CHEMISTRY OF MATERIALS

Article

pubs.acs.org/cm

Statistical Analysis of Coordination Environments in Oxides

David Waroquiers,[†] Xavier Gonze,[†] Gian-Marco Rignanese,[†] Cathrin Welker-Nieuwoudt,[‡] Frank Rosowski,[¶,§] Michael Göbel,[∥] Stephan Schenk,[⊥] Peter Degelmann,[⊥] Rute André,[#] Robert Glaum,[▽] and Geoffroy Hautier*[,†]

[†]Institute of Condensed Matter and Nanosciences, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

[‡]BASF SE, Catalyst Division, Carl-Bosch-Str. 38, 67056 Ludwigshafen, Germany

[¶]BasCat - UniCat BASF Joint Lab, Technische Universität Berlin, 10623 Berlin, Germany

[§]BASF SE, Chemicals Research and Engineering, Carl-Bosch-Str. 38, 67056 Ludwigshafen, Germany

[∥]BASF SE, Process Research and Chemical Engineering - New Technologies, Carl-Bosch-Str. 38, 67056 Ludwigshafen, Germany

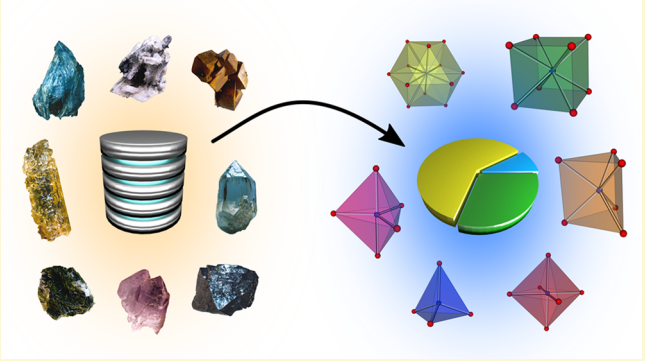
[⊥]BASF SE, Digitalization in R&D, Carl-Bosch-Str. 38, 67056 Ludwigshafen, Germany

[#]BASF SE, Advanced Materials & Systems Research, Formulation of Biological Actives, Carl-Bosch-Str. 38, 67056 Ludwigshafen, Germany

[▽]Institut für Anorganische Chemie, Rheinische Friedrich-Wilhelms-Universität, 53113 Bonn, Germany

Supporting Information

ABSTRACT: Coordination or local environments (e.g., tetrahedra and octahedra) are powerful descriptors of the crystalline structure of materials. These structural descriptors are essential to the understanding of crystal chemistry and the design of new materials. However, extensive statistics on the occurrence of local environment are not available even on common chemistries such as oxides. Here, we present the first large-scale statistical analysis of the coordination environments of cations in oxides using a large set of experimentally observed compounds (about 8000). Using a newly developed method, we provide the distribution of local environment for each cation in oxides. We discuss our results highlighting previously known trends and unexpected coordination environments, as well as compounds presenting very rare coordinations. Our work complements the know-how of the solid state chemist with a statistically sound analysis and paves the way for further data mining efforts linking, for instance, coordination environments to materials properties.



INTRODUCTION

The search for and development of new materials with specific properties are of utter importance in the design of innovative or improved devices and components. The process of materials discovery can be substantially accelerated using computer-aided analysis techniques such as data-mining and machine-learning.[1] Recently, there has been an important growth of available materials data collections relying both on experiments (e.g., Inorganic Crystal Structure Database (ICSD),[2] Pearson's Crystal Data,[3] and Crystallography Open Database[4,5]) and on computations (e.g., the MaterialsProject,[6,7] Aflowlib,[8,9] the NoMaD repository,[10] and OQMD[11,12]). This move toward more structured materials databases is opening the possibility to use knowledge discovery techniques to explore them and reveal structure−property relationships that ultimately will speed up the materials design process. Several groups have already started to use such materials informatics approaches based on both experimental and computed data.[13−23]

The discovery of structure−property relationship requires the development of powerful and simple ways to describe crystal structures. In solid state chemistry and materials science, this is often achieved through *coordination* or *local environments* (e.g., cations coordinated by anions). These clearly simplify the representation of crystal structures. For instance, quartz can be depicted as a network of corner-sharing $SiO_{4/2}$ tetrahedra while ABO_3 perovskites can be described as a network of corner-sharing $BO_{6/2}$ octahedra with the A atoms occupying the empty space between the octahedra and being hence typically 12-coordinated.

Such coordination environments are most commonly used for visualizing and describing crystal structures, but they also directly influence materials properties. For instance, the tetrahedral or

Received: July 3, 2017
Revised: September 5, 2017
Published: September 6, 2017

ACS Publications © 2017 American Chemical Society

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346−8360

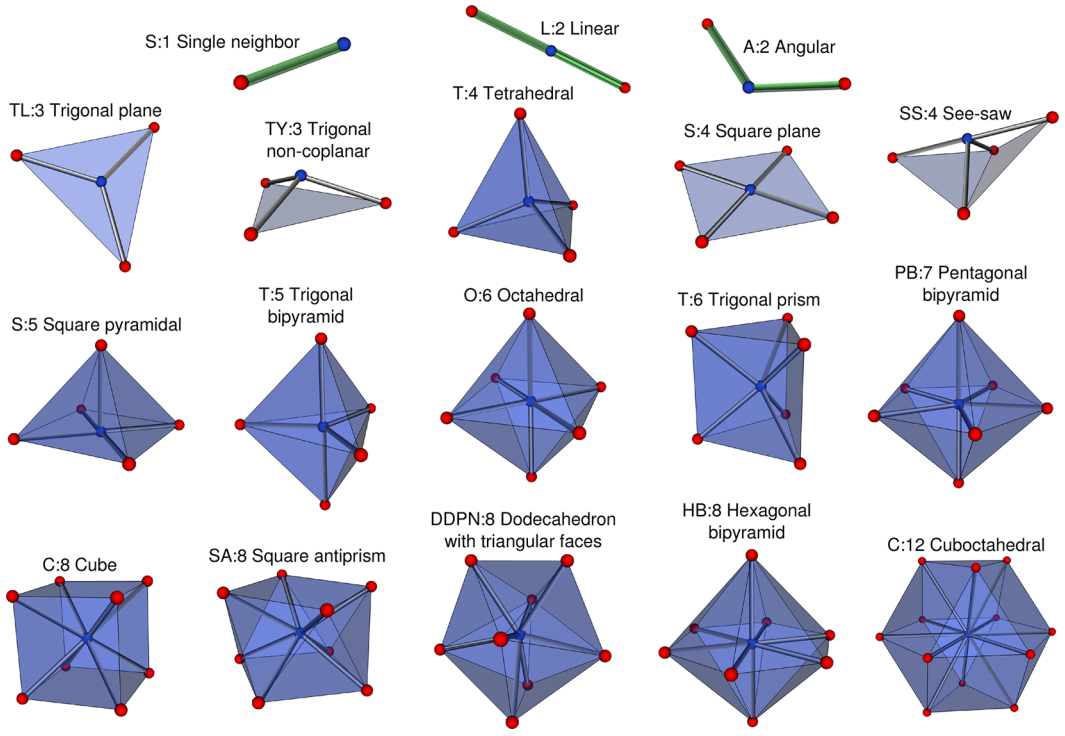


Figure 1. Most commonly observed coordination environments as well as coordination environments discussed in the present paper. Their symbol (e.g., TL:3, HB:8), name (e.g., trigonal prism, square antiprism), and a schematic representation (the central atom is in blue while the neighbors are in red) are shown.

octahedral coordination of Mn^{2+} influences the formation of polarons that limit electrical conductivity,[24] and the preferred local environment of a multivalent ion has been recently linked to ion mobilities in battery cathodes.[25] Vanadyl(IV) groups $(VO)^{2+}$ with 4 or 5 further oxygen atoms coordinated to the V^{4+} ion and linkage of these polyhedra are thought to be key structural features in catalyst materials for selective oxidation of hydrocarbons.[26−28] Without any doubt, coordination environments are very effective structural descriptors.

Previous studies have mainly focused on structure types, inferring coordination environments from the type-defining structure prototype,[29,30] or on simple statements of preferential coordination numbers[31] based on very simplistic rules such as the maximum gap rule and the packing factor of the structures.[32,33] Moreover, all the methods used in the above-mentioned studies are very sensitive to small distortions in the structure. As a matter of fact, no systematic and robust way to automatically identify local environments has been developed so far. This impedes the addition of local environment data to crystal structure databases and all the potential data mining studies that could result. Even the simple question of a statistically robust assessment of the distribution of coordination environments in a given chemistry (e.g., prevailing coordination polyhedra in oxides) has not been answered yet. While the experienced solid state chemist often has a sense of the occurrence of local environments for certain ions, there is no database available indicating a rigorously and statistically sound local environment preference for certain ions.

In this work, the coordination environments of about 8000 oxides present in the ICSD (about 3500 different crystal structure types) have been identified using an automatic and robust detection algorithm. From these results, we collect the statistics of occurrence of local environments for each cation in oxides. We present these distributions and discuss them, highlighting chemical trends but also exceptionally rare occurrences of certain coordination environments. To our knowledge, such a statistically sound study on local environment distribution has never been performed and will of be great interest for the large community of scientists working with crystalline solids. In addition, the provided raw data of coordination environments for oxides present in the ICSD provides a first step toward data mining studies linking materials properties (e.g., from computational databases) to cation local environments.

METHODS

Identification of Coordination Environments. The description of a crystalline structure based on coordination environments is a common practice among chemists and crystallographers. The expert eye of a chemist or solid state scientist can easily assert that a given atom is in a given coordination environment for a given crystal structure. However, this is not feasible when large structures are involved or when thousands of structures are investigated in large materials databases. One thus needs to find an automatic way to relate a real (distorted) environment with a model (perfect) environment. This should be performed using not only the most known coordination environments, for example, tetrahedral, octahedral, cubic, but also less common environments. There is a long list of model coordination environments as reported by the International Union of Crystallography (IUCr)[34] and by the International Union of Pure and Applied Chemistry (IUPAC).[35] Figure 1 represents the coordination environments that are most commonly found as well as those discussed in the present paper.

We have developed an automatic and robust tool for automatically identifying the coordination environments of atoms in any material. The software is integrated as a subpackage called ChemEnv in the pymatgen package and is open source.[36] The identification is based on the sole consideration of the geometrical knowledge of the structure, that is, analysis of the coordinates of the atoms in the structure. The general idea of our approach is to compare the distorted local environments present in real solids to the library of perfect environments and to assess the ones

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346−8360

to which the real environments are the "closest". A simplified version of the identification procedure for a given atom can be summarized in three steps (developed further hereafter):

1. An extended list of neighbors of the atoms is obtained through a Voronoï approach similar to what was proposed by O'Keefe.[37]
2. Some of these neighbors are discarded from the list based on a distance parameter, an angle parameter, or any other condition (e.g., if only cation−anion bonds should be considered or if bonds between atoms of the same type should not be considered), leading to a restricted set of neighbors.
3. The Continuous Symmetry Measure (CSM)[38] is used as a measure of similarity to determine which model coordination environment is the closest to this set of neighbors.

The Voronoï analysis[39] allows one to split the space into regions that are closer to one atom than to any other one. In the standard Voronoï approach for determining the neighbors of a given atom X, all the atoms $\{Y_1, ..., Y_n\}$ in the regions that are contiguous to the region of atom X are considered to be bonded to atom X. Inspired by the work from O'Keefe,[37] two additional cutoff parameters can be added. The first parameter is a cutoff that excludes some of the neighbors on the basis of the distance. The second cutoff parameter is based on the solid angle defined by the face common to the two Voronoï regions. A schematic drawing of these two parameters is shown in Figure 2.

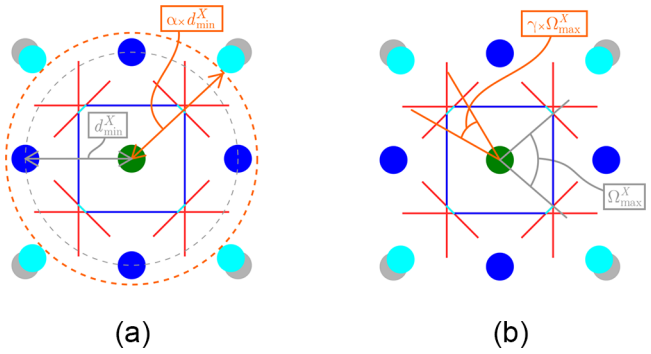


Figure 2. Schematic 2-dimensional view of the distance and angle cut-offs used in the Voronoï procedure for determining the neighbors of a given atom. (a) Distance cutoff α. Any atom that is at a distance larger than α times the distance, $d^X_{\min}$, to the nearest neighbor is not considered as a neighbor. Typical values for α range between 1.05 and 2.0. (b) Angle cutoff γ. For each neighbor Y of a given atom X, a solid angle corresponding to the face common to the Voronoï cells associated with atoms X and Y can be defined. Any neighbor whose solid angle is smaller than γ times the largest solid angle, $\Omega^X_{\max}$, is not considered as a neighbor. Typical values for γ range between 0.05 and 0.8.

In addition to the distance and angle cut-offs, some of the neighbors can be excluded in order to satisfy other conditions. Many of the compounds in inorganic chemistry are ionic. This leads to the separation between anions and cations. Therefore, cations surrounding a given cation are not considered as neighbors. The distinction between a cation and an anion is based on the oxidation state. The latter can be obtained using a bond-valence analysis[40] as implemented in pymatgen.[36] The procedure presented here is similar to other previously published methods such as the *Domains* method.[41]

Once the set of neighbors has been defined, it can be compared to each of the model polyhedra through a measure of similarity. For the latter, we use the concept of Continuous Symmetry Measure (CSM) introduced by Pinsky and Avnir.[38] The CSM can be understood as a *distance to a shape* and can take values between 0.0 (if a given environment is perfect) and 100.0 (if a given environment is very distorted). The environment of the atom is then the model polyhedron for which the similarity is the highest, that is, for which the CSM is the lowest.

It is worth noting that the distance and angle cut-offs used to determine the set of neighbors can strongly influence the environment. While a unique set of reasonable parameters (e.g., $K = 1.4$ and $\gamma = 0.3$) already allows for a fair description of environments in many cases, the environment decision is not very robust to small changes in parameters in certain situations. As an example, for a perfect octahedron which is smoothly distorted by moving away one atom from the central atom, the environment of the atom switches abruptly from the octahedral to the square pyramidal environment when the distance cutoff is exceeded.

In order to ensure robustness, we have established a procedure consisting of analyzing how sensitive the results are to such changes in the cutoff parameters. In practice, several sets of neighbors are obtained for different values of distance/angle cut-offs. The choice of a *strategy* is then used to analyze the results. It is thus possible to use different strategies depending on the user's purpose. This flexibility constitutes a valuable asset for our algorithm. The chosen strategy allows for a broader view of the local environments. The latter can be seen either as one unique coordination environment or as an intermediate between two (or more) coordination environments. They are thus described as a set of model coordination environments, with an assigned fraction or percentage. More technical details about the strategy and, in particular, the way in which the fractions are obtained, are given in SI. Two examples of application of the strategy are shown in Figure 3 for an octahedron in which one atom is moved away and for a triangular bipyramid environment that is smoothly distorted toward a square-pyramid environment.

As shown above and unlike other algorithms and software available, our tool is robust with respect to small distortions of the structures. For example, the Domains method[41] is very sensitive to small changes in the positions of the atoms as a fixed cutoff is used to determine whether an atom is considered as a neighbor or not. In the example shown in Figure 3a, the coordination would change abruptly from 6 to 5 with the latter method while our procedure allows for a smooth and continuous transition between octahedral and square-pyramidal.

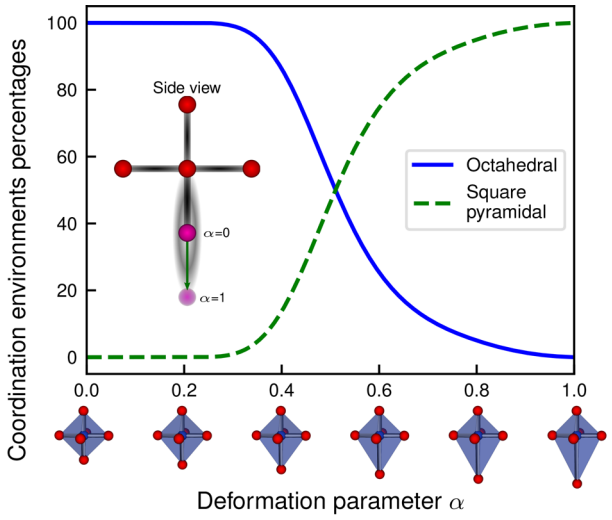
As for the identification of coordination polyhedra for a given set of neighbors, Shevchenko et al.[42] have compared three other methods for the square-pyramidal and triangular bipyramid environments. The three methods are based on (a) the topology of the Voronoï polyhedron, (b) maximizing the volume of the intersection of polyhedra, and (c) the comparison of angular fingerprints. In all three cases, the coordination environment obtained is very sensitive to small distortions. The authors suggest that the coordination environment should be considered not assigned if the three methods do not agree (which happens in more than 23% of the cases in their data set), while our procedure allows us to describe such cases as intermediates between two or more polyhedrons. Furthermore, these methods have been implemented for a limited number of model coordination polyhedrons, whereas our tool is capable of identifying all the environments referenced by the IUPAC and IUCr. Other studies have mainly focused on structure types, inferring coordination environments from the type-defining structure prototype,[29,30] or on simple statements of preferential coordination numbers for each atom (without any percentage or statistics) with little or no concern for the environment's shape.[31] Other schemes based on effective coordination number, close-packed spheres, and cation to anion ratios[32] are also very sensitive to small distortions.

As a matter of fact, while most of the above-mentioned previous approaches are systematic (i.e., can be applied universally for any structure and on a large scale), they do not take into account distortions, preventing the description of a given local environment as an intermediate between two or more coordination polyhedra in the case of highly distorted environments. Our method remedies this problem of robustness and provides further flexibility, allowing one to use different strategies depending on the user's purpose.

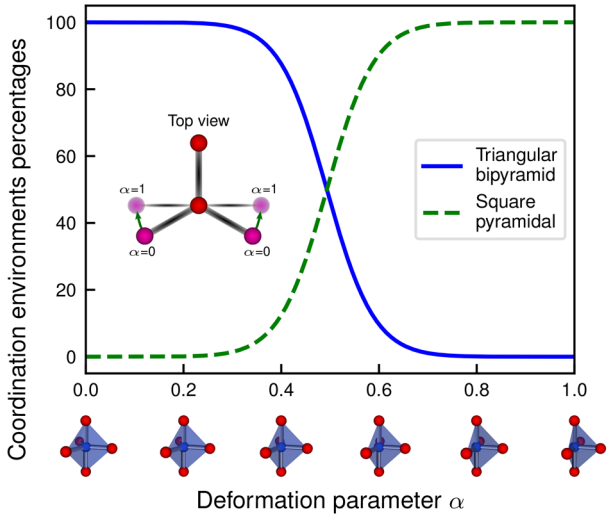
Statistical Analysis. Hereafter, we use the procedure described above to identify the coordination environments for all the cation sites of the oxides present in the Materials Project database and coming from the Inorganic Crystal Structure Database (ICSD). The collected data is then used to perform a statistical analysis of the occurrences of each environment.

Evaluation of the Probability Distributions of the Coordination Environments. The probability $p_{M^{n+}}(C_{env})$ is defined as the probability for a given cation M^{n+} to form the coordination environment C_{env}. These

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346−8360



(a) Deformation of an octahedron.



(b) Deformation of a triangular bipyramid to a square-pyramid.

Figure 3. Illustration of the evolution of the percentages of (a) octahedral and square-pyramidal environments for an octahedron that is smoothly distorted (the bottom atom is moved away to a distance equal to twice the initial distance to the central atom as shown by the green arrow in the inset) and (b) triangular bipyramidal (T:5) and square-pyramidal environments (S:5) during the distortion from T:5 to S:5 (two atoms in the triangular plane are smoothly displaced along the two arcs shown by the green arrows in the inset). In both cases, the deformation is schematically shown in the inset with the initial and final positions respectively given by the values 0 and 1 of the deformation parameter α.

probabilities are estimated using the statistics of the local environments determined for the structures of a given compound database.

$$p_{M^{n+}}(C_{env}) = \frac{\sum_{S \ni M^{n+}} f_{M^{n+},C_{env}}(S)}{N_{S \ni M^{n+}}} \quad (1)$$

where $N_{S \ni M^{n+}}$ is the number of structures containing cation M^{n+}. The fraction $f_{M^{n+},C_{env}}(S)$ of cations M^{n+} in local environment C_{env} to the total number $N_{M^{n+}}(S)$ of cations M^{n+} in structure S is defined as

$$f_{M^{n+},C_{env}}(S) = \frac{N_{M^{n+},C_{env}}(S)}{N_{M^{n+}}(S)} \quad (2)$$

where $N_{M^{n+}}(S)$ is the number of cations M^{n+} in structure S.

Analysis of the Coordination Environment Distributions. Comparisons can be performed between the probability distributions of different cations using statistical tools.

In order to compare the probability distributions of two different cations, a distance between these distributions is needed. One of the popular statistical distances used is based on the Jensen–Shannon[43] divergence $D_{JS}(p_1, p_2)$ or total divergence to the average of two probability distributions p_1 and p_2

$$D_{JS}(p_1 \| p_2) = \frac{1}{2} D_{KL}(p_1 \| m) + \frac{1}{2} D_{KL}(p_2 \| m) \quad (3)$$

where m is the average distribution $\frac{p_1 + p_2}{2}$ and $D_{KL}(q_1 \| q_2)$ is the Kullback–Leibler divergence of q_2 from q_1

$$D_{KL}(q_1 \| q_2) = \sum_i q_1(i) \log_2 \frac{q_1(i)}{q_2(i)} \quad (4)$$

The Jensen–Shannon distance $d_{JS}(p_1 \| p_2)$ between the two distributions p_1 and p_2 is then defined as the square root of the Jensen–Shannon divergence

$$d_{JS}(p_1 \| p_2) = \sqrt{D_{JS}(p_1 \| p_2)} \quad (5)$$

This statistical distance can take values between 0 and 1. It indicates whether the two distributions are similar (the value is close or equal to 0) or very different (the value is close or equal to 1). It can thus be used to identify cations that are found in similar environments in a rigorous manner by looking at the Jenson–Shannon distance between the probability distributions of their coordination environments. In the Results section, this distance is used to construct *heat map* figures. These heat maps represent the matrix of distances between each pair of considered cations. The actual value of the distance (i.e., the similarity) is represented using a color scale. Such a graphical representation allows us to quickly identify trends and groups of cations forming similar coordination environments.

RESULTS

We have analyzed oxides of the Materials Project database, which originally stem from the ICSD (i.e., discarding any prediction of structure from the analysis). Oxidation states have been obtained from a bond-valence analysis and from data present in the ICSD in the cases where the latter analysis failed. While all the compounds have been reported experimentally, we used their *ab initio* relaxed structure as presented in the Materials Project to avoid possible errors (e.g., wrong crystallographic position assignments) as well as to exclude duplicate structures (e.g., alumina Al_2O_3 in its corundum phase and silica SiO_2 in its α-quartz phase appear, respectively, ∼80 and ∼140 times in the ICSD). We excluded high pressure phases, compounds that are highly unstable (for which the calculated energy above hull[44] is larger than 100 meV/atom), and compounds for which the *ab initio* relaxed structure was unusually different from the experimentally reported one in terms of bond length and unit cell volume.[45,46] In order to facilitate the interpretation of the results, the analysis has been restricted to compounds without hydrogen (as the location or even presence of the latter is not always accurately reported in crystallographic databases) and with only oxygen as anion (excluding oxysulfides, for instance). We also considered only materials without partial occupancies. In total, 7982 oxides were considered in our study, which corresponds to a large representative data set (∼80%) of the structures in the ICSD with similar exclusions.

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346–8360

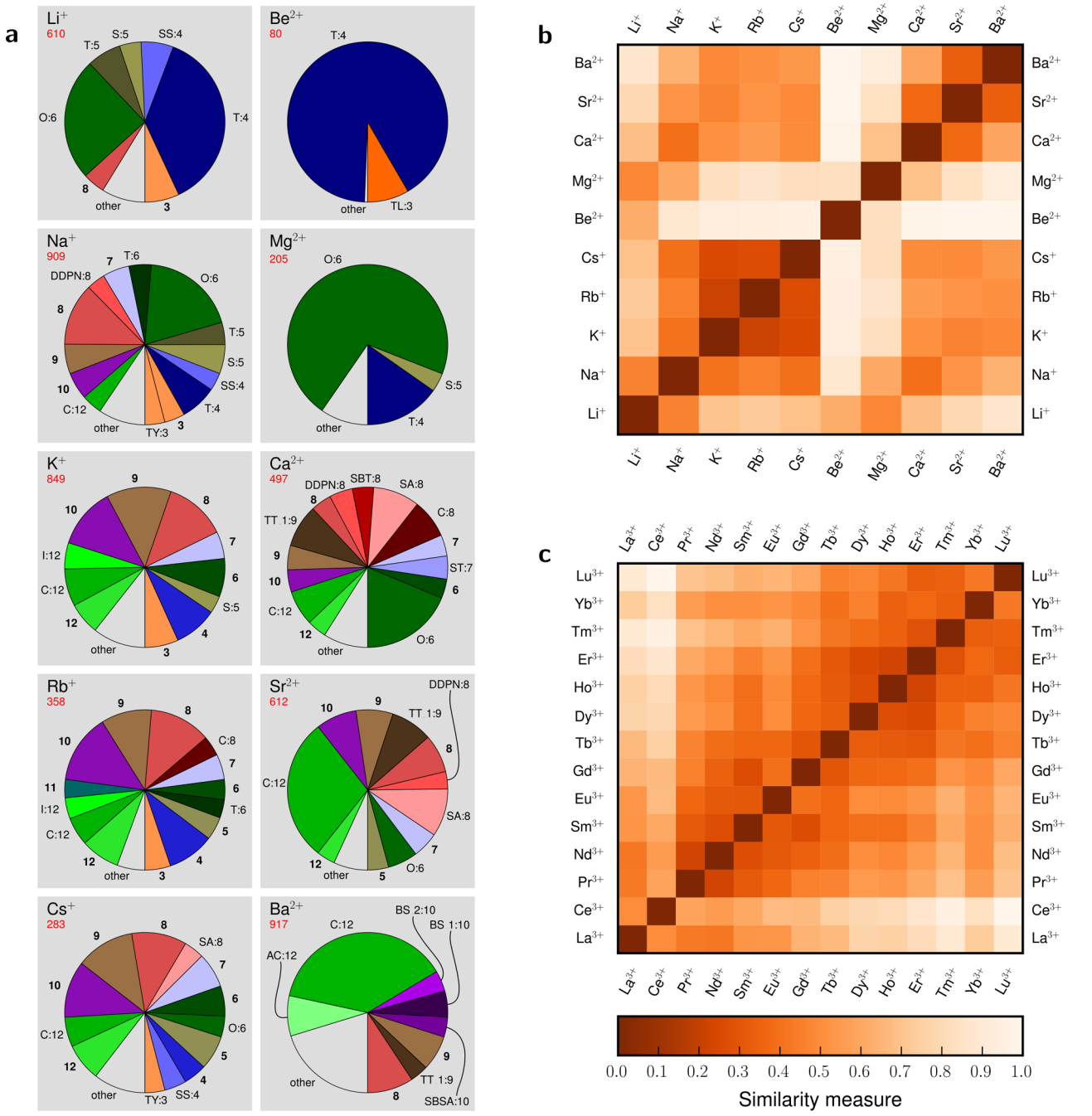


Figure 4. (a) Statistics of occurrences of coordination environments for the alkali and alkaline-earth metal cations. The cation is indicated to the upper left of each pie chart. The number of representatives (i.e., the number of structures in our data set containing at least one occurrence of the cation under consideration) is also indicated just below the cation (in red). (b) Measure of similarity between coordination environment distributions of the alkali and alkaline-earth metals. The Jensen−Shannon (JS) distance[43] between probability distributions has been used to compare the coordination environments of different cations. This allows us to provide a *measure of similarity* between two distributions. Two identical distributions will give a value of 0.0 (dark orange) for the JS distance while two completely different distributions will lead to a value close to 1.0 (white). (c) Measure of similarity between coordination environments distributions of the rare earth cations.

We used our coordination environment detection algorithm presented in the Methods section to assign coordination environments to any cation in every oxide in this data set. The idea of the algorithm is to compare the actual cation coordination with oxygen anions (obtained from a Voronoï analysis) to a set of about 60 perfect polyhedra (e.g., octahedron, tetrahedron, square pyramidal). Our algorithm is robust with respect to distortions and can assign several local environments to a given cation. The full list of coordination environment assignments to compounds is directly accessible in SI. This data is used to provide a statistical distribution of the local environments for each cation, which we present here as pie charts. Cations with rare oxidation states (that are observed in less than 15 structures in our data set) are not shown in the pie charts (results and full statistics for all the cations are provided in SI). Each local environment is identified by a shorter name consisting of a series of 1 to 4 capital letters followed by a colon then by the coordination number, CN (e.g., T:4 for tetrahedron, O:6 for octahedron, L:2 for linear, C:12 for

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346−8360

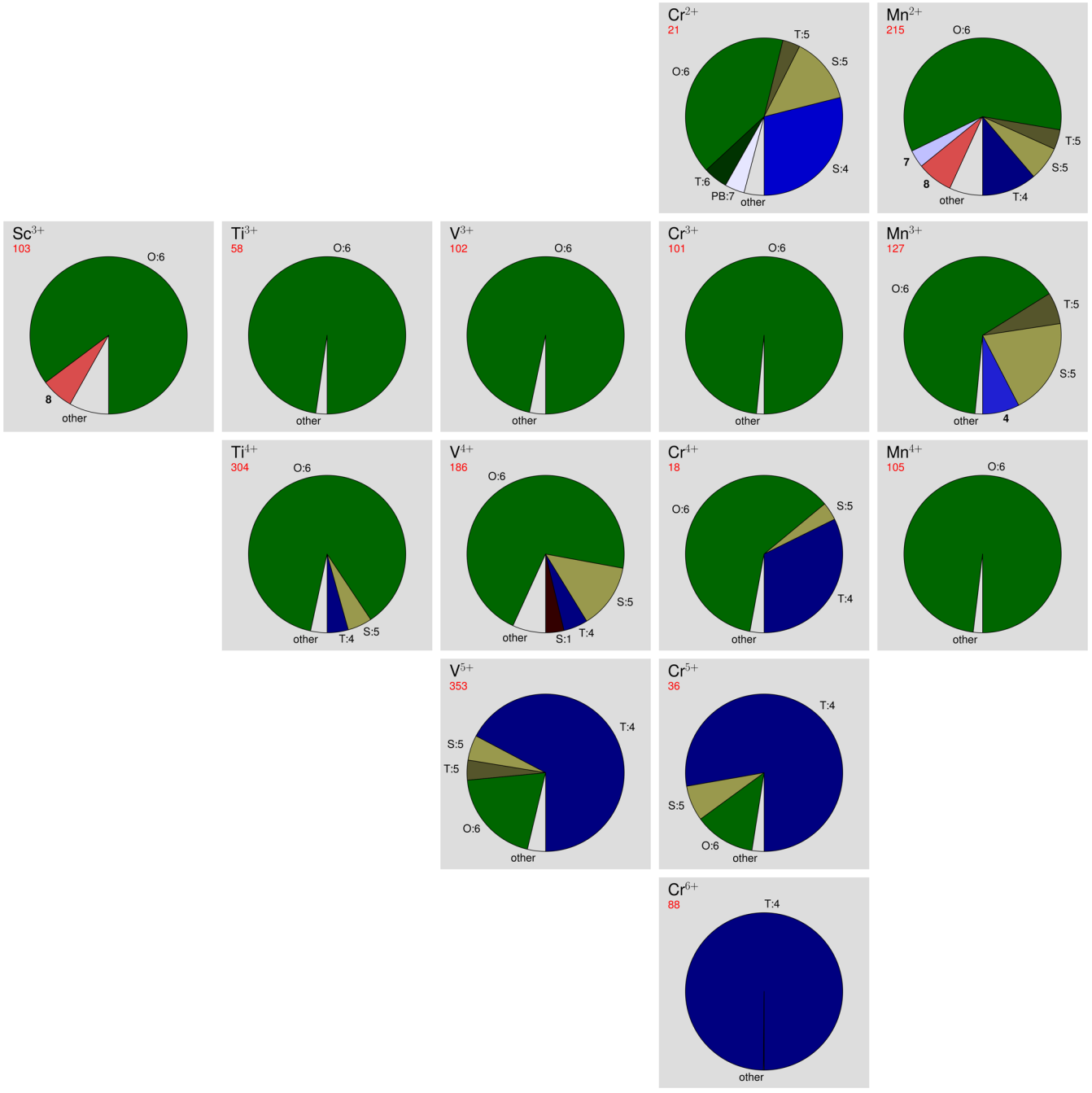


Figure 5. Statistics of occurrences of coordination environments for 3d transition metal cations, groups III to VII. See Figure 4 for a detailed explanation of content of the figure.

cuboctahedron, see SI for a complete list of reported environments).

For visual effectiveness, only the most likely coordination environments (with a probability larger than 3.5%) are explicitly given, while the remaining ones are gathered either by coordination numbers or in one single group (*other*). The full statistics is provided in SI, as well as all the data about the coordination environments of each compound. In the following sections, we will present our statistics and discuss the results highlighting usual and unusual local environments for each cation. We also highlight trends in the periodic table. For the sake of clarity, we divided our analysis in four sections: alkali and alkaline-earth, transition metals, main group, and rare-earth elements.

Alkali and Alkaline-Earth Metal Cations. The statistics for the alkali and alkaline-earth oxides are provided in Figure 4a. Low oxidation states (+1 and +2) are most commonly adopted by these cations. The number of suboxides (e.g., Cs_4O, Cs_7O, $Cs_{11}O_3$) is very small and thus statistically insignificant for the reasoning adopted in the present study. As the size of the alkali increases when going down in the periodic table, larger local environments are preferred. While Li^+ is found in the majority of the cases in tetrahedral environments, the larger Na^+ already prefers a much more diverse set of environments, from the relatively rare tetrahedral (e.g., Na_2O, $NaAg_3O_2$, Na_3CuO_2) to octahedral (e.g., Na_2CO_3), trigonal prismatic (e.g., $NaCoO_2$), or even hexagonal bipyramidal (e.g., $Na_6V_6O_{11}$). Compounds containing small alkali cations are critical for intercalation

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346−8360

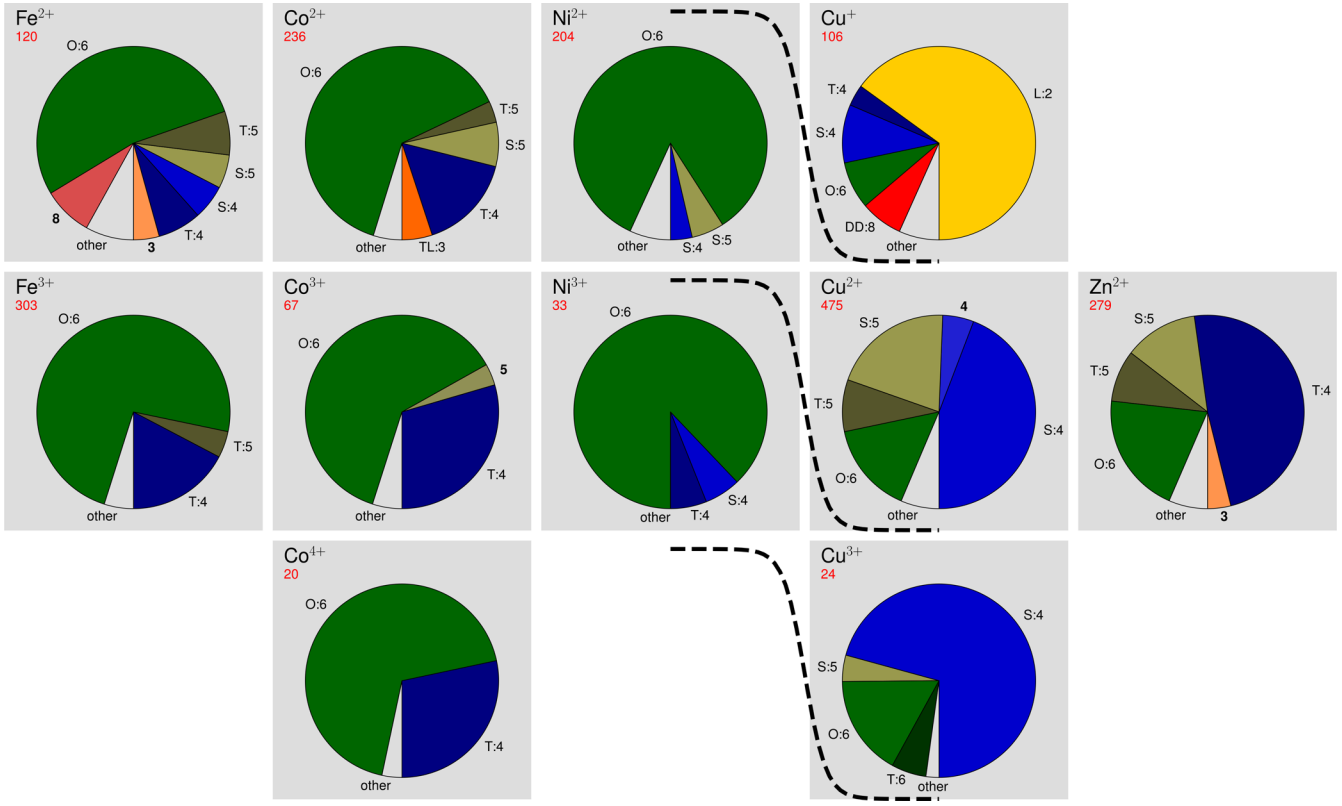


Figure 6. Statistics of occurrences of coordination environments for 3d transition metal cations, groups VIII to XII. See Figure 4 for a detailed explanation of content of the figure. Dashed lines are used as a guide to indicate isovalent species.

batteries. Our analysis identifies a somewhat different local environment behavior for Li^+ and Na^+. Indeed, Na^+ is less prone to fill tetrahedral sites than Li^+. This leads to very different alkali diffusion mechanisms.[47] The larger K^+, Rb^+, and Cs^+ favor mainly highly coordinated (CN > 8) large environments. A similar trend is observed for the alkaline-earth metals with the larger Ca^{2+}, Sr^{2+}, and Ba^{2+} favoring higher coordination environments (e.g., 12-coordinated cuboctahedral environment in perovskite structures such as $CaTiO_3$, $SrFeO_3$ or $BaZrO_3$). The smaller Mg^{2+} strongly favors octahedral environments, and the smallest Be^{2+} is found in the vast majority of the cases in tetrahedral environments with a notable occurrence of trigonal-planar environments (e.g., $SrBe_3O_4$ in which two Be atoms are trigonal-planar and 4 Be atoms are tetrahedral; $K_2Na_4Be_2O_5$ in which all Be atoms are trigonal-planar coordinated).

A more direct comparison of the alkali and alkaline-earth cations can be obtained through an estimate of how comparable cations are in terms of their local environment distribution. Figure 4b compares all alkali and alkaline-earth metals using a heat map representing the Jensen−Shannon distance (described in the Methods section) between all pairs of atoms. The elements in this heat map appear in column order of the periodic table, that is, group 1 first with increasing periods, then group 2 with increasing periods.

Several interesting features appear in Figure 4b. First, there is a clear splitting in the heat map as Be^{2+} is dissimilar to almost any other alkali and alkaline-earth cation. Mg^{2+} is somewhat close to Li^+ and Na^+ but is very dissimilar to all the other cations. These two cations separate the heat map into four regions.

The first region in the lower left part of the heat map shows that the environments of K^+, Rb^+ and Cs^+ are rather similar. This can be easily understood in terms of their sufficiently large ionic radii (ranges for the ionic radii reported by Shannon,[48] 1.37−1.64 Å, 1.52−1.83 Å, and 1.67−1.88 Å, respectively, for K^+, Rb^+, and Cs^+), as well as their similar electronic configuration. Na^+ is also somewhat close to these latter cations, but its lower ionic radius (0.99−1.39 Å[48]) is probably the source of deviation from K^+, Rb^+, and Cs^+.

A similar observation applies for Ca^{2+}, Sr^{2+}, and Ba^{2+} as observed in the upper right region of the heat map. While there is a clear similarity between Ca^{2+} and Sr^{2+}, as well as between Sr^{2+} and Ba^{2+}, Ca^{2+} and Ba^{2+} are more dissimilar. Indeed, both Sr^{2+} and Ba^{2+} show 12-coordinated environments, while very few are observed for Ca^{2+}. Moreover, Ca^{2+} and Sr^{2+} both appear in 8- and 9-coordinated environments in about 35% of the cases, while Ba^{2+} is less often found in these coordinations.

There is also a rather good mapping between Li^+ and Mg^{2+} and between Na^+ and Ca^{2+}. This reflects the well-known *diagonal relationship* and can be understood in terms of the similar ionic radii of Li^+ and Mg^{2+} and of Na^+ and Ca^{2+} (ranges for the ionic radii reported by Shannon,[48] 0.59−0.92 Å, 0.57−0.89 Å, 0.99−1.39 Å, and 1.00−1.34 Å, respectively, for Li^+, Mg^{2+}, Na^+ and Ca^{2+}). More globally, the upper left region of the heat map (which is the same as the lower right one) shows that the environment distributions of Ca^{2+}, Sr^{2+}, and Ba^{2+} are relatively close to those of Na^+, K^+, Rb^+, and Cs^+. These cations can indeed all be found in very diverse large environments.

Transition Metal Cations. Transition metal elements can take many different oxidation states, in particular those in the middle of the transition metal rows (groups VI, VII, and VIII). For clarity, we have separated the analysis into the 3d, 4d, and 5d transition metals.

3d Transition Metal Cations. Oxidation states +2, +3, and +4 are frequently realized. In any case, all the cations of 3d transition

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346−8360

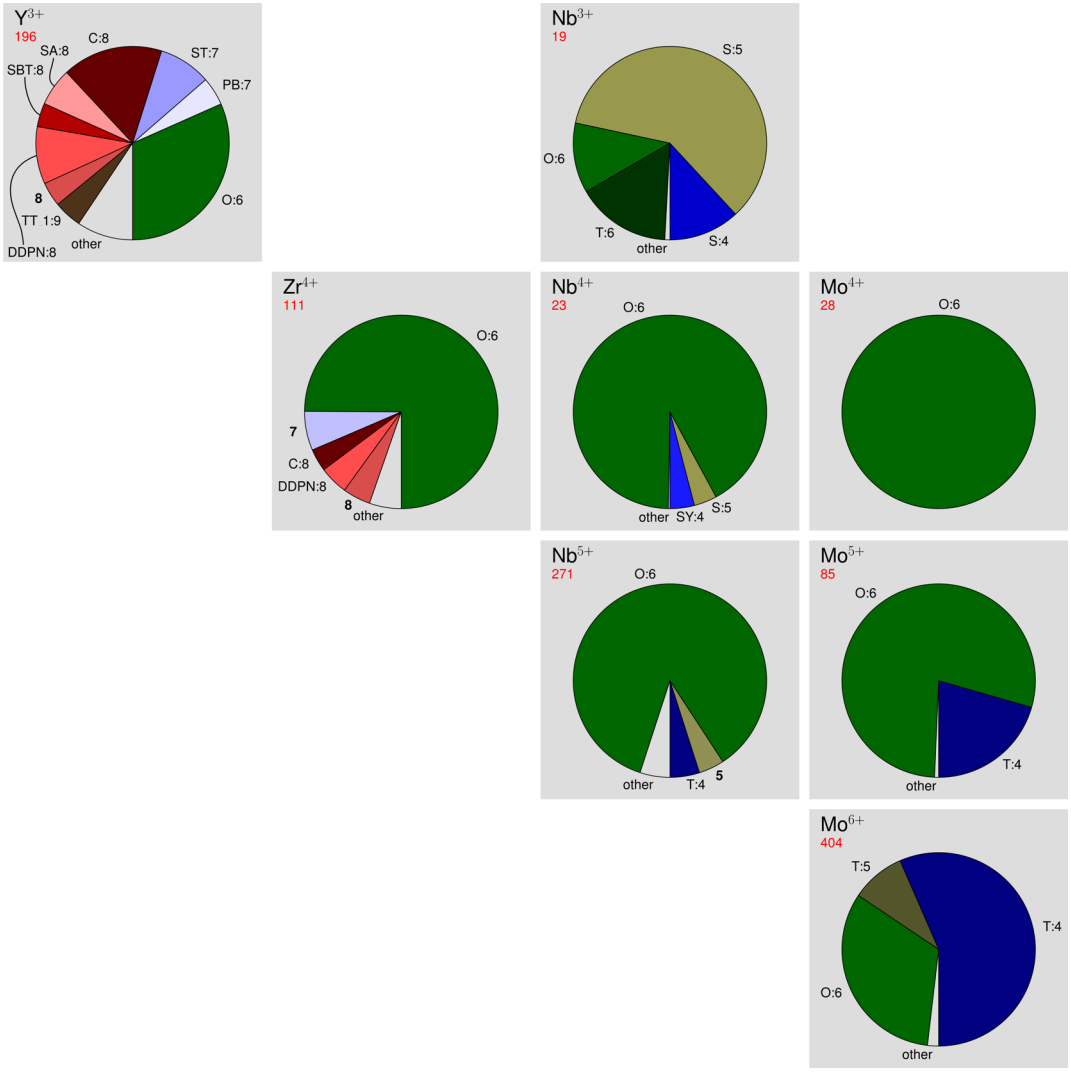


Figure 7. Statistics of occurrences of coordination environments for 4d transition metal cations, groups III to VII. See Figure 4 for a detailed explanation of content of the figure.

metals have an empty 4s orbital, and the 3d levels are partially filled with the outer electrons. The statistics for the 3d transition metal cations are provided in Figures 5 and 6.

The vast majority of transition metal ions adopt octahedral environments (in green in the pie charts). This is especially true for lower atomic number +3 (e.g., Cr^{3+}, V^{3+}, Ti^{3+}, Co^{3+}, and Sc^{3+}) or +2 ions (Co^{2+}, Ni^{2+}, and Fe^{2+}). When d-electrons are stripped away and higher oxidation states are reached, tetrahedral environments become often more favorable. This is the case for the Co^{2+}−Co^{4+}, Fe^{2+}−Fe^{3+}, Ni^{2+}−Ni^{3+}, Cr^{3+}−Cr^{6+}, Ti^{3+}−Ti^{4+}, and V^{3+}−V^{5+} series. The tetrahedral preference is particularly strong for the smallest 3d cations with high oxidation states. V^{5+} shows a large number of compounds with tetrahedral coordinations with some octahedral, square pyramidal, and trigonal bipyramidal coordinations, and Cr^{6+} forms exclusively tetrahedral environments. This trend toward more tetrahedral environments with higher oxidation states is also followed by Mn. It forms octahedral (e.g., MnO) and tetrahedral environments (e.g., $MnCr_2O_4$) when +2, almost exclusively octahedral coordination when oxidized to +4 (e.g., MnO_2), and exclusively tetrahedral environment when highly (+5, + 6, and +7) oxidized (see SI). Only one oxide with tetrahedral Mn^{4+} is found in our data set, namely $MnBi_{12}O_{20}$.

Cu is an outlier among the 3d metals in terms of coordination environment. Monovalent 3d transition metals are not easily accessible except for Cu. This is due to its unique electronic configuration with the 3d states filled before the 4s. Cu^{+} forms in the vast majority of the cases linear environments (e.g., $LaCuO_2$). Cu^{2+} and Cu^{3+} do not show a strong preference for either tetrahedral or octahedral environments: Cu^{2+} (e.g., $CaCuO_2$) and Cu^{3+} (e.g., $KCuO_2$) favor square planar environments.

Jahn−Teller distortions are known to strongly affect transition metals crystal chemistry of d^4 and d^9 configurations. These distortions will push perfect octahedral environments toward square pyramidal or square environments. Besides Cu^{2+} which is a d^9 ion that we already discussed, our data is consistent with the presence of strong Jahn−Teller effects in d^4 cations such as Mn^{3+} and Cr^{2+}. Mn^{3+} can appear in perfect octahedra (e.g., the $REMnO_3$ perovskites with RE a rare-earth cation) but also commonly in distorted environments such as square pyramidal (e.g., $REMn_2O_5$ with RE a rare-earth cation) or square planar (e.g., $CaMn_7O_{12}$). Jahn−Teller distortions are also strong in Cr^{2+} with a large proportion of square planar (e.g., $CaCr(Si_2O_5)_2$) and square pyramidal (e.g., $CaCrP_2O_7$) compounds.

4d Transition Metal Cations. The 4d transition metal cations do not show the same oxidation states as 3d ones and the +2 and

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346−8360

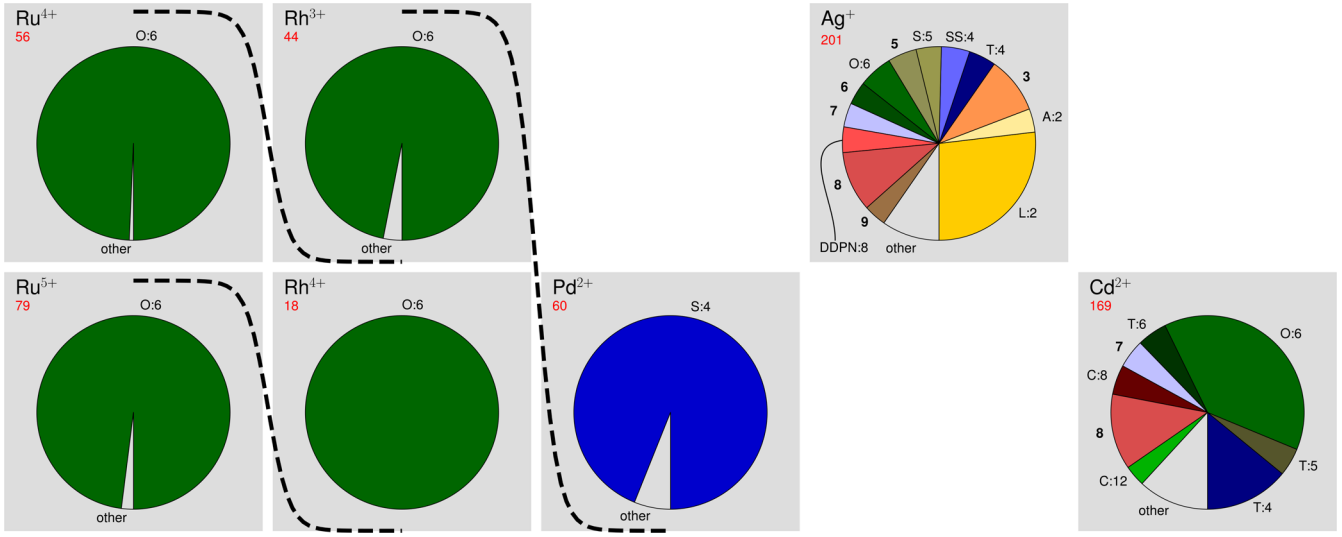


Figure 8. Statistics of occurrences of coordination environments for 4d transition metal cations, groups VIII to XII. See Figure 4 for a detailed explanation of content of the figure. Dashed lines are used as a guide to indicate isovalent species.

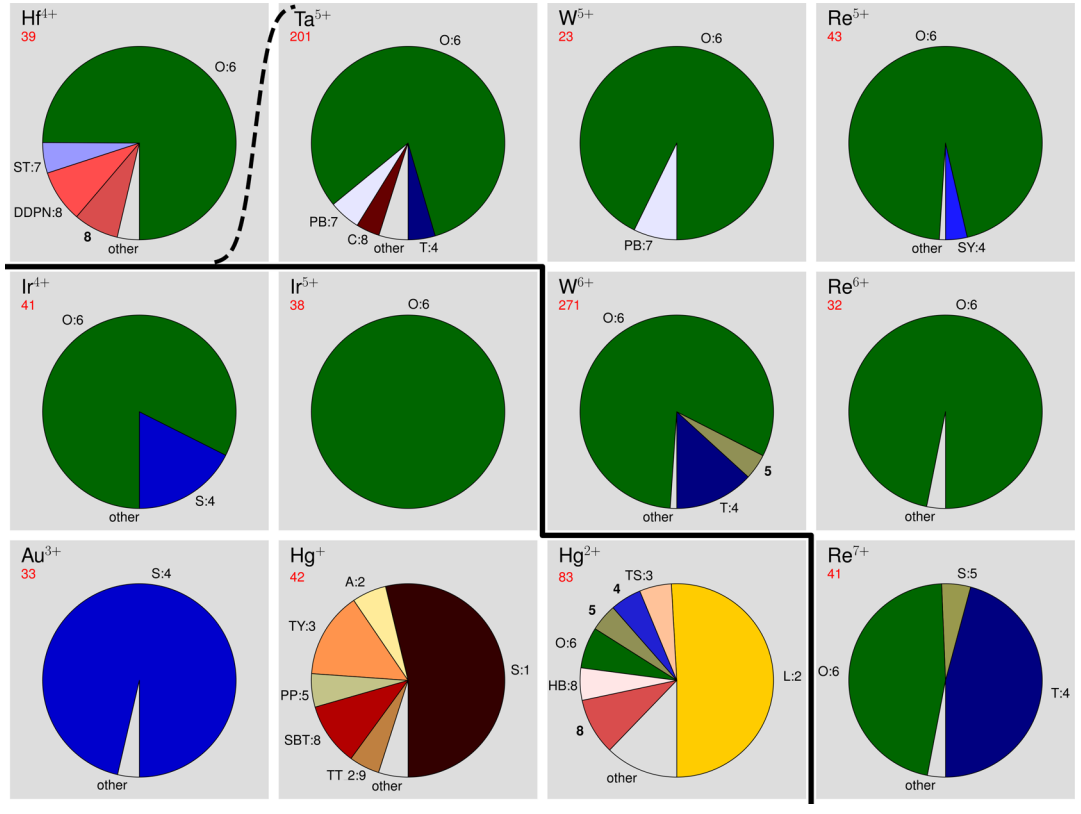


Figure 9. Statistics of occurrences of coordination environments for 5d transition metal cations. See Figure 4 for a detailed explanation of content of the figure. Elements of groups IV to VII (Hf, Ta, W, and Re) are shown in the upper right of the thick line while Ir, Au, and Hg are shown in the lower left. The dashed line is used to separate Hf^{4+} from the isovalent Ta^{5+}, W^{5+}, and Re^{5+}.

+3 oxidation states are less frequent. The statistics for the 4d transition metal cations are provided in Figures 7 and 8.

Similar to 3d elements, the octahedral environment is prevalent for most 4d transition metal cations (Zr^{4+}, Nb^{4+}, Nb^{5+}, Mo^{4+}, Mo^{5+}, Ru^{3+}, Rh^{3+}, Ru^{4+}, Rh^{4+}, Ru^{5+}). Y^{3+} behaves differently than other 4d transition metal cations. In addition to octahedral environments (e.g., Y_2O_3), it forms a large variety of larger environments such as cubic or square antiprismatic. Yttrium crystal chemistry appears closer to that of rare-earth than transition metals. Cd^{2+} also shows an extremely diverse crystal chemistry with many possible environments: octahedral (e.g., $CdSO_4$, V_2CdO_6), tetrahedral (e.g., K_2CdSiO_4), 6-fold trigonal prismatic (e.g., $Mn_3Cd_2O_8$), etc.

Pd^{2+} is found almost exclusively in a square planar environment (e.g., La_2PdO_4), similar to the isoelectronic Cu^{3+}. Similar to Cu^{+} in the 3d elements, Ag^{+} adopts linear environments (e.g., $AgCrO_2$) but also many other environments from triangular

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346−8360

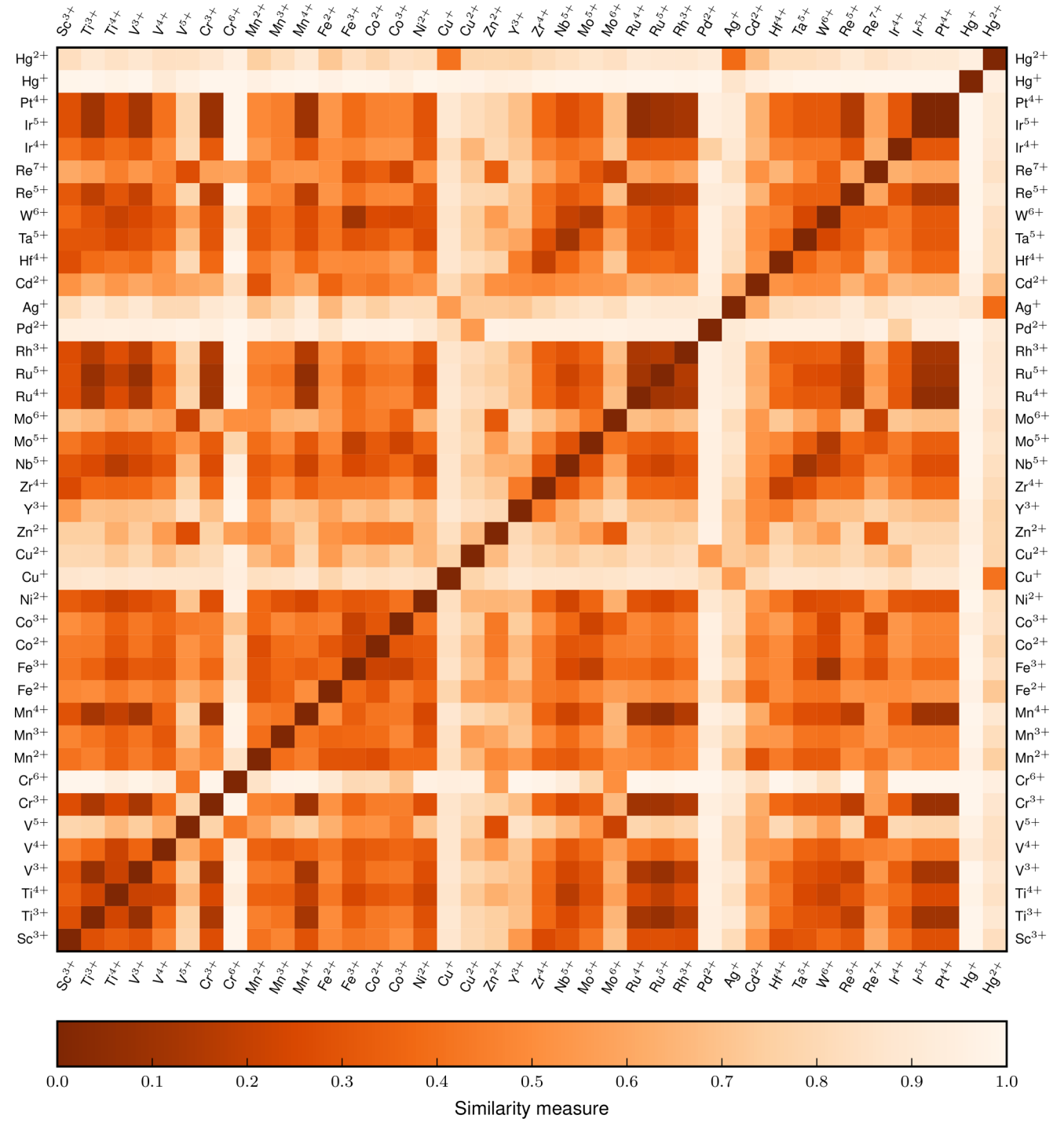


Figure 10. Measure of similarity between coordination environment distributions of the transition metal cations. See Figure 4 for a detailed explanation of the content of the figure.

planar (e.g., Ag_2BiO_3) to tetrahedral (e.g., $ZnAg_2GeO_4$) and octahedral (e.g., Ag_2MoO_4).

The highest oxidation states for Nb (+4 and +5) show mainly octahedral environments with a rare occurrence of tetrahedral Nb^{5+} (e.g., $RENbO_4$ with RE a rare-earth such as Pr or Yb). On the other hand, Nb^{3+} is less common but has a very strong preference for square pyramidal environments by oxygen with metal−metal bonding via the sixth vertex of the octahedral environment (e.g., $TiNb_3O_6$, $Ba_2Nb_5O_9$, $BaNb_8O_{14}$). Molybdenum strongly favors octahedral coordinations for its +4 and +5 oxidation states but favors tetrahedral coordination when +6. This is similar to the isoelectronic Cr^{6+}, but Mo^{6+} offers more diversity and accepts not so rarely to be octahedral (e.g., MoO_3) or even trigonal bipyramidal (e.g., $La_4Cu_3MoO_{12}$), while Cr^{6+} is exclusively tetrahedral.

5d Transition Metal Cations. The oxidation states for 5d transition metal elements of groups IV to IX (Hf to Ir) are +4 or more, while Au can be +1 or +3 and Hg +1 or +2. The statistics for the 5d transition metal cations are provided in Figure 9.

With the exception of Hg, most 5d transition metal cations are commonly observed in octahedral environments. Ir^{4+} shows additionally a preference for square planar environments (e.g.,

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346−8360

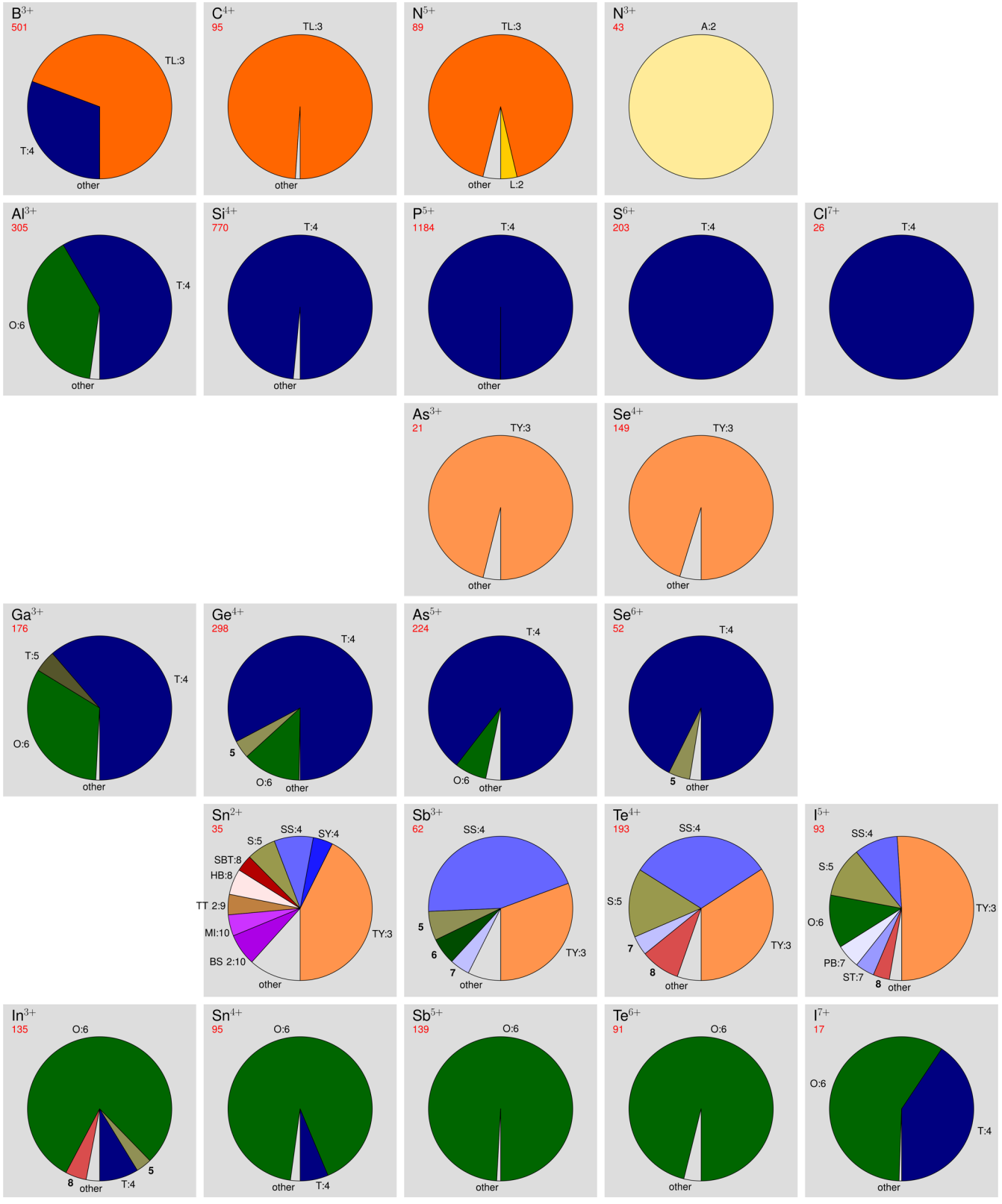


Figure 11. Statistics of occurrences of coordination environments for the main group elements from the 2nd to the 5th period. See Figure 4 for a detailed explanation of content of the figure.

K_4IrO_4, Na_4IrO_4) and Re^{7+} shows both tetrahedral (e.g., $(VO)(ReO_4)_2$ and $AgReO_4$) and octahedral environments (e.g., Li_5ReO_6 and La_3ReO_8). Hf^{4+} forms preferentially in octahedral environments (e.g., HfO_2) but also in higher coordinations such as dodecahedra (e.g., $HfSiO_4$). Ta^{5+} also favors octahedral environments with only rare occurrences of tetrahedral (e.g., $RETaO_4$) or pentagonal bipyramidal environments (e.g., $LaTa_3O_9$).

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346−8360

Mercury is extremely different from the other 5d metals. Hg behaves even very differently from the elements in the same column (Zn and Cd). Hg^{+} shows preferentially a single oxygen neighbor (e.g., Hg_3PO_4, $CrHg_5O_6$). In the case of Hg^{+}, Hg−Hg bonds are very common and the coordination by anions will show just a single neighbor. Hg^{+} also forms triangular non-coplanar environments (e.g., Hg_3AsO_4). Hg^{2+} on the other hand forms a very large proportion of linear environments (e.g., $BaHgO_2$, Hg_3TeO_6), but other environments are possible (e.g., octahedral $HgCrO_4$).

The case of W^{6+} is also interesting. While Mo^{6+} and W^{6+} are in the same column, with the same number of d electrons, their slightly different ionic radii lead to quite different environment behavior. Because Mo^{6+} polarizes O atoms much more than W^{6+}, it strongly favors tetrahedral environments, while W^{6+} mainly forms octahedral environments (e.g., WO_3, $PbWO_4$). Tetrahedral environments are not uncommon for W^{6+} (e.g., K_2WO_4, $Al_2(WO_4)_3$) but not as prevalent as for Mo^{6+}. This illustrates that effects more subtle than ionic size and number of d electrons affect local environment preferences.

Comparison of Transition Metal Cations. As for the alkali and alkaline-earth metal cations, it is possible to carry out a more rigorous comparison of transition metal cations using the JS distance between distributions of environments. Figure 10 show a heat map of the similarity between coordination environments distributions of the 50 most observed transition metal cations. The elements are ordered by their atomic numbers (contrary to the alkali and alkaline-earth metals where the Mendeleev number was used).

The first interesting observation is that one recovers the periodicity of the periodic table in the heat map. Indeed, the three rows of the transition metal oxides are clearly identified in the map with Cu/Zn and Pd/Ag/Cd separating the 3d from the 4d transition metals and the 4d from the 5d transition metals, respectively. In fact, this study provides an unprecedented statistical basis for many assumed or long-known relationships. For example, there is a clear similarity between the $Ti^{4+}/Nb^{5+}/Ta^{5+}$, $Cr^{3+}/Mn^{4+}/Ru^{5+}$, and V^{5+}/Mo^{6+} series due to their respective comparable electronic configurations (isoelectronic species). Another striking similarity relationship exists for isovalent species, for example, $Sc^{3+}-Ti^{3+}-V^{3+}-Cr^{3+}$, $Mn^{2+}-Fe^{2+}-Co^{2+}-Ni^{2+}$, and $Zr^{4+}-Nb^{4+}-Mo^{4+}$.

Singular transition metal cations that do not compare to any other can also directly be identified by their light line in the heat map. This includes Cr^{6+} because of its exclusive tetrahedral preference (other highly oxidized transition metals such as $Mn^{5+/6+/7+}$ and Fe^{6+} also show tetrahedral coordination, but the number of representatives for these cations is quite low, see SI). Nevertheless, the similarity of V^{5+}, Re^{7+}, Co^{3+}, and Zn^{2+} shows up nicely. Further cations with peculiar coordinations are Cu^{+} and Ag^{+} for their linear environments, Pd^{2+} for its very strong preference for square planar, and Hg^{+} for its unique metal−metal bond preference.

Main Group Elements. The main group elements consist of all the elements from groups XIII to XVII on the right of d-block part of the periodic table. The statistics for the cations of the main group are shown in Figures 11 and 12. When comparing the isoelectronic B^{3+}, C^{4+}, and N^{5+}, there is a common behavior between C^{4+} and N^{5+}, both favoring trigonal planar environments in CO_3^{2-} and NO_3^{-} polyanionic groups (e.g., $CaCO_3$, K_2CO_3, $NaNO_3$, $Pb(NO_3)_2$). Carbon only forms triangular planar environment, and only a few alkali nitrates form tetrahedral NO_4^{3-} groups (e.g., Na_3NO_4). On the other hand,

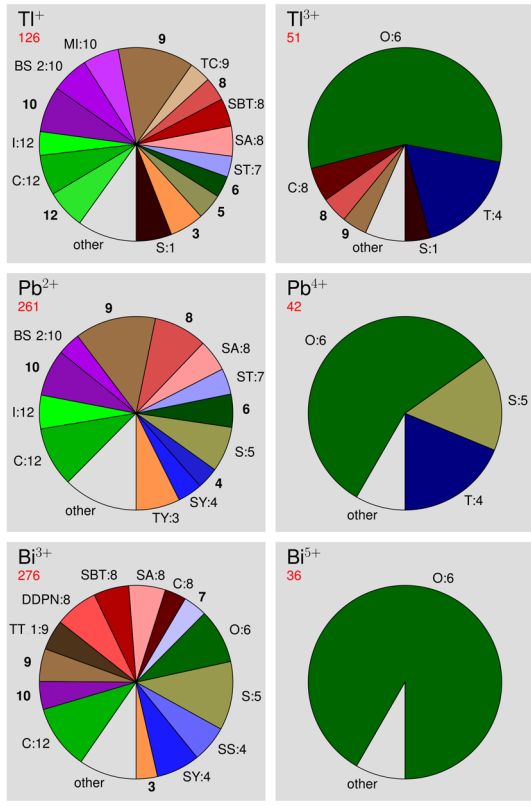


Figure 12. Statistics of occurrences of coordination environments for main group elements from the 6th period. See Figure 4 for a detailed explanation of content of the figure.

B^{3+} occurs in the vast majority of the cases in triangular environments (BO_3^{3-} polyanions), but it also appears as tetrahedrally coordinated (e.g., HfB_2O_5 or $NbBO_4$). Borates often mix tetrahedral and triangular environments in the same compound.(e.g., $Bi(B_3O_6)$).

For the less common N^{3+} oxidation state, angular NO_2^{-} groups are preferred (e.g., $K_2Cd(NO_2)_4$). $(NO)^{+}$ cations are removed from the list due to the presence of other anions than oxygen.

For the main group elements from the third row, most of the ions are isoelectronic. Si^{4+}, P^{5+}, S^{6+}, and Cl^{7+} are found almost exclusively in tetrahedral environment. In our data set, only very few compounds show an octahedral Si^{4+} in an ambient pressure phase (e.g., SiP_2O_7, $Rb_2SiP_4O_{13}$, and $Si^{o}_3[Si^{t}_2(PO_4)_6]$ in which both tetrahedral Si^{t} and octahedral Si^{o} are observed in the latter). Al^{3+} on the other hand favors tetrahedral (e.g., $AlPO_4$, $RbAlO_2$) but also commonly occurs in octahedral environments (e.g., Al_2O_3).

The isoelectronic series $Ga^{3+}/Ge^{4+}/As^{5+}/Se^{6+}$ shows a clear trend toward lower coordinations when the atomic number is increased. Se^{6+} is almost exclusively tetrahedral with rare occurrence of trigonal bipyramidal (e.g., Li_4SeO_5, $K_6Se_2O_9$), square pyramidal (e.g., Na_4SeO_5), or octahedral environments (e.g., $Na_6Se_2O_9$). As^{5+}, Ge^{4+}, and Ga^{3+} show on the other hand a larger fraction of octahedral environments. In other oxidation states, As^{3+} and Se^{4+} form almost exclusively in triangular non-coplanar environments (e.g., $Mn(SeO_3)_2$, $PbSe_2O_5$, $NaAsO_2$ or $Sc_2(SeO_3)_3$).

For the main group elements of the fifth row of the periodic table, the isoelectronic series $In^{3+}/Sn^{4+}/Sb^{5+}/Te^{6+}$ is mainly found in octahedral environment. The smaller ions, In^{3+} and Sn^{4+} can still present a small proportion of tetrahedral environments

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346−8360

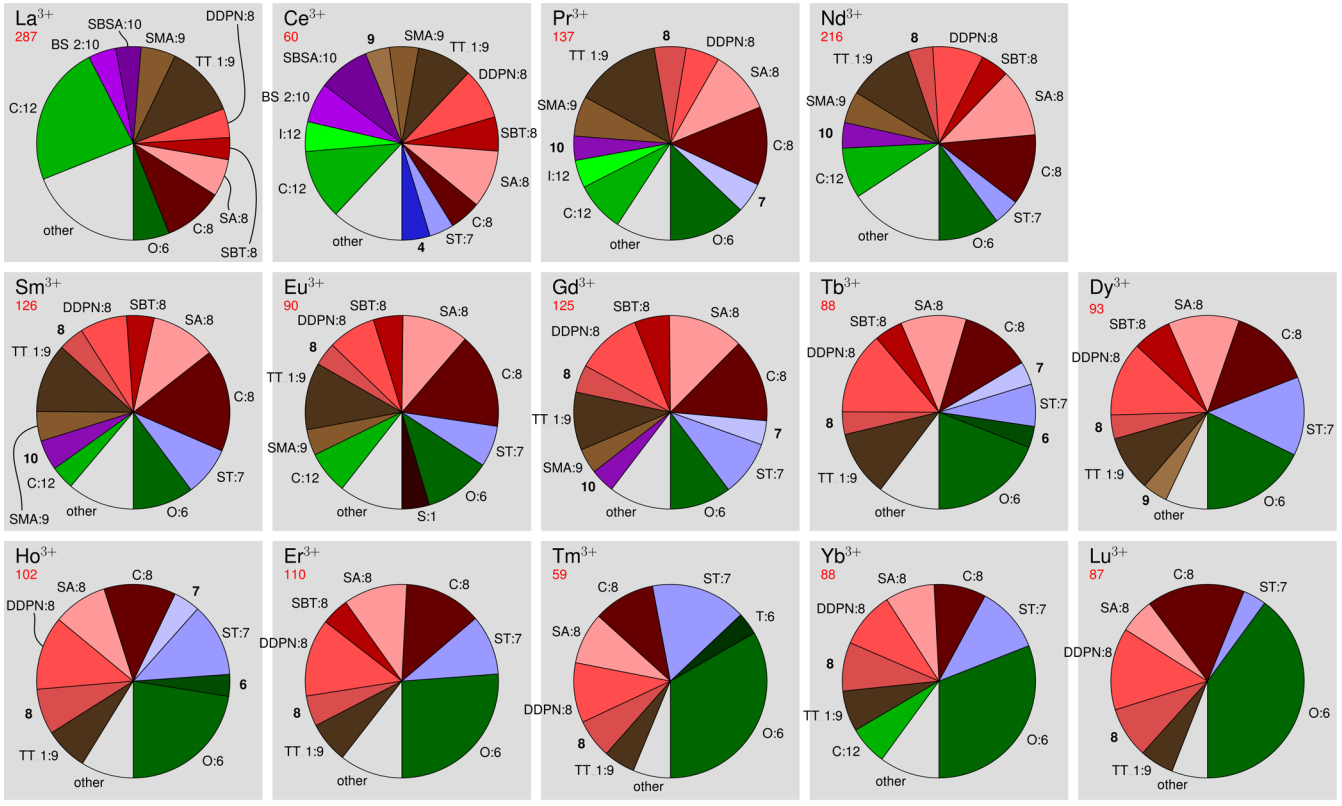


Figure 13. Statistics of occurrences of coordination environments for the rare earth cations. See Figure 4 for a detailed explanation of content of the figure.

(especially when mixed with alkali or alkali-earth such as in Li_5InO_4, Rb_3InO_3, Ca_2InO_5, $Na_6Sn_2O_7$ or K_4SnO_4 or $K_{14}In_4O_{13}$).

The environments of the other isoelectronic series $Sn^{2+}/Sb^{3+}/Te^{4+}/I^{5+}$ are much more diversified with a large preference for triangular out-of-plane and seesaw environments. These specific highly asymmetric environments are indicative of the presence of lone-pair electrons.[49]

The distribution of environments for Th, Pb, and Bi are shown in Figure 12. As observed in the pie charts, the lower oxidation states (Tl^{+}, Pb^{2+}, and Bi^{3+}) show very diversified local environments as in large alkali or alkali-earth. Lone-pairs are present as in the $Sn^{2+}/Sb^{3+}/Te^{4+}/I^{5+}$ series but the larger ionic size leads to larger coordination numbers (>7) especially for Pb^{2+} and Bi^{3+}. Octahedral environments are only very rarely observed (e.g., $BaBiO_3$ perovskite, or $Ag_5Pb_2O_6$)

In contrast, the cations Tl^{3+}, Pb^{4+}, and Bi^{5+} with higher oxidation states have no lone-pair and are in the majority octahedrally coordinated. Bi^{5+} is almost exclusively octahedral while Tl^{3+} or Pb^{4+} also present a large proportion of tetrahedral (e.g., Li_2PbO_3) and square pyramidal environments (e.g., Cs_2PbO_3).

Rare Earth Cations. The rare earth cations or *lanthanides* are most commonly observed in the +3 oxidation state, while a few of them can adopt +2 (Nd, Sm, Eu, Tm, Yb) or +4 (Ce, Pr, Tb) oxidation states. The environment distributions for the rare earth cations are shown in Figure 13.

There is a large diversity of environments for all rare earth cations and larger environments (CN > 7) are usually favored for many of them, in agreement with their larger ionic size. Nevertheless, the proportion of octahedral coordination environments is globally increasing from ~6% for La^{3+} (e.g., La_2O_3, $KLaO_2$) to ~40% for Lu^{3+} (e.g., Lu_2O_3, $RbLuO_2$), with the exception of Ce^{3+}, which is never found octahedrally coordinated. This higher preference for octahedral environments with the atomic number is in agreement with the lanthanide contraction.

Figure 4c shows a heat map of the similarity measure between environment distributions of the rare earth elements. With the exception of Ce^{3+}, which is very dissimilar to all other rare earth elements, the heat map appears mostly as dense-diagonal, that is, similarities are higher for pairs of elements that are closer to the diagonal. This means that the environments distributions slightly change from one atomic number to the next one in a rather continuous manner.

■ CONCLUSIONS

Coordination environments are critical to the understanding of crystal chemistry and structure−property relationships. Using the Inorganic Crystal Structure Database (ICSD), we statistically studied the distribution of environments across cations in about 8000 oxides. We discussed the chemical origin of certain preference in local environments and outlined trends in cation coordination environments. Our work complements the important empirical know-how from experienced solid state chemists and materials scientists. It provides the first statistically based study of coordination environments in oxides. Besides the importance of this statistical analysis for researchers working on crystalline solids, the data and algorithm provided constitute a stepping-stone toward the data mining of coordination environment−properties relationships.

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346−8360

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.chemmater.7b02766.

List of all considered compounds and their coordination environments (XLS)
Complete statistics for all considered cations (TXT)
Coordination environments for all compounds (ZIP)
Name and nomenclature of available coordination environments, technical details for their identification, descriptions and definitions for other files (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: geoffroy.hautier@uclouvain.be.

ORCID

David Waroquiers: 0000-0001-8943-9762
Robert Glaum: 0000-0001-5805-1466
Geoffroy Hautier: 0000-0003-1754-2220

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was performed during a research project supported by BASF SE. Computational resources have been provided by the supercomputing facilities of the Université catholique de Louvain (CISM/UCL) and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CÉCI) funded by the Fond de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under convention 2.5020.11. We would like to thank Gérard Philippe for providing his collection of minerals for the pictures of the crystals in the graphical TOC.

■ REFERENCES

(1) Rajan, K. *Informatics for Materials Science and Engineering: Data-driven Discovery for Accelerated Experimentation and Application*; Elsevier Science, 2013.
(2) Inorganic Crystal Structure Database. http://www2.fiz-karlsruhe.de/icsd_home.html (accessed September 21, 2017).
(3) Pearson's Crystal Data. http://www.crystalimpact.com/pcd/ (accessed September 21, 2017).
(4) Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A. Crystallography Open Database - an open-access collection of crystal structures. *J. Appl. Crystallogr.* **2009**, *42*, 726−729.
(5) Crystallography Open Database. http://www.crystallography.net/cod/ (accessed September 21, 2017).
(6) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.
(7) Materials Project. https://www.materialsproject.org/ (accessed September 21, 2017).
(8) Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R. H.; Nelson, L. J.; Hart, G. L. W.; Sanvito, S.; Buongiorno-Nardelli, M.; Mingo, N.; Levy, O. AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **2012**, *58*, 227−235.
(9) Aflowlib. http://www.aflowlib.org/ (accessed September 21, 2017).
(10) NoMaD Repository. https://nomad-repository.eu/ (accessed September 21, 2017).
(11) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65*, 1501−1509.
(12) OQMD: The Open Quantum Materials Database. http://www.oqmd.org/ (accessed September 21, 2017).
(13) Hautier, G.; Fischer, C.; Ehrlacher, V.; Jain, A.; Ceder, G. Data Mined Ionic Substitutions for the Discovery of New Compounds. *Inorg. Chem.* **2011**, *50*, 656−663.
(14) Pilania, G.; Wang, C.; Jiang, X.; Rajasekaran, S.; Ramprasad, R. Accelerating materials property predictions using machine learning. *Sci. Rep.* **2013**, *3*, 2810.
(15) Yang, L.; Ceder, G. Data-mined similarity function between material compositions. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *88*, 224107.
(16) Hautier, G.; Miglio, A.; Ceder, G.; Rignanese, G.-M.; Gonze, X. Identification and design principles of low hole effective mass p-type transparent conducting oxides. *Nat. Commun.* **2013**, *4*, 2292.
(17) Meredig, B.; Agrawal, A.; Kirklin, S.; Saal, J. E.; Doak, J. W.; Thompson, A.; Zhang, K.; Choudhary, A.; Wolverton, C. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 094104.
(18) Hautier, G.; Miglio, A.; Waroquiers, D.; Rignanese, G.-M.; Gonze, X. How Does Chemistry Influence Electron Effective Mass in Oxides? A High-Throughput Computational Analysis. *Chem. Mater.* **2014**, *26*, 5447−5458.
(19) Faber, F.; Lindmaa, A.; von Lilienfeld, O. A.; Armiento, R. Crystal structure representations for machine learning models of formation energies. *Int. J. Quantum Chem.* **2015**, *115*, 1094−1101.
(20) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114*, 105503.
(21) Isayev, O.; Fourches, D.; Muratov, E. N.; Oses, C.; Rasch, K.; Tropsha, A.; Curtarolo, S. Materials Cartography: Representing and Mining Materials Space Using Structural and Electronic Fingerprints. *Chem. Mater.* **2015**, *27*, 735−743.
(22) Jain, A.; Hautier, G.; Ong, S. P.; Persson, K. New opportunities for materials informatics: Resources and data mining techniques for uncovering hidden relationships. *J. Mater. Res.* **2016**, *31*, 977−994.
(23) Wagner, N.; Rondinelli, J. M. Theory-Guided Machine Learning in Materials Science. *Front. Mater.* **2016**, *3*, 28.
(24) Peng, H.; Ndione, P. F.; Ginley, D. S.; Zakutayev, A.; Lany, S. Design of Semiconducting Tetrahedral $Mn_{1-x}Zn_xO$ Alloys and Their Application to Solar Water Splitting. *Phys. Rev. X* **2015**, *5*, 021016.
(25) Liu, M.; Rong, Z.; Malik, R.; Canepa, P.; Jain, A.; Ceder, G.; Persson, K. A. Spinel compounds as multivalent battery cathodes: a systematic evaluation based on ab initio calculations. *Energy Environ. Sci.* **2015**, *8*, 964.
(26) Mizuno, N. *Modern Heterogeneous Oxidation Catalysis: Design, Reactions and Characterization*; Wiley, 2009.
(27) Schlögl, R. Active Sites for Propane Oxidation: Some Generic Considerations. *Top. Catal.* **2011**, *54*, 627.
(28) Karpov, A.; Dobner, C.-K.; Glaum, R.; Schunk, S.; Rosowski, F. Catalytic Properties of Silver Vanadium Phosphates in n-Butane Oxidation - Considerations on the Impact of the $[V_xO_y]$ Substructure. *Chem. Ing. Tech.* **2011**, *83*, 1697−1704.
(29) Villars, P.; Cenzual, K.; Daams, J.; Chen, Y.; Iwata, S. Data-driven atomic environment prediction for binaries using the Mendeleev number: Part 1. Composition AB. *J. Alloys Compd.* **2004**, *367*, 167−175.
(30) Dshemuchadse, J.; Steurer, W. Some statistics on intermetallic compounds. *Inorg. Chem.* **2015**, *54*, 1120−1128.
(31) Blatov, V. A. Voronoi-dirichlet polyhedra in crystal chemistry: theory and applications. *Crystallogr. Rev.* **2004**, *10*, 249−318.
(32) Hoppe, R. The Coordination Number − an "Inorganic Chameleon". *Angew. Chem., Int. Ed. Engl.* **1970**, *9*, 25−34.
(33) Walsh, A.; Da Silva, J. L. F.; Wei, S.-H. Interplay between Order and Disorder in the High Performance of Amorphous Transparent Conducting Oxides. *Chem. Mater.* **2009**, *21*, 5119−5124.

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346−8360

(34) Lima-de Faria, J.; Hellner, E.; Liebau, F.; Makovicky, E.; Parthé, E. Nomenclature of inorganic structure types. Report of the International Union of Crystallography Commission on Crystallographic Nomenclature Subcommittee on the Nomenclature of Inorganic Structure Types. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **1990**, *46*, 1−11.
(35) Hartshorn, R. M.; Hey-Hawkins, E.; Kalio, R.; Leigh, G. J. Representation of configuration in coordination polyhedra and the extension of current methodology to coordination numbers greater than six (IUPAC Technical Report). *Pure Appl. Chem.* **2007**, *79*, 1779−1799.
(36) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314−319.
(37) O'Keefe, M. A Proposed Rigorous Definition of Coordination Number. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1979**, *35*, 772−775.
(38) Pinsky, M.; Avnir, D. Continuous Symmetry Measures. 5. The Classical Polyhedra. *Inorg. Chem.* **1998**, *37*, 5575−5582.
(39) Aurenhammer, F.; Klein, R.; Lee, D.-T. *Voronoi Diagrams and Delaunay Triangulations*; World Scientific, 2013.
(40) Brown, I. D. Recent Developments in the Methods and Applications of the Bond Valence Model. *Chem. Rev.* **2009**, *109*, 6858−6919.
(41) Blatov, V. A. A method for topological analysis of rod packings. *Struct. Chem.* **2016**, *27*, 1605−1611.
(42) Shevchenko, A. P.; Blatov, I. A.; Kitaeva, E. V.; Blatov, V. A. Local Coordination versus Overall Topology in Crystal Structures: Deriving Knowledge from Crystallographic Databases. *Cryst. Growth Des.* **2017**, *17*, 774−785.
(43) Endres, D.; Schindelin, J. A new metric for probability distributions. *IEEE Trans. Inf. Theory* **2003**, *49*, 1858−1860.
(44) Glossary of Terms in the Materials Project. https://materialsproject.org/wiki/index.php/Glossary_of_Terms (accessed September 21, 2017).
(45) Volume Change Error manual. https://materialsproject.org/wiki/index.php/Volume_Change_Error_manual (accessed September 21, 2017).
(46) Bond length Change Error manual. https://materialsproject.org/wiki/index.php/Bond_length_Change_Error_manual (accessed September 21, 2017).
(47) Kim, S.; Ma, X.; Ong, S. P.; Ceder, G. A comparison of destabilization mechanisms of the layered Na_xMO_2 and Li_xMO_2 compounds upon alkali de-intercalation. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15571−15578.
(48) Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1976**, *32*, 751−767.
(49) Walsh, A.; Payne, D. J.; Egdell, R. G.; Watson, G. W. Stereochemistry of post-transition metal oxides: revision of the classical lone pair model. *Chem. Soc. Rev.* **2011**, *40*, 4455−4463.

DOI: 10.1021/acs.chemmater.7b02766
Chem. Mater. 2017, 29, 8346−8360