Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig

The chemical space from which the periodic system arose

by

Wilmer Leal, Eugenio J. Llanos, Peter F. Stadler, Jürgen Jost, and Guillermo Restrepo

Preprint no.: 68 2019

The chemical space from which the periodic system arose

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

August 21, 2019

[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] Fundación Instituto de Inmunología de Colombia (FIDIC), Avenida 50 No. 26-20, 111321 Bogota, Colombia
[4] Corporación SCIO, Calle 57b 50-50 bloque d22 of. 412, 111321 Bogota, 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] The Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, New Mexico 87501

Aiming at presenting mid 19th century chemical knowledge to chemistry students, Meyer and Mendeleev embarked on writing textbooks from which their periodic systems arose [1]. These systems, as stated by Mendeleev [2], were based on similarity and order relationships among elements, which underlie the mathematical structure of every possible system [3]. Meyer and Mendeleev addressed similarity and order using the available chemical compounds by the 1860s, which constitute the chemical space at that time. Here we explore the size and diversity of this space and its influence upon the periodic system.

Chemical space: provider of order and similarity

In the 1860s chemists ordered elements by their atomic weights. Mendeleev, in his second 1869 publication about the system, discussed the pros and cons of selecting other properties as ordering criterion, e.g. electrochemical properties, relative affinities and valency [4]. Mendeleev settled on atomic weight, mainly based upon its invariability across all substances containing a particular element. Compounds were central for atomic weight determinations, which resulted from finding the smallest common weight of a large set of compounds containing the reference element. Similarity among elements was addressed through chemical resemblance, which for Meyer was captured through common valency in compounds. Mendeleev commented upon physical properties such as optical, electrical or magnetic ones and vapour densities [4], but discarded them for their variability over compounds of the same element. He was after a property related to the essential presence of an element in its compounds, which moved him to analyse chemical properties such as acidity or alkalinity of oxides. However, they were ruled out because of the several amphoteric oxides. Mendeleev ended up taking substance composition as the proxy for similarity, which boils down to the valency, selected by Meyer. "Thus the fluorine group contains elements which preferentially combine with a single atom of hydrogen, the oxygen group with two, the nitrogen group with three, and the carbon group with four atoms of hydrogen or chloride" [4], Mendeleev claimed. Thus, compounds provided order and similarity to the system.

The chemical space by 1869

As compounds were central for developing the system, the question that arises is which compounds and how many were known by the time the system arose, which prompted us to explore the size of the chemical space and its diversity.

How large was the chemical space?

Compounds can be either extracted from plants, animals and other sources, or synthesized, or both. They are then reported in the scientific literature, along with several of their properties. Reaxys©[1] is a large database of chemical information built from the Gmelin and Beilstein Handbooks and the Patent Chemistry Database, which is a suitable source of information for historical studies of chemistry [5] as it contains detailed data on reactions and compounds since 1771, along with information on the associated publications.

Meyer's and Mendeleev's first publications on the system date back to 1864 and 1869, respectively. After Mendeleev's one [2] (March 1st in the Gregorian calendar, February 17th in the Russian old style calendar) [6], Meyer published a paper updating his 1864 results [7]. As the spread of the 1869 literature was not that rapid at those times, and the production of new substances rose exponentially [5], scientists were hardly aware of the most recent results. For example, Mendeleev wrote in his 1869 extended publication about the system: "So far, bismuth has not been combined with hydrogen as have the elements similar to it" [4]. However, BiH_3 had been synthesized in 1843 [8]. For Mendeleev, elements similar to Bi were U, Sb, As, P and N (Table 1), which by 1869 had known hydrides of similar compositions: SbH_3 [9], AsH_3 [10], PH_3 [11] and NH_3 (UH_3

[1]Copyright © 2019 Elsevier Limited except certain content provided by third parties. Reaxys is a trademark of Elsevier Limited. Used under license via the Elsevier R&D Collaboration Network.

was reported in 1949 [12]). Moreover, Mendeleev acknowledged little knowledge about In [4], but there were already reports of more than 20 of its compounds.

The question that arises is how knowledge, or ignorance, of the chemical space may have affected the periodic system. We addressed the question by retrieving information from Reaxys from 1771 up to 1868 (included), which amounts to 26,502 single-step reactions and 11,484 substances, mainly reported in *Gmelins Handbuch der anorganischen Chemie* and gathered from leading 19th century journals, such as *Justus Liebigs Annalen der Chemie*, *Journal für praktische Chemie*, *Annalen der Physik*, *Annales de Chimie*, *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences*, *Jahresbericht über die Fortschritte der Chemie und Verwandter Theile Anderer Wissenschaften* and *Journal of the Chemical Society* among others. These compounds span 60 elements: H, Li, Be, B, C, N, O, F, Na, Mg, Al, Si, P, S, Cl, K, Ca, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Br, Rb, Sr, Zr, Nb, Mo, Ru, Rh, Pd, Ag, Cd, In, Sn, Sb, Te, I, Cs, Ba, La, Ce, Ta, W, Os, Ir, Pt, Au, Hg, Tl, Pb, Bi, Th, U. Meyer's 1870 paper does not include H, La, Ce, Th and U [7], whereas Mendeleev's first 1869 publication included them plus Er, Yt and Di [2]. Er and Yt, along with In, were elements whose identity was questioned by Mendeleev and expressed as ?Er, ?Yt and ?In in his table. Yt was the symbol used until 1920 for Y [13] and the first Y (or Yt) reaction is from 1872. Thus, neither Meyer nor Mendeleev had clear information about the element. Er was also problematic. By 1868 it was ignored that Er was actually a mixture of an element later (1878) coined Er and Yb, which were separated one year later into the current Er, Ho and Tm; and Sc and a Yb, respectively [14]. In 1907 the 1879 Yb was found to be a mixture of the current Lu and Yb [14]. Moreover, there was confusion between erbium and terbium on the basis of their spectra [14] and it is now known that Di, reported by Mendeleev as a chemical element, is a mixture of Pr and Nd. Therefore, we excluded Er, Yt and Di from our analysis and all the remaining study is based on our findings for the aforementioned 60 elements.

To know the extent of the chemical space populated by the different elements, we counted the number of compounds containing each element (Figure 1a; Supplementary, Table 1), where the enormous amount of O, H, C and N compounds is evident. Were O compounds absent, only 26% of the space would remain. If H substances were also removed, the available space would be 11.4%. By ignoring compounds containing O, H or C, the spanned space would be only 6% of that of 1869, which drops to 5% if N substances were disregarded. The importance of these organogenic elements contrasts with the percentage of space spanned if other quartets of elements were deleted. For example, removing halogen compounds would have much less dramatic effects, as 64% of the space would remain. Disregarding elements with few compounds such as Ta, Ce, Th and La would allow access to 96% of the space.

Thus, the chemical space was rather unevenly explored by 1869, oxygen and hydrogen being the elements with most compounds. Chemists were aware of this, as Mendeleev wrote: "The most widely distributed substances in nature possess the smallest atomic weights" [4] and "The higher atomic weights belong to elements that are rarely encountered in nature, which do not form large deposits, and which have, as a consequence, been little studied" [4].

As elements populate the space through combinations with others, we wondered how distributed the combinations of elements were.

Diversity of the chemical space: combinations of elements

By a combination we mean the elements present in a compound and arranged in lexicographic order, e.g. HOS for H_2SO_4 [5]. We found 3,022 combinations and analysed the distribution of their sizes (number of elements). Was the space populated with compounds of several elements? Or with substances of just few elements? For 60 elements, there are 1.15×10^{18} possible combinations (Methods). We found that chemists reported compounds of few elements: from two up to eight. Moreover, about three quarters of the combinations were of three up to five elements (Figure 1b). The differences between experimental and theoretical number of combinations across the different sizes (Figure 1b), show that there was still plenty of opportunity to populate the space with new compounds within the observed range.

To assess whether the space was also concentrated on some few combinations, we analysed the percentage of chemical space spanned by combinations weighted by their compounds (Figure 1c). The left hand side of the plot shows that combinations associated to one compound only, account for about 13% of the space. The right hand side shows the most frequent combinations: CHO and CHNO, which account for 8% and 7.5% of the space, respectively. We found that 37% of the combinations (1,120) account for 80% of the compounds and only 172 combinations spanned about half of the space. Strikingly, only 6% of the combinations gathered 50% of the reported substances. The top-20 of these combinations (Figure 1c) account for 30% of the space and their compositions show that the most populated combinations are based on organogenic elements including C, H, O, N, Cl, S, Br; which are among the elements with most compounds (Figure 1a). Notoriously, top-20 combinations are C and H based. The minor metallic span of the space is evident through the most populated metal combination: CHAgNO, which only gathers 0.8% of the space (Figure 1c). This is consistent with recent results on composition and growth of the chemical space [5].

The number of combinations per element follows a similar trend to that of compounds per element shown in Figure 1a (Supplementary, Figure 1; Spearman correlation of 0.99). The question arising is on the ratio between the number of compounds and of combinations for each element. We calculated it (Figure 1a, inset) and found that organogenic elements have much more compounds than combinations, when compared with other elements. C has six times more compounds than combinations and H, O and N, 4.6, 3.8 and 3.9, respectively. For La and Ta, each one of their combinations correspond to a single compound.

We wondered whether compounds of an element are concentrated in few of its combinations or spread over them. Figure 1d shows that organogenic elements concentrate most of their compounds in very few combinations: H, C, N and O gather three quarters of their compounds in about 1% of their combinations. This contrasts with the spread of other elements such as V and Au, which concentrate three quarters of their compounds in an ample range of combinations; 60% of V combinations and 67% of Au ones.

To explore the role of C combinations and to contrast them with some non-C combinations, we analysed how often some molecular fragments appear over time in known substances. The fragments studied were sulphate and nitrate anions as typical inorganic non-C ensembles. Carbonate anion was studied as an example of an inorganic C-fragment. Monosubstituted benzenes; primary amines and carboxylic acids were the organic C-ensembles studied. The organometallic junction C-M, M being any of the following elements: Zn, Sb, As, Hg, Tl, Bi, Pb, Rh, Co, Pt, Li, Be, Al, Fe, Si, Ge, was considered (Methods). Figures 1e-f show that before 1830 the number of compounds containing each one of the fragments was rather low, not surpassing 29 compounds per year. However, after 1830 organic fragments surged, while inorganic and organometallic fragments grew but not as dramatically as their organic counterparts. Although we showed elsewhere [5] that synthetic compounds steadily became the way to reach new substances well before Wöhler's 1828 synthesis of urea, the blossoming of organic compounds observed after 1830 and therefore the bias of the chemical space towards C-combinations may be caused by Wöhler's invigorating influence upon synthesis [15]. Before, C-combinations resulted from a balance of organic, organometallic and inorganic compounds. Therefore, Meyer and Mendeleev had at their disposal a chemical space with inorganic tradition and with about 40 years of very rapidly growing organic colonies.

Similarities arising from the chemical space

Granted the ordering of elements by atomic weight, we only need to determine the similarities among chemical elements arising from the available chemical space by 1869 to obtain the periodic system at that time [3]. We addressed similarity using Meyer's and Mendeleev's approach based on compounds. Although we keep the spirit of our previous approaches [16], here we consider key features of chemical similarity that have been disregarded before. As Mendeleev stated: "The elements, which are most chemically analogous, are characterised by the fact of their giving compounds of similar form RX_n" [17], an example of such common formula for similar elements is R_2O, where R is an alkali metal. This similarity statement can be interpreted as the degree of replaceability of element x in the formulae of compounds of y and the converse, whenever x and y are similar. From compounds of the chemical space, we extracted their different formulae. To quantify the similarity of element x regarding element y, we took each formula containing x and replaced x by the symbol A. The resulting formula was arranged by lexicographic order. Likewise, y was replaced by A in its formulae, which were lexicographically ordered (Figure 2). The similarity of x regarding y was quantified as the fraction of arranged formulae that x shares with y. This is an asymmetric relation, as shown in Figure 2, where Be is more similar to Mg ($s(\text{Be} \rightarrow \text{Mg})$) than the converse ($s(\text{Mg}\rightarrow\text{Be})$). This occurs as four of the six Be arranged formulae are shared with Mg, whereas four of the seven Mg formulae are shared with Be. In fact, by considering the actual space by 1869, the asymmetry strengthed as $s(\mathrm{Be} \rightarrow \mathrm{Mg})$=24/49=0.49, while $s(\mathrm{Mg} \rightarrow \mathrm{Be})$=24/296=0.081. Thus, in half of Be formulae, the element can be replaced by Mg and the resulting formulae are part of the space; whereas in less than 10% of Mg formulae, Be can replace Mg.

Figure 3 shows the most similar element for each element according to the 1869 chemical space. There are eight components of similar elements, labelled according to the oxidation state of their elements. The type of compounds making elements of each component similar is provided. The yellow component contains alkali metals and Ag, Tl and H (oxidation state I), which are similar because of their arsenates, sulphates, nitrates, carbonates, chlorides and iodides. Alkaline-earth metals group together with other elements having oxidation state II (blue component), except Si and Zr. Disregarding Si and Zr, with oxidation state IV, elements of this component form fluorides, chlorides and sulphides. Elements with oxidation state III forming chlorides gather in the red component, which contains pnictogens, except N. Halogens (oxidation state -I) show up together with N and V in the pink component, which form oxygenated carbon compounds: $CH_2(OH)V$, CH_3NCO and acetyl halides. The white component groups chalcogens and C (oxidation state -II) that are similar because of their ammonia derivatives: CH_3NH_2, $NH_3{\cdot}H_2O$, $NH_4{\cdot}HS$, $NH_4{\cdot}HTe$ and $NH_4{\cdot}HSe$. Transition metals are divided into three components, the green one gathering ferrous metals, Al, In and some other elements with oxidation state III sharing oxides. Elements of the orange component have oxidation states IV, while those of the purple component a mixture of oxidation states, represented in several types of common compounds.

Note that even if elements of a component are connected by sequences of similarities, it does not follow that two non-adjacent elements of a component hold a high degree of similarity. This is a consequence of the lack of transitivity of similarity relations [18]. For example, in the blue component Zr is similar to Si, as they form compounds where their oxidation state is IV, e.g. XO_2, XCl_4. Likewise, Si and Ti are similar, as well as Ti and Sn. However, Ti and Sn are also similar because of their compounds with oxidation state II. If we keep following the similarity arrows, the presence of oxidation state II becomes stronger, while that of IV winds down.

Of the similarity relations (Figure 3), those of organogenic elements are the weakest ones, as these elements only share less than 3% of their large sets of arranged formulae with their most similar elements. The strongest similarities occur for Rb, Cs, La, Se, Br, Li, Be, Pd and Tl, which share more than 50% of their arranged formulae with their most similar elements. In fact, for only less than half of the formulae of 85% of the elements, they can be replaced by their most similar element. Hence, the chemical space by 1869 made most of the elements to be dissimilar from each other. In such a space of tiny similarities, chemists, nevertheless, were able to recognise several of them. Was it the result of chemical genius? Or a consequence of the organogenic bias of the space?

To assess whether deep knowledge of the space was needed to gauge similarities among elements, or, on the contrary, whether partial knowledge sufficed to detect similarities, we took random samples of the formulae in the space of different sizes and analysed how often the similarities given by the whole formulae in the space (Figure 3) were present in the samples (Methods). After analysing the 9,752 formulae spanning the space, we found that there are stable similarities, still evident in tiny portions of the space, e.g. Br→Cl, Na→K, O→S, Cl→Br, S→O, K→Na, Se→S, As→P (shown at the top of Figure 4). For example, Br→Cl is a similarity often observed with even 10% of the space (about 970 formulae). In contrast, similarities that are only evident by considering large portions of the formulae are Zn→Mg, Cs→Na, Ca→Ba and those at the bottom of Figure 4. For instance, Ru→Mo only appears in more than half of the samples if the formulae analysed account for more than 95% of the known ones.

Figure 4 shows that most of the stable similarities occur for main group elements, while transition metal ones require deep knowledge of the space. This is a consequence of the small number of formulae associated to transition metals, formulae that are easily discarded by random sampling (Supplementary, Table 3). Moreover, by taking less than half of the formulae (4,876), about the same proportion of similarities is obtained; mainly among main group elements, having by far more formulae of few combinations than transition metals.

Meyer's and Mendeleev's similarities for chemical elements are shown in Table 1 [19, 2, 4, 7]. Mendeleev explained that in his 1869 table "in certain parts of the system the similarity between members of the horizontal rows will have to be considered, but in other parts, the similarity between members of the vertical columns" [4]. By contrasting Table 1 with Figures 3 and 4, we found that Meyer's and Mendeleev's similarities match to a large extent with the similarities arising from the chemical space, especially for main group elements. Their mismatches occur for transition metals, where deep knowledge of the space was necessary to detect the subtle differences among these elements.

To conclude, our results show the enormous bias of the chemical space by 1869 towards organic compounds, which populated the exponentially growing space in the last 40 years and tilted the scale of compounds in their favour. Before evident relevance of organic compounds, the space had a more homogeneous mixture of inorganic and organic compounds [5].

The size and diversity of the space shows that with knowledge of organogenic elements, chemists would have gauged a representative part of the space. In fact, random knowledge of about half of the space would have led to about half of the similarities resulting from the space, mainly for main group elements. This justifies the almost invariable resemblances among chemical elements reported by Meyer and Mendeleev and others [20] and the difficulty in finding stable similarities for transition metals [16, 21, 22], characterised by few compounds, therefore requiring a more complete knowledge of the space.

Meyer's and Mendeleev's idea of a book concentrating the most salient features of mid 19th century chemical knowledge, from which the periodic system arose, could have concentrated their content on oxoacid salts, halides, sulphides, simple carbon oxygenated compounds, ammonia derivatives and oxides, which

Table 1: Similarities reported by Meyer and Mendeleev in their 1869-1870 publications. Bold face sets indicate similarities found using the chemical space by 1869.

Meyer	Mendeleev	
B, **Al, In**, Tl	**Pt, Ir**, Os	**Si, Ti, Zr, Sn**
C, **Si, Sn, Pb**	Cu, Ag	**Mg, Zn, Cd**
Ti, Zr	**K, Rb, Cs**	B, Al
N, **P, As, Sb, Bi**	**Te, Se, S**	**Ca, Sr, Ba**
V, Nb, Ta	**Pb, Ba, Sr, Ca**	**Li, Na, K**
O, S, Se, Te	**Tl and alkali metals**	**Bi, Sb, As, P**, N, U
Cr, **Mo, W**	**Cl, Br, I**	Pd, Rh, Ru
F, Cl, Br, I	**O, S, Se, Te**	N, **P, As, Sb**
Mn, Fe, Co, Ni	Ag, Pb, Hg	**P, As, Sb**
Ru, Rh, Pd	V, **Mo, W**	Ta, **Sn, Ti**
Os, **Ir, Pt**	V, **Nb, Sb**	Fe, Ce, Pd, Pt
Li, Na, K, Rb, Cs	C, B, Si, Al	**Ba, Pb**, Tl
Cu, Ag, Au	V, Cr, Nb, **Mo, Ta, W**	U, B, Al
Be, Mg, Ca, Sr, Ba	**H***	**Ni, Co**
Zn, Cd, Hg		

*H is only slightly similar to K, only about 2% of its arranged formulae are shared with K.

are the carriers of most of the similarities among chemical elements. The challenge is writing the book of chemistry for the contemporary available space and to explore the influence of the current space upon the periodic system. Is it still there?

Acknowledgements

W.L. acknowledges support from the German Academic Exchange Service (DAAD): Forschungsstipendien-Promotionen in Deutschland, 2017/2018 (Bewerbung 57299294).

Contributions of authors

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., E.J.L., P.F.S., J.J. and G.R. discussed the results; G.R. wrote the original draft; W.L., E.J.L., P.F.S., J.J. and G.R. reviewed and edited the original draft.

Supplementary information

- Figure 1: Number of combinations per element
- Figure 2: Structures of the fragments studied

Methods

Number of combinations.

It was calculated as $\sum_{i=2}^{60}\binom{60}{i}$. This is a rough upper bound disregarding valency and compound stability.

Molecular fragments.

Search of these fragments was performed by exploring the connection tables of the compounds. A connection table is a "listing of atoms and bonds, and other data, in tabular form" [23]. For the sake of clarity, C-A means that at least a bond between C and A is reported on the table. It does not necessarily mean that C and A are bonded by a single covalent bond. The structures of the analysed molecular fragments are shown in Figure 2 of Supplementary Material.

Sampling the space.

We randomly took s% of the space and determined the most similar element of each element. This experiment was carried out 100 times. For each similarity $A \to B$ resulting for the whole space (Figure 3), we counted its frequency of appearance in the 100 experiments of size s%. The fractions of the space analysed ranged from 95%, 90%, 85%,..., 5%. The higher the frequency of appearance of the similarities shown in Figure 3 in the 100 experiments of each size s, the more stable the similarity regarding size s is. Moreover, the higher the frequency for different values of s, the more size independent the similarity is.

References

[1] 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`$_4$.

[2] 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).

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

[4] 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).

[5] Llanos, E. J. *et al.* Exploration of the chemical space and its three historical regimes. *Proceedings of the National Academy of Sciences* (2019). URL `https://www.pnas.org/content/early/2019/06/10/1816039116`.

[6] Gordin, M. D. The table and the word: Translation, priority, and the periodic system of chemical elements. *Ab Imperio* 53–82 (2013).

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

[8] Meurer, F. Ueber die bei der Anwendung des Marsh'schen Apparates gemachte Bemerkung, dass auch Wismuth, ferner Schwefelarsen und Schwefelantimon im Wasserstoff löslich und durch Verbrennen desselben wieder abgeschieden weden können. *Archiv der Pharmazie (Weinheim, Germany)* **86**, 33 – 33 (1843). URL `https://doi.org/10.1002/ardp.18430860110`.

[9] Vogel, A. Ueber arsenikhaltige phosphorige Säure und über Antimonwasserstoffgas. *Journal für Praktische Chemie* **13**, 55–60 (1838). URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/prac.18380130105`.

[10] Proust, J. L. *J. Phys. Chim. Delam'etherie* **51**, 173–184 (1801).

[11] de Grotthuss, M. T. Expériences sur la combinaison du phosphore ave les métaux et leurs oxides pas la voie humide; suivies de l'examen d'un gaz provenant d'une décomposition particulière de l'alcool. *Annales de chimie* **64**, 19–41 (1807).

[12] Spedding, F. H. *et al.* Uranium hydride; preparation, composition and physical properties. *Nucleonics* **4**, 4–15 (1949).

[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 (2019).

[14] Evans, C. *Episodes from the History of the Rare Earth Elements* (Boston : Kluwer Academic Publishers, 1996).

[15] Partington, J. R. *A history of chemistry* (London: Macmillan, 1964).

[16] 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).

[17] 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).

[18] Restrepo, G. & Mesa, H. Chemotopology: beyond neighbourhoods. *Current Computer-Aided Drug Design* **7**, 90–97 (2011).

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

[20] Scerri, R. E. *The periodic table: Its story and its significance* (New York: Oxford University Press, 2007).

[21] Bernal, A., Llanos, E., Leal, W. & Restrepo, G. Similarity in chemical reaction networks: Categories, concepts and closures. In Basak, S. C., Restrepo, G. & Villaveces, J. L. (eds.) *Advances in Mathematical Chemistry and Applications*, 24 – 54 (Bentham Science Publishers, 2015). URL `http://www.sciencedirect.com/science/article/pii/B9781681080536500028`.

[22] Restrepo, G. The periodic system: A mathematical approach. In Scerri, E. & Restrepo, G. (eds.) *Mendeleev to Oganesson: A Multidiciplinary Perspective on the Periodic Table*, chap. 4, 80–103 (Oxford University Press, New York, 2018).

[23] Warr, W. A. Representation of chemical structures. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**, 557–579 (2011). URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.36`.

W.L. acknowledges support from the German Academic Exchange Service (DAAD): Forschungsstipendien-Promotionen in Deutschland, 2017/2018 (Bewerbung 57299294). G.R. is grateful to Michael Gordin for his comments about Mendeleev by 1869.

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.R. (email: restrepo@mis.mpg.de).

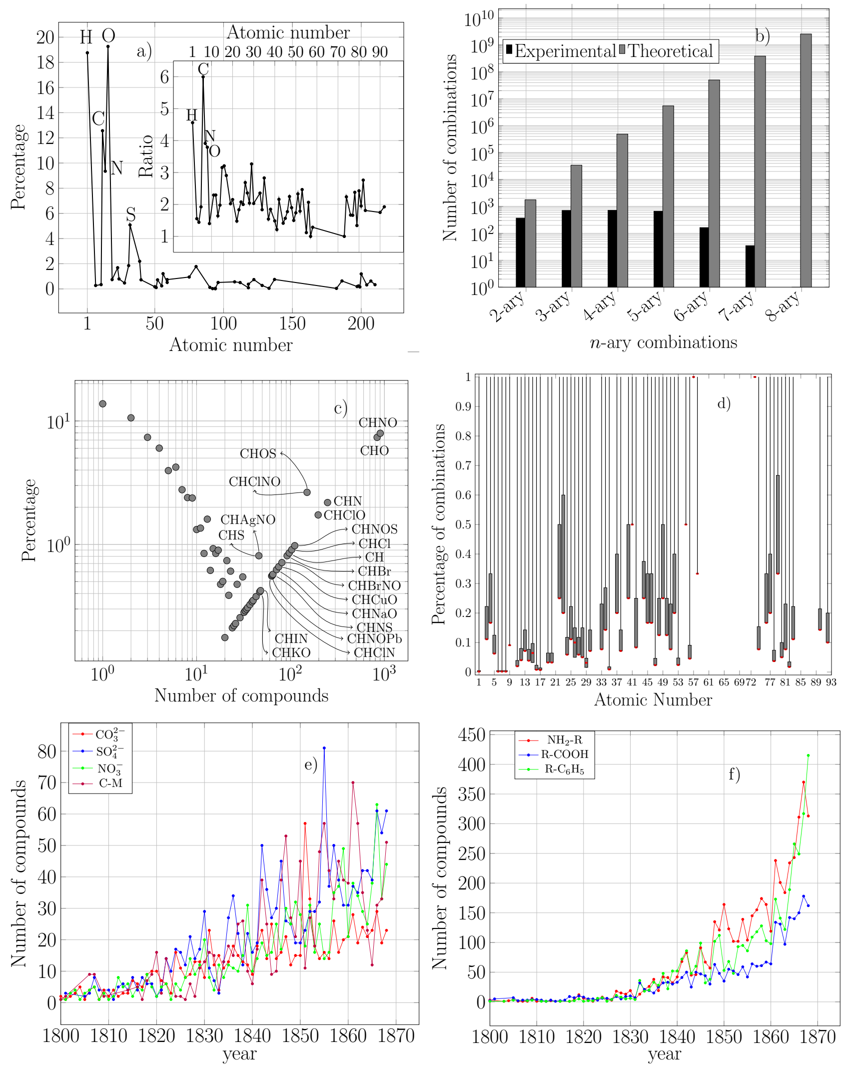

Figure 1: Size and diversity of the chemical space: a) Chemical elements known by 1869 and the percentages of compounds in which they take part. These percentages are non-additive as a single compound adds to each one of its elements, e.g. H_2O contributes to both H and O counting. Inset) Ratio between the number of compounds and of combinations for each element. b) Size distribution of combinations. Number of n-ary reported combinations of elements (black) and theoretical bounds (gray) (Methods). c) A pair (x, y) indicates that $y = cx$ percentage of the chemical space is spanned by c combinations accounting for x compounds. d) Box plots of the distribution of compounds per combination for each element, with whiskers indicating combinations with least and most populated compounds; red line shows the median of each distribution. e) Temporal distribution of several typical inorganic, organometallic and f) organic molecular fragments. C-M stands for the bond C-metal, with M={Zn, Sb, As, Hg, Tl, Bi, Pb, Rh, Co, Pt, Li, Be, Al, Fe, Si, Ge}.

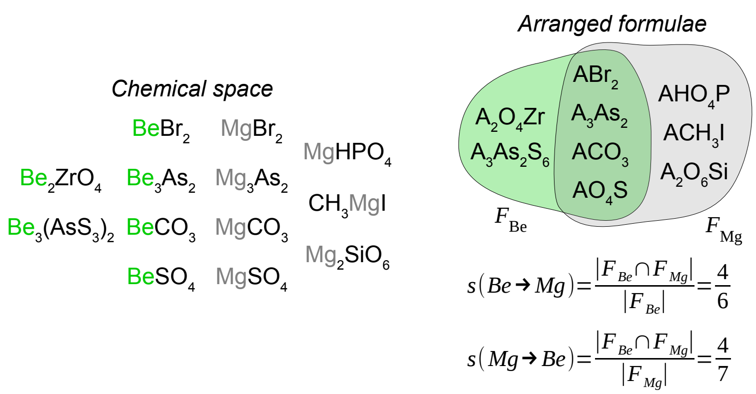

Figure 2: Similarity among chemical elements. A toy-chemical space of 13 formulae. Each formula provides an arranged formula for an element in the given formula. For example, A_2O_4Zr is the arranged formula of Be from formula Be_2ZrO_4. Arranged formulae of element x are gathered in F_x. Similarity of element x regarding element y is given by $s(x \rightarrow y)$.

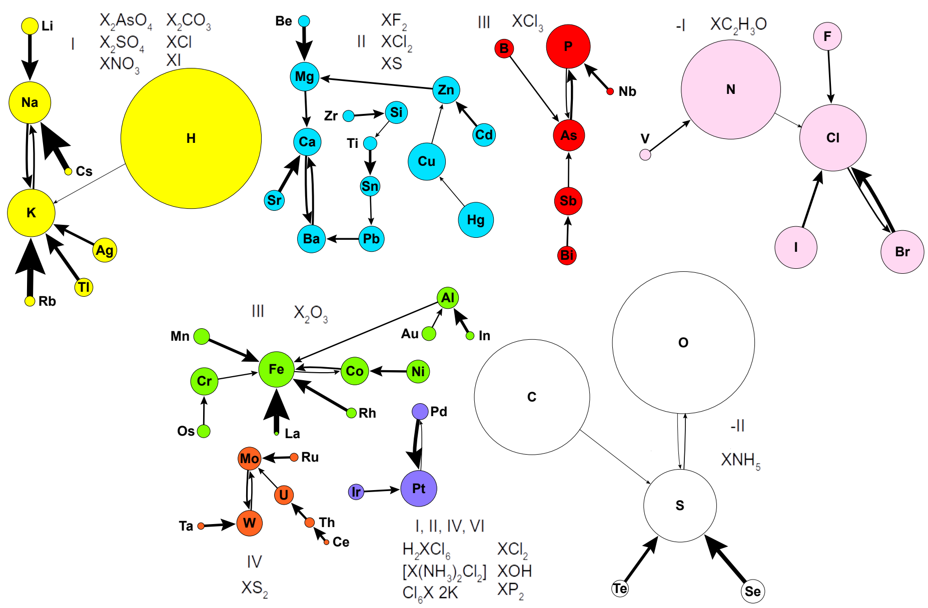

Figure 3: Most similar elements by 1869. Arrows $A \to B$ indicate that A is most similar to B. Node size is proportional to the number of formulae in which the element represented by the node is involved. Arrow size is proportional to the similarity $s(A \to B)$. Shared formulae for elements belonging in a component (connected set of elements) are shown. Formulae of the blue component disregard Zr and Si. Likewise, disulfide of elements in the orange component does not apply to Th and Ce.

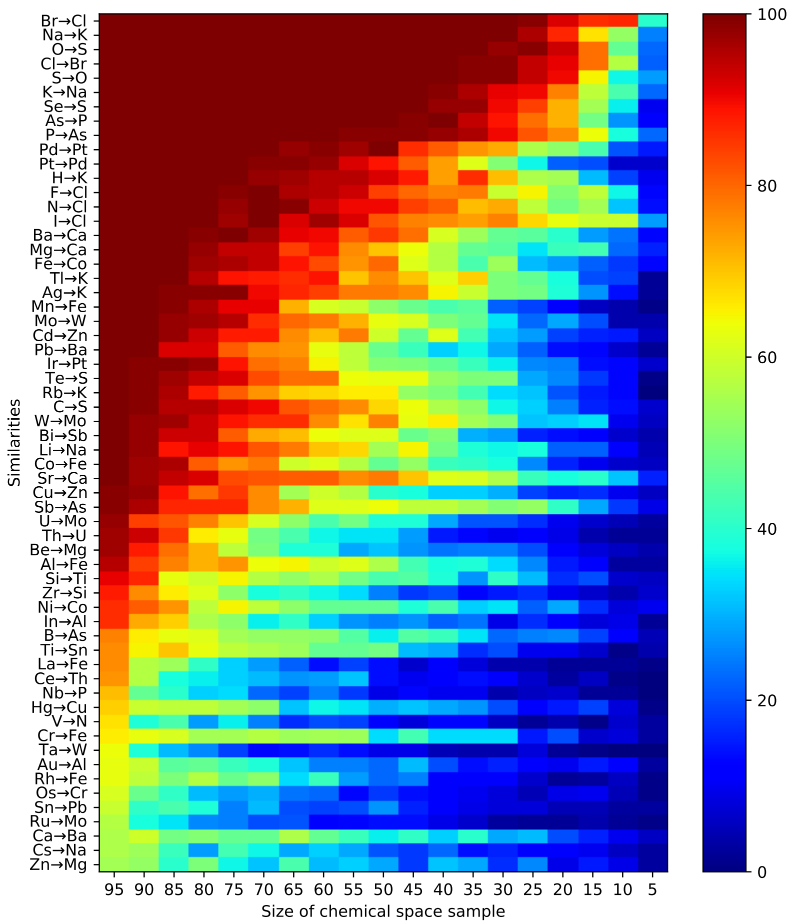

Figure 4: Stability of similarities regarding chemical space size. Most similar element for each element obtained by considering the whole space and their frequency of appearance (colour scale) in random samples of the space of different sizes (ranging from 5% to 95% of the space).