New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG IOP Institute of Physics

PAPER • OPEN ACCESS

The optimal one dimensional periodic table: a modified Pettifor chemical scale from data mining

To cite this article: Henning Glawe *et al* 2016 *New J. Phys.* **18** 093011

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
18 March 2016

REVISED
9 August 2016

ACCEPTED FOR PUBLICATION
10 August 2016

PUBLISHED
6 September 2016

PAPER

The optimal one dimensional periodic table: a modified Pettifor chemical scale from data mining

Henning Glawe[1], **Antonio Sanna**[1], **E K U Gross**[1] **and Miguel A L Marques**[2,3]

[1] Max-Planck Institut für Microstrukture Physics, Weinberg 2, 06120 Halle, Germany
[2] Institut für Physik, Martin-Luther-Universität Halle-Wittenberg, D-06099 Halle, Germany
[3] Institut Lumierère Matière, UMR5306 Université Lyon 1-CNRS, Université Lyon, F-69622 Villeurbanne Cedex, France

E-mail: sanna@mpi-halle.mpg.de **and** miguel.marques@physik.uni-halle.de

Keywords: Pettifor Scale, chemical similarity, atomic substitution, crystal structures

Supplementary material for this article is available online

Abstract

Starting from the experimental data contained in the inorganic crystal structure database, we use a statistical analysis to determine the likelihood that a chemical element A can be replaced by another B in a given structure. This information can be used to construct a matrix where each entry (A, B) is a measure of this likelihood. By ordering the rows and columns of this matrix in order to reduce its bandwidth, we construct a one-dimension ordering of the chemical elements, analogous to the famous Pettifor scale. The new scale shows large similarities with the one of Pettifor, but also striking differences, especially in what comes to the ordering of the non-metals.

1. Introduction

The organization of the chemical elements in a 'table' has fascinated and motivated scientists for the best part of two centuries. The traditional representation of the periodic table has a two-dimensional structure, with elements arranged in periods and groups. This arrangement not only puts into evidence the chemical similarity between atoms, but also reflects the basic quantum-mechanical character that rules atomic physics. Since the seminal works of Lothar Meyer and Dimitri Mendeleev, hundreds of such two- (or even higher-) dimensional representations have been put forward, featuring spirals, circles, cubes, etc. Moreover, one can arrange the elements according to their atomic properties, or choose to put in evidence other molecular or solid-state properties.

It is true that the best description of the relationship between the chemical elements requires two (or more) dimensions. However, in many practical cases, one requires a much simpler, one-dimensional ordering where elements that are chemically similar occupy neighboring positions. In this paper we are concerned with one such ordering, already studied by Pettifor 30 years ago [1], and that is used extensively in the modern fields of accelerated materials design and high-throughput calculations (see, e.g., [2–5]).

Pettifor's original interest was in the structural stability of binary AB compounds [1]. Binary compounds crystallize in 34 different structure types. If we assign a different symbol to each structure type and plot it for each A and B we obtain a so-called structure map [6]. The problem that Pettifor tried to solved was how to order the chemical elements in order to achieve the best structural separation within such two-dimensional plot. There had been several previous attempts to achieve such separation using properties like the core radius, the electronegativity, the number of valence electrons, etc. Unfortunately these approaches not only led to high-dimensional structure maps (difficult to plot and visualize), but also to a rather disappointing structural separation. Pettifor's solution was rather elegant, but also quite radical. He neglected all theoretical considerations and constructed a fully phenomenological one-dimensional ordering of the elements that provided a near-perfect structural separation of the AB binaries. Further work showed that this was also true for other binary A_xB_y systems [7].

Pettifor had at his disposal 574 binary AB compounds plus a few hundred other binaries phases. Today we have available the experimental crystal structures for at least two orders of magnitude more compounds. This information can be found in several databases such as the inorganic crystal structure database (ICSD) database [8], or the crystallography open database [9]. In this paper we show that we can use Pettifor's idea that chemical similarity manifests itself in the formation of similar structures and use the wealth of new information to improve its original scale.

We will assume that the structural information in ICSD is statistically significant, and we measure how often the substitution of an constituent element in a known material leads to another known material with the same structure type. For example, the non-superconductor AlB_2 is related to the superconductor MgB_2 by such a substitution, accompanied by a slight adjustment of the interlayer distance. This measure, properly normalized, can be seen as giving a quantitative value for the probability that an element can be replaced by another, and therefore of the 'chemical similarity' between the two elements.

Having obtained a measure of chemical similarity we can construct a Pettifor-like scale by asking that similar chemical elements occupy neighboring positions in such a scale. This task is performed by borrowing some numerical tools from the field of optimization of sparse linear equations. One should stress that, as for the original periodic table, a 'chemical similarity' scale is not only a convenient tool for plotting quantities. In fact, structure maps using this scale can reveal, at a fast glance, trends and outliers allowing for an intuitive understanding of the data. They can also form the basis for simple extrapolations of data, and even for predicting new materials, in line, e.g., with the heuristic rules of Pauling [10] (relation between ionic radii and structure in ionic crystals) or of Hume-Rothery [11] (relation between valence electrons per atom and the crystal structure).

The rest of this paper is organized as follows. In section 2 we construct a function that measures the similarity between two chemical elements. Then, in section 3 we show how to mathematically define and construct a modified Pettifor scale, that is then obtained in section 4. Finally, we present our conclusions and a brief outlook in section 5.

2. Chemical element substitutions

The question we will try to answer in this section is how to measure the degree of chemical similarity between a pair of elements (A, B). Our working hypothesis is that the chemical elements A and B are similar if when mixed with other elements of the periodic table they crystallize in the same structure prototype. For example, the elements Li, Na, K, Rb, Cs, etc crystallize in the rocksalt structure when combined with Cl, and are therefore, by our definition, similar. Of course, if two chemical elements crystallize in the same structure they have to be, in a certain sense, related chemically.

We have to start the discussion by our definition of a structure prototype. Each such prototype is defined by a space group and by the set of occupied Wyckoff positions. All other information concerning, e.g., the lattice constants or the chemical elements occupying the Wyckoff positions, is discarded. Two materials I and J are substitution partners (SP), i.e. they are related by element substitution, if (i) they share a structure prototype and (ii) they differ only by the substitution of one chemical element A by B. We only consider substitutions involving a single pair of chemical elements. Furthermore, elemental solids are only considered regarding partial substitutions, i.e. the structure prototype is required to have more than one occupied Wyckoff position and the substitution only occurs in one of the sublattices.

We analyzed the whole ICSD [8], excluding the database entries for which we have only incomplete information, alloys (materials with Wyckoff positions randomly occupied by more than one element or only fractionally occupied), and duplicated entries. ICSD is the largest available database for completely identified inorganic crystal structures, containing about 173000 entries. SPs were detected in 20500 materials, corresponding to 44% of the usable set, which demonstrates that such relation between different materials is a rather common phenomenon. The vast majority of materials are found to have only few SP; less then 100 examples are found for any SP count above 30. Moreover only a hand full of materials are related to 45 and more materials by single-element substitutions. These are the family of binary selenides and tellurides in the rocksalt structure.

We define $\delta^{IJ}_{AB} = 1$ if the two materials I and J are SP for the replacement of A by B and 0 otherwise. Furthermore, we define $\delta^{I}_{A} = 1$ if material I is a partner with any other material by substituting element A. The basic data for the further steps of our analysis is simply the number of non-duplicate SP materials related by substitution of element A by B:

$$S_{AB} := \sum_{I,J \neq I} \delta^{IJ}_{AB}, \tag{1}$$

where I and J run over all non-duplicate materials. By construction, this quantity is symmetric $S_{AB} = S_{BA}$. For a fixed element A, the quantity S_{AB} reveals, at a quick glance, which substitutions B are available for A. It can also

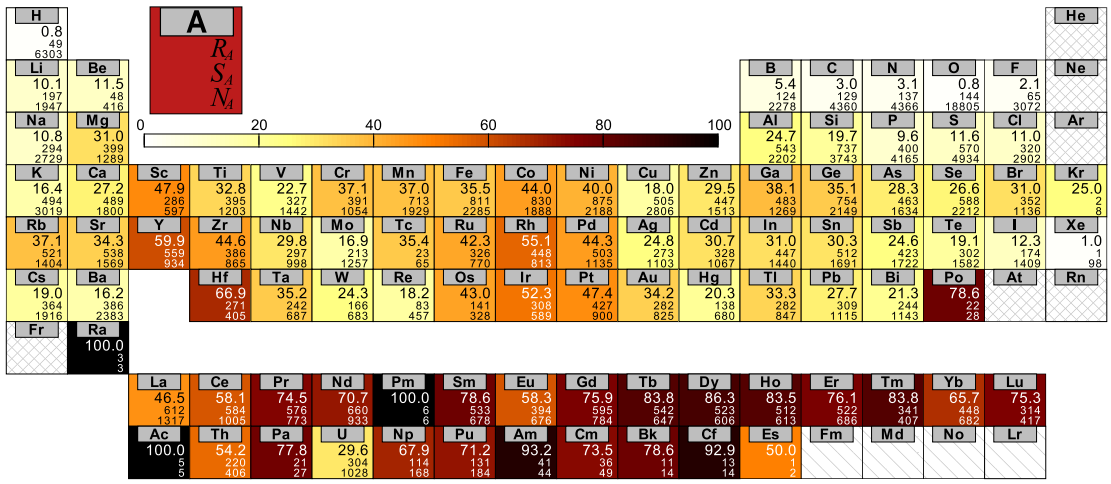

Figure 1. Periodic table reporting for each element A its (percentual) replacibility R_A (defined in equation (3)), its substitution count S_A (equation (2)) and N_A the total number of materials containing element A. The colorscale highlights the repleacibility.

be used to bias the search for new materials using computational high-throughput techniques. Due to their general usefulness, we include as Supplemental Information plots of S_{AB} in the form of periodic tables for each chemical element A.

In order to establish a general picture of the element substitution properties within ICSD, we computed the number of materials having any SP materials by substitution of element A

$$S_A := \sum_I \delta_A^I \tag{2}$$

figure 1 displays S_A, together with the *element replaceability* that we define as the ratio

$$R_A := \frac{S_A}{N_A} \tag{3}$$

with N_A the total number of materials containing A. If we assumed that ICSD contains all possible materials, this ratio could be interpreted as a probability to reach a new, stable crystal structure by the substitution of element A; however, as ICSD is restricted to the materials that have been published and included in the database, the numbers should be interpreted as trends.

The highest replaceabilities are found for the lanthanides and actinides, with ratios above 60% for most elements; substitutions are most frequently observed among elements in the same series, in agreement with the chemical similarity [12, 13] observed in these series. On the other hand, the lowest ratios are observed for hydrogen and the first-row p elements. This finding is in agreement with the so-called 'first-row anomaly' [14], that states that properties of elements in this row are significantly different from properties of other elements in the same group. This is true not only for the atomic radii, electronegativity, etc but also for the bonding behavior [15] of the elements. Among the remaining main group elements, SP rates lie between 10% and 38%, while for d transition metals lie between 16% and 67%.

In the limit of a sufficiently large and uniformly sampled set of materials, the relative element-pair substitution rate can serve as an approximation for the corresponding (conditional) probability of replacing an atom A by B while conserving the crystal structure:

$$P_{A\to B} = \frac{S_{AB}}{\sum_{B'} S_{AB'}}. \tag{4}$$

However the ICSD does not fulfill the required criteria: chemical elements occur at rather different frequencies, certain structure types and compositions have received more attention than others, and the total number of materials is far from the required limit; summarizing, the data is noisy and biased. For this reason for the scope of our investigation it is better to work with a symmetrized version of equation (4) that partially compensates for the database limited statistics:

$$P_{AB} = \sqrt{\frac{S_{AB}^2}{\left(\sum_{A'} S_{A'B}\right)\left(\sum_{B'} S_{AB'}\right)}}. \tag{5}$$

In addition we use a noise reduction scheme by a applying a threshold on S_{AB} (equation (1)). As always in such cases, the choice of threshold presents a tradeoff between noise reduction and loss of information.

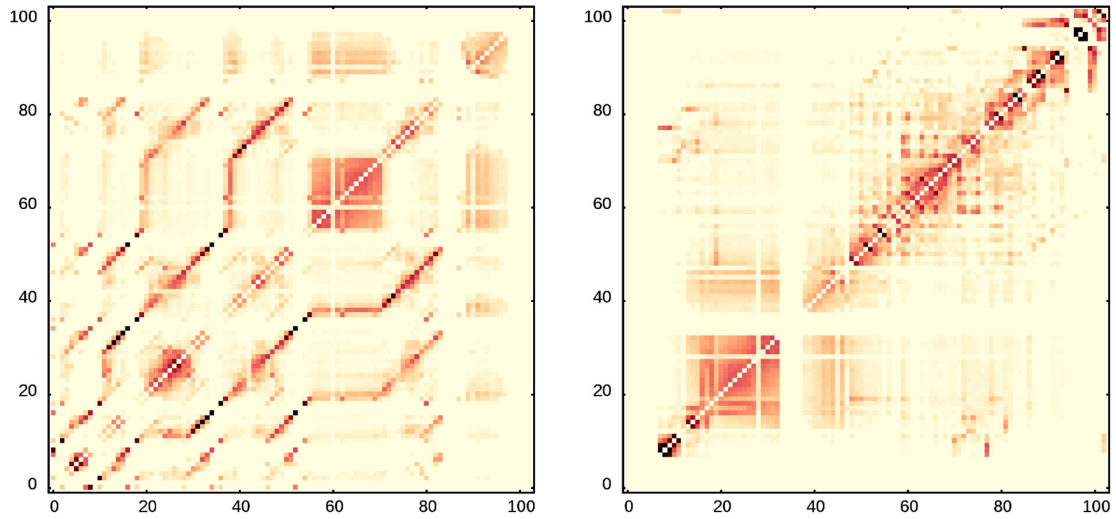

Figure 2. The probability matrix P_{AB} using the atomic number (left) and the Pettifor scale (right) to order the chemical elements. The color range goes from white (zero entry) to black (entry >= 0.3). The maximum entry is ≈0.62 for the pair Br–Cl.

Cross-validation considering a set of test elements suggests the conservative choice of 3 elements counts, i.e. any pair detected only once or twice is classified as noise and removed from the set; this process shrinks the original set of pairs by 32%. Due to its nature as a product of probabilities, $0 \leqslant P_{AB} \leqslant 1$, where the upper bound is reached in case A and B substitute each other exclusively.

3. A mathematical definition of the (modified) pettifor scale

Let us remember that Pettifor constructed his scale by trying to separate the different crystal structures of binary compounds AB in a binary diagram [1]. Besides the knowledge of the crystal structures of AB compounds, his main tools were his formidable chemical intuition and trial and error. In our work we give a step beyond, and use the statistical analysis of section 2 to perform the task of creating a chemical scale. This has the advantage of being completely unbiased with respect to possible (human) prejudices, and of assuring the optimal (or at least a very good) ordering based on the totality of the available data.

Having obtained the matrix describing the probability of a successful substitution of an element A by an element B, we can now proceed to the construction of a new Pettifor map. The idea is very simple: If two chemical elements are similar, then it is probable that one can substitute one by another in a given crystal structure. This will lead to a large entry in the matrix element (A, B). If we can order the chemical elements such that the large matrix elements are close to the diagonal, this implies that similar elements will occupy neighboring positions in the chemical scale.

The left panel of figure 2 shows the matrix P_{AB} using as ordering of the chemical elements the atomic number. The first striking evidence is that the matrix has a very geometrical structure and it contains very large entries far away from the diagonal. In fact, this is due to the two-dimensional nature of the Mendeleev periodic table [16], and the large off-diagonal entries that form lines that run through the matrix simply reflect the large similarity between elements in the same group. There are also some empty rows and columns, due to chemical elements for which there is no structural information in ICSD. This is true for most noble gases that do not usually form any stable compounds in normal conditions (He, Ne, Ar), and for the radioactive Rn, Fr, At and some of the actinides (Es, Fm, Md, and No). The large brown square is due to the lanthanides that are well-know to be quite similar chemically and to a large degree interchangeable in many crystal structures.

Clearly, using the Pettifor scale (see table 1) to order the chemical elements yields a matrix that is in a much more diagonal form (see right panel of figure 2). Now, not only the lanthanides form a clear structure, but several other groups are also clearly visible across the figure. It turns out that the Pettifor scale is already a very good solution to our ordering problem (we will show actual quantitative evidence below).

Finding a numerical framework to make the matrix more diagonal is very simple as soon as we recognize that our problem is similar to the reduction of the bandwidth of a sparse matrix. As this plays a very important role in the solution of large linear systems, it has been studied intensively since the original Cuthill-McKee algorithm in 1969 [17]. This problem is also related to the famous traveling salesman problem. In fact, the traveling salesman has to find a path through a certain number of cities (i.e., an ordering of the cities) that minimizes the total travel

Table 1. The different chemical scales: **Z** is the atomic number; **P** is the Pettifor scale; **GA** is the scale stemming from the genetic algorithms; $\mathbf{P_m}$ is the modified Pettifor scale proposed in this work.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Z	H	He	Li	Be	B	C	N	O	F	Ne	Na	Mg	Al	Si	P	S	Cl	Ar	K	Ca	Sc	Ti	V	Cr	Mn	Fe
P	He	Ne	Ar	Kr	Xe	Rn	Fr	Cs	Rb	K	Na	Li	Ra	Ba	Sr	Ca	Yb	Eu	Y	Sc	Lu	Tm	Er	Ho	Dy	Tb
GA	He	Ne	Ar	At	Rn	Fr	Es	Fm	Md	No	Lr	Kr	Xe	Pm	Cs	Rb	K	Na	Li	Ra	Ba	Sr	Ca	Eu	Yb	Lu
$\mathbf{P_m}$	He	Ne	Ar	Kr	Xe	Rn	Fr	Cs	Rb	K	Na	Li	Ra	Ba	Sr	Ca	Eu	Yb	Lu	Tm	Y	Er	Ho	Dy	Tb	Gd
n	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
Z	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te
P	Gd	Sm	Pm	Nd	Pr	Ce	La	Lr	No	Md	Fm	Es	Cf	Bk	Cm	Am	Pu	Np	U	Pa	Th	Ac	Zr	Hf	Ti	Nb
GA	Tm	Y	Er	Ho	Dy	Tb	Gd	Sm	Nd	Pr	Ce	La	Ac	Am	Cm	Bk	Cf	Pu	Np	U	Th	Pa	Sc	Zr	Hf	Ti
$\mathbf{P_m}$	Sm	Pm	Nd	Pr	Ce	La	Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr	Sc	Zr	Hf	Ti	Ta
n	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78
Z	I	Xe	Cs	Ba	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu	Hf	Ta	W	Re	Os	Ir	Pt
P	Ta	V	Mo	W	Cr	Tc	Re	Mn	Fe	Os	Ru	Co	Ir	Rh	Ni	Pt	Pd	Au	Ag	Cu	Mg	Hg	Cd	Zn	Be	Tl
GA	Nb	Ta	V	Cr	Mo	W	Re	Tc	Os	Ru	Ir	Rh	Pt	Pd	Au	Ag	Cu	Ni	Co	Fe	Mn	Mg	Zn	Cd	Hg	Be
$\mathbf{P_m}$	Nb	V	Cr	Mo	W	Re	Tc	Os	Ru	Ir	Rh	Pt	Pd	Au	Ag	Cu	Ni	Co	Fe	Mn	Mg	Zn	Cd	Hg	Be	Al
n	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	
Z	Au	Hg	Tl	Pb	Bi	Po	At	Rn	Fr	Ra	Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr	
P	In	Al	Ga	Pb	Sn	Ge	Si	B	Bi	Sb	As	P	Po	Te	Se	S	C	At	I	Br	Cl	N	O	F	H	
GA	Al	Ga	In	Tl	Pb	Sn	Ge	Si	B	C	N	P	As	Sb	Bi	Po	Te	Se	S	O	I	Br	Cl	F	H	
$\mathbf{P_m}$	Ga	In	Tl	Pb	Sn	Ge	Si	B	C	N	P	As	Sb	Bi	Po	Te	Se	S	O	At	I	Br	Cl	F	H	

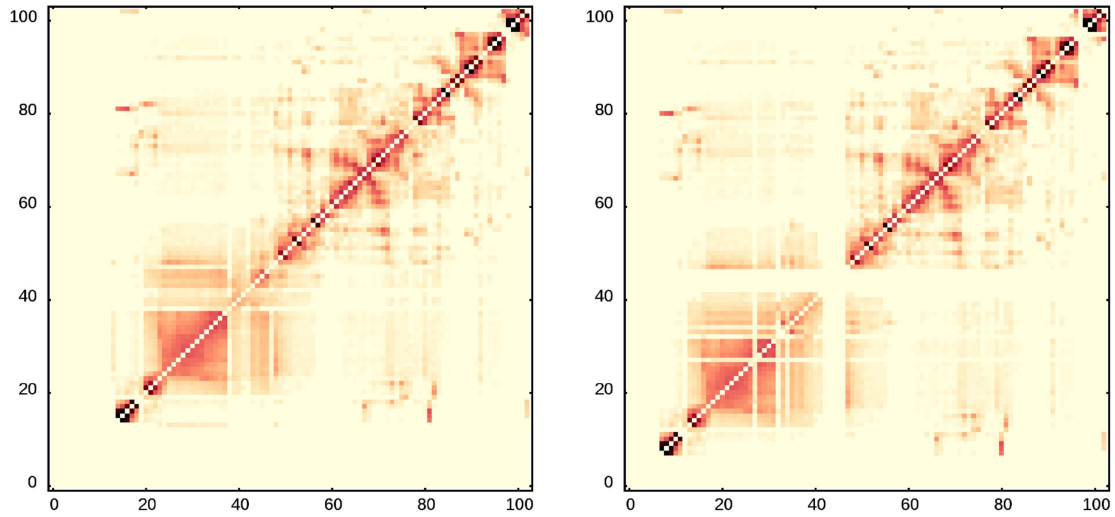

Figure 3. The probability matrix P_{AB} using our chemical scale optimized with genetic algorithms (left) and our proposed modified Pettifor scale (right) to order the chemical elements.

distance. We have to find a path through the chemical element space that maximizes the diagonal character of the matrix.

The traveling salesman problem is a hard problem (NP-complete), and the time to find the optimal solution grows exponentially with the number of cities. However, many strategies have appeared over the years to obtain good solutions. We decided to use genetic algorithms, mainly due to their simplicity. The first step in using genetic algorithms is defining the objective function to be optimized. Following several numerical experiments we selected the following function:

$$\mathcal{F} = -\sum_{A,B\neq A} \frac{P_{AB}}{|i_A - i_B|}, \tag{6}$$

where P_{AB} is defined by equation (5), i_A is the position of element A in the ordering, and the sum runs through all pairs such that $A \neq B$. This choice gives increased weight for entries close to the diagonal, while not penalizing too much small entries far from the diagonal. Obviously, the function $\mathcal{F}$ has to be minimized.

The second step is to define a gene. We take simply a list of 103 entries with the natural numbers from 1 to 103, indicating the order in which the chemical elements should be arranged. For the crossover operator we first select randomly a segment of one of the parents that is passed to the same position to the child, and then fill the voids using the gene of the second parent by removing the entries already contained in the child gene. We tried as mutation operations: (i) swapping two random elements in the gene or (ii) moving an element from one random position in the gene to another. We found that the second choice was greatly superior in our simulations. The mutation rate was set to 20%. We note that, as our matrix has a relatively low dimensionality, we did not have to use more sophisticated and efficient genetic algorithms such as the ones from [18].

To solve the problem of the elements for which no information exists in ICSD we moved them all to the beginning of our gene. Furthermore, and in order to have an easier comparison with the Pettifor scale we decided to fix the two end-points of our scale to be Kr (the first rare-gas for which we have data) and H (number 103 in the Pettifor scale). We checked that this arbitrary choice does not have a significant impact in the value of the minimum objective function.

4. Results

We run a series of simulations using a pool of 200 genes in our population that were evolved for around 500 generations. We used as starting points random genes, the Pettifor scale, and the ordering by atomic number. Our best result is shown in the left panel of figure 3 and in table 1.

We can use the numerical value of equation (6) in order to have a quantitative assessment of the quality of our chemical scale. A random ordering of the chemical elements leads to a value of $\mathcal{F}$ typically between -3 and -4. Using the atomic number to order the matrix (see left panel of figure 2), i.e. taking into account the similarity of elements along a period of the periodic table, improves this value to $\mathcal{F} = -7.68$. Using the Pettifor scale (right panel of figure 2) yields $\mathcal{F} = -15.47$. As we can see this an excellent improvement. This value decreases

further to $\mathcal{F} = -15.62$ by eliminating the elements for which there are no entries in ICSD. Finally, the optimal ordering coming out of our genetic algorithms (see figure 3) yields $\mathcal{F} = -16.91$.

Not only our approach led to a better quantitative results, but we can clearly see a qualitative improvement of the matrix. Now, in the upper right corner of figure 3 we can clearly identify two blocks (darker squares) that represent subgroups of chemical elements that are much similar between themselves than to any other element of the periodic table (we stress that this structure was absent from Pettifor's original scale). The first of these blocks includes the elements H, F, Cl, Br, I. It is interesting to notice that H appears together with the halogens, which is justified by the fact that most of the H substitutions present in ICSD are with F. This supports the argument that H, F, and Cl form a triad, and that should therefore be placed in the same group of the periodic table [19].

Then comes another group containing O, S, Se, Te, Po, Bi, Sb, As, P and N, with a clear subgroup formed by the chalcogens. The next group contains only C and B, which are known to be somewhat special elements of the periodic table. Then there is a group of 9 elements containing Be and the remaining members of the boron group (Al, Ga, In, and Tl) and of the carbon group (Si, Ge, Sn, and Pb). We note that this group is considerably less well-defined than the previous ones, with several possible substitutions outside itself. The next group is constituted by transition metals (plus Mg), followed by the actinides and then by the lanthanides. The end of the table is quite well defined and it is essentially unchanged from the original Pettifor scale. It includes the rest of the alkali Earth metals (Ca, Sr, Ba, and Ra), then the alkali metals (Li, Na, K, Rb, and Cs), and the noble gases (Kr and Xe). The radioactive rare-Earth Pm appears in between the alkali metals and the noble gases, which is strange chemically, but can be understood due to the very small number of entries in ICSD containing this element, leading to very poor statistics. Finally we find all elements for which there is no entry in ICSD, namely He, Ne, Ar, At, Rn, Fr, Es, Fm, Md, No, and Lr.

Our modified Pettifor scale follows mostly the order of the groups of the periodic table, but there are several cases where it follows the period, or even a diagonal. Note that a relationship is well-known to exist between certain pairs of diagonally adjacent elements, as trends moving down the periodic table are usually the exact opposite of the trend moving across. From the significant diagonal relationships known to exist (Li/Mg, Be/Al, and B/Si), only Li/Mg is not present in our scale.

We have one last task left in order to have a complete chemical scale similar to the one of Pettifor: to reorganize the elements for which there is little or no data in ICSD. We performed the following operations: (i) we restored the normal ordering of the noble gases (He, Ne, Ar, Kr, Xe, Rn); (ii) we inserted At next to I; (iii) we moved Pm to between Nd and Sm; (iii) as the statistics for the actinides is very limited, we restored their normal (atomic number) ordering; (iv) Nb and Ta turn out to be basically interchangeable without changing significantly the value of the objective function. Therefore, we decided to swap their positions to restore the normal group ordering. The resulting scale has $\mathcal{F} = -16.70$.

Our final modified Pettifor scale (P_m) is given in given in table 1 and the corresponding matrix is shown in the right panel of figure 3. In spite of the fact that we use a far larger and chemically more diverse statistical set, the overall structure of the scale is quite similar to the original Pettifor scale, supporting the universality of the chemical similarity concept. However, there are a few significant differences between our scale and the original one of Pettifor that can be understood by looking at the substitutional tables given in the Supplemental Material. Let us discuss a few striking examples.

Let us start with Sc. In the original Pettifor table it is between Y and Lu. However, from the table we see that Sc is actually chemically closer to Zr than to Y or Lu (a diagonal similarity apparently overlooked up to now). On the other side, Lu is actually a better choice, however, our algorithm chooses to include it among the other lanthanides and actinides, which is also a very reasonable choice (again by looking at the substitution tables). The new choice of Lr as the other neighbor of Sc certainly does not seem optimal, but it is a compromise to keep the chemical consistency of the lanthanides and actinides.

The sequence containing Mn and Fe in the old Pettifor table was Re-Mn-Fe-Os. From the substitutional tables it is clear that Mn and Fe have to be neighbors, however the connection to Re and Os is much weaker. Our algorithm preferred to keep the logical sequence of magnetic metals Ni-Co-Fe-Mn. This 'local' organization required a 'global' movement of the Mn and Fe to a different place in the scale.

Finally, N, and to some extent the other non-metals. In the original table, N was between O and Cl. The connection with O is acceptable, but we basically did not find any system where N could be replaced by Cl. In the new ordering, all the halogens are grouped together (which makes a lot of chemical sense), while N is in the sequence B-C-N-P, which is the most logical sequence in view of the substitutional tables. Again, such local reorganizations required global, large movements of elements across the scale.

5. Conclusion and outlook

In conclusion, we performed a statistical study of the possible substitutions of a chemical element A by another B in all known crystal structures. This was possible by using a data mining approach performed on the inorganic crystal structure database. With these data we constructed a function P_{AB} that quantifies the chemical similarity between the elements A and B. We showed that the structure of the periodic table of elements can be reconstructed from a visual inspection of P_{AB}, in a similar manner as in Mendeleev's original work (solely based on statistics, and before the discovery of quantum mechanics).

Having access to a measure of chemical similarity, we were able to propose a mathematical construction for a one-dimensional chemical scale, analogous to the famous Pettifor scale, where similar elements are found in neighboring positions. This way we have obtained a new scale that, while showing a overall structure similar to the original Pettifor scale, corrects it in several aspects and, most importantly, encompasses all available information on the crystal structure of materials, and not only on binary phases.

We believe that our proposed 'modified Pettifor scale' can be of use not only for the representation of structure maps, but also as a tool for both theorists and experimentalists to study possible chemical substitutions in the quest for new materials with tailored properties.

Acknowledgments

MALM acknowledges financial support from the French ANR project ANR-12-BS04-0001-02.

References

[1] Pettifor D 1984 A chemical scale for crystal-structure maps *Solid State Commun.* **51** 31–4
[2] Curtarolo S, Hart G L W, Nardelli M B, Mingo N, Sanvito S and Levy O 2013 The high-throughput highway to computational materials design *Nat. Mater.* **12** 191–201
[3] Hart G L W, Curtarolo S, Massalski T B and Levy O 2013 Comprehensive search for new phases and compounds in binary alloy systems based on platinum-group metals, using a computational first-principles approach *Phys. Rev.* X **3** 041035
[4] Armiento R, Kozinsky B, Fornari M and Ceder G 2011 Screening for high-performance piezoelectrics using high-throughput density functional theory *Phys. Rev.* B **84** 014103
[5] Levy O, Hart G L W and Curtarolo S 2010 Structure maps for hcp metals from first-principles calculations *Phys. Rev.* B **81** 174106
[6] Villars P, Mathis K and Hulliger F 1989 Chapter i — environment classification and structural stability maps *The Structures of Binary Compounds* (*Cohesion and Structure* vol 2) (Amsterdam: North-Holland) pp 1–03
[7] Pettifor D 1986 The structures of binary compounds: i. Phenomenological structure maps *J. Phys. C: Solid State Phys.* **19** 285
[8] Belsky A, Hellenbrandt M, Karen V L and Luksch P 2002 New developments in the inorganic crystal structure database (ICSD): accessibility in support of materials research and design *Acta Crystallographica section* B **58** 364–9
[9] Gražulis S, Daškevič A, Merkys A, Chateigner D, Lutterotti L, Quirós M, Serebryanaya N R, Moeck P, Downs R T and Le Bail A 2012 Crystallography open database (cod): an open-access collection of crystal structures and platform for world-wide collaboration *Nucleic Acids Research* **40** D420–7
[10] Pauling L 1929 The principles determining the structure of complex ionic crystals *J. Am. Chem. Soc.* **51** 1010–26
[11] Hume-Rothery W and Raynor G 1956 The structure of metals and alloys (London: Institute of Metals)
[12] Moeller T 2013 *The Chemistry of the Lanthanides: Pergamon Texts in Inorganic Chemistry* (Oxford: Elsevier)
[13] Morss L R, Edelstein N M, Fuger J and Katz J J 2007 *The Chemistry of the Actinide and Transactinide Elements (Volumes 1–5): Vol. 1* vol 1 (Berlin: Springer Science & Business Media)
[14] Miessler G and Tarr D 2004 *Inorganic Chemistry* (London: Pearson)
[15] Dunning T H Jr, Woon D E, Leiding J and Chen L 2012 The first row anomaly and recoupled pair bonding in the halides of the late p-block elements *Acc. Chem. Res.* **46** 359–68
[16] Leicester H M 1948 Factors which led mendeleev to the periodic law *Chymia* **1** 67–74
[17] Cuthill E and McKee J 1969 Reducing the bandwidth of sparse symmetric matrices *Proceedings of the 1969 24th National Conference* (New York, USA: ACM '69) 157–72 ACM, 1969
[18] Pop P, Matei O and Comes C- A 2014 Reducing the bandwidth of a sparse matrix with a genetic algorithm *Optimization* **63** 1851–76
[19] Scerri E 2008 The role of triads in the evolution of the periodic table: past and present *J. Chem. Educ.* **85** 585