Oxidation states of binary oxides from data analytics of the electronic structure

S. Posysaev[1], O. Miroshnichenko[1], M. Alatalo[1], D. Le[2*], and T. S. Rahman[2,3]

[1]Nano and Molecular Systems Research Unit, Faculty of Science, P. O. Box 3000, FIN-90041, University of Oulu, Finland

[2]Department of Physics, University of Central Florida, Orlando, FL 32816, USA.

[3]Department of Applied Physics, Aalto University, Aalto, Finland.

22.10.2018

*Corresponding author (email: Duy.Le@ucf.edu)

Abstract

A connection between the oxidation state (OS) and Bader charge has been missing so far. To our knowledge, all previous work tried to connect OS with Bader charges only with few compounds. The aim of this work was to find a dependency between OS and Bader charge, using a large number of compounds from an open database. We show that a correlation indeed exists between OSs and Bader charges using the simplest machine learning algorithm, linear regression. The applicability of determining OS by Bader charges in mixed-valence compounds and surfaces is considered.

Keywords

Bader charge; electronic structure; oxidation state

1. Introduction

The concept of valency, formulated in the middle of the19th century,[1] forced researchers to think about underlying reasons for complexities in interactions between atoms. It gave rise to theoretical concepts such as oxidation state, Lewis structure, and molecular orbitals. However, the term valency is not widely used anymore and one of the reasons is that valency cannot indicate the loss or the gain of electrons. A valency is always a positive number. For example, the valency of both hydrogen and chlorine in the HCl molecule is 1, in contrast to oxidation state (OS) which indicates the number of electrons an atom has lost or gained (in HCl OS of hydrogen is +1, OS of chlorine

1

is -1). It is thus not surprising that the concept of OS is used in many fields of chemistry and physics. What comes to mind first is the calculation of reduction-oxidation reactions, which should be balanced by changes in reduced and oxidized atoms, but the concept of OS goes far beyond that. Changes in OS lead to changes in the structure of solids along with their physical/chemical properties.

A chemical element can exist in different OSs in different compounds. Furthermore, all elements in the compound can be in the same OS or in several OSs, called single valence (SV) and mixed valence (MV), respectively. Impact of organic MV compounds on modern functional materials chemistry has been summarized by Heckmann and Lambert,[2] while historical background of the field has been collected in the work by Day et al.[3] These publications give a glimpse of the unique properties of MV materials.

Despite its utility, the definition of OS varies from field to field depending on the task at hand. Jansen and Wedig[4] write *'to minimize these risks, the concepts (of OS) need to be defined as precisely as possible, and these definitions must always be kept in mind during applications'*. International Union of Pure and Applied Chemistry (IUPAC) has collected definitions of OS among scientific fields and made comprehensive descriptions with sets of rules for defining OS, which are being constantly renewed.[5] Karen et al. formulated the definition as "*The OS of a bonded atom equals its charge after ionic approximation*".[6]

Perhaps the most accurate definition of OS comes from the experimental techniques that identify it directly: X-ray photoelectron spectroscopy (XPS), which utilizes the fact that the binding energy of an electron in an atom (core level shift) changes with its OS. XPS, however, is most sensitive to surfaces (as they are extended). Defining OS in MV compounds, molecules, clusters, near vacancies and defects is a challenging task for XPS because of the limited size of the measured area. Current computational techniques for determining OS vary from the calculation of the peak shift in the electron density of states to full-scale XPS calculations. The former may not always provide an accurate measure, and the latter is a demanding calculation, available in a limited

number of computational codes. An independent theoretical tool for OS determination can thus provide a good alternative or at least a benchmark for XPS results.

It is interesting to note that the widely used computational method for evaluation of electronic structure, density functional theory (DFT), has not been designed to calculate oxidation states directly, although DFT data is routinely used for extraction of atomic charges.[7–10] For example, Bader analysis[7] based on DFT, utilizes electron density for estimating the net charge on an atom. A volume of a cell is separated in Bader volumes when: $\nabla\rho(\mathbf{r}) = 0$, where $\rho(\mathbf{r})$ is the electron density, and the gradient is taken with respect to the three-dimensional coordinate r. This zero flux condition effectively yields distinct regions (Bader volumes) for each atom in a system and the integral of density inside them is assigned as a net charge of an atom. For example, charge densities of titanium atoms in +1 and +4 OS are depicted in *Figure 1*. Applying Bader charge analysis leads to 0.75 e charge on titanium in +1 OS and 2.4 e in +4 OS in agreement with Koch and Manzhos[11] where Bader charges have been obtained for Ti in bulk anatase. It worth to note that Koch and Manzhos[11] are suggested that OS of titanium in titanium dioxide is +3 but in Walsh et. al.[12] view this is an unphysical conclusion. Bader charges typically are not integer numbers, and therefore it is not a trivial task to connect them with the concept of OS.
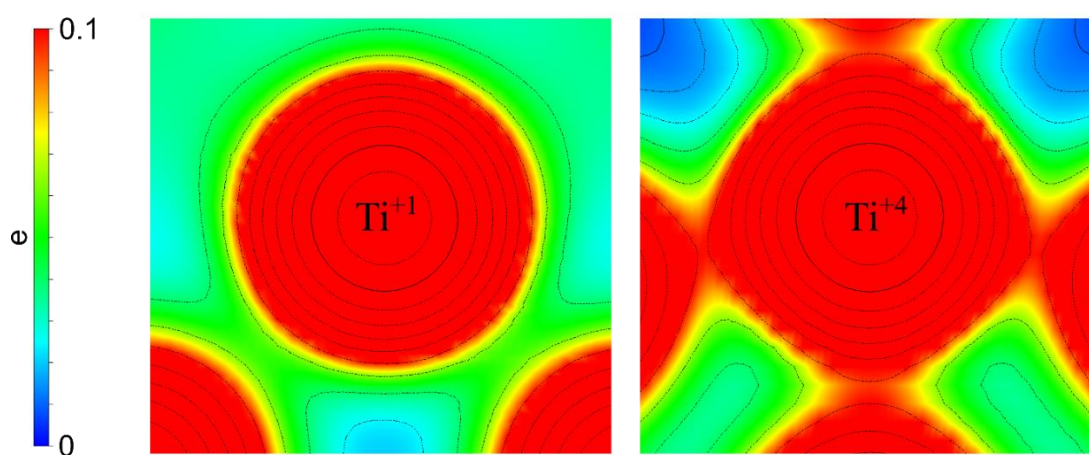
Figure 1 Electron density for titanium atoms in +1 (left) and +4 OS (right).

At the same time, the increasing popularity of the DFT approach[13] has given rise to continuously growing open material databases for millions of compounds. Examples of such databases are:

AFLOW,[14,15] Materials Project,[16] Engineering Virtual Organization for Cyber Design,[17] and NOMAD.[18] Such databases have been used for combining machine learning techniques with physics and show that it is possible to predict material properties by intelligent algorithms that learn from the large number of known data[19] as well as for the inverse task - to find a material which possesses the desired property.[20] Recent statistical analysis of the occurrence of oxidation states in inorganic compounds[21] using material databases showed that a large number of cations are only likely to exhibit certain oxidation states in combination with particular anions. The AFLOW database, one of the largest of material databases, contains also computed Bader charges along with other material properties. There is thus the possibility to utilize available DFT data, particularly Bader charges, for predicting OS.

A direct connection between OS and Bader charge has been missing so far, although a number of attempts have been made. One reason could be that that the charge difference between two consecutive OS (between Ru(II) and Ru(III))[22] is rather marginal (smaller than unity), as pointed by Reeves and Kanai. It is natural to look for a method, which gives an integer number for OS, but is this a possibility in nature? Can we say that all 8 valence electrons of osmium in OsO_4 go entirely to the oxygen atoms? It is well known that all ionic bonds have some features of a covalent bond. Jansen and Wedig[4] are confident that a connection between OS and Bader charge exists and they highlight *"the need for factors other than calculated charges to be taken into consideration in order to obtain a meaningful connection with oxidation states"*. Despite these proposals, to our knowledge, all previous work tried to connect OS with Bader charges only in a few compounds, and not look for general trends. Our aim is to find a correlation between OS and Bader charge, using a large number of compounds from an open database, and to provide guidelines which may facilitate the determination of the oxidation state from DFT calculations that are routinely carried out. We make a hypothesis that a change in the charge density should reflect a change in the OS. With the definition of OS in terms of electron density, a measure of the probability of an electron to be in the specific region in space, we search the AFLOW database for correlations, limiting our

focus to binary oxides, compounds of two elements, one of which is oxygen. They are chosen by the following criteria: simplicity, they contain only two species; abundance on Earth; they are among the most calculated compound types in material databases.

In the rest of this work, we present a correlation between OSs and Bader charges and applicability of determining OS by Bader charges in MV compounds. More specifically, after the description of methodology in *section 2*, we examine the trends in the correlation between Bader charges and OS for SV compounds in *section 3*, followed by the applicability of our method for MV compounds in *section 4*, and finally, the correlation between OS and Bader charges for surfaces is discussed in *section 5*.

2. Methods

2.1. Data set

The data set was downloaded from the AFLOW database,[14,15] a materials properties repository with a defined standard for high-throughput materials science *ab initio* calculations.[23] The database currently consists of over 1 million compounds. The REpresentational State Transfer–Application Programming Interface (REST–API) [24] was used for exploring the repository and retrieving the data needed.

Note that, herein we refer to binary oxide as "binary". Binaries are all oxides in the database with specific elements. For example, titanium oxides refer to all records (or compounds) of the type Ti_xO_y available in the database. Each Ti_xO_y record will be referred to as compound or record. Note also that only records with calculated Bader charges were examined. And, since the database is being constantly updated, our results are based on the version from May 2017. The properties of all binary oxides in the AFLOW database were calculated using density functional theory based electronic structure code VASP[25,26] with the PAW formalism[27,28] and PBE+U[29] exchange-correlation functional (full parameters for DFT calculations were chosen in accordance with the AFLOW standard[23]). In total 69 types of binaries of the following elements were taken into

account: Li, Na, K, Rb, Cs, Be, Mg, Ca, Sr, Ba, Sc, Y, Ti, Zr, Hf, V, Nb, Ta, Cr, Mo, W, Mn, Tc, Re, Fe, Ru, Os, Co, Rh, Ir, Ni, Pd, Pt, Cu, Ag, Au, Zn, Cd, Hg, B, Al, Ga, In, Tl, C, Si, Ge, Sn, Pb, P, As, Sb, Bi, S, Se, Te, La, Ce, Pr, Nd, Pm, Eu, Gd, Tb, Dy, Ho, Er, Yb, Lu. We did not consider nitrogen and halogen oxides since they have small charge transfer between atoms, possibly because of their high electronegativity, which complicates the determination of OS. Noble gases and hydrogen are also not considered in this work.

Throughout this work, we use the following notation for elements in binary: cation - which has a positive charge and an anion - which has a negative charge (oxygen and anion are used interchangeably below). Furthermore, to classify materials by their OS we need to formulate an accurate definition of the cation's OS. By stating that the sum of the OSs in a compound has to be neutral, that oxygen has an OS of -2, and that the OS of the anion is a known quantity, we calculate the OS per compound and the OS of the cation as follows:

$$OS \text{ per compound} = OS \text{ of anion} * \text{number of anions (1)},$$

$$OS \text{ of cation} = \frac{OS \text{ per compound}}{\text{number of cations}}. (2)$$

Note that the OS of the anion for the binaries under consideration is always -2.

In data analysis, proper classification is of major importance for establishing and understanding relationships among variables of a database. Of course, as is true of all databases, there are outliers, which complicate the establishment of dependencies. A thorough explanation of the procedure for curing and splitting the database into two sets, one with SV and another with MV compounds is presented in the Appendix. The resulting data set includes 697 compounds and consists of 10743 atoms.

2.2. DFT calculations

As will be seen in *sections 3.1.2* and *3.3,* we validated results for several titanium compounds using a number of exchange-correlation (XC) functionals, and carried out calculations for several

oxide surfaces (binary oxides in the AFLOW data base consist mostly of bulk materials as well as some portion of molecules). Here we provide some essentials of our DFT calculations. In particular, we compared results for titanium binaries using the LDA, LDA+U, PBE,[30] PBE+U and HSE06[31] XC functionals, in accordance with the AFLOW standard.[23] The LSDA+U approach (further referred as PBE+U), introduced by Dudarev el al.[29] was used with parameters $l = 2$, $J = 0$ and $U_{EFF} = 4.4$ for titanium. Structural optimization was performed for each functional. Numbers of k-points for LDA and PBE (for HSE06 in parentheses) for OS +1, +2, +3 and +4 of the Ti atom are 19x19x10 (9x9x5), 16x16x16 (8x8x8), 10x10x10 (5x5x5), and 9x9x13 (7x7x11) respectively. The number of k-points for hybrid HSE06 functional was decreased for computational feasibility. Since we are interested in trends, this approximation did not lead to qualitative differences in the results. Bader charge analysis was performed using the code by the Henkelman group.[32]

For the DFT calculations for surfaces, pristine rutile $TiO_2(110)$ along with surfaces with adatom and vacancy were relaxed before Bader charge analysis was performed. The rutile (110) surface was modelled as a slab with a 4x2 cell, which consisted of five O-Ti-O layers. Two bottom layers were fixed. The surface with one oxygen vacancy thus contained 239 atoms, while the surface with a Ti adatom contained 241 atoms and the pure surface consisted of 240 atoms. 10 Å of vacuum was added in the direction normal to the surface to avoid interaction between periodic slabs. All systems were relaxed with 2x2x1 k-points.

3. Results

3.1. Single valence compound

Here we present results of the trends that we find in the correlation between the OS and Bader charge for SV compounds. As we shall see, there is a linear dependence between Bader charges and OSs for these compounds. At the end of the section, we discuss the applicability of the results to calculations with other exchange-correlation functionals.

As shown in Figure 2, Bader charges form intervals, which do not intersect between oxidation states inside each binary. For example, Bader charge intervals for niobium oxides (Nb group 5) are: for + 2 OS - [1.09, 1.55], + 3 - [1.86, 1.87], +4 - [2.37, 2.4], +5 - [2.72, 2.81].
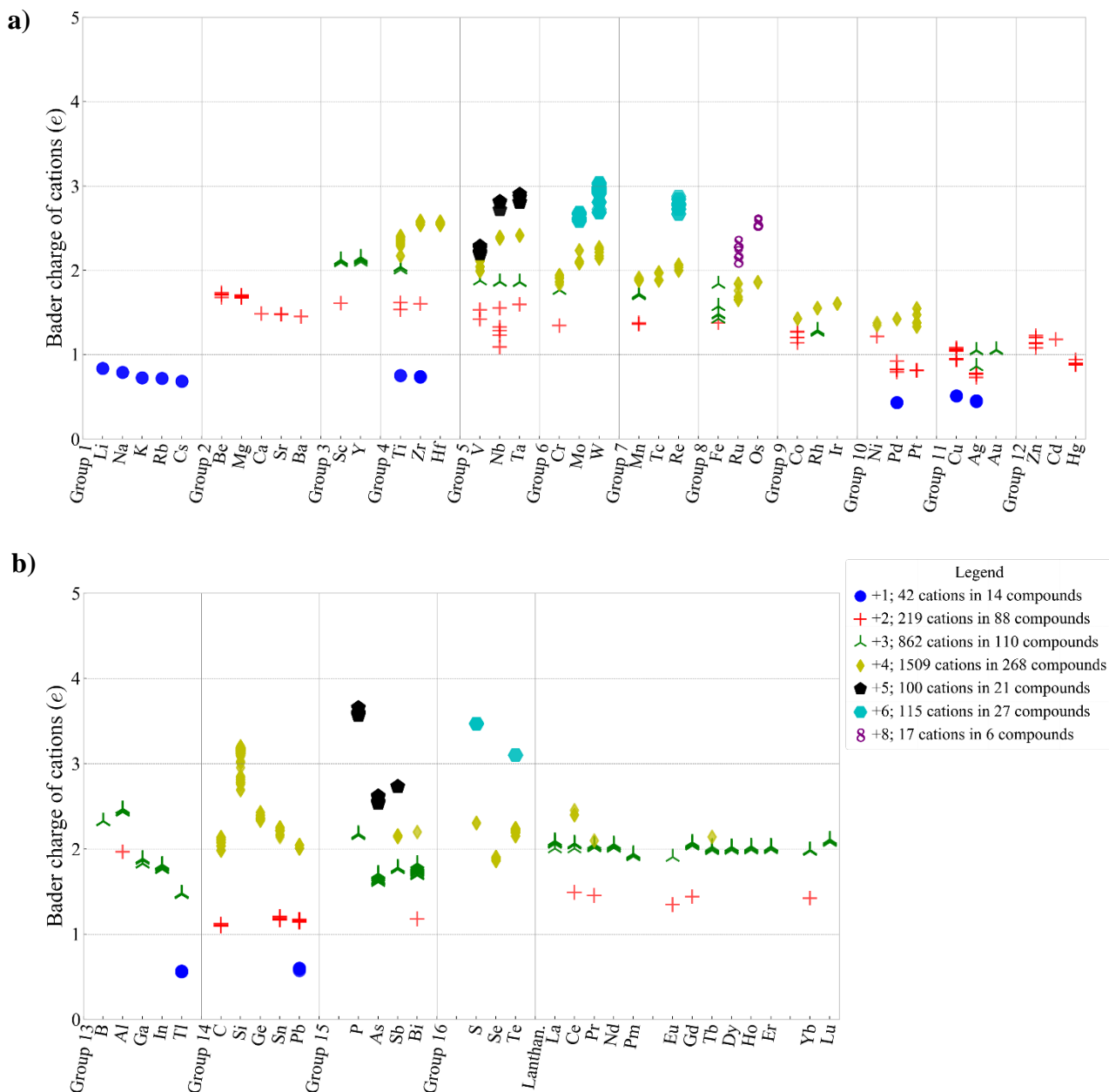
a)

b)

Figure 2. Bader charges for cations. Each point on the graph is a Bader charge for one atom. The numbers of atoms are in the legend. Atoms are grouped according to the Mendeleev table. a. Groups 1-12; b. Groups 13-16 and lanthanoids.

It is worthwhile to analyze Figure 2 in more detail. From group 2 to group 11 one can see a descending trend in Bader charges for OSs +2, +3 and +4. The same maximum OS can be noted between group 3 and 13, 4 and 14, 5 and 15, 6 and 16, as cations in this group have the same number of valence electrons but different open d orbitals in groups 3-6 and p orbitals in groups 13-16. As a result of almost the same bonding energy between the core and open f orbital electrons, lanthanides have almost the same Bader charge intervals across the series.

In most cases, the Bader charges of cations in one compound differ by less than 0.03 e. In the few compounds in which the difference is higher, it is still less than 0.1 e (Appendix: Figure 17). Of major interest is the charge difference between two consecutive OSs for a binary. How much should the charges differ for cations to be in different OS? The smallest Bader charge difference is between +4 and +5 OSs of vanadium oxides, which is equal to 0.025 e. It indicates that a difference of 0.025 e is enough for cations to be in consecutive OSs, even though in one compound charges may vary up to 0.1 e.

Bader charge intervals per OS for all binaries are presented in Figure 3. Intervals for each oxidation state naturally shift upwards from +1 to +5 OS. The trend starts to change from the interval for +6 OS, still having a higher minimum than for +5 OS, but a lower maximum. Surprisingly, the maximum and minimum of the interval for +8 OS are smaller than for +6 OS. No records were found with cations in +7 OS in the AFLOW database.
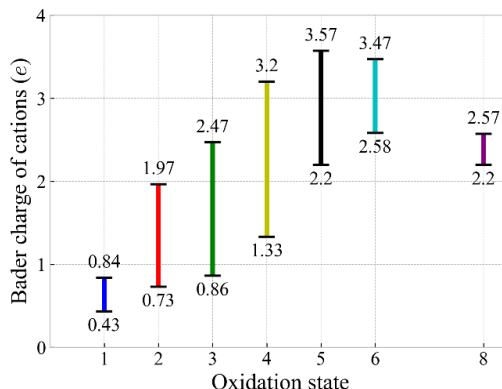

Figure 3. Bader charge intervals per OS for all binaries.

3.1.1. Linear dependency in SV compounds

Linear regression is able to describe the OS vs Bader charge trend, allowing one to predict the OS of cations in compounds based on Bader charges. As can be seen from the linear regression graphs for Nb and Ti binaries presented in *Figure 4*, the slope of lines varies among binaries. We also find the mean squared error to be lower than 0.1 for all binaries except Pr oxides. From *Figure 4*a we can predict that Nb in +1 OS has 0.8 e. The rest of the graphs for binaries which have at least 3 OS in our data set are included in the Supporting Information (SI) (Figure S1).
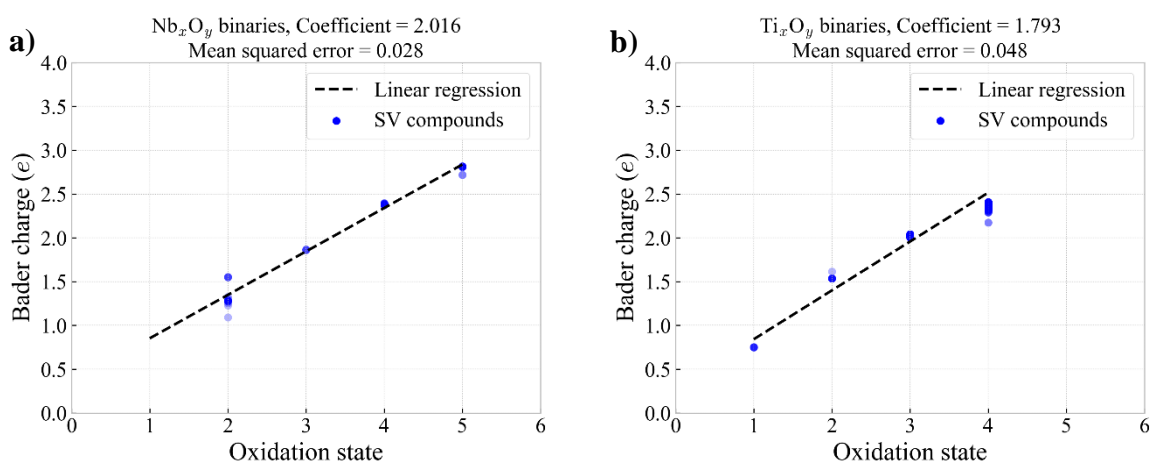
Figure 4. Linear regression for Nb and Ti oxides.

3.1.2. Applicability to calculations with other parameters

The relationship between OS and Bader charge that we show above for the compounds from the AFLOW library were based on calculations performed with a specific exchange-correlation functional (PBE+U). A question arises whether we can observe the same results with other functionals?

To address the above question, we downloaded structures from the AFLOW database in which Ti atoms are present in +1, +2, +3 and +4 OS (ICSDs successively are #99784, #77692, #9648, and #23697). Structures were chosen with the Bader charges lying in the middle of intervals per OS. Titanium oxides were chosen as one of the most studied materials nowadays due to their abundance and promising physical properties, for example, their application in water splitting.[33] Our

10

calculations show that the Bader charges of titanium inside each compound differ less than 0.03 e for all functionals. The discrepancy in percentages from the referenced value (PBE+U) is presented in Figure 5. The closest result is given by the LDA+U approach with a discrepancy of 2%. Calculations with the HSE06 hybrid potential have been performed with fewer k-points because of its computational complexity, giving the same discrepancy around 2%. PBE and LDA resulted in up to 7% and 9% discrepancy, correspondingly.
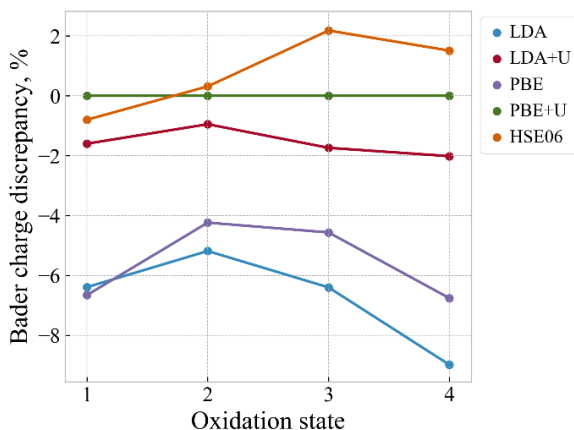
Figure 5. The discrepancy of cations' Bader charges in percentages for LDA, LDA+U, PBE, PBE+U and HSE06 functionals from the referenced value (PBE+U).

While we have not performed calculations for all binaries examined here, it is reasonable to assume that dependence of the results on the functionals for the rest of the compounds would be close to what we found above for titanium oxides (2-9% discrepancy from the PBE+U value).

3.2. Mixed valence compounds

In this section, we examine MV compounds. We characterize the OS of the cations in MV compounds based on unique intervals for each binary, which show different behavior from OSs in SV compounds analyzed above, as we shall see. We also pay attention to compounds with the OS of the cation lower than +1.

Modeling methods commonly predict OS only with certain accuracy. Bader charge analysis is not an exception. Robin-Day[34] made a classification of MV compounds and subdivided them into three

groups: 1. Valence electrons are localized; 2. Valence electrons have some localization, but activation energy is needed for localization or delocalization to occur; 3. Valence electrons are delocalized, and OS is impossible to determine. It shows that defining OS in MV compounds can be done only for first two groups of Robin-Day's classification. We will see further that strong separation of Bader charges between cations can show different OS, but in some cases, the separation is not strong enough.

The OS of the cation lower than +1 OS (for example Ti_6O) appeared in 12 compounds (Figure 6, pink dashes), OS of compounds vary from 0.12 e to 0.537 e. These records lie in agreement with the oxidation state/charge trend, which was shown in *section 3.1*. Cations in these records have smaller Bader charges than cations in +1 state (or other lowest state available in the database). Charge in compounds is mostly spread among cations, and we cannot distinguish cations of 0 OS and +1 OS. We assume that cations are in some fractional OS lower than +1 and thus leave them out from further investigation.
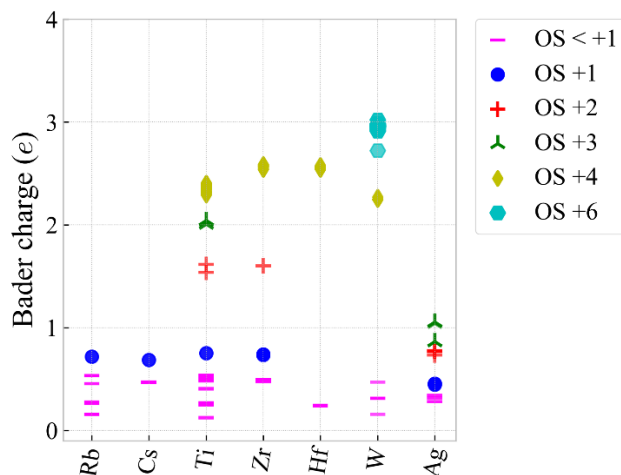

Figure 6. Compounds with OS of cations lower than +1 and SV compounds

For further characterization of cations in MV compounds, bisected intervals of Bader charges are needed for each binary, where each interval relates to a specific OS. The intervals are based on

Bader charges of each OS (Figure 2). A thorough explanation of the obtaining and adjusting the bisected intervals with an example is presented in section 3 of SI.

The workflow of the characterization of cations in MV compounds is depicted in Figure 7. First, we take one specific binary, which has MV compounds and pick one. The OS of cations are assumed to be defined if all cation's Bader charges of the MV compound lie in earlier defined bisected intervals and the sum of the defined OSs is equal to the OS per compound. Otherwise, it is not defined, and the next compound is analyzed.
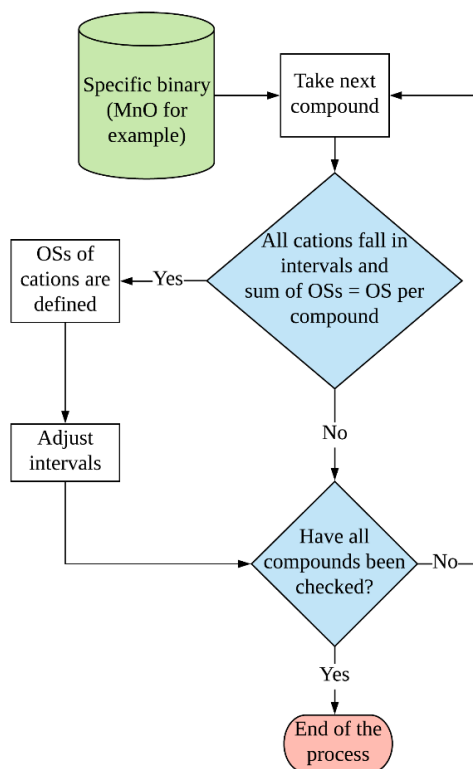
Figure 7. The workflow of characterization of cations in MV compounds

Further, we examine some of the MV binaries using the workflow described in Figure 7.

3.2.1. Ti_xO_y

The curated data set contains 42 Ti_xO_y compounds, from which 18 (consisting of 136 atoms) are MV compounds (Figure 8). Cations of 12 MV compounds (consisting of 100 atoms) are defined using the workflow depicted in Figure 7; they consist of cations in +3 and +4 states. The number

of cations inside compounds varies from six up to twelve atoms. Final intervals include cations for SV and defined MV compounds.
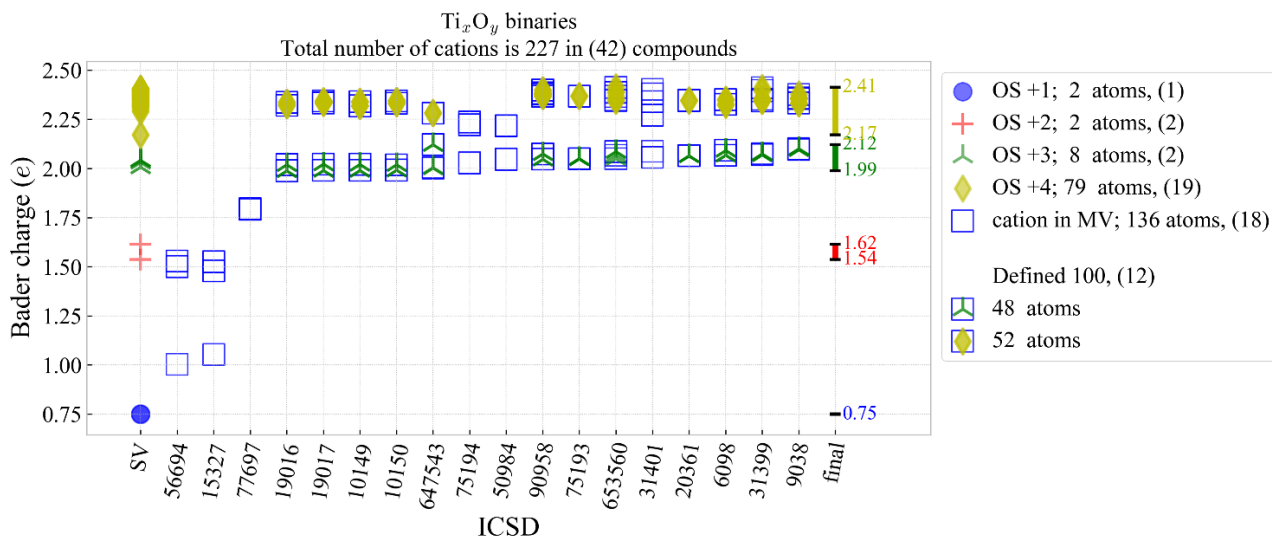

Figure 8. Titanium binaries. ICSD is a unique identifier in Inorganic Crystal Structure Database

For the remaining six compounds, the sums of the defined OSs are not equal to the OS per compound (Table 1). As mentioned earlier, for the OS of the cation to be defined, not only Bader charges should lie in well-defined intervals, but also the sum of the defined OSs should be equal to the OS per compound. The number of cations inside undefined compounds varies from 6 up to 10 atoms. We can conclude that Bader charges are not always evenly correlated with the OSs of the cations, the results agree within 1-2 OSs.

ICSD	The sums of the defined OSs	OS per compound by formula (2)
56694	9	10
15327	9	10
77697	8	10
75194	22	20
50984	22	20
31401	37	36

Table 1. The sums of the defined OSs and OS per compound for titanium binaries

14

3.2.2. V_xO_y

There are 35 vanadium compounds in the AFLOW data set; 12 compounds (consisting of 100 atoms) are MV compounds, from which OSs for 9 compounds (consisting of 76 atoms) are defined, they consist of cations in +3, +4 and +5 states (Figure 9).
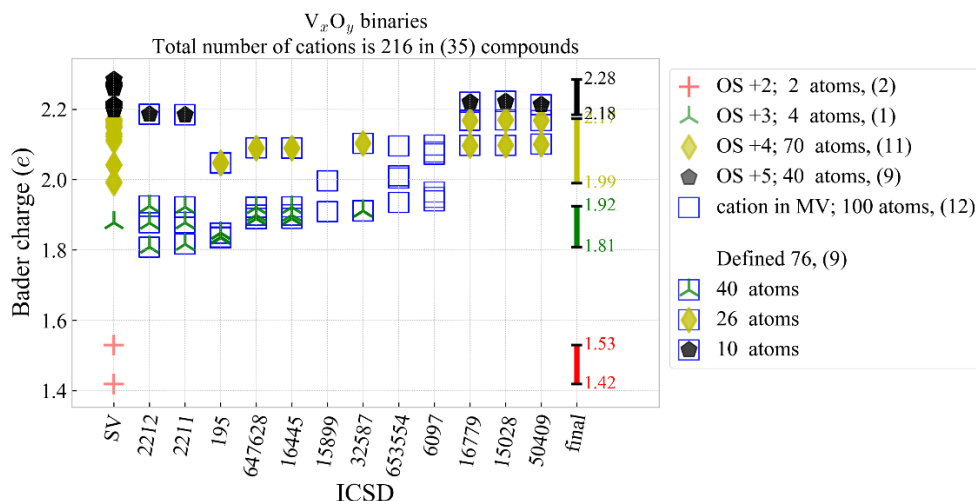

Figure 9. Vanadium binaries

3.2.3. Fe_xO_y

Binaries with iron show a completely different picture (Figure 10). None of the cations in MV compounds can be characterized. We conclude that the characterization of the OS of the cations in iron oxides cannot be done using Bader charge intervals. There is a possibility that another approach (not PBE+U) with other XC potential can be more suitable for iron oxides. Even a small change in the geometry of the cell would give different Bader volumes, and consequently different Bader charges.
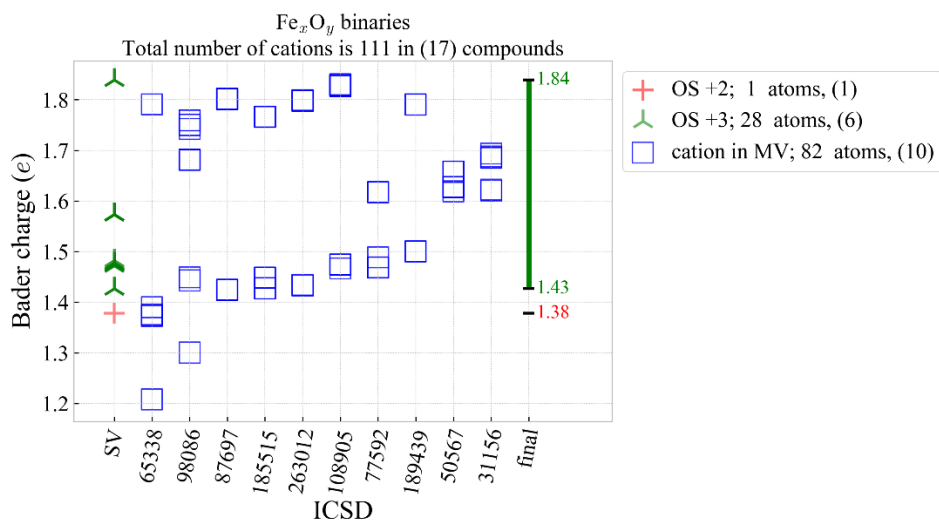
Figure 10. Iron binaries

The rest of the graphs with OSs in MV compounds are presented in the SI (Figures S3-21). The summary was collected in Table S1. All cations in 49 MV compounds were defined, for 7 compounds we have not enough data (the database does not include compounds with the cation in certain OS) and 31 compounds were undefined.

3.3. Application: $TiO_2(110)$ surface

Although most of the data from AFLOWLIB is for periodic bulk compounds, we can use our classification presented above for other types of systems. For example, to illustrate possible applications of our finding to surfaces, we study the OSs of Ti atoms on the TiO_2 (110) surface, in which few types of Ti atoms are present with different OS. For surfaces, Bader charges do not necessarily have the same connection with OS because the environment of the uppermost atoms on the surface is different from the environment in bulk. Here, we examine only rutile TiO_2 (110) surface (titanium atoms in bulk have the OS of +4) mainly due to the availability of experimental and theoretical results for titanium OSs on surfaces with oxygen vacancy and titanium adatom (Figure 11). A Rutile (110) slab consisting of 240 ± 1 atoms was relaxed using DFT, and Bader analysis was performed. The OSs of titanium atoms were determined from bisected Bader charge intervals, defined after analysis of SV and MV titanium binaries (Figure 8).
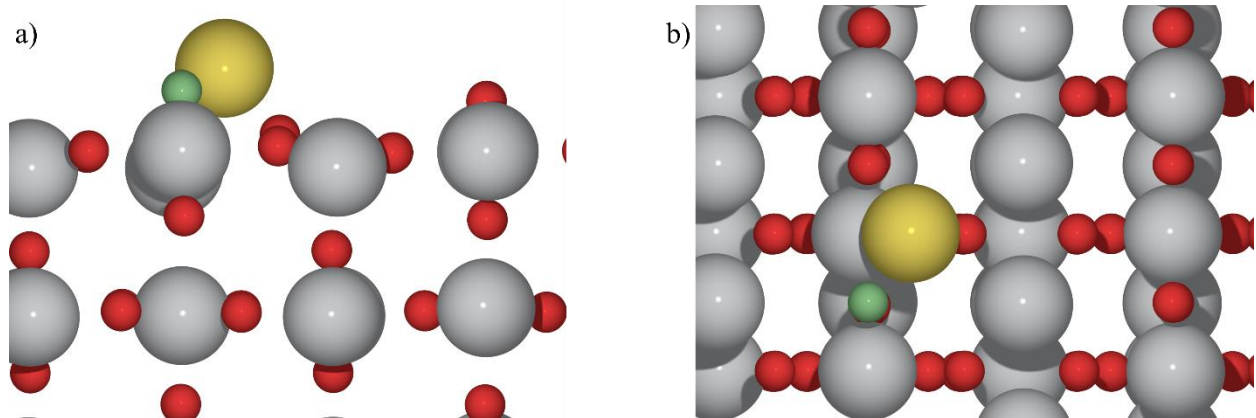
16

Figure 11. (110) surface of TiO_2: a) Side view; b) Top view. Ti atoms in grey, O atoms in red, Ti adatom in yellow, and O vacancy in green.

3.3.1. Pure rutile (110)

Bader charges for pure (110) rutile surface vary from 2.329 e up to 2.374 e. It shows maximal Bader charge difference of 0.045 e. According to Figure 12, the corresponding OS state of Ti atom is +4.

3.3.2. Rutile (110) with Ti adatom

According to Ref.[35] titanium adatom on top of rutile (110) surface is present in +2 OS and the neighboring titanium atom is present in +3 OS. The calculated Bader charge for the adatom is 1.686 e, falling in +2 interval and the Bader charge of the neighboring atom is found to be 2.132 e, falling in the interval for +3 state (Figure 12). The rest of the titanium atoms whose Bader charges are ranging from 2.316 e to 2.379 e, fall in the interval for +4 OS.

3.3.3. Rutile (110) with vacancy

We have calculated Bader charges for the rutile (110) surface with one oxygen vacancy. We found Bader charges for the neighboring Ti atoms near the vacancy to have the smallest charges of 2.303 e but they fall in the interval for +4 OS (Figure 12). This result is not in agreement with Ref.[36], where Ti atoms near the vacancy are reported to be in +3 OS. Bader charges of all other titanium

17

atoms are from 2.323 e to 2.402 e, falling in the interval for +4 OS (Bader charges vary up to 0.1 e). In this example, Bader charges give only a hint to use more advanced methods for defining OSs.
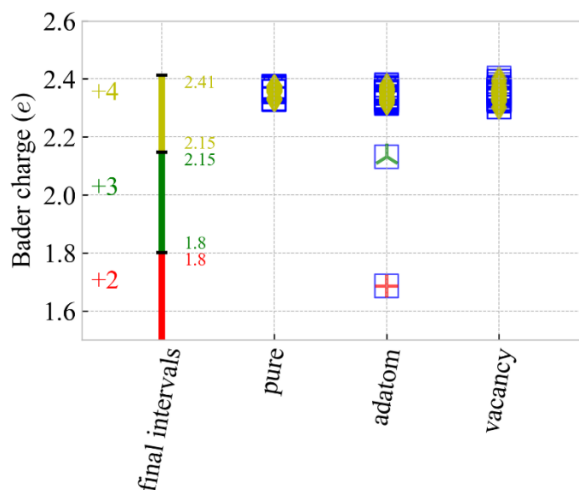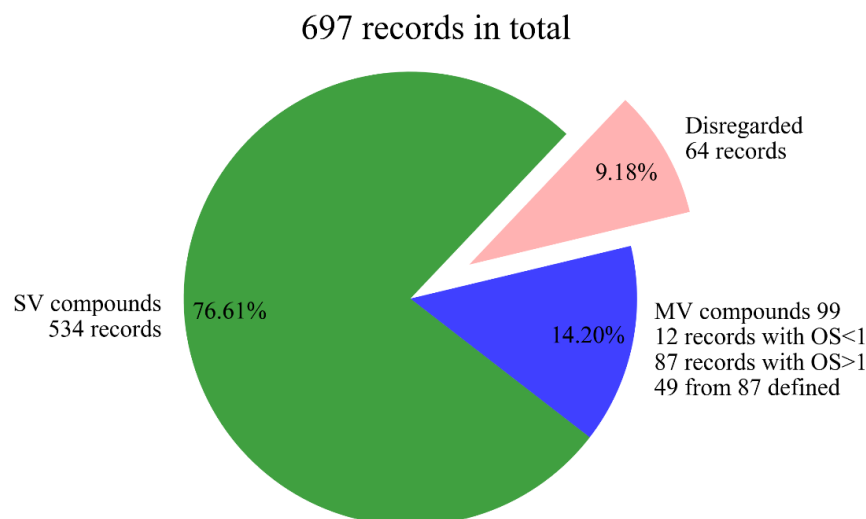

Figure 12. Bisected intervals of Bader charges for titanium oxides; Bader charges for pure rutile (110) surface; for the surface with titanium adatom; surface with oxygen vacancy

4. Discussion

Total summary of the number of compounds in each group presented in Figure *13*. Using techniques described in *section 3.2.,* the OS of cations was successfully defined in 49 compounds from the rest of 87 MV compounds with OS > 1. Some undefined MV compounds could have fractional OS. We assume that cations with OS < 1 also have a fractional OS and cannot be separated in cations with 0 and +1 OSs.

697 records in total

Disregarded
64 records
9.18%

SV compounds
534 records
76.61%

MV compounds 99
14.20%
12 records with OS<1
87 records with OS>1
49 from 87 defined

Figure 13. Distribution of records by groups

Distribution of Bader charges of oxygen (Figure 14) shows a variance of values in the interval [-2, -0.53] e for oxygen in -2 OS (data sets with SV and MV compounds together). We analyze outliers and draw Bader charges for anions where minimal O-O bond is less than 1.65 Å, indicating single or double O-O bond. Oxygen in -1 and 0 OS states have Bader charges in the interval [-0.96, 0.05] e.
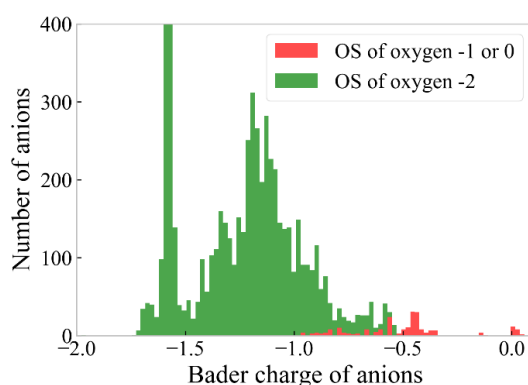
Figure 14. Distribution of Bader charges of oxygen

The distribution of maximal Bader charge difference of cations for both data sets is presented in Figure 15. The difference is defined as the highest Bader charge minus the lowest Bader charge of

19

the cations in one compound. Knowing that 91% of SV compounds and only 3% of MV compounds in binaries have maximal charge difference less than 0.03 e allows us to make a 'rule of thumb' such as 'With high probability, cations of the same element in binaries are in the same OS, if maximal Bader charge difference is less than 0.03 e'. A second rule can be expressed as 'With high probability, cations of the same element in binaries are in different OS if maximal Bader charge difference is more than 0.1 e'. It is based on the result that 1% of SV compounds and 89% of MV compounds have maximal Bader charge difference more than 0.1e. The probability of error is large, and we recommend using another method for defining OS if the initial guess of OS of cations in compounds with maximal Bader charge difference lies in the interval [0.03, 0.1] e.
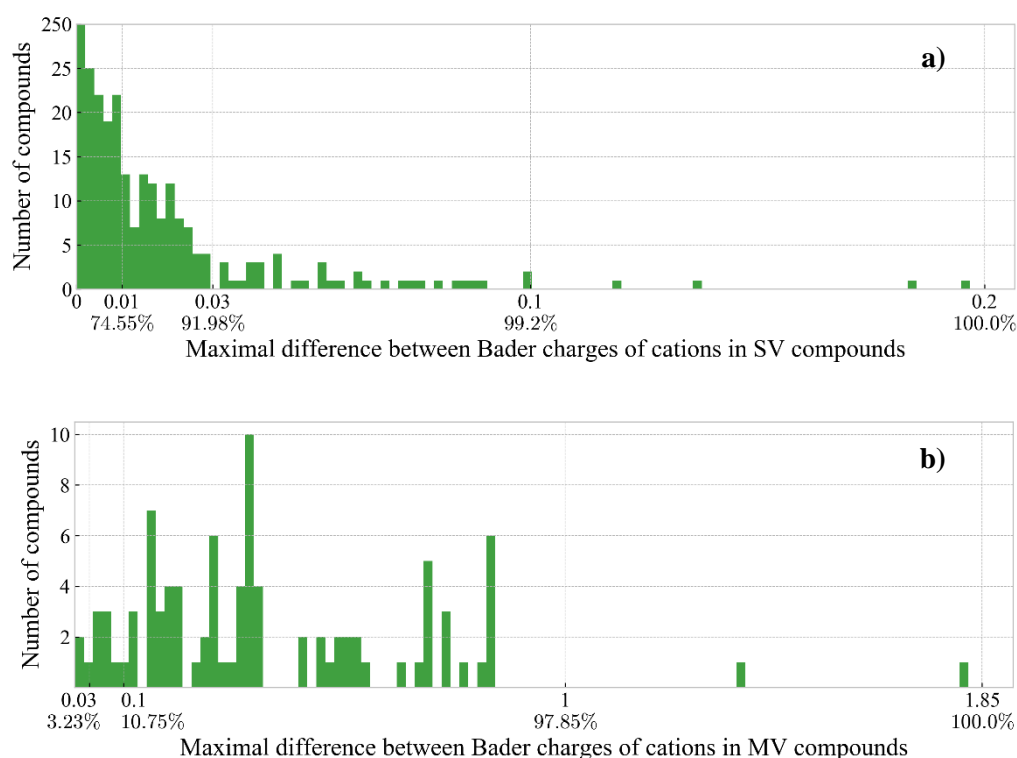
Figure 15. Distribution of maximal Bader charge difference of cations in a) SV compounds; b) MV compounds (MV compounds with OS lower than 1 are excluded from the histogram). The percentages written under the numbers on the horizontal axes show how many compounds have smaller or equal to the maximal difference between Bader charges of cations specified by the numbers. Please note that the scales on both axes are not the same.

20

5. Conclusions

In this work, we have demonstrated a correlation between OS and Bader charge for a large set of simple and mixed valence systems. Of course, there are a good number of cases for which our proposed correlation fails. This failure could be attributed to issues that there is no unambiguous way to partition the electron density between atoms as well as DFT has with its validity for systems with strong electron correlations, which is likely the case for oxides such as those composed of Fe. In fact, it is to the credit of the PBE+U functional that it still provides reasonable results for the correlation between OS and Bader charge for such a large number of oxides.

The analysis of rutile surfaces showed that the Bader charges for the uppermost cations differ less than 0.045 e from the cations in the bulk. The surface with an oxygen vacancy displays a Bader charge difference of 0.1 e, which does not exclude the possibility of Ti atoms to be present in the differing OS. In the case of the surface with the adatom, Bader charges successfully predict titanium OSs.

Furthermore, we present here a technique which can give at least a hint about the OS of the cations and predict OS in MV compounds. For surfaces, clusters and more complex systems than binary - charges can differ. As we show, the results do depend, albeit weakly, on the choice of the exchange-correlation functionals. The results presented above for the materials of interest can be summarized as below:

- Bader charge is a good parameter for predicting OS
- There is a linear relationship between Bader charges and OS
- The results in Figure 2 may be used for predicting OS in binaries
- Our approach fails to predict OS of iron in MV compounds
- Bader charge intervals are not equal between OSs and tend to shrink with higher OS
- Oxygen in -1 or 0 OS can have Bader charge down to -0.96 e, oxygens with lower Bader charge are in -2 OS

21

- With a high probability cations of the same element in binaries are in the same OS, if the maximal Bader charge difference is less than 0.03 e

- With a high probability cations of the same element in binaries are in different OS, if the maximal Bader charge difference is more than 0.1 e

- Methods beyond DFT are needed if maximal Bader charge difference lies is in the interval [0.03, 0.1] e

- The results can be compared with those for other exchange-correlational functionals, but an estimation of the discrepancy is needed.

Supplementary material

The linear regression graphs for 12 binaries, an example of adjusting bisected intervals, determining oxidation states of MV binaries. The Python scripts to explore trend between Bader charges and oxidation states for binaries with AFLOW database and data set used in this article can be found here github.com/Posysaev/BADERvsOS.

Appendix: Data curation

Based on the information available in the AFLOW database, the following geometrical values were added to each record of the dataset: 1) minimal anion-cation bond; 2) minimal distance between oxygen atoms; 3) first and second coordination number of anions around each cation; 4) first and second coordination number of cations around each cation. This data was calculated taking into account the periodicity of the structures. Coordination number displays the number of

closest atoms around a particular atom. Atoms are treated as closest if they fall inside the sphere, named the first coordination sphere, with a radius equal to the distance to the closest atom +5% of this distance. Additionally, coordination numbers are calculated for the second closest atoms excluding all atoms in the first coordination sphere.

The AFLOW consortium consists of more than 10 research groups, which add compounds to the database. Duplicates can therefore occur. Each record was compared to others and two records were defined as duplicates, when all the following rules were met:

1) Both compounds have the same number of atoms of the same elements.

2) The size and angles of unit cell differ less than 1%.

3) First and second coordination numbers of cations are the same. Due to periodic boundary conditions, atoms of the same crystal structure can have different coordinates, but their environment should be the same. Therefore, coordination numbers of each cation are compared rather than their coordinates.

1.1. Outlier selection rules.

Since our goal is to discover the possible relationship between OS and Bader charges, it is crucial to know the exact OS per compound on which the accuracy of our prediction depends. We consider records where oxygen is present in OS of -2 with high probability. The formula presented in Eq. (2) would give the wrong OS, if oxygen is present in -1 or 0 OS. To ensure that all atoms of oxygen are present in -2 state we disregard compounds using the following rules:

Rule 1.a: if they have the shortest bond between cation and anion larger than 130% of the sum of the covalent radii,[37] to ensure that each atom in a compound has a covalent bond (Figure 16a).

Rule 1.b: if there is an O-O bond length smaller than 1.65 Å. The OS of the oxygens in a O-O bond can be 0 or -1 depending on the multiplicity of the bond (double or single). The length of a double O=O bond is 1.21 Å, the length of a single O-O bond is 1.48 Å.[38] To take into account DFT errors,[39] the distribution of minimal O-O bond lengths for all compounds has been drawn in Figure

16b. The value of 1.65 Å was chosen due to the empty interval between bond lengths 1.65 and 1.95 Å.

Rule 1.c: if the calculated OS of the cation is higher than the number of its valence electrons. For example, K has only 1 valence electron, but the OS of K in a KO compound should be +2.

Rule 1.d: if the enthalpy of cell formation is positive. These compounds are energetically unfavorable, and they would not exist in the real world under normal conditions.

Rule 1.e: because of errors in the AFLOW database: three records in which some atoms have Bader volume equal to zero which have the following unique ICSD identifiers: #621706, #161062, and #173014.
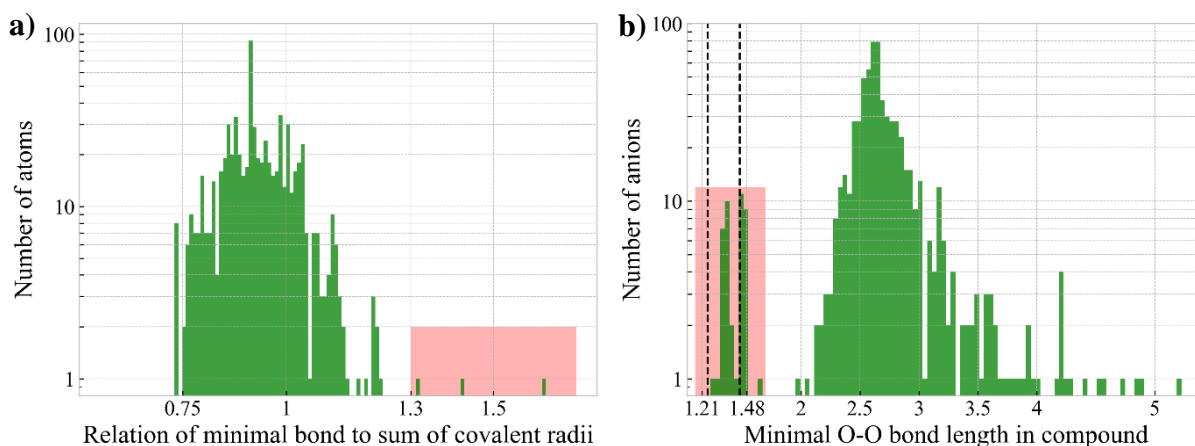
Figure 16. Distributions of the examined data set of binaries. Both vertical axes have a logarithmic scale. Records falling in the red region are disregarded from further examination: a) Distribution of the relation of minimal anion-cation bond to the sum of covalent radii (histogram consist of 100 bins); b) Distribution of minimal oxygen-oxygen bond length across compounds. Dashed lines are experimental values for double O=O and single O-O bonds (100 bins)

1.2. MV selection rules.

We group the MV compounds in a separate data set by the rules discussed below.

Rule 2.a: Compounds were regarded as MV compounds if the oxidation state of a cation is fractional. For example, Mn_3O_4 and Fe_3O_4 consist of cations in +2 and in +3 OS, resulting in the OS of +2.67 according to formula (2).

Rule 2.b: Compounds were regarded as MV compounds if the OS of a cation is integer and cations have a different environment. Formula (2) gives the correct OS for cations in compounds when the surrounding structure of all cations in the cell is the same. In contrast, the OS of cations in one compound can vary, when cations have a different environment. Formula (2) can give wrong results for compounds with an equal number of cations, which differ by two OSs (OSs equal to *n* and *n+2*). For example, in the case of antimony (III, V) oxide,[40] where the number of cations in +3 and +5 OSs is the same, the cations are at +4 OS according to formula (2). Coordination number can help in finding such compounds. We calculate separately the first and second coordination numbers of anions and cations around each cation, resulting in 4 coordination numbers for each cation in the compound. A compound is treated as an MV compound if any of the 4 coordination numbers for cations is not the same for all cations in the compound. Note that our calculated coordination numbers do not always reflect the strength of the bond or even the existence of the chemical bond between atoms. Therefore, in some cases even the cations with different coordination could be in the same OS. Compounds with integer OS but different coordination numbers for cations are discussed further below. Moreover, the classification of some compounds is redefined based on Bader charge difference.

To further refine *Rule 2.b*, we have calculated maximal Bader charge difference (that between the highest and the lowest for the cations in one compound) for each SV compound. Figure 17 representing the distribution of Maximal Bader charge difference of cations among SV compounds (compounds with only one cation in a compound were excluded) indicates that the maximal Bader charge difference is lower than 0.1e, and for 99.5% of records it is lower than 0.03e.
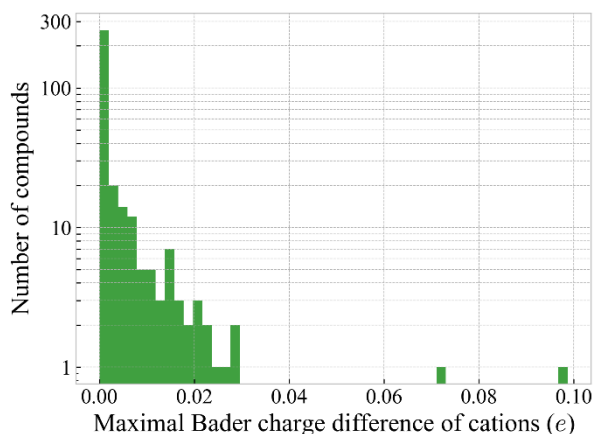
Figure 17. Distribution of maximal Bader charge difference of cations among SV compounds. The vertical axis has a logarithmic scale.

Using this result, we can update *Rule 2.b*. Even though in some compounds cations have different coordination numbers, they can be in the same OS if their maximal Bader charge difference is less than 0.1e. As mentioned earlier, coordination numbers do not necessarily have chemical meaning, in which case we treat the compounds as SV compounds. The updated rule allows redefining 131 out of 152 earlier defined MV compounds as SV.

To confirm that the remaining 21 MV compounds are MV compounds with cations of OS *n* and *n+2*, a deeper look should be taken at Bader charges of SV compounds plotted in Figure 17. As can be seen, Bader charges form intervals, which do not intersect between oxidation states inside each binary. For example, Bader charge intervals for niobium oxides (Nb group 5) are the following: for + 2 OS - [1.09, 1.55], + 3 - [1.86, 1.87], +4 - [2.37, 2.4], +5 - [2.72, 2.81].

We have searched for minimal Bader charge difference between OSs that differ by 2 for each element, to find that nickel oxides have Bader charge difference of 0.133e between +2 and +4 OS. The rest of the binaries have larger minimal Bader charge differences between n and n+2 OS. Moreover, 21 remaining MV compounds have minimal Bader charge difference between cations that is greater than 0.133e. Therefore, they were confirmed to be MV compounds with the exception of four compounds with the highest OS for cation (Ti, Si in +4, W in +6 and Ru in +8).

If the cations in the four compounds had OS differing by 2 it would mean that Ti, Si would be in +3 and +5, W in +5 and +7, Ru in +7 and +9, which is unrealistic. In our work, these exceptions are assumed to be SV compounds, and their Bader charges are plotted in Figure 3 along with the other SV compounds.

The above sets of selection rules thus divide the data set into three parts, yielding the workflow shown in Figure 18. The first part is disregarded from further examination, eliminating 9% of all records. The second part consists of MV compounds. Finally, the third part consists of SV compounds.
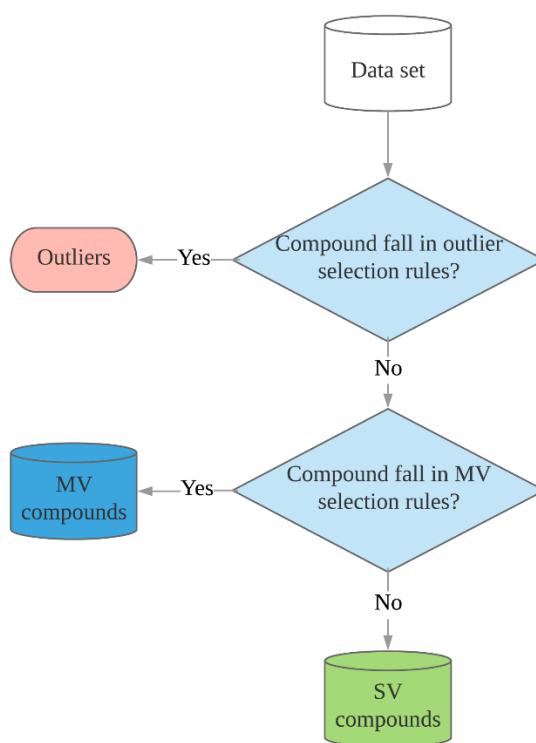
Figure 18. Workflow for dividing the data set

References

1. Russell, C. A. *History of Chemistry A History of the Concept of Valency to 1930. By W. G. Palmer. Pp. viii + 178. London: Cambridge University Press, 1965. 42s. The British Journal for the History of Science* **3,** (Cambridge University Press, 1966).

2. Heckmann, A. & Lambert, C. Organic mixed-valence compounds: A playground for electrons and holes. *Angewandte Chemie - International Edition* **51,** 326–392 (2012).

3. Day, P., Hush, N. S. & Clark, R. J. . Mixed valence: origins and developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **366,** 5–14 (2008).

4.   Jansen, M. & Wedig, U. A piece of the picture - Misunderstanding of chemical concepts. *Angewandte Chemie - International Edition* **47,** 10026–10029 (2008).

5.   Karen, P., McArdle, P. & Takats, J. Comprehensive definition of oxidation state (IUPAC Recommendations 2016). *Pure Appl. Chem.* **88,** 831–839 (2016).

6.   Karen, P., Mcardle, P. & Takats, J. IUPAC Technical Report Toward a comprehensive definition of oxidation state (IUPAC Technical Report). *Pure Appl. Chem* **86,** 1017–1081 (2014).

7.   Bader, R. F. W. *Atoms in Molecules: A Quantum Theory. Oxford University Press* **36,** (1990).

8.   Mulliken, R. S. Electronic Population Analysis on LCAO[Single Bond]MO Molecular Wave Functions. I. *J. Chem. Phys.* **23,** 1833 (1955).

9.   Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* **97,** 10269–10280 (1993).

10.   Löwdin, P. O. On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. *J. Chem. Phys.* **18,** 365–375 (1950).

11.   Koch, D. & Manzhos, S. On the Charge State of Titanium in Titanium Dioxide. *J. Phys. Chem. Lett.* **8,** 1593–1598 (2017).

12.   Walsh, A., Sokol, A. A., Buckeridge, J., Scanlon, D. O. & Catlow, C. R. A. Electron Counting in Solids: Oxidation States, Partial Charges, and Ionicity. *Journal of Physical Chemistry Letters* (2017). doi:10.1021/acs.jpclett.7b00809

13.   Jones, R. O. Density functional theory: Its origins, rise to prominence, and future. *Rev. Mod. Phys.* **87,** (2015).

14.   Curtarolo, S. *et al.* AFLOW: An automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58,** 218–226 (2012).

15.   Curtarolo, S. *et al.* AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58,** 227–235 (2012).

16.   Jain, A. *et al.* Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials* **1,** (2013).

17.   Haupt, T., Sukhija, N. & Horstemeyer, M. F. Cyberinfrastructure support for engineering virtual organization for CyberDesign. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **7204 LNCS,** 161–170 (2012).

18.   NOMAD Repository. Available at: http://nomad-repository.eu/. (Accessed: 22nd May 2018)

19.   Isayev, O. *et al.* Universal fragment descriptors for predicting properties of inorganic crystals. *Nat. Commun.* **8,** 15679 (2017).

20.   Isayev, O. *et al.* Materials Cartography: Representing and Mining Materials Space Using Structural and Electronic Fingerprints. *Chem. Mater.* **27,** 735–743 (2015).

21.   Davies, D. W., Butler, K. T., Isayev, O. & Walsh, A. Materials discovery by chemical analogy: role of oxidation states in structure prediction. *Faraday Discuss.* (2018). doi:10.1039/C8FD00032H

22.   Reeves, K. G. & Kanai, Y. Theoretical oxidation state analysis of Ru-(bpy)3: Influence of water solvation and Hubbard correction in first-principles calculations. *J. Chem. Phys.*

141, (2014).

23.   Calderon, C. E. *et al.* The AFLOW standard for high-throughput materials science calculations. *Comput. Mater. Sci.* **108,** 233–238 (2015).

24.   Taylor, R. H. *et al.* A RESTful API for exchanging materials data in the AFLOWLIB.org consortium. *Comput. Mater. Sci.* **93,** 178–192 (2014).

25.   Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B - Condens. Matter Mater. Phys.* **54,** 11169–11186 (1996).

26.   Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented - wave method. *Phys. Rev. B* (1999). doi:10.1103/PhysRevB.59.1758

27.   Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* (1994). doi:10.1103/PhysRevB.50.17953

28.   Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* (1996). doi:10.1016/0927-0256(96)00008-0

29.   Dudarev, S., Botton, G., Savrasov, S., Humphreys, C. & Sutton, A. Electron-energy-loss spectra and the structural stability of nickel oxide: An LSDA+ U study. *Phys. Rev. B* **57,** 1505–1509 (1998).

30.   Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77,** 3865–3868 (1996).

31.   Krukau, A. V., Vydrov, O. A., Izmaylov, A. F. & Scuseria, G. E. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *J. Chem. Phys.* **125,** 224106 (2006).

32.   Yu, M. & Trinkle, D. R. Accurate and efficient algorithm for Bader charge integration. *J. Chem. Phys.* **134,** (2011).

33.   Ma, Y. *et al.* Titanium dioxide-based nanomaterials for photocatalytic fuel generations. *Chemical Reviews* **114,** 9987–10043 (2014).

34.   Robin, M. B. & Day, P. Mixed valence chemistry. A survey and classification. *Adv. Inorg. Chem. Radiochem.* **10,** 247–422 (1967).

35.   Nolan, M. *et al.* Electronic structure of point defects in controlled self-doping of the TiO2 (110) surface: Combined photoemission spectroscopy and density functional theory study. *Phys. Rev. B - Condens. Matter Mater. Phys.* **77,** 235424 (2008).

36.   Lu, G., Linsebigler, a & Yates, J. T. Ti3+ Defect Sites on TiO2(110): Production and Chemical Detection of Active Sites. *J. Phys. Chem.* **98,** 11733–11738 (1994).

37.   Cordero, B. *et al.* Covalent radii revisited. *Dalt. Trans.* 2832 (2008). doi:10.1039/b801115j

38.   Lide, D. R. CRC Handbook of Chemistry and Physics. *eBook* 3485 (2003). doi:978-1466571143

39.   Lejaeghere, K., Van Speybroeck, V., Van Oost, G. & Cottenier, S. Error estimates for solid-state density-functional theory predictions: An overview by means of the ground-state elemental crystals. *Critical Reviews in Solid State and Materials Sciences* **39,** 1–24 (2014).

40.   Amador, J., Gutiérrez Puebla, E., Monge, M. A., Rasines, I. & Ruíz Valerol, C. Diantimony tetraoxides revisited. *Inorg. Chem.* **27,** 1367–1370 (1988).

Supplementary material

Oxidation states of binary oxides from data analytics of the electronic structure

S. Posysaev[1], O. Miroshnichenko[1], M. Alatalo[1], D. Le[2*], and T. S. Rahman[2,3]

[1]Nano and Molecular Systems, Faculty of Science, BOX 3000, FIN-90041, University of Oulu, Finland

[2]Department of Physics, University of Central Florida, Orlando, FL 32816, USA.

[3]Department of Applied Physics, Aalto University, Aalto, Finland.

22.10.2018

*Corresponding author (email: Duy.Le@ucf.edu)

1. Linear dependency in SV compounds

The linear regression graphs for 12 binaries which have least 3 OS in our data set are presented in Figure S 1, the slope of lines varies among binaries.
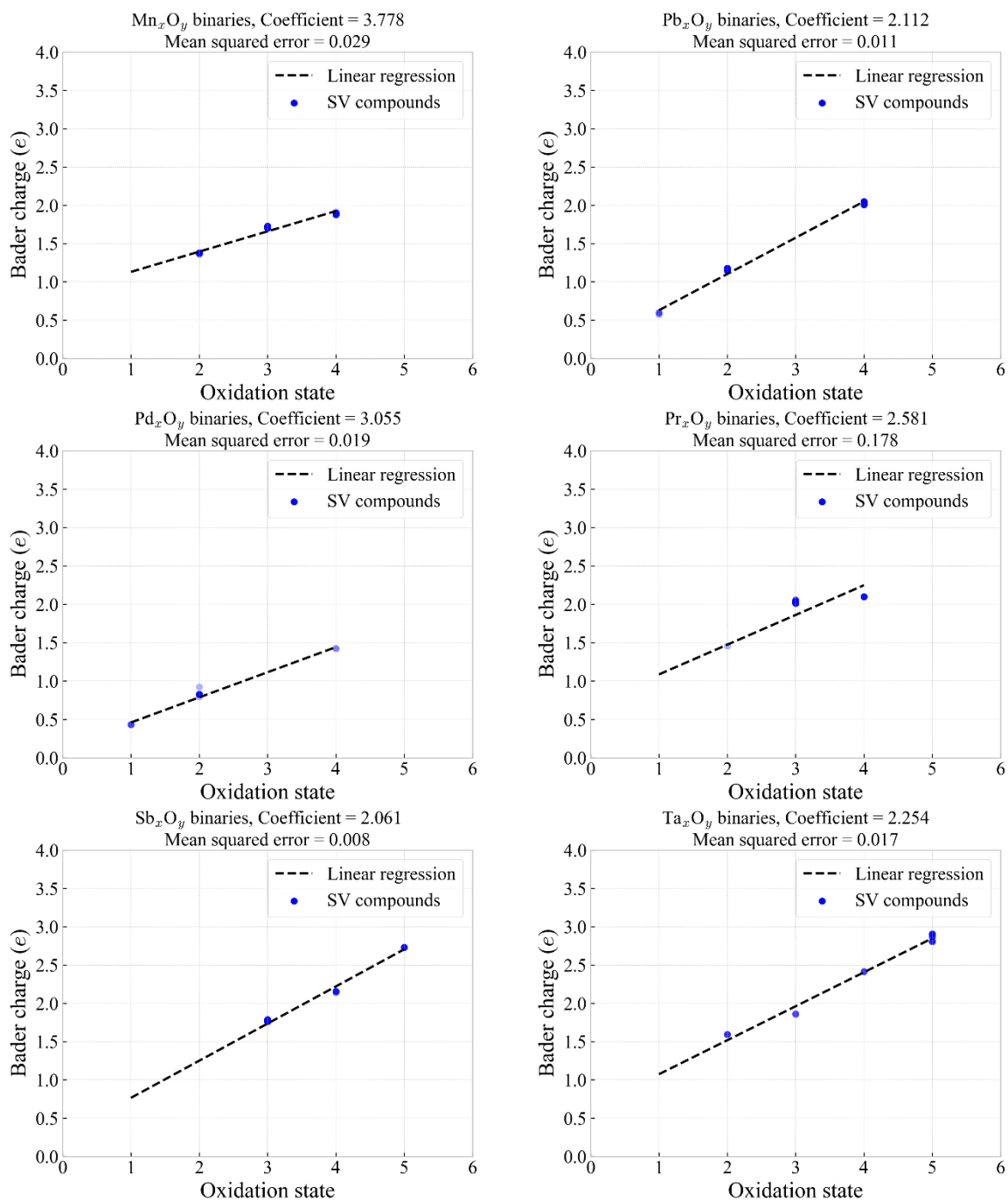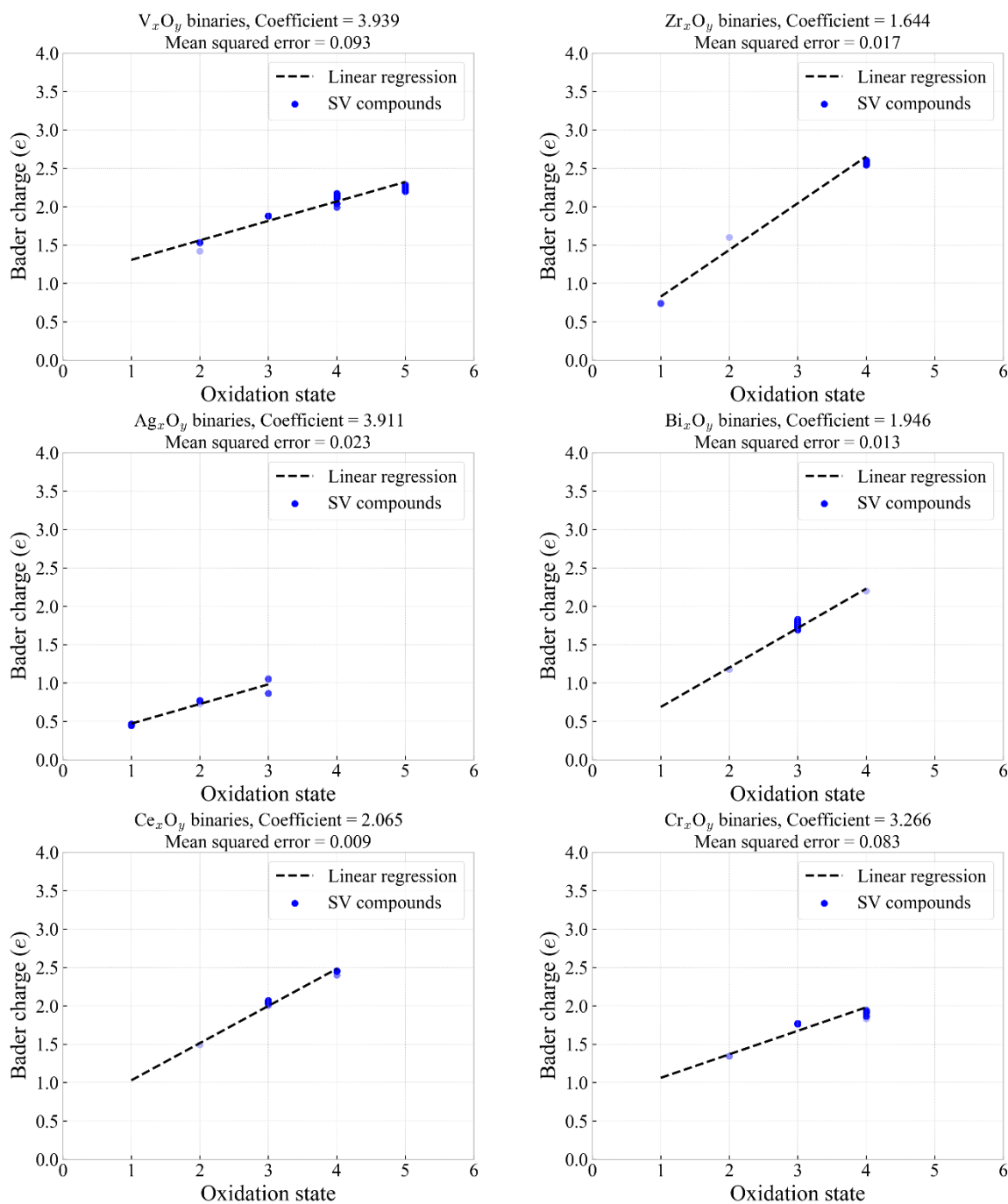
Figure S 1 Linear regression for Zr, V, Ta, Cr, Mn, Pd, Ag, Pb, Sb, Bi, Ce, Pr.

2. Example of adjusting bisected intervals for manganese oxides

In Figure S2 we present an example of the process of adjusting bisected intervals for manganese oxides. In the first column of Figure S2, one can see Bader charges for SV manganese oxides. The charges cover only part of the Bader charge space. For MV binaries charges can differ, but intervals of different OSs should not overlap between each other. In order to further refine intervals, Bader charge space is divided into bisected intervals based on charges from SV binaries and defined as follows: Bader charge space is divided in the middle (dashed lines in Figure S2) between interval maximum for n OS and interval minimum for $n+1$ OS (for example, middle value between maximum charge for OS +2 and minimum charge for OS +3).
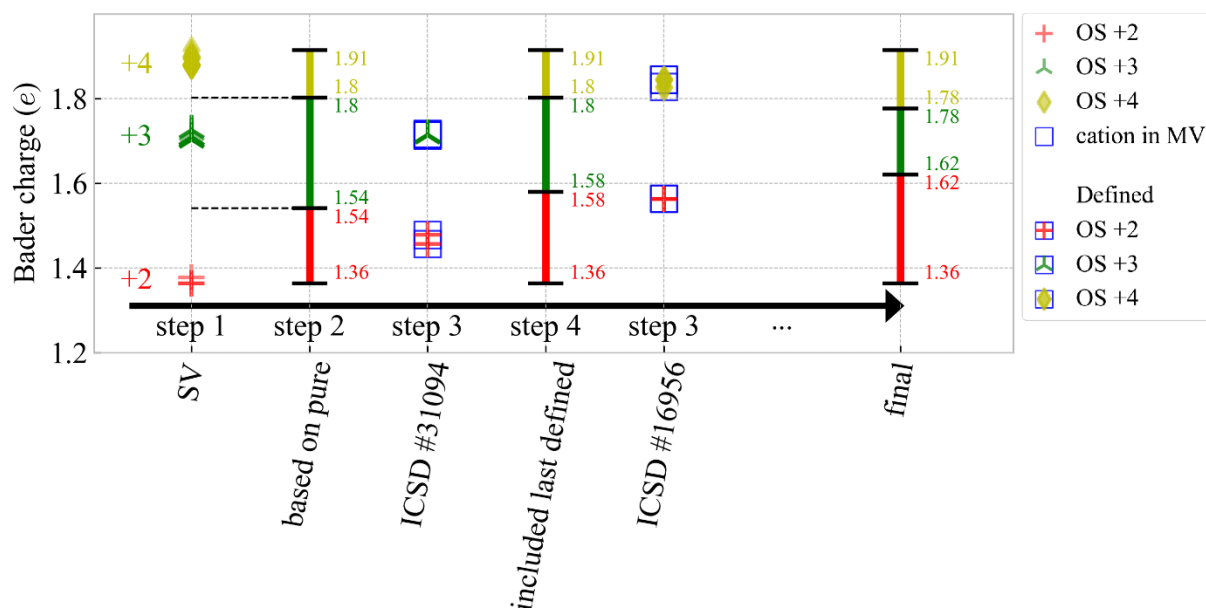
Figure S2. Workflow of adjusting bisected intervals for manganese oxides.

In the case where OSs were correctly defined, new bisected intervals are made taking into account Bader charges for the last defined compound. For example for manganese binary (Figure S2), we found the following Bader charges for SV compounds: for OS +2 [1.36, 1.38] e, for OS +3 [1.70, 1.73] e, for OS +4 [1.88, 1.91] e (step 1); the resulting bisecting intervals are: +2: [1.36, 1.54] e, +3: [1.54, 1.80] e, +4: [1.80, 1.91] e (step 2). We take one MV compound with ICSD #31094 from the group 2, in which case Bader charges for cations are the following: [1.46, 1.48, 1.72, 1.71, 1.71, 1.71] e (step 3). The OS of cations by formula (2) is 2.66, the sum of OS is equal to 16 by formula (1). All Bader charges lie in the defined bisected intervals, 1.46, 1.48e lie in the interval for +2 OS, 1.72, 1.71, 1.71, 1.71e lie in the interval for +3OS. The total sum of defined OSs is equal to the sum of OS calculated by the formula (1). We assume that OSs are correctly defined. The next step is to redefine bisected intervals since now we have a new Bader charge maximum for OS +2 (step 4). Previously undefined MV compounds will be checked with new bisected intervals. Making new intervals specifies the range of Bader charges for each OS, which improves the defining of OS of cations in the rest of MV compounds. The cycle repeats until no new compounds are found.

3. Determining oxidation states of MV binaries

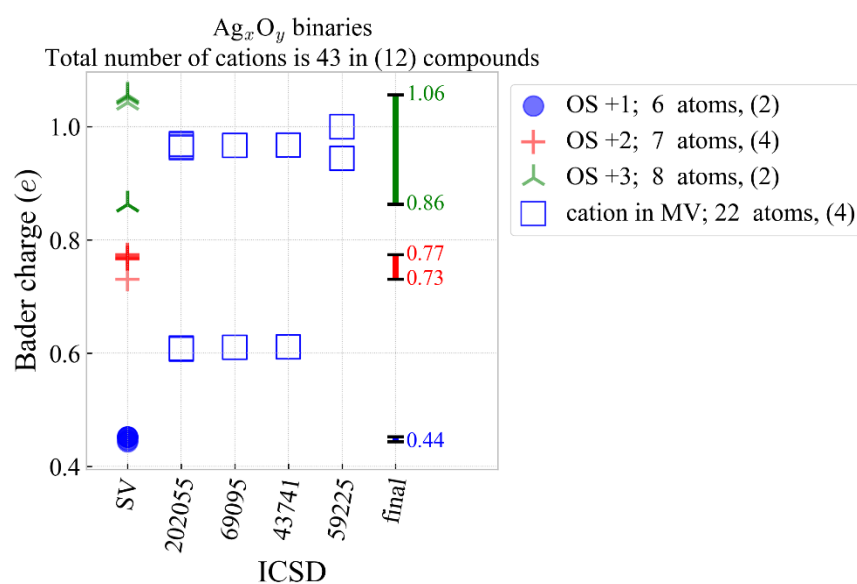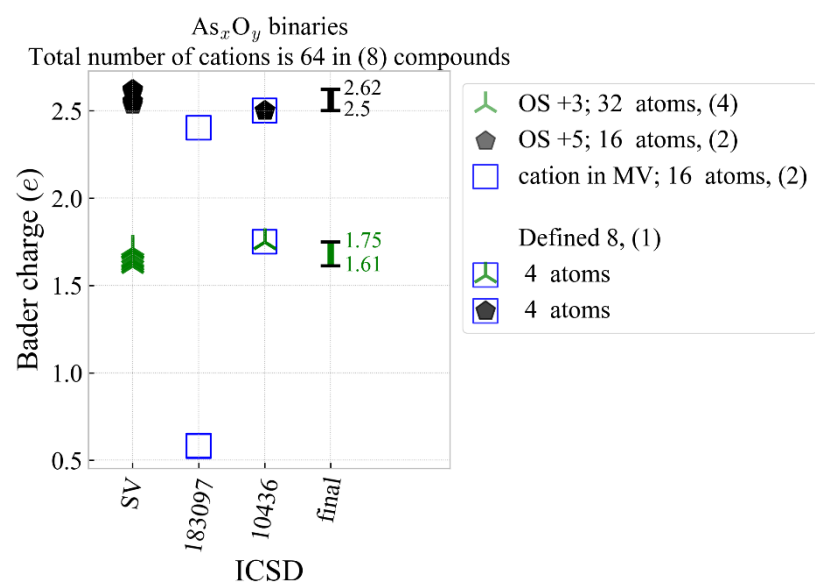Below presented results for determining oxidation states in MV binaries.
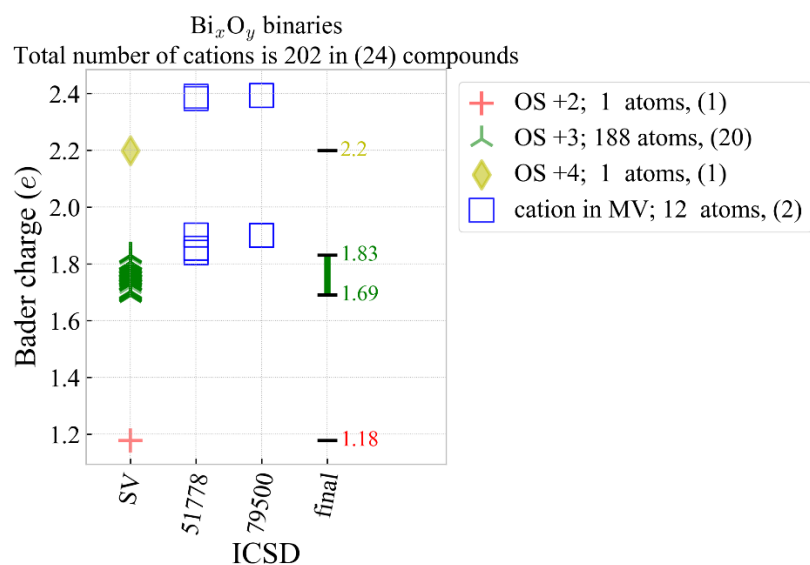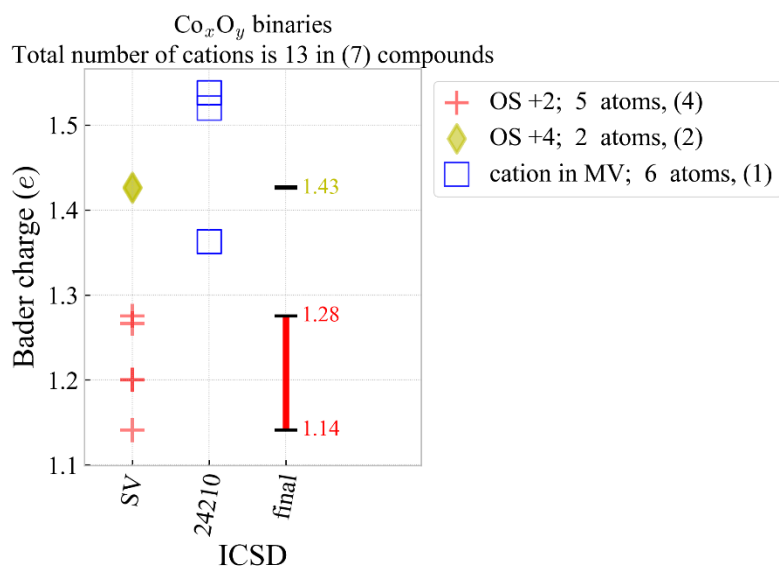

Figure S3

Figure S4

Bi$_x$O$_y$ binaries
Total number of cations is 202 in (24) compounds

OS +2; 1 atoms, (1)
OS +3; 188 atoms, (20)
OS +4; 1 atoms, (1)
cation in MV; 12 atoms, (2)

Figure S5

Co$_x$O$_y$ binaries
Total number of cations is 13 in (7) compounds

OS +2; 5 atoms, (4)
OS +4; 2 atoms, (2)
cation in MV; 6 atoms, (1)

Figure S6

Cr$_x$O$_y$ binaries
Total number of cations is 39 in (11) compounds

OS +2; 1 atoms, (1)
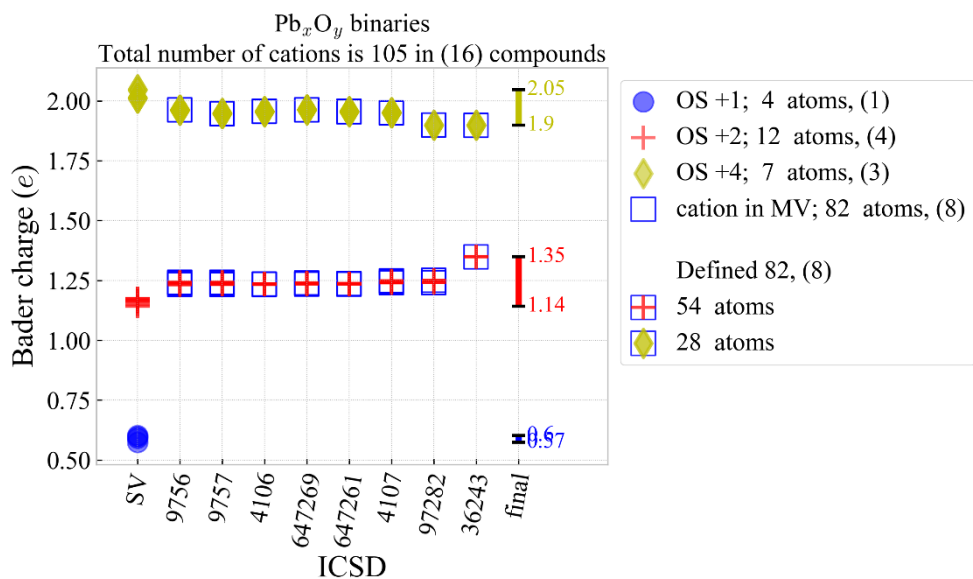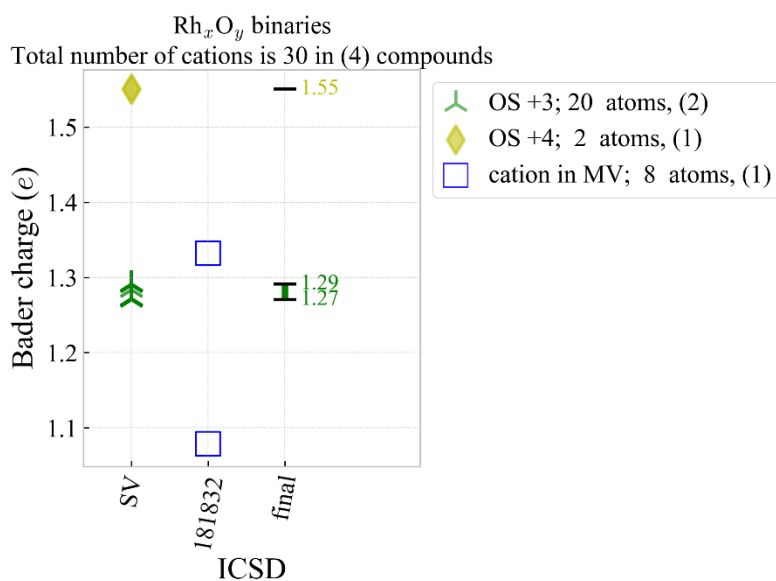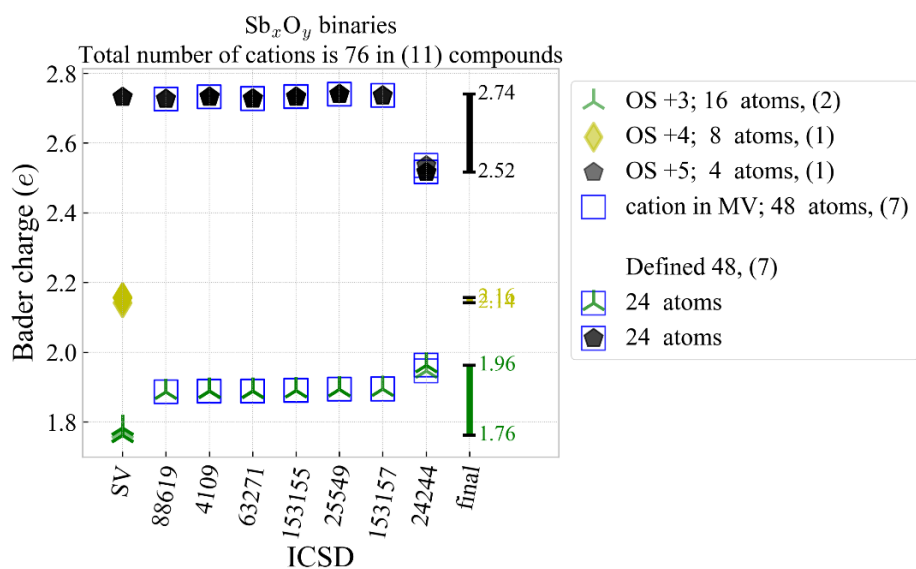OS +3; 4 atoms, (1)
OS +4; 15 atoms, (6)
cation in MV; 19 atoms, (3)

Figure S7

36
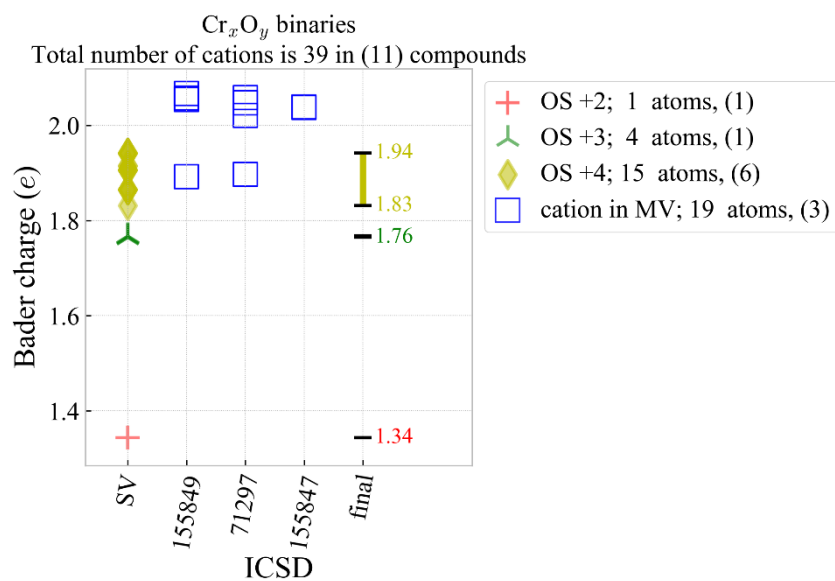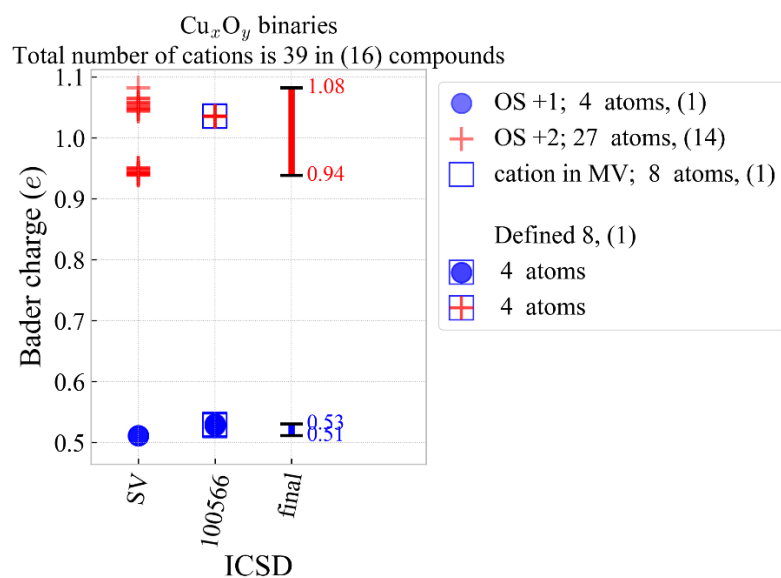
Figure S8

Figure S9

Figure S10

37

Nb$_x$O$_y$ binaries
Total number of cations is 37 in (10) compounds

Figure S11

P$_x$O$_y$ binaries
Total number of cations is 32 in (4) compounds

Figure S12

Pb$_x$O$_y$ binaries
Total number of cations is 105 in (16) compounds

Figure S13

38

Pr_xO_y binaries
Total number of cations is 33 in (6) compounds

OS +2; 1 atoms, (1)
OS +3; 24 atoms, (3)
OS +4; 1 atoms, (1)
cation in MV; 7 atoms, (1)

Figure S14

Pt_xO_y binaries
Total number of cations is 22 in (10) compounds

OS +2; 3 atoms, (2)
OS +4; 10 atoms, (6)
cation in MV; 9 atoms, (2)

Figure S15

Rh_xO_y binaries
Total number of cations is 30 in (4) compounds

OS +3; 20 atoms, (2)
OS +4; 2 atoms, (1)
cation in MV; 8 atoms, (1)

Figure S16

39

Figure S17

Figure S18

Figure S19

40
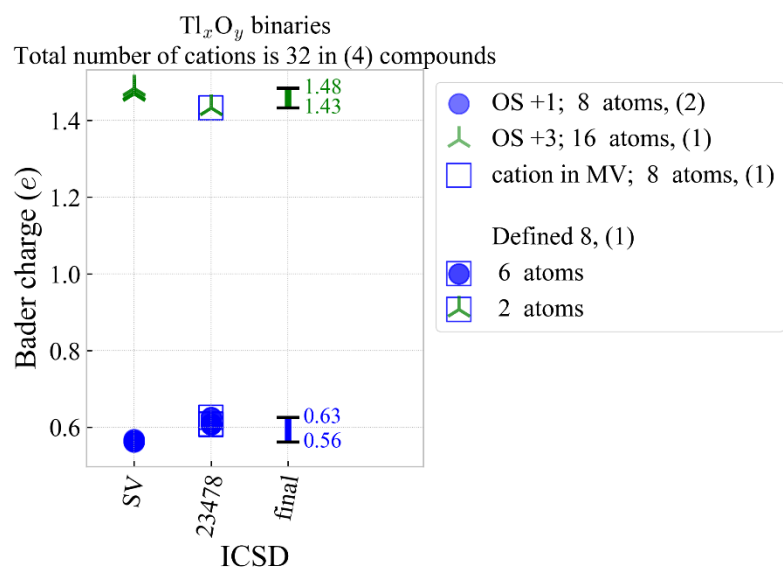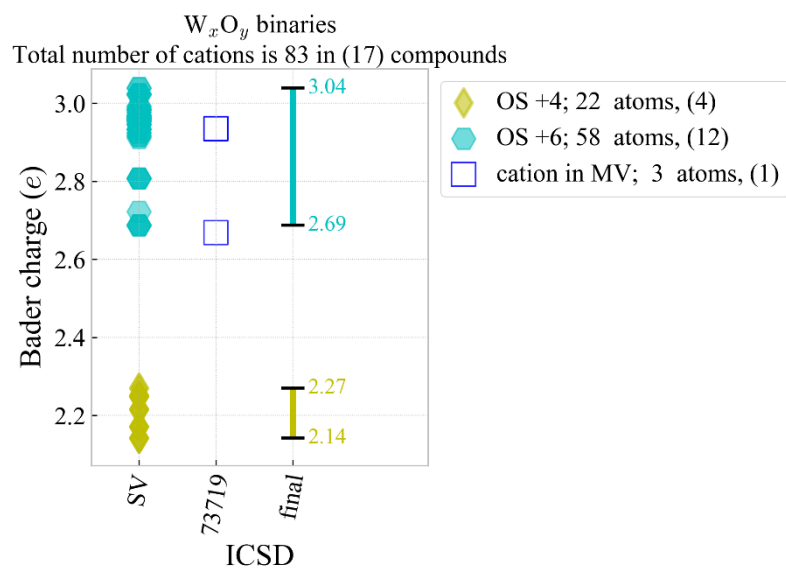
Figure S20

Figure S21

Oxide	Total MV compounds	Defined	Undefined	Not enough data
Ti	18	12	6	
V	12	9	3	
Fe	10		10	
Pb	8	8		
Sb	7	7		
Mn	5	5		
Ag	4		4	
Tb	3		3	
Cr	3			3
Pt	2		2	
As	2	1		1
Te	2	2		
Bi	2			2
W	1	1		
Rh	1			1
Mo	1		1	
P	1	1		
Nb	1	1		
Cu	1	1		
Co	1		1	
Pr	1		1	
Tl	1	1		
Total	**87**	**49**	**31**	**7**

Table S1 Summary of MV compounds.