The structure and composition statistics of 6A binary and ternary structures

Alon Hever,1 Corey Oses,2,3 Stefano Curtarolo,2,3 Ohad Levy,2,3,4 and Amir Natan1,3,5,*

1Department of Physical Electronics, Tel-Aviv University, Tel-Aviv 69978, Israel
2Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708, United States
3Center for Materials Genomics, Duke University, Durham, North Carolina 27708, United States
4Department of Physics, NRCN, P.O.Box 9001, Beer-Sheva 84190, Israel
5The Sackler Center for Computational Molecular and Materials Science, Tel-Aviv University, Tel-Aviv 69978, Israel

(Dated: March 14, 2017)

The fundamental principles underlying the arrangement of the elements into solid compounds with an enormous variety of crystal structures are still largely unknown. This study presents a general overview of the structure types appearing in an important subset of the solid compounds, i.e., binary and ternary compounds of the 6A column oxides, sulfides and selenides. It contains an analysis of these compounds, including the prevalence of various structure types, their symmetry properties, compositions, stoichiometries and unit cell sizes. It is found that these compound families include preferred stoichiometries and structure types, some of which have highly specific atomic compositions that may reflect both their specific chemistry and research bias in the available empirical data.

I. INTRODUCTION

The creation of novel materials with optimal properties for diverse applications requires a fundamental understanding of the factors that govern the formation of crystalline solids from various mixtures of the elements. Compounds of the non-metallic elements of column 6A, oxygen, sulfur and selenium, are of particular interest. They serve in a large variety of applications in diverse fields of technology, e.g., chemistry, catalysis, optics, gas sensors, electronics, thermoelectrics, piezoelectrics, topological insulators, spintronics and more. [1–8]. Given the very large number of possibilities, many of the alloy systems of these elements have not been fully investigated, some of them even not at all.

In recent years, high-throughput computational techniques based on ab-initio calculations have emerged as a potential route to bridge these experimental gaps and gain understanding of the governing principles of compound formation [9], leading to the creation of large databases of computational materials data [10, 11]. Yet, these computational approaches are practically limited by the number and size of structures that can be thoroughly analyzed, and fundamental issues that limit the applicability of standard semi-local DFT for non-metallic compounds. The sought-after governing principles are thus still largely unknown. Nevertheless, the considerable body of experimental data that is already available, although by no means complete, is a useful basis for large-scale data analysis. This experimental data is usually presented in compendiums that lack statistical analysis, making it difficult to identify trends and gain insights regarding why some compounds are there and others are not, and also which missing compounds should be expected to be revealed by targeted experimental or theoretical probing. Presenting this data in a structured manner may be conducive for gaining such insights into the essential factors that determine structure formation, and may help to provide material scientists with the necessary foundation for rational materials design.

Analyses recently carried out for the intermetallic binaries [12] and ternaries [13] have uncovered interesting details regarding their distributions among the different Bravais lattices and an unexpected large prevalence of unique structure types. Here we extend the analysis to the binary and ternary compounds of the 6A non-metals and discuss trends, as well as special phenomena, across these families of oxides, sulfides and selenides. This analysis reveals the following interesting observations:

- Considerable overlap exists between the sulfides and selenides, of about a third of the total number of structure types presented in both compound families. In contrast, the overlap between the oxides and the other two families is rather small.

- The majority of structure types in all of these compound families are represented by only one compound in the experimental database.

- High symmetry lattices, e.g. the orthorhombic face centered, orthorhombic body centered and cubic lattices are relatively rare among these compounds. This reflects the spatial arrangement of the compound forming orbitals of the 6A non-metals, whose chemistry does not favor these structures.

- Oxide compounds tend to have higher oxygen content, for both binaries and ternaries, than the sulfur and selenium content in the corresponding compounds of these elements.

- The prevalence of different compound stoichiometries in the sulfide and selenide families is very similar to each other but different from that of the oxides. Some stoichiometries are abundant in the oxides but are almost absent in the sulfides or the
TABLE I. Data extraction numerical summary.

<table>
<thead>
<tr>
<th></th>
<th>compounds</th>
<th>unique compounds</th>
<th>structure types</th>
</tr>
</thead>
<tbody>
<tr>
<td>total</td>
<td>88,373</td>
<td>50,294</td>
<td>13,324</td>
</tr>
<tr>
<td>unary</td>
<td>1752</td>
<td>499</td>
<td>197</td>
</tr>
<tr>
<td>binary</td>
<td>27,487</td>
<td>10,122</td>
<td>1,962</td>
</tr>
<tr>
<td>binary oxides</td>
<td>3,256</td>
<td>844</td>
<td>538</td>
</tr>
<tr>
<td>binary sulfides</td>
<td>1,685</td>
<td>495</td>
<td>270</td>
</tr>
<tr>
<td>binary selenides</td>
<td>1,050</td>
<td>332</td>
<td>168</td>
</tr>
<tr>
<td>ternary</td>
<td>37,907</td>
<td>23,398</td>
<td>4,409</td>
</tr>
<tr>
<td>ternary oxides</td>
<td>10,350</td>
<td>5,435</td>
<td>2,079</td>
</tr>
<tr>
<td>ternary sulfides</td>
<td>3,190</td>
<td>2,041</td>
<td>784</td>
</tr>
<tr>
<td>ternary selenides</td>
<td>1,786</td>
<td>1,256</td>
<td>521</td>
</tr>
<tr>
<td>quaternary</td>
<td>15,138</td>
<td>11,050</td>
<td>3,855</td>
</tr>
<tr>
<td>5 atoms</td>
<td>4,638</td>
<td>3,899</td>
<td>2,053</td>
</tr>
<tr>
<td>6 atoms</td>
<td>1,219</td>
<td>1,101</td>
<td>682</td>
</tr>
<tr>
<td>7 atoms</td>
<td>212</td>
<td>201</td>
<td>154</td>
</tr>
<tr>
<td>8 atoms</td>
<td>20</td>
<td>20</td>
<td>12</td>
</tr>
</tbody>
</table>

selenides, and vice versa. These discrepancies reflect the uniqueness of oxygen chemistry compared to sulfur and selenium.

- Most of the known compounds have small or medium size unit cells, with the maxima of unit cell distributions obtained for less than 30 atoms per cell. These distributions have very long tails with just one compound with a given unit cell size in most cases. However, notable exceptions, that may indicate preferable arrangements of cluster-based structures, appear near 80 atoms and 200 atoms per unit cell in the binary and ternary distributions, respectively, of the three compound families. Additional peaks occur near 300 and 600 atoms per unit cell for the binary and ternary oxides, respectively.

In the analysis presented here, we adopt the ordering of the elements by Mendeleev numbers as defined by Pettifor [14, 15], and complement it by investigating the crystallographic properties of the experimentally reported compounds. Pettifor maps constructed for these compound families exhibit similar separation between different structure types as the classical Pettifor maps for binary structure types [14, 15]. For some atoichiometries, the structure types show similar patterns in the maps of the three compound families, suggesting that similar atoms tend to form these stoichiometries with all the three elements. Such similarity of patterns is more common between the sulfides and selenides than between either of them and the oxides.

II. DATA METHODOLOGY

The ICSD [16] includes approximately 169,800 entries (as of August 2016). For this study we exclude all entries with partial or random occupation and those that do not have full structure data. The remaining set of structures has been filtered using the AFLOW software [17–24], which uses an error checking protocol to ensure the integrity of each entry. AFLOW generates each structure by appropriately propagating the Wyckoff positions of the specified spacegroup. Those structures that produce inconsistencies, e.g., overlapping atoms or a different stoichiometry than the structure label are ignored. If atoms are detected to be too close (≤ 0.6˚A), alternative standard ITC (International Table of Crystallography) [25] settings of the spacegroup are attempted. These settings define different choices for the cell’s unique axes, possibly causing atoms to overlap if not reported correctly. Overall, these considerations reduce the full set of ICSD entries to a much smaller set of 88,373 “true” compounds. These entries are contained in AFLOW Database [26–29]. For the analysis presented here we identify all the binary and ternary compounds included in this set, 27,487 binary entries and 37,907 ternary entries. From these, we extract all the entries that contain oxygen, sulfur or selenium as one of the components. Of the binaries, we find 3,256 oxides, 1,685 sulfides and 1,050 selenides. 10,530 oxides, 3,190 sulfides and 1,786 selenides are found among the ternaries. Duplicate entries representing different experimental reports of the same compound, i.e., the same elements, stoichiometry, space group and Pearson designation, are then eliminated to obtain a list in which every reported compound is represented by a single entry. This reduces our list of binaries to 844 oxides, 495 sulfides and 332 selenides, and the list of ternaries to 5,435
oxides, 2,041 sulfides and 1,256 selenides. These results are summarized in Table I. Throughout the rest of the paper, we will refer to these sets of binary and ternary compounds. We choose not to discuss multi-component structures with four or more elements since their relative scarcity in the database most probably indicates incomplete experimental data rather than fundamental issues of their chemistry.

In the next stage, we identify unique structure types. Structure types are distinguished by stoichiometry, space group, and Pearson designation, without consideration of the specific elemental composition. This implicit definition of structure type is common in the literature [30, 31], and we use it throughout the manuscript as providing a good balance of clarity and simplicity. However, it should be noted that there are a few rare cases of complex structures where a given structure type under this definition includes a few sub-types (see Figure S1 in the Supporting Information). Examples exist of more complex definitions of structure types, formulated to define similarities between inorganic crystals structures [32].

The binary structure type lists contain 538 oxides, 270 sulfides and 168 selenides. The ternary lists contain 2,079 oxides, 784 sulfides and 521 selenides. This means that 64% of the binary oxides, 55% of the sulfides and 51% of the selenides are distinct structure types. The corresponding ratios for the ternaries are 38% of the oxides, 38% of the sulfides and 41% of the selenides. All the other entries in the compound lists represent compounds of the same structure types populated by different elements. Differently put, this means that there are on average about 1.6 compounds per structure type in the binary oxides, 1.8 in the binary sulfides and 2 in the binary selenides. Among the ternaries, the corresponding numbers are 2.6 compounds per structure type in the oxides, 2.6 in the sulfides and 2.4 in the selenides. These numbers may be compared to the intermetallics, where there are 20,829 compounds of which 2,166, about 10%, are unique structure types [12]. There are about seven compounds per structure type in the binary intermetallics and about nine in the ternaries. The number for binary intermetallics is considerably larger than for ternary oxides, sulfides or selenides. Together with the higher proportion of unique structure types in the latter, this reflects the limits on materials chemistry imposed by the presence of one of those 6A elements.

It should be noted that this structure selection procedure produces lists that partially overlap, i.e., certain structure types may appear in more than one list, since there might be oxide structure types that are also represented among the sulfide or selenide structures, and vice versa. 11% of the oxide binary structure types also appear in the sulfides binary list and 8% are represented in the selenides binary list. 33% of the sulfide binaries are also represented in the selenides list. The total number of binary oxides, sulfides and selenides structure types is 976, which is reduced by 16%, to 818 structure types, by removing all overlaps. The corresponding overlap ratios for the ternaries are 10% for the oxides and sulfides, 6% for the oxides and selenides and 31% for the sulfides and selenides. The total number of entries in the ternary oxides, sulfides, and selenides structure type lists is 3,384, which is reduced to 2,797 structure types by removing all overlaps, a 17% reduction. Therefore, the overlaps between these three compound families are similar for the binaries and ternaries. In both, the overlap between the oxides and the other two families is rather small, whereas the overlap between the sulfides and selenides represents about a third of the total number of structure types.

![FIG. 1. Distributions of the compounds among structure types for binary (inset) and ternary compounds. Oxides are shown in blue, sulfides in yellow and selenides in green.](image)

The sequence of Mendeleev numbers includes all 103 elements, from hydrogen to lawrencium with numbers 1-6 assigned to the noble gases, 2-16 to the alkali metals and alkaline earths, 17-48 to the rare earths and actinides, 49-92 to the metals and metalloids and 93-103 to the non-metals. Of these, noble gases are not present in compounds and artificial elements (metals heavier than uranium) have very few known compounds. We are thus left with 86 elements, of which the above compounds are composed. That means there are about ten times more oxide binaries than element-oxygen combinations, about six times more sulfides than element-sulfur combinations and four times more selenides than element-selenium combinations. Oxides are much more common than sulfides and selenides. The corresponding numbers for the ternaries are much lower. There are about 1.6 times more ternary oxides than two-elements-oxygen ternary possible systems, about 0.6 times less ternary sulfides and about 0.4 times less ternary selenides than the corresponding two-element combinations. The ternaries are relatively quite rare, more so as we progress from oxides to sulfides and then to selenides. A similar analysis of the intermetallic binaries in Reference [12] shows that of the 20,829 intermetallics, 277 are unaries (about three times more than possible metal elements), 6,441 are bi-
TABLE II. Distribution of the oxide, sulfide and selenide compounds and structure types among the 14 Bravais lattices.

<table>
<thead>
<tr>
<th></th>
<th>O</th>
<th>S</th>
<th>Se</th>
<th>O</th>
<th>S</th>
<th>Se</th>
<th>O</th>
<th>S</th>
<th>Se</th>
</tr>
</thead>
<tbody>
<tr>
<td>binary</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>structure types</td>
<td>12</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>compounds per</td>
<td>1.3</td>
<td>1.5</td>
<td>1.6</td>
<td>1.3</td>
<td>1.5</td>
<td>1.6</td>
<td>1.3</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td>structure type</td>
<td>358</td>
<td>79</td>
<td>60</td>
<td>219</td>
<td>56</td>
<td>40</td>
<td>1.7</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>O</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>S</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Se</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>aP</td>
<td>123</td>
<td>82</td>
<td>48</td>
<td>123</td>
<td>82</td>
<td>48</td>
<td>123</td>
<td>82</td>
<td>48</td>
</tr>
<tr>
<td>cP</td>
<td>55</td>
<td>46</td>
<td>21</td>
<td>120</td>
<td>104</td>
<td>49</td>
<td>255</td>
<td>240</td>
<td>71</td>
</tr>
<tr>
<td>hR</td>
<td>46</td>
<td>12</td>
<td>4</td>
<td>46</td>
<td>12</td>
<td>4</td>
<td>46</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>oI</td>
<td>46</td>
<td>12</td>
<td>4</td>
<td>46</td>
<td>12</td>
<td>4</td>
<td>46</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>oF</td>
<td>46</td>
<td>12</td>
<td>4</td>
<td>46</td>
<td>12</td>
<td>4</td>
<td>46</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>oP</td>
<td>46</td>
<td>12</td>
<td>4</td>
<td>46</td>
<td>12</td>
<td>4</td>
<td>46</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>mP</td>
<td>39</td>
<td>24</td>
<td>11</td>
<td>39</td>
<td>24</td>
<td>11</td>
<td>39</td>
<td>24</td>
<td>11</td>
</tr>
<tr>
<td>mF</td>
<td>39</td>
<td>24</td>
<td>11</td>
<td>39</td>
<td>24</td>
<td>11</td>
<td>39</td>
<td>24</td>
<td>11</td>
</tr>
<tr>
<td>mS</td>
<td>68</td>
<td>31</td>
<td>22</td>
<td>68</td>
<td>31</td>
<td>22</td>
<td>68</td>
<td>31</td>
<td>22</td>
</tr>
<tr>
<td>mP</td>
<td>39</td>
<td>24</td>
<td>11</td>
<td>39</td>
<td>24</td>
<td>11</td>
<td>39</td>
<td>24</td>
<td>11</td>
</tr>
</tbody>
</table>

The distribution of the binary and ternary compounds among the corresponding structure types is shown in Figure 1. Detailed data for the most common structure types is presented in Tables S1 – S6 in the Supporting Information.

About 84% of the binary oxide structure types represent a single compound, characterizing the tail end of the binary oxide distribution. They include about 53% of the binary oxide compounds. The most common structure type represents 29 compounds, 3.4% of the oxide compounds list. Among the binary sulfides, 76% of the structure types represent a single compound. They include 41% of the binary sulfide compounds. The most common structure type represents 32 compounds, 6.5% of the sulfide compounds list. Among the binary selenides, 76% of the structure types represent a single compound. They include 39% of the binary selenide compounds. The most common structure type represents 31 compounds, 9.3% of the selenide compounds list.

In all three binary lists the most common structure type is rock salt (NaCl). The binary oxide structure type distribution has a much longer tail than the sulfides and selenides, i.e., more oxide compounds have unique structure types. The most common structure type in these three distributions represents a similar number of compounds but a smaller proportion of the corresponding compounds in the oxides. The middle regions of the distributions are very similar (inset Figure 1). This means that the much larger number of binary oxide compounds, compared to the sulfides and selenides, is expressed at the margin of the distribution, in the long tail of unique compounds.

This discrepancy between the three binary distributions is much less apparent among the ternary compounds. 64% of the ternary oxide structure types represent a single compound. They include 24% of the ternary oxide compounds. The two most common structure types, pyrochlore and perovskite, represent 116 and 115 compounds, respectively, about 2% each of the entire compounds list. Among the ternary sulfides, 70% of the structure types represent a single compound. They include 34% of the ternary sulfide compounds. The most common structure type, delafossite, represents 65 compounds, 4% of the entire compounds list. Among the ternary selenides, 62% of the structure types represent a single compound. They include 26% of the ternary selenide compounds. The most common structure type, again delafossite, represents 51 compounds, 4% of the ternary selenides.

In contrast to the binaries, the larger count of ternary oxides, compared to the sulfides and selenides, is expressed by a thicker middle region of the structure type distribution, whereas the margins have a similar weight in the distributions of the three compound families.
FIG. 2. Number of compounds (a and b) and structure types (c and d) for each Bravais lattice. Binaries are on the left (a and c) and ternaries on the right (b and d).

IV. SYMMETRIES

The distribution of the compounds and structure types among the 14 Bravais lattices is presented in Table II and Figure 2. It is interesting to note that in all six cases (binary and ternary oxides, sulfides and selenides) the distribution is double peaked, with the majority of the compounds belonging to the monoclinic and orthorhombic primitive lattices, and a smaller local maximum at the hexagonal and tetragonal lattices. All distributions exhibit a local minimum for the orthorhombic face and body centered lattices. The high symmetry cubic lattices are also relatively rare. This reflects the complex spatial arrangement of the compound forming electrons of oxygen, sulfur and selenium, which does not favor the high symmetry cubic structures or the densely packed face and body centered orthorhombic structures.

Figure 3 shows a more detailed distribution of the compounds among the different space groups. The binary compounds show a distinct seesaw structure, with a few local peaks near the highest symmetry groups of each crystal system. The corresponding ternary distributions have three sharp peaks in the triclinic, monoclinic and orthorhombic systems, and much smaller peaks in the hexagonal and cubic groups. It is interesting to note that the three compound families, oxides, sulfides and selenides, exhibit distributions of very similar structure. The oxide distributions are the densest, simply due to the existence of more oxide compounds in the database, and become sparser in the sulfide and selenide cases. The compounds of all these families are distributed among a rather limited number of space groups, with most space groups represented by just a single compound or not at all.
V. UNIT CELL SIZE

The distributions of unit cell sizes (i.e., the number of atoms per unit cell) for the six compound families we discuss are shown in Figure 4. All of these distributions have strong dense peaks at small cell sizes and decay sharply at sizes above a few tens of atoms. However, the details of the distributions differ quite significantly from group to group. Among the binaries, the oxides exhibit the highest and widest peak with its maximum of 102 oxide binary compounds located at 12 atoms per cell. 90% of the oxide binaries have less than 108 atoms in the unit cell and 50% of them have less than 24 atoms. The sulfides distribution has a lower and narrower peak of 70 compounds at 8 atoms. The distribution of the selenides has a still lower peak of 60 compounds at 8 atoms. Detailed data for these dense parts of the distributions is tabulated in table S12 the Supporting Information (SI). The oxides distribution exhibits the longest tail of the binaries, with the largest binary oxide unit cell including 576 atoms. The largest binary sulfide and selenide unit cells include 376 and 160 atoms, respectively.

The distributions of the ternary compounds have higher, wider peaks and longer tails than their binary counterparts. The relative differences between the oxide, sulfide and selenide distributions remain similar to the distributions of the binaries. The oxide ternaries exhibit a high and wide peak. Its maximum of 465 compounds is located at 24 atoms per cell, and 90% of the compounds have less than 92 atoms in the unit cell and 50% of the compounds have less than 32 atoms. As in the binary case, the distribution of the ternary sulfides has a lower and narrower peak than the oxides, where the maximum of 190 compounds at 28 atoms and 90% of the compounds have less than 72 atoms in the unit cell. The distribution of the selenides has a still lower and narrower peak, where the corresponding numbers are 130 compounds at 28 atoms and 90% of the compounds having less than
28 atoms in the unit cell. Detailed data for these dense parts of the distributions is shown in table S13 of the SI. The ternary oxides distribution exhibits the longest tail of the three types, with the largest oxide ternary unit cell having 1,080 atoms. The largest ternary sulfide and selenide unit cells have 736 and 756 atoms, respectively.

It should be noted that large unit cells, within the tails of all distributions, tend to have very few representatives, with just one compound with a given unit cell size in most cases. Notable exceptions are local peaks near 80 atoms per unit cell in the binary distributions and near 200 atoms per unit cell in the ternary distributions. The oxide distributions exhibit additional peaks, near 300 atoms per unit cell for the binaries and near 600 atoms per unit cell for the ternaries. These minor peaks may indicate preferable arrangements of cluster-based structures.

VI. STOICHIOMETRIES

The stoichiometries of the binary oxide, sulfide and selenide compounds are examined in Figure 5. We define the binaries as $A_xB_y$, where $B$ is O, S or Se. and the ternaries as $A_xB_yC_z$, with $C$ representing O, S or Se. The prevalence of various compositions in terms of the component ratios, $y/x$ for the binaries and $z/(x+y)$ for the ternaries are shown in the respective frames of Figure 5. The binary distributions have sharp peaks at integer ratios, 1, 2 and 3, with the maximum located at $y/x = 2$ for the oxides and at $y = x$ for the sulfides and selenides. The selenide and sulfide distributions are nearly identical. For the oxides, an additional major peak appears at $y/x = 1.5$, and a few more minor peaks are found at one-half values in all three distributions. There are almost no compounds with ratios larger than three. For $y/x < 1.5$, there are more gaps in the plot (missing stoichiometries) for the oxides compared to the sulfides and selenides, while for $y/x > 1.5$ there are more gaps in the sulfides and selenides, this behavior is shown in detail in Table S10 in the SI. The distributions of the ternaries are, as might be expected, slightly more complex, with maxima at $z/(x+y) = 1.5$ for the oxides and $z = y+x$ for the sulfides and selenides. The major peaks still appear at integer and half integer values, but with more minor peaks at intermediate values. The ternary selenide and sulfide distributions are also nearly identical, and there are almost no compounds with ratios larger than three in the oxides or larger than two in the sulfides and selenides.

Figure 6 shows the abundance of different stoichiometries of the ternary oxide, sulfide and selenide compounds, $A_xB_yC_z$, where $C$ is O, S or Se, respectively. The biggest circle in each diagram denotes the prevalence of the most common stoichiometry, which is $718$ ($x = 1$, $y = 1$, $z = 3$) for oxides, $242$ ($x = 1$, $y = 1$, $z = 2$) for sulfides, and $145$ ($x = 1$, $y = 1$, $z = 2$) for selenides. The smaller circles in each plot are normalized to the corresponding highest prevalence. The plots show the most common stoichiometries for each class of ternaries.

These diagrams highlight the similarities as well as important differences between the three families of compounds. In all three cases, the most common stoichiometries appear on the symmetry axis of the diagram, i.e., at equal concentrations of the $A$ and $B$ components, or very close to it. For the oxides, they are concentrated near 0.5-0.6 fraction of oxygen, representing the $A_1B_1O_2$ and $A_1B_1O_3$ stoichiometries, respectively, and form a very dense cluster with many similar reported stoichiometries of lower prevalence. Outside this cluster, the occurrence of reported compositions drops sharply, and other regions of the diagram are very sparsely populated, in particular near the vertices of the $B$ and $O$ components.

The sulfide and selenide diagrams also exhibit prominent clusters on the $AB$ symmetry axes, but they appear at a lower S or Se concentration of about 0.5, i.e., $A_1B_1C_2$ stoichiometry. They are considerably more spread out and include a significant contribution at the $ABC$ stoichiometry. In both sulfides and selenides, an additional minor cluster appears closer to the $A$ vertex, $A$ few members of this cluster are ternary oxides, reflecting the high electronegativity, and high Mendeleev number (101). The $B$ and $C$ vertex regions are still sparsely populated, but less so than in the oxides case. Overall, the sulfide and selenide diagrams are very similar to each other and different from that of the oxides. They are more spread out and less $AB$ symmetric than the oxide diagram and less tilted towards high $C$-component concentration. This discrepancy may reflect some uniqueness of oxygen chemistry compared to sulfur and selenium, or rather simply reflect the oxygen rich environment in which naturally formed compounds are created in the atmosphere. The number of stoichiometries and the differences in the $C$-component concentration are summarized in Table III.

Another interesting observation is that while some stoichiometries are abundant in the oxides they are almost absent in the sulfides or the selenides. For example, there are 299 compounds with the $A_2B_3O_7$ stoichiometry (ignoring order between $M_A$ and $M_B$), but only two $A_2B_2S_7$ compounds and no $A_2B_2Se_7$ compounds. Also, there are 71 $A_1B_3O_5$ compounds but no $A_1B_3S_9$ and $A_1B_3Se_9$ compounds. On the other hand, there are no $A_1B_11X_{22}$ oxides, but 20 sulfides and 8 selenides. If we require that $M_A > M_B$ - there are no oxides of the $A_3B_2X_2$ stoichiometry, but 25 sulfides and 7 selenides.

VII. COMPOSITION AND MENDELEEV MAPS

The occurrence of each element in the binary and ternary compound lists has been counted and tabulated. The results are described in Figure 7. For the binary oxides a very prominent peak appears at $M = 85$, the Mendeleev number of silicon. It represents the 185 different silicon oxide structures types (s.t.) reported in the ICSD database for just a single stoichiometry, SiO$_2$. Smaller peaks appear for $M = 51$ (titanium, 42 s.t., 14
FIG. 4. Unit cell size distribution for all compounds. Binaries are on the left (a, c and e) and ternaries on the right (b, d and f). Oxides are at the top (a and b), sulfides in the middle (c and d) and selenides at the bottom (e and f). The insets show the compounds with up to 50 atoms per unit cell in each case.

stoichiometries, leading stoichiometry is TiO$_2$ with 14 s.t., $M = 54$ (vanadium, 42 s.t., 18 stoichiometries, lead-
FIG. 5. Distribution of the stoichiometric ratios between oxygen (top; a and b), sulfur (middle; c and d) and selenium (bottom; e and f) and the other components in their respective binary (left; a, b and c) and ternary (right; b, d and f) compounds and structure types.

ing stoichiometry is VO$_2$ with 10 s.t.), $M = 56$ (tungsten, 24 s.t., 9 stoichiometries, leading stoichiometry is WO$_3$
FIG. 6. Prevalence of stoichiometries among the ternary (a) oxide, (b) sulfide and (c) selenide compounds. (d) shows, for reference, all the possible stoichiometries with up to 12 atoms of each component per unit cell. In each figure, the smaller circles are normalized to the biggest one, which denotes the highest prevalence, i.e., 718 for oxides, 242 for sulfides, and 145 for selenides. The $x$ and $y$ axes denote the atomic fractions in the ternaries $A_{(1-x)}B_xC_y$, where $C$ is O, S or Se, respectively. $A$ and $B$ are ordered by Mendeleev number where $M_A > M_B$.

<table>
<thead>
<tr>
<th></th>
<th>Oxygen</th>
<th>Sulfur</th>
<th>Selenium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of stoichiometries</td>
<td>585</td>
<td>282</td>
<td>206</td>
</tr>
<tr>
<td>$C$-rich stoichiometries ratio</td>
<td>0.85</td>
<td>0.67</td>
<td>0.66</td>
</tr>
<tr>
<td>$C$-rich compound ratio</td>
<td>0.92</td>
<td>0.77</td>
<td>0.73</td>
</tr>
</tbody>
</table>

TABLE III. Ternary stoichiometry data - $A_xB_yC_z$. “$C$-rich” refers to stoichiometries where $z > x + y$.

with 13 s.t.), and $M = 45$ (uranium, 22 s.t., 9 stoichiometries, leading stoichiometry is $\text{UO}_2$ and $\text{U}_3\text{O}_8$ with 6 s.t. each). Unlike the silicon peak which is composed of a single stoichiometry, the other leading peaks evidently include multiple stoichiometries, reflecting the different chemistry of those elements. These differences also carry over into the ternary oxide compounds involving those elements. For example, the stoichiometry distribution of silicon oxide ternaries is more tilted towards the silicon poor compounds compared to the corresponding distributions of vanadium and titanium ternary oxides, as is shown in Figure S3 in the SI.

The distribution of the sulfides is generally much lower than that of the oxides, due to the much smaller total number of known binaries, but is also more uniformly structured. It has one major peak for $M = 76$ (zinc, 40 s.t., 2 stoichiometries, leading stoichiometry is $\text{ZnS}$ with 39 s.t.), and quite a few smaller ones such as $M = 51$ (ti-
tanium, 16 s.t., 5 stoichiometries, leading stoichiometry is TiS$_2$ with 9 s.t.), $M = 61$ (iron, 18 s.t., 5 stoichiometries, leading stoichiometry is FeS with 6 s.t.), $M = 67$ (nickel, 16 s.t., 6 stoichiometries, leading stoichiometry is NiS$_2$ with 8 s.t.), $M = 90$ (phosphorus, 13 s.t., 8 stoichiometries, leading stoichiometries are P$_2$S$_7$, P$_4$S$_9$, P$_4$S$_6$, P$_4$S$_5$ and P$_4$S$_3$ with 2 s.t. each). The $M = 8−33$ region also exhibits a minor concentration of participating elements. The selenides distribution is yet smaller than that of the sulfides, and even more uniform. Several peaks appear, $M = 51$ (titanium, 13 s.t., 9 stoichiometries, leading stoichiometry is TiSe with 3 s.t.), $M = 52$ (niobium, 15 s.t., 8 stoichiometries, leading stoichiometry is NbSe$_2$ with 8 s.t.), $M = 53$ (tantalum, 15 s.t., 4 stoichiometries, leading stoichiometry is TaSe$_2$ with 10 s.t.) and $M = 79$ (indium, 14 s.t., 5 stoichiometries, leading stoichiometry is In$_2$Se$_3$ with 6 s.t.). All distributions cover most of the elements except two obvious gaps, one at $M < 9$, which includes the noble gases and the two heaviest alkali metals, cesium and francium, and another at $34 ≤ M ≤ 42$ which represents the heavy actinides. Another gap appears in the sulfide and selenide distributions at $91 ≤ M ≤ 97$, which reflects the rarity of polonium and astatine compounds and shows that the elements of the 6A column, except oxygen, do not coexist, in the known compounds, with each other or with the heavier halogen iodine.

The element occurrence distributions for the oxide, sulfide and selenide ternaries exhibit greater similarity than the corresponding binary distributions. The most apparent difference, however, is the identity of the most common component, which is sulfur, $M = 90$, in the oxides, but oxygen itself, $M = 101$, in the sulfides and selenides. The sulfide and selenide distributions are almost the same, except for generally lower numbers in the selenides (due to the smaller total number of compounds) and an apparent lower participation of the lanthanides $M = 17−35$.

Mendelev maps for the ternaries are shown in Figures 8-10. Figure 8 shows the cumulated maps for all stoichiometries reported for the respective ternary family. They reflect the same major gaps as the binary distributions. The maps show that most of the reported compositions are represented by one or two compounds with just a few hotspots that include up to 20 compounds in the oxides and 10 compounds in the sulfides and selenides. The oxides map is obviously denser, reflecting the much richer chemistry of the oxides compared to the other two elements. The chemistry becomes more constrained as we proceed down the periodic table column from oxygen to sulfur and then to selenium.

Next, we examine maps of specific stoichiometries. Maps of a few notable oxide stoichiometries and their leading structure types are shown in Figure 9. These maps reflect the dominant features of the full oxide ternaries map (Figure 8), but with significant new additional gaps of absent compounds. These gaps are naturally wider for less prevalent stoichiometries, i.e., the map of the most prevalent stoichiometry, $A_1B_1O_3$, is denser than the three other maps in Figure 9. Different structure types in all stoichiometries tend to accumulate at well defined regions of the map. The separation between them is not perfect, but is similar to that exhibited by the classical Pettifor maps for binary structure types [14, 15]. A similar picture is obtained for the sulfide and selenide structure types, Figure 10, which populate maps that are, as expected, more sparse, but show similar degree of separation between different structures. It is interesting to note that the maps of, e.g., the ($A_1B_2C_4$, $C = O, S, Se$) show similar patterns in the map for oxides (Figure 9) and sulfides and selenides (Figure 10) — suggesting that similar atoms tend to form this stoichiometry in all the three elements. In the same manner, the 2:1:1 stoichiometry $A_2B_1C_3$...
FIG. 8. Mendeleev maps of ternary (a) oxide \( A_x B_y O_z \), (b) sulfide \( A_x B_y S_z \) and (c) selenide \( A_x B_y Se_z \) compounds. It is assumed that \( x \geq y \) with the \( x \)-axis indicating \( M_A \) and the \( y \)-axis \( M_B \). If the stoichiometry is such that \( x = y \), the compound is counted as \( 0.5A_xB_yO_z + 0.5B_xA_yO_z \).

VIII. SUMMARY

We present a comprehensive analysis of the statistics of the binary and ternary compounds of oxygen, sulfur and selenium. This analysis and the visualization tools presented here are valuable to finding trends as well as exceptions and peculiar phenomena.

Oxygen has a higher electronegativity (3.44) than sulfur (2.58) and selenium (2.55), which are similar to each other. Therefore, one can expect that oxygen will form compounds with a stronger ionic character. In addition, oxygen is by far more abundant than sulfur and selenium, leading to many more reported oxygen compounds. These important differences are reflected in our analysis by the significantly larger fraction of oxygen rich compounds compared to those that are sulfur or selenium rich. Structure type classification also shows that there is little overlap between the oxygen structure types to sulfur or selenium structure types, while sulfur and selenium present a much higher overlap. Comparison of the three different 6A elements binary and ternary compounds shows significant differences but also some similarities in unit cell size and symmetry distributions among the various Bravais lattices and their corresponding space groups. In particular, the majority of structure types in all three families have a few or single compound realizations.

An important observation is the existence of different
gaps (missing stoichiometries) in the stoichiometry distribution of the oxide binary compounds compared to the sulfides and selenides. Specific elements tend to present very different stoichiometry distributions, for example, silicon forms only one oxide stoichiometry (SiO$_2$) while transition metals such as titanium and vanadium present 14 and 18 different stoichiometries respectively. These differences clearly reflect the different chemistry of those elements, while the large number of reported SiO$_2$ structures might reflect research bias into silicon compounds.

A Mendeleev analysis of the common structure types of these families shows accumulation of different structures at well defined regions of their respective maps, similar to the well-known Pettifor maps of binary structure types. Furthermore, at least for some of the stoichiometries, similarity of the maps for a given stoichiometry is demonstrated across all three elements. These maps should therefore prove useful for predictive purposes regarding the existence of yet unknown compounds of the corresponding structure types.

ACKNOWLEDGEMENTS

AN acknowledges financial support from the Israeli National Nanotechnology Initiative (INNI, FTA project). AN and OL acknowledge financial support from the Pazi foundation. CO acknowledges support from the National
Science Foundation Graduate Research Fellowship under Grant No. DGF1106401. SC and CO acknowledge support by DOD-ONR (N00014-17-1-2090).


