, H, C, N and S. This facilitated the detection of similarities among organogenic elements even with low fractions of the chemical space, but hindered the recognition of some similarities among metals, which required large fractions of the chemical space to be observed. This explains why chemists in the nineteenth-century struggled to detect similarities among metals but readily observed those among organogenic elements.

All sets of atomic weights proposed across the nineteenth-century ultimately converged towards the same similarity patterns among the elements, much faster than previously thought. Disagreements regarding atomic weights turn out to be less critical than it seems: it is consistency within the system that matters, rather than its absolute accuracy.

By 1837-1845 a large core of stable similarities arose, which was to remain until Meyer and Mendeleev's formulation of their SCEs. This "golden period" turns out to be a ripe moment to devise SCEs similar to theirs, which challenges the traditional account that the 1860s were the proper time for the periodic system to arise [7, 1]. Here we have shown that, in terms of the material system of chemistry to which the chemical space belongs [8], the ripe moment was in the 1840s. In fact, this potential materialised to a significant extent in Gmelin's V-shaped SCE [23]. As the SCE had to wait about three further decades to be formulated, the question arises of whether the social and semiotic conditions for its formulation were still not ripe.

This study shows the interplay between the chemical space and the SCE and highlights the difficulties of formulating it before the 1840s as a consequence of the many changes of the chemical space of those times. Likewise, it justifies post-1840s SCEs supported by a stable chemical space and by sharper and accurate atomic weights.

Regarding Meyer and Mendeleev, both chemists were active in times of a mature chemical space and of rather stable atomic weights, which led to their SCEs. Although their systems coincide to a large extent with the possible SCEs of their times, the underlying reasons they had to finally arrange elements as they did cannot be reduced solely to the chemical space. For instance, both chemists regarded Cu and Ag as very similar (Figure S7), which does not agree with the chemical space of their time. Mendeleev believed in Tl and Cs similarity, a thought Meyer shared until 1869/70, when he arranged a new family containing Tl (current group 3). Yet, the SCEs resulting from Meyer and Mendeleev's weights show that Tl is most similar to K. Likewise, by 1869 the similarity between Pd and Pt was detectable with only 5% of the chemical space, but Mendeleev did not include it in his system, nor commented upon it, while Meyer' system included it. This indicates that both chemists were not guiding their inquiries exclusively by the chemical space, as it is actually observed in the physical properties of substances they made reference to [20, 27]. How was the interplay between chemical and physical evidence leading them to the details of their systems?

We believe the data driven approach presented here to study the evolution of the SCE opens the door to further studies on the current status of the periodic system. As we have shown elsewhere [6], the chemical space of the twentieth- and early twenty first-centuries kept the emphasis on organic substances. Does it mean that the current SCE is very similar to that of 1869? Or do we rather have a very different system?

From a more general perspective, this study contributes to the ongoing computational approaches to the history of science and the evolution of knowledge [41, 8]. We suggest the methods presented here can complement and be complemented by more conventional approaches to the history of science.

Data

We retrieved 21,521 single-step reactions with publication year before 1869 from Reaxys. These reactions had 11,451 associated substances with their respective formulae. Some of them were curated and others discarded (Supporting Information), leading to 11,356 substances. We associated each of these substances with its earliest publication year (in a chemical reaction) and with its molecular formula.

Theoretical combinations

For n known elements in a given year, its theoretical number of combinations of size s is $\texttt{theo}(n, s) = \binom{n}{s}$. Hence, the theoretical number of combinations of n elements is $\texttt{theo}(n) = \sum_{s=2}^{n} \binom{n}{s}$. This is a rough upper bound disregarding valency and compound stability. The percentage of theoretical combinations actually observed (Figure 1f) corresponds to $\texttt{exp}(n, s)/\texttt{theo}(n, s)$, where $\texttt{exp}(n, s)$ is the number of reported substances with n elements whose combinations size is s.

Backbone of the system of chemical elements

Figure 3d depicts similarities $i \rightarrow j$ appearing in more than 60% of the SCEs containing i and j. This percentage is computed as $(f(i \rightarrow j)/1868 - y) \times 100$, where $f(i \rightarrow j)$ is the number of SCEs containing $i \rightarrow j$ and y is the first year in which i and j appear in a SCE. The normalisation factor $1868 - y$ represents the time window where the similarity $i \rightarrow j$ could have been observed.

Similarity between systems of chemical elements

A system of chemical elements is devised as described in Figure 2 and stored as a collection N of pairs of elements (e_i, e_j), indicating the similarity of element e_i with respect to element e_j ($e_i \rightarrow e_j$ in Figure 3). Each year x has an associated network N_x. We quantify the relative fraction of similarities of N_x observed in another network N_y as $s(x, y) = |N_x \cap N_y|/|N_x|$, where $|N_x|$ indicates the number of pairs (e_i, e_j) in N_x. Whenever a network N_y is calculated from a chemical space approximated with a tolerance τ (see retrodictive approach), the similarity of such a network regarding the corresponding network to 1868 is given by $s_\tau(y, 1868) = |N_y^\tau \cap N_{1868}|/|N_{1868}|$.

Sampling the chemical space

For each year we randomly took $s\%$ of the space and determined the most similar element(s) for each element. This experiment was carried out 100 times. For each similarity $x \rightarrow y$ resulting for the whole space of that year, we counted in how many of the 100 experiments $x \rightarrow y$ appeared. The higher this number, the more stable the similarity is. We carried out this analysis for 19 sample sizes (95%, 90%, 85%, ..., 5%). The higher these numbers for different values of $s\%$, the higher the ubiquity of $x \rightarrow y$ in the chemical space (Figure 4b).

Chemical spaces from atomic weights

As contemporary atomic weights are related by simple fractions with atomic weights of different chemists (Table S4), we adjusted a contemporary chemical formula $F = X_x Y_x \ldots Z_z$ to $F_A = X_{x f_{A(X)}} Y_{y f_{A(Y)}} \ldots Z_{z f_{A(Z)}}$ [eq. 1]. Here, X, Y,..., Z are chemical elements and $x, y, \ldots, z$ their stoichiometric coefficients in F; $f_{A(X)}, f_{A(Y)}, \ldots, f_{A(Z)}$ are the respective coefficients modifying $x, y, \ldots, z$ to yield the formula F_A, as an approximation to that regarded by chemist A. Coefficients f_A are calculated as follows: knowing the current ($W(e)$) and chemist's ($A(e)$) atomic weights of element e (Table S4), as well as the respective values for hydrogen ($W(H)$ and $A(H)$), we calculate the ratios $(W/A)(e) = (W(e)/W(H))/(A(e)/A(H))$ and $(A/W)(e) = (A(e)/A(H))/(W(e)/W(H))$. Our aim is determining the simplest fraction f approximating either $(W/A)(e)$ or $(A/W)(e)$. As these ratios either fall in the real interval $(0, 1]$ or correspond to figures of the form $\alpha + \beta$, where α is an integer and β is a real number in the interval $(0, 1]$, we need to find a $0 < f \leq 1$ that best approximates either β or the ratio falling in the interval $(0, 1]$. The best f corresponds to a fraction of a Farey sequence [42] (Supporting Information) minimising the relative error of the approximation ($error(r, f) = |r - f|/r$, with r either $(W/A)(e)$ or $(A/W)(e)$). We allowed 20 different error tolerances τ for the approximation, from 1 to 20% of relative error, in such a manner that for each τ, the selected fraction f always approximates r with an $error \leq \tau$. Hence, for a given τ a fraction f is found, which corresponds to the coefficient $f_{A(e)}$ in eq. 1. By applying this algorithm to each element of the contemporary formula F, the respective fractions are found and the adjusted formula F_A of chemist A is found (Further details in the Supporting Information). By applying this method it is found, for instance, that contemporary Fe_2O_3 corresponds to FeO_3 according to Berzelius' table of atomic weights of 1819 (Table S4).

Author information

Contributions

G.R. conceived the idea; W.L. and G.R. designed the research; W.L. dumped and analysed data; W.L., E.J.L., P.F.S. and G.R. devised similarity measure; W.L. and E.J.L. computed and analysed similarities; W.L. and G.R. designed method to adjust chemical space to given atomic weights. A.B. and E.J.L. implemented data visualisations in the Interactive Information; W.L., E.J.L., A.B., P.F.S., J.J. and G.R. discussed the results; G.R. wrote the original draft; W.L., E.J.L., A.B., P.F.S., J.J. and G.R. reviewed and edited it.

Corresponding author

Correspondence to Guillermo Restrepo (email: restrepo@mis.mpg.de).

Ethics declarations

Competing interests

The authors declare no competing interests.

References

[1] Scerri, R. E. *The periodic table: Its story and its significance* (New York: Oxford University Press, 2019).

[2] Leal, W. & Restrepo, G. Formal structure of periodic system of elements. *Proceedings of the Royal Society A* **475** (2019).

[3] Restrepo, G. Challenges for the periodic systems of elements: Chemical, historical and mathematical perspectives. *Chemistry - A European Journal* **25**, 15430–15440 (2019). URL `https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/chem.201902802`. `https://chemistry-europe.onlinelibrary.wiley.com/doi/pdf/10.1002/chem.201902802`.

[4] Restrepo, G. Compounds bring back chemistry to the system of chemical elements. *Substantia* **3**, 115 – 124 (2019). URL `https://riviste.fupress.net/index.php/subs/article/view/739`.

[5] Rocke, A. J. *Chemical atomism in the nineteenth century* (Columbus: Ohio University Press, 1984).

[6] Llanos, E. J. *et al.* Exploration of the chemical space and its three historical regimes. *Proceedings of the National Academy of Sciences* **116**, 12660–12665 (2019). URL `https://doi.org/10.1073/pnas.1816039116`.

[7] Bensaude-Vincent, B. Mendeleev's periodic system of chemical elements. *The British Journal for the History of Science* **19**, 317 (1986).

[8] Restrepo, G. & Jost, J. A formal setting for the evolution of chemical knowledge. Preprint (2020). URL `https://www.mis.mpg.de/publications/preprints/2020/prepr2020-77.html`.

[9] Gordin, M. D. *A well-ordered thing: Dmitrii Mendeleev and the shadow of the periodic table* (New York: Basic Books, 2004).

[10] Leal, W., Llanos, E. J., Stadler, P. F., Jost, J. & Restrepo, G. The chemical space from which the periodic system arose (2019). URL `https://chemrxiv.org/articles/preprint/The_Chemical_Space_from_Which_the_Periodic_System_Arose/9698888`.

[11] Gordin, M. D. *The textbook case of a priority dispute: D. I. Mendeleev, Lothar Meyer, and the periodic system*, 59–82 (Palgrave Macmillan US, New York, 2012). URL `https://doi.org/10.1057/9780230338029_4`.

[12] Mendeleev, D. On the relation of the properties to the atomic weights of the elements. In Jensen, W. B. (ed.) *Mendeleev on the periodic law: Selected writings, 1869-1905*, chap. 1, 16–17 (Dover, New York, 2002).

[13] Coplen, T. B. & Peiser, H. S. History of the recommended atomic-weight values from 1882 to 1997: A comparison of differences from current values to the estimated uncertainties of earlier values (technical report). *Pure and Applied Chemistry* **70**, 237–257 (1998).

[14] Evans, C. *Episodes from the history of the rare earth elements* (Boston : Kluwer Academic Publishers, 1996).

[15] Klein, U. *Experiments, Models, Paper Tools: Cultures of Organic Chemistry in the Nineteenth Century* (Stanford University Press, Stanford, 2003). URL `http://www.sup.org/books/title/?id=1917`.

[16] Mendeleev, D. The grouping of the elements and the periodic law. In Jensen, W. B. (ed.) *Mendeleev on the periodic law: Selected writings, 1869-1905*, chap. 13, 253–314 (Dover, New York, 2002).

[17] Mestres, J. & Maggiora, G. M. Putting molecular similarity into context: Asymmetric indices for field-based similarity measures. *Journal of Mathematical Chemistry* **39**, 107–118 (2006). URL `https://doi.org/10.1007/s10910-005-9007-3`.

[18] Leal, W., Restrepo, G. & Bernal, A. A network study of chemical elements: from binary compounds to chemical trends. *MATCH communications in mathematical and in computer chemistry* **68**, 417–442 (2012).

[19] Mendeleev, D. On the periodic regularity of the chemical elements. In Jensen, W. B. (ed.) *Mendeleev on the periodic law: Selected writings, 1869-1905*, chap. 3, 38–109 (Dover, New York, 2002).

[20] Mendeleev, D. On the correlation between the properties of the elements and their atomic weights. In Jensen, W. B. (ed.) *Mendeleev on the periodic law: Selected writings, 1869-1905*, chap. 2, 18–37 (Dover, New York, 2002).

[21] Rayner-Canham, G. *The Periodic Table* (World Scientific, 2020). URL https://worldscientific.com/doi/abs/10.1142/11775. https://worldscientific.com/doi/pdf/10.1142/11775.

[22] Gordin, M. D. The organic roots of Mendeleev's periodic law. *Historical Studies in the Physical and Biological Sciences* **32**, 263–290 (2002).

[23] Gmelin, L. *Handbuch der Chemie* (Karl Winter, Heidelberg, 1843). URL //catalog.hathitrust.org/Record/010055907. 8 v. in 9.

[24] Lenßen, E. Ueber die Gruppirung der Elemente nach ihrem chemisch - physikalischen Character. *Justus Liebigs Annalen der Chemie* **103**, 121–131 (1857). URL https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/jlac.18571030202. https://chemistry-europe.onlinelibrary.wiley.com/doi/pdf/10.1002/jlac.18571030202.

[25] Meyer, L. *Die modernen Theorien der Chemie und ihre Bedeutung für die chemische Statik* (Breslau: Verlag von Maruschke & Berendt, 1864).

[26] Meyer, L. Kritik einer Abhandlung von A. Kekulé. *Zeitschrift für Chemie* **8**, 250–254 (1865).

[27] Meyer, L. Die Natur der chemischen Elemente als Function ihrer Atomgewichte. *Ann. Chem. Pharm.* **VII Supplementband**, 354–364 (1870).

[28] Odling, W. On the proportional numbers of the elements. *Quarterly Journal of Science* **1**, 642–648 (1864).

[29] Hinrichs, G. *Atomechanik* (Iowa, 1867).

[30] Thyssen, P. & Binnemans, K. *Mendeleev and the rare-earth crisis*, 155–182 (Springer Netherlands, Dordrecht, 2015). URL https://doi.org/10.1007/978-94-017-9364-3_11.

[31] Rocke, A. J. Atoms and equivalents: The early development of the chemical atomic theory. *Historical Studies in the Physical Sciences* **9**, 225–263 (1978). URL http://www.jstor.org/stable/27757379.

[32] Bernadette, B.-V. Languages in chemistry. In Mary Jo, N. (ed.) *The Cambridge History of Science*, vol. 5, 174–190 (Cambridge University Press, Cambridge, 2003).

[33] Dalton, J. *A new system of chemical philosophy, part 2* (Manchester, 1810).

[34] Thomson, T. On the Daltonian theory of definite proportions in chemical combinations. *Annals of Philosophy* **2**, 32–52 (1813).

[35] Berzelius, J. J. *Essai sur la théorie des proportions chimiques et sur l'influence chimique de l'électricité* (Paris, 1819).

[36] Berzelius, J. J. Ueber die Bestimmung der relativen Anzahl von einfachen Atomen in chemischen Verbindungen. *Annalen der Physik* **83**, 397–416 (1826). URL https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.18260830802. https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.18260830802.

[37] Berzelius, J. J. Ueber die Bestimmung der relativen Anzahl von einfachen Atomen in chemischen Verbindungen. *Annalen der Physik* **84**, 1–24 (1826). URL https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.18260840902. https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.18260840902.

[38] Berzelius, J. J. Ueber die Bestimmung der relativen Anzahl von einfachen Atomen in chemischen Verbindungen. *Annalen der Physik* **84**, 177–190 (1826). URL https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.18260841006. https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.18260841006.

[39] Rex, F. Zur Erinnerungen an Felix Hoppe-Seyler, Lothar Meyer und Walter Hückel. *Bausteine zur Tübinger Universitätsgeschichte* **8**, 103–130 (1997).

[40] Rocke, A. J. Lothar Meyer's pathway to periodicity. *Ambix* **66**, 265–302 (2019). URL https://doi.org/10.1080/00026980.2019.1677976. https://doi.org/10.1080/00026980.2019.1677976.

[41] Laubichler, M. D., Maienschein, J. & Renn, J. Computational perspectives in the history of science: To the memory of peter damerow. *Isis* **104**, 119–130 (2013). URL https://doi.org/10.1086/669891. https://doi.org/10.1086/669891.

[42] Farey, J. On a curious property of vulgar fractions. *The Philosophical Magazine and Journal* **47**, 385–386 (1816).