AB_2X_6 Compounds and the Stabilization of Trirutile Oxides

Emily C. Schueller,[†] Yuzki M. Oey,[†] Kyle D. Miller,[‡] Kira E. Wyckoff,[†]

Ruining Zhang,[¶] William Zhang,[¶] Stephen D. Wilson,[†] James M. Rondinelli,[‡] and

Ram Seshadri[∗,†,§]

[†]*Materials Department and Materials Research Laboratory*

University of California, Santa Barbara, Santa Barbara, CA, 93106

[‡]*Department of Materials Science and Engineering, Northwestern University, Evanston,*

Illinois 60208, USA

[¶]*Materials Research Laboratory*

University of California, Santa Barbara, Santa Barbara, CA, 93106

[§]*Department of Chemistry and Biochemistry*

University of California, Santa Barbara, Santa Barbara, CA, 93106

E-mail: seshadri@mrl.ucsb.edu

Abstract

The properties of crystalline materials tend to be strongly correlated with their structures, and the prediction of crystal structure from only the composition is a coveted goal in the field of inorganic materials. However, even for the simplest compositions, such prediction relies on a complex network of interactions, including atomic or ionic radii, ionicity, electronegativity, position in the periodic table, and magnetism, to name only a few important parameters. We focus here on the AB_2X_6 (AB_2O_6 and AB_2F_6) composition space with the specific goal of finding new oxide compounds in the trirutile family, which is known for unusual 1D antiferromagnetic behavior. Through machine learning methods, we develop an understanding of how geometric and bonding constraints determine the crystallization of compounds in the trirutile structure as opposed to other other ternary structures in this space. In combination with density functional theory (DFT) calculations, we predict 18 candidate, previously unreported trirutile oxides. We successfully prepare one of these and show it forms in the disordered rutile structure, under the preparation conditions adopted here.

Introduction

Machine learning is an emerging method in materials science which has been used to predict properties of materials such as hardness,[1,2] band gap,[3] and drive the search for new materials with desirable properties such as high performing transparent conductors.[4] For these predictive applications, machine learning models can outperform *ab initio* methods such as density functional theory (DFT) because the low computational cost of running a machine learning algorithm allows for fast screening of thousands of compounds. Machine learning is consequently a promising method for filtering an initial large pool of candidates into a sample size on which it is feasible to perform DFT or other computationally intensive simulations.[5] Machine learning has also been used to build on fundamental understanding of well known materials; for example, determination of factors which drive crystallization

in the Heusler as opposed to the inverse Heusler and related structures.[6]

The prediction of crystal structure from composition has been a longstanding goal in inorganic materials science hearkening back to 1929, when Linus Pauling published the now eponymous rules to predict the coordination environment and bonding of many families of inorganic materials.[7] These deceptively simple rules, which take into account relative atomic radii, valence, and mixed-cation behavior, explain many of the fundamental interactions that determine crystal structure formation. This work was extended by Slater,[8] and later Shannon,[9] in the determination of atomic and ionic radii for atoms in crystals. For valence precise binary semiconductors, Phillips and Van Vechten[10] formulated the more specific and sophisticated rules for the separation of rock-salt, zinc blende, and wurtzite crystal structures. Since these early developments, the understanding of crystal structure has been explored for several structural motifs.[11,12] A resurgence of interest in the topic occurred with the development of high-throughput DFT-based modeling in conjunction with improvements in computing capabilities,[13–15] and crystal structure prediction (or perhaps more accurately, screening) studies have since been performed on both specific families as well as general classes of materials. The relationship between a physical property and its descriptive parameters has also been studied in an attempt to determine how to choose a meaningful descriptor.[16] Completely unbiased (by any knowledge of crystal structures) approaches, such as the use of genetic algorithms,[17] and particle swarm optimization,[18,19] have also become available, but these methods, while very successful in specific instances, tend to be computationally exceedingly expensive. Therefore, we chose a machine learning method over genetic algorithm alternatives because machine learning allows us to gain chemical understanding about the ternary composition space.

In this work, we use machine learning and DFT to study crystal structure formation in AB_2X_6 oxides and fluorides composition space, which is a complex space containing many material families. In particular, we explore the underlying factors which determine whether an AB_2X_6 compound will crystallize in the trirutile structure with the goal of

predicting new tirutile oxides. The trirutile family of materials with formula AB_2X_6 are known for unusual low dimensional magnetism,[20] typically in the form of 1D antiferromagnetic chains along the [110] direction,[21–23] which is the closest metal-metal distance in the structure. The trirutile structure is an ordered supercell of the rutile unit cell with A-B-B ordering in the c-direction. Because of the similarity to the rutile family, which hosts various canonical insulator-to-metal transition materials such as VO_2,[24] it is possible that some trirutiles could have a similar structural instability to the one which leads to the insulator-to-metal in rutile materials.[25] However, even in the high symmetry ideal trirutile structure and in the absence of dimerization, all known oxide trirutiles appear to be insulators, possibly due to the disruption of conduction pathways along the edge sharing octahedra from electron localization or charge disproportionation on A and B ions. Many trirutile compounds with unpaired d-electrons such as $CuSb_2O_6$ and $NiTa_2O_6$ have been suggested as exhibiting Mott or charge-transfer insulating properties.[26] Through machine learning methods we find that, consistent with factors determining crystallization in other structural families, geometric and bonding constraints are the most important features determining the formation of a trirutile structure. In particular, the trirutile structure is preferred over others when both the A and B atoms are relatively small, and less electronegative.

Starting from 461 novel AB_2O_6 compositions, we predict 53 new candidate trirutile materials via machine learning. From DFT calculations, we find 18 of the 53 have a formation energy in the trirutile structure that is less than $50\,\text{meV/atom}$ above the formation energy of the constituent binary oxides. From these 18 compounds, we prepared two novel AB_2O_6 compounds, $TiTa_2O_6$ and $CrSb_2O_6$ and find that they form, but in a disordered rutile $(A_{1/3}, B_{2/3})O_2$ under the conditions employed, rather than in an ordered trirutile structure. Under the appropriate preparative conditions, it is possible that ordered compounds would indeed be accessible. The work points to the importance of accounting for synthesizability beyond simple stability calculations, in the prediction new compounds. None of the

predicted AB_2O_6 compounds we attempted formed non-rutile ternary structure types, indicating the relative success of the machine learning model in predicting preferred local coordination environments based only on composition. This suggests the model has identified the most important chemical features which drive the selection of the trirutile crystal structure.

Methods

Computational Methods

Data on existing AB_2X_6 oxide and fluoride compounds was collected from the Inorganic Crystal Structure Database (ICSD).[27] The data was filtered to remove disordered materials and duplicate entries. For polymorphic structures the stable room temperature structure was chosen, and in cases where multiple stable room temperature structures had been reported, the structure belonging to a known structure type was selected. The data was annotated with binary targets (1 = trirutile, 0 = other) in order to set up a classifier algorithm. The cleaned data contained 293 compounds of which 36 were trirutiles, so the dataset was imbalanced with only approximately 12% positives.

A brief summary is presented here, and full details of the model, hyperparameters, feature selection, and data scaling and cleaning can be found in in the Supplementary Information, as well as comparison of our model with other models from the SCIKIT-LEARN package.[28] Featurization of the starting materials was performed in MATMINER[29] using composition-based features such as atomic radius and electronegativity, as well as features based on the guessed oxidation states from the chemical formula, such as the number of valence electrons. The featurized matrix was cleaned, removing features for which data was missing for more than 3% of samples, using the AUTOMATMINER pipeline.[30] The features were downselected for the most important features in multiple steps. First, features which were correlated >95% were removed. Then features which sum to <1% impor-

tance by weight within a random forest algorithm were removed. This brought the initial
>600 features down to 46. The starting matrix with 46 features was input into a genetic
algorithm for further preprocessing and machine learning optimization, as an alternative
to grid-based cross validation methods.[30,31] Within the algorithm, half of the 46 features
were removed with recursive feature elimination via an extra trees classifier. Tree-based
feature reduction methods provide an advantage over principal component analysis (PCA)
because they allow retention of the feature names which provide insight into the most
important features for predicting within the machine learning model. Polynomial features
up to order 2 were implemented, so a "feature" in the model could be a multiplication of
two features or a single feature squared. We chose the F_1 score, a combination of recall
and precision, as an optimization metric due to the imbalanced nature of the dataset. Ad-
ditionally, we focused on the recall capabilities of the model, because with more negatives
than positives, models tend to predict false negatives and still perform relatively well. The
machine learning algorithm was trained on 75% of the data in a stratified manner to ob-
tain a similar number of positives in the training and test set. Selection of the machine
learning model and hyperparameters was performed in AUTOMATMINER using nested cross
validation to obtain the model with the best and most reproducible F_1 score. The best
model was an extra trees classifier with a F_1 score on both training and test data of 82%,
indicating the model was not over-fitting the data.

For the prediction of new oxide trirutile candidates, common valences of the elements
(excluding rare earth elements) were tabulated and combinations of $A + 2 \times B$ which
combined to $12(=6 \times 2)$ were input in a matrix. Compounds which existed already were
removed and the model was implemented on 461 new AB_2O_6 formulas. 53 were predicted
as possible trirutile candidates, similar to the ratio of positives in known materials.

For the 53 predicted candidates, density functional theory (DFT) calculations were
used to calculate the energy of the ternary compounds in the trirutile structure com-
pared with the energy of their reported binaries. All electronic structure calculations

were performed using the Vienna Ab–Initio Simulation Package (VASP)[32] with projector-augmented-waves[33,34] within the Perdew–Burke–Ernzerhof generalized gradient approximation.[35] All calculations used an automatically generated Γ–centered k–point grid with a density of 50 and an energy cutoff of 550 eV with a ferromagnetic initialization on non d^0 transition metals. In binary compounds where the magnetism is known, the nearest collinear magnetic configuration was used to initialize the calculations. Structural relaxations with symmetry on were performed for all materials in a 3 step process in which the unit cell and ion positions were allowed to relax until convergence, then just the ion positions, and then a static energy calculation. Gaussian smearing with a sigma of 0.1 was used for the structural relaxations, and the Blöchl tetrahedron method was used for the final static energy calculations to obtain the most accurate total energy. We used the Phonopy package with unit-cell structures for calculation of the phonon spectrum at Γ.[36]

Experimental Methods

$TiTa_2O_6$ powder was obtained by solid state reaction of TiO and Ta_2O_5 powders, and $CrSb_2O_6$ was obtained by reaction of Cr_2O_3, Sb_2O_3, and Sb_2O_5. Stoichiometric amounts totaling around 500 mg of the starting materials were ground with an agate mortar and pestle and pressed into a 6 mm pellet under 950 MPa of pressure. The pellets were placed in Al_2O_3 crucibles capped with Ta foil to prevent reaction with the fused silica tubes. The samples were sealed in silica ampoules under vacuum with a 0.25 atm Ar to suppress volatilization. Both samples were slowly ramped to 1100°C over 3 days, reacted for 5 to 7 days, and allowed to cool in the furnace. Laboratory Cu-Kα X-ray diffraction (XRD) was acquired on a PANalytical Empyrean diffractometer. $CrSb_2O_6$ formed a phase-pure disordered rutile structure, while $TiTa_2O_6$ formed the same but with a small Ta_2O_5 impurity. Slow cooling and re-annealing at lower temperatures was performed in an attempt to order the cations into a trirutile structure but a tripling of the unit cell was never observed within lab resolution. X-ray fluorescence measurements were performed on a Rigaku

ZSF Primus IV spectrometer on pressed pellets of the powder material to confirm composition in $TiTa_2O_6$. To measure temperature–dependent and field–dependent magnetization in $TiTa_2O_6$, a Quantum Design Superconducting Quantum Interface Device (SQUID) Magnetic Property Measurement System (MPMS) with a vibrating sample magnetometer (VSM) was used. 7.4 mg of sample power was packed into a plastic sample holder and loaded into a brass sample holder rod. Temperature– dependent magnetization data were acquired between 2 K and 380 K under a field of $H = 200$ Oe under zero-field cooled and field-cooled conditions, and under a field of $H = 2000$ Oe under field-cooled conditions. A magnetization *vs.* field hysteresis loop was acquired at 2 K in a field between 5 T and –5 T. The sample was measured using a Thermo Fisher Escalab Xi+ XPS equipped with a monochromatic Al anode (E = 1486.7 eV). A survey scan was measured at 100 eV pass energy and high-resolution scans were measured in the Ti 2p and Ta 4d regions at 20 eV pass energy. The spectra were referenced to adventitious carbon at 284.8 eV. CasaXPS was used to fit the data using Shirley backgrounds and SGL peak shapes. High resolution scans of Ti and Ta were fit using appropriate spin-orbit splitting and peak area ratios.

Results

The composition space AB_2X_6 oxides and fluorides, with approximately 293 known compounds, is distributed over a variety of structure types. Figure 1 highlights some representative structure types in this space. While the trirutile family is the most common, with 36 members representing approximately 12% of AB_2X_6 compounds, no one type is preferred within this composition region. There is varying coordination of the A and B cations although octahedra are a common motif. Two types that appear similar to the trirutile structure are the Na_2SiF_6 and the columbite structure, in which both A and B atoms are octahedrally coordinated in alternating fashion in 3 dimensions.

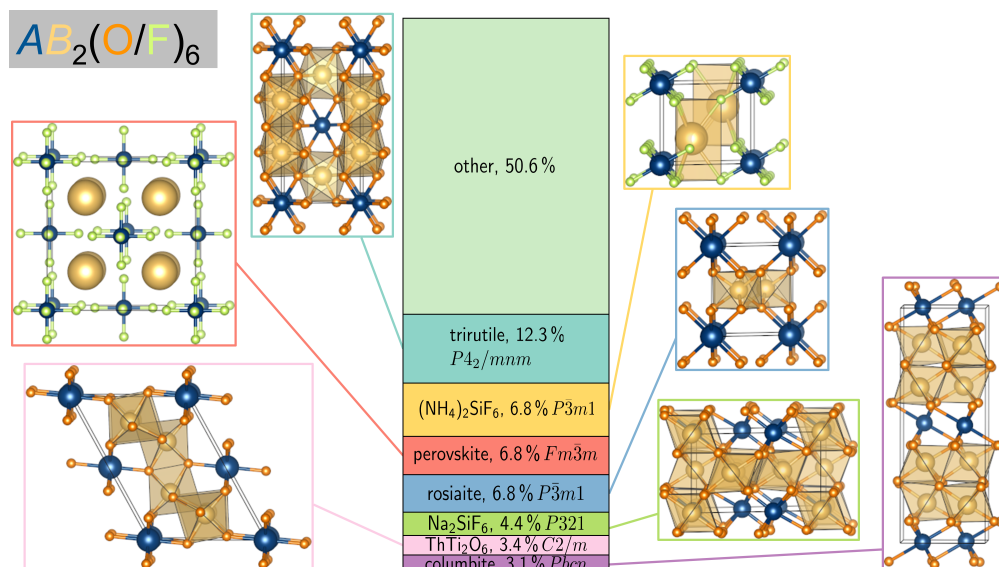Starting from the 293 known AB_2X_6 oxide and fluoride compounds from the ICSD, we

Figure 1: The most common structure types in AB_2X_6 oxide and fluoride composition space. No one structure type dominates the landscape in this region. Most structures are characterized by octahedral coordination of the A or B sites, and layered compounds feature prominently both in the common structure types and in the "Other" class.

trained a machine learning model to classify the compounds into trirutile or not trirutile. The process flow is shown in figure 2. Details of the machine learning process and model can be found in the methods and the SI. The best performing model was an extra trees classifier, a form of decision tree, with a cross validation F_1 score on the training data of 82% and on the test data of 82%. The F_1 score, a combination of precision and recall, is a good metric for unbalanced datasets where negatives outweigh positives (in our data only 12% are positives). The model was used to classify 461 unknown AB_2O_6 materials and predicted 53 should be trirutiles, similar to the percentage of trirutiles in known compounds. Because the machine learning model has no thermodynamic information, we used DFT to evaluate whether the trirutile candidates could feasibly be synthesized. A simple proxy for stability is the formation energy of the product relative to the starting materials. For all 53 candidates, the energy of the reactant binary oxides were compared to the energy of the product trirutile using density functional theory (DFT) calculations, and a conservative threshold of $50\,\mathrm{meV\,atom^{-1}}$ above the binary energy was set as the limit of possible stability. When these compounds were additionally screened for inconvenient starting materials
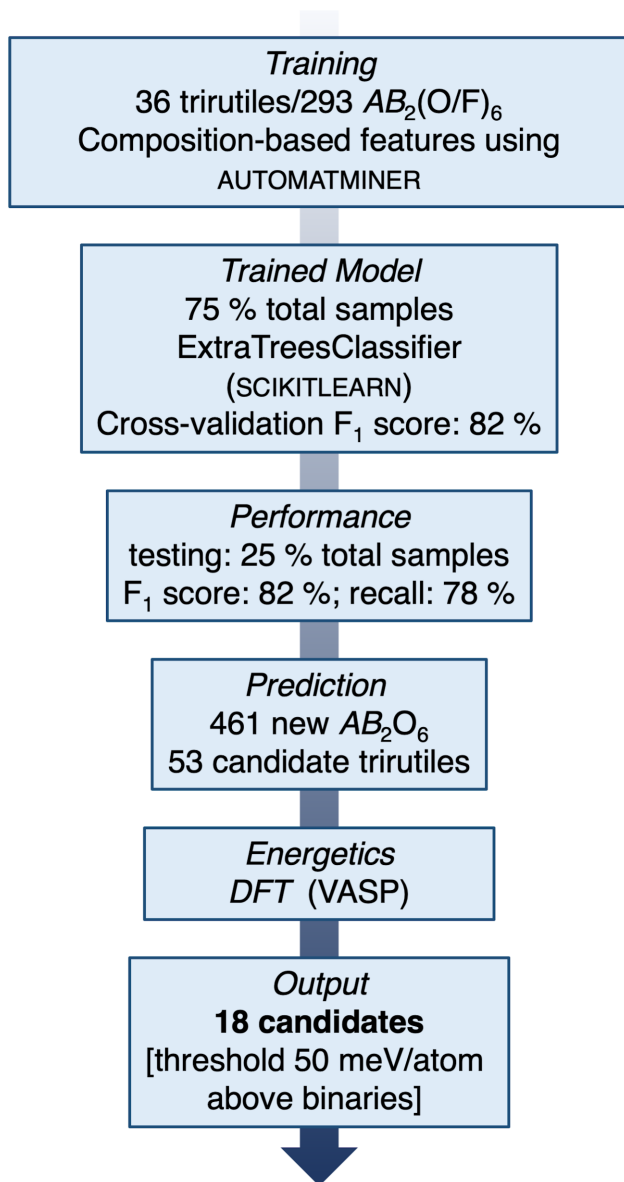
Figure 2: The process flow for predicting new trirutile materials. Using only composition-based features through automatminer, we trained a machine learning model to classify compounds into trirutile or not trirutile. The model was then used to classify 461 novel AB_2O_6 compositions and 53 were potential trirutile candidates. Density functional theory was employed to calculate the energy of the ternary trirutile relative to the oxide binaries as a proxy for stability, and 18 compounds emerged as promising new trirutile oxides. These candidates are further screened for 0K dynamical stability and against competing structure types.

(such as OsO_4), 18 materials emerged as promising novel trirutile candidates.
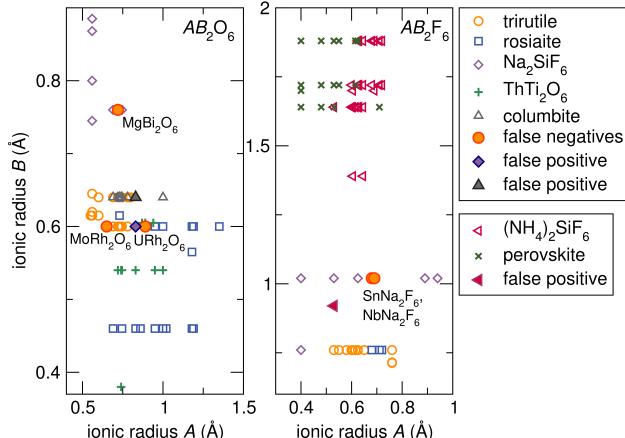
Figure 3: Plotting the ionic radius of the B-site vs the atomic radius of the A-site performs reasonably well at clustering families in both the AB_2O_6 and AB_2F_6 space. The machine learning model is able to correctly predict even overlapping compositions on these axes, indicating it includes more complex factors. From the model, both Na-containing fluoride trirutiles are predicted as false negatives, likely due to the scarcity of Na trirutiles. In the oxide trirutiles, one outlying Bi compound and two Rh-containing trirutiles are false negatives.

In addition to using the model to predict new trirutile materials, we aim to understand the chemistry that governs crystallization in the AB_2X_6 composition space. A common method for clustering structure types in ternary composition space is to plot the ionic radii of the A and B atoms.[37] In figure 3, we plot the ionic radius of the B-site against the ionic radius of the A-site for all common structure types in the AB_2X_6 space. This performs reasonably well at clustering structural families, but there is still significant overlap, notably between the trirutiles and the columbite and rosiaite families, indicating more complex features are necessary to understand crystallization in these materials. However, the machine learning model is able to correctly predict overlapping compounds in this view, indicating chemical features used by the model are able to differentiate trirutile compositions from others with similar ionic radii. The clearest flaw in the machine learning model is its inability to predict Na-containing fluoride trirutiles. This is likely because only two Na trirutiles have been reported while there are several ANa_2F_6 compounds not in the trirutile structure. The only false positive in the fluorides is $MnLi_2F_6$ which is chemically

11

similar to known trirutiles $CrLi_2F_6$ and VLi_2F_6. In the oxides, two Rh-containing trirutiles are predicted incorrectly, likely because there are only two Rh-containing oxides reported in the training data. Interestingly, both oxides falsely predicted to be trirutiles, $MnTa_2O_6$ and $MnSb_2O_6$, have reported metastable trirutile phases.[38,39]
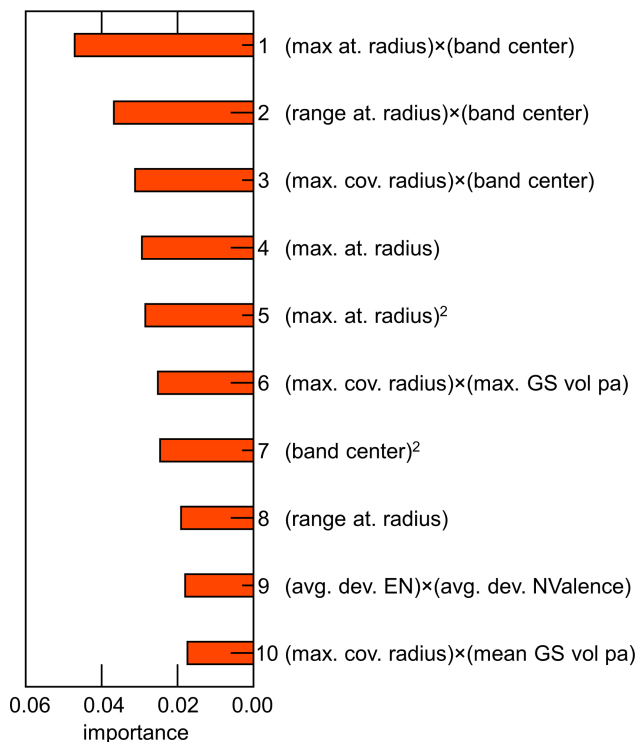

Figure 4: The 10 most important features for classification of a trirutile composition in this model

.

From the model, we extract the 10 most important features, denoted by their weights in the model, that are used to decide whether the composition is a trirutile, as shown in figure 4(a). The most important feature (feature 1) includes both geometric and bonding factors – the maximum atomic radius present in the composition and the band center, which is related to the mean electronegativity of the elements of the compound. In order to determine the most relevant features for further study, a correlation matrix was generated for the top 10 features, shown in the Supplementary Information. The most important feature is strongly correlated with 7 others of the top 10. This indicates that a combination of geometry and bonding is by far the most important determinant of whether the trirutile
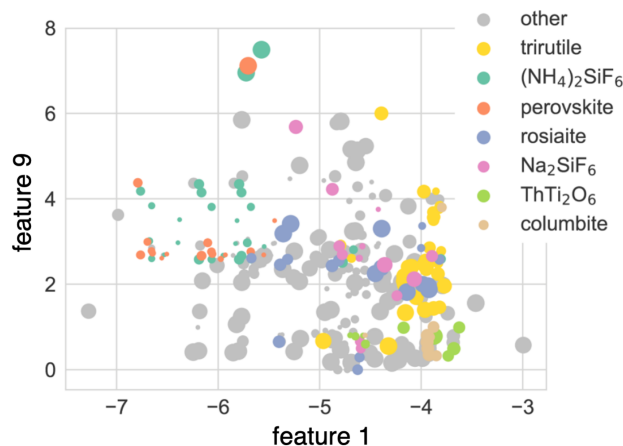
12

Figure 5: A plot of feature 9 vs feature 1 shows how these two features start to cluster the trirutiles. The size of the points is proportional to feature 7, which also helps to distinguish the trirutiles from neighboring compounds. Interestingly, these features appear to cluster other structure types such as vacancy-ordered double perovskites and the $(NH_4)_2SiF_6$ compounds. This indicates that similar features may govern crystal structure formation across many families.

structure will form. Feature 7, which is not as correlated with the others, is the square of the band center, again showing the importance of bonding. Feature 9, which is the other unique feature, is a combination of the average deviation of the electronegativity of the atoms and the average deviation of the number of valence electrons. Figure 4(b) shows feature 9 plotted versus feature 1. The trirutiles cluster at the less negative end of the x-axis and the medium-to-high end of the y-axis. The x-axis trend indicates that the trirutile structure is formed preferentially when the band center is less negative (closer to 0, indicating the A and B atoms have lower electronegativity) and when the maximum atomic radius in the compound is relatively small. The y-axis trend shows that once these criteria are met, A and B atoms with a larger difference in valence electrons or a larger difference in electronegativity are preferred. Interestingly, these features also seem to cluster the perovskites, $(NH_4)_2SiF_6$ compounds, and columbites, indicating similar features may govern crystal structure formation across this composition space.

By evaluating trends in the values of the most important features, it is possible to generate a simplified human-readable decision tree to determine whether a given composition
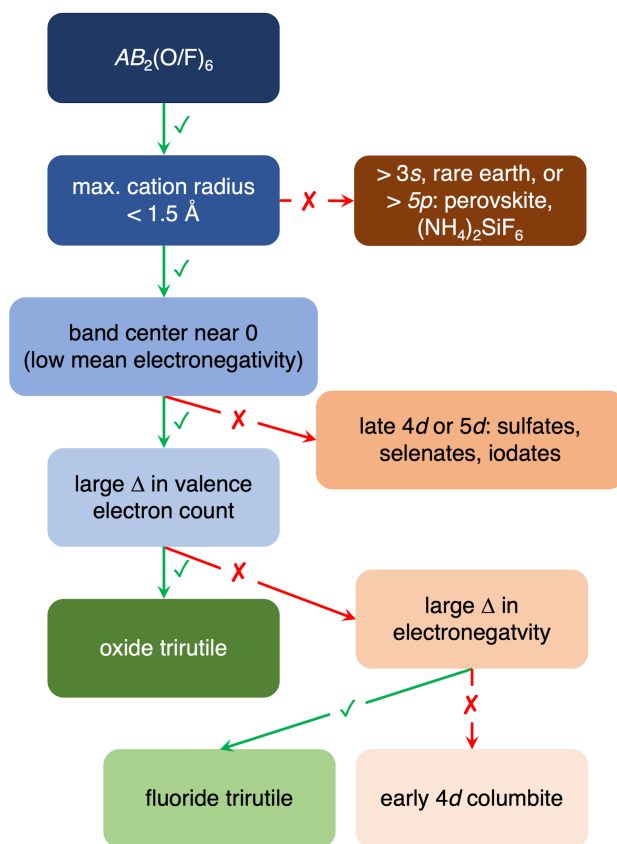
Figure 6: A simplified decision tree for the formation of the trirutile structure. The trirutile structure prefers small A and B-site atoms, likely due to the compact edge-sharing octahedral network. Additionally, low electronegativity A and B atoms are preferred relative to other compositions in the region (such as sulfates with the formula $A(SO_3)_2$). Finally, the trirutile structure is preferred in oxides when the number of valence electrons of the A and B atoms are very different (*i.e.* $3d$ and $5d$ or $5p$) and in fluorides when the bonding is very ionic.

will form in the trirutile structure, shown in figure 6. While this does not capture the full complexity of the family, these features describe most of the known trirutiles and exclude most other AB_2X_6 compounds. The first factor favoring trirutile formation is a maximum atomic radius in the compound of < 150 pm, at least for the oxide trirutiles. This excludes most of the alkali and alkaline earth elements, as well as rare earths, and lead or bismuth compounds. Most compounds with very large atoms in this composition seem to prefer the perovskite and $(NH_4)_2SiF_6$ structure types. The preference for a low maximum atomic radius can be explained by the compact nature of the trirutile structure, which is characterized by alternating edge-sharing octahedra. There are no sites like the B site in the double perovskite which can be occupied by a larger atom. Once the geometry has been satisfied, a band center near 0 is desirable (this is related to a low mean electronegativity of the elements). This bonding restriction excludes sulfates, and all compounds where the A or B site is a halogen or chalcogen, as well as discourages late $4d$ and $5d$ elements. The last conditions are more flexible and less immediately intuitive. For oxide trirutiles a large difference in the number of valence electrons across the composition, as denoted by the average deviation of the NValence feature from the Magpie database, is preferred. This criterion is related to the way that the NValence feature is calculated. For $5d$ atoms, the f electrons are counted in the valence electrons. Therefore, the preference for a large deviation in NValence for oxide trirutiles highlights the predisposition for $3d$ (or $3s$) and $5d$ or $5p$ combinations over $3d$ and $4d$ combinations, which are common in the columbites. For fluoride trirutiles, a large deviation of electronegativity across the composition is preferred. Most fluoride trirutiles are characterized by extremely ionic bonding and include small alkali metals like Li and Na. An understanding of the most important features allows us to gain insight into the mechanisms governing crystal structure formation in the trirutiles as well as other structure types within the region.

Now that we have established some guidelines related to trirutile formation, we turn to the results of DFT calculations on predicted novel trirutile oxides. The energies of the most

Table 1: DFT-based stability calculations for predicted trirutiles. ΔE refers to the energy of the trirutile relative to its constituent binary oxides.

Composition	ΔE (meV/atom)	Composition	ΔE (meV/atom)
PtV_2O_6	-6.9	$MnPt_2O_6$	25.4
$TiTa_2O_6$	-5.3	$GeSn_2O_6$	25.4
$ReMn_2O_6$	-3.1	$GeMn_2O_6$	26.8
$IrCr_2O_6$	0.1	WAl_2O_6	30.3
$CrSb_2O_6$	1.9	$RuRe_2O_6$	31.5
IrV_2O_6	2.2	$ReFe_2O_6$	41.1
$IrMn_2O_6$	16.9	$TiIr_2O_6$	46.0
$BeSb_2O_6$	18.0	$MnSn_2O_6$	48.4
ReV_2O_6	24.5	WNb_2O_6	50.0

stable candidates relative to their constituent binary oxides are shown in Table 1. Among the top candidates, we find many in which the AO_2 or BO_2 binaries are known to form in the rutile structure. For example, the most stable predicted candidate relative to the starting materials is PtV_2O_6, and its binaries, PtO_2 and VO_2, are both rutiles. Additionally, we find many compounds which are chemically similar to the known trirutiles. $TiTa_2O_6$ is similar to many of the known Ta-based trirutiles such as $FeTa_2O_6$ or $MgTa_2O_6$. $ReMn_2O_6$ is similar to $ReCr_2O_6$, and $CrSb_2O_6$ is similar to $CrTa_2O_6$. The similarity of predicted compounds to known compounds suggests that the model is picking up on underlying chemical trends in the trirutiles.

To further evaluate the machine learning predictions, we used DFT to calculate both dynamic stability *via* 0 K phonon dispersions and lattice stability relative to the four most common structure types observed in the AB_2O_6 space: rosiaite, Na_2SiF_6, $ThTi_2O_6$, and columbite. Figure 7 compares the energies of the ideal trirutile structure to those of the four alternative structures as well as to the distorted trirutile phases. Supporting the predictions of the machine learning model, our calculations predict that for 16 of the 18 chemistries examined, the trirutile phase (distorted or ideal) is competitive for the room temperature equilibrium structure. More information about the distorted trirutile phases is available in Figure 2 of the Supporting Information.
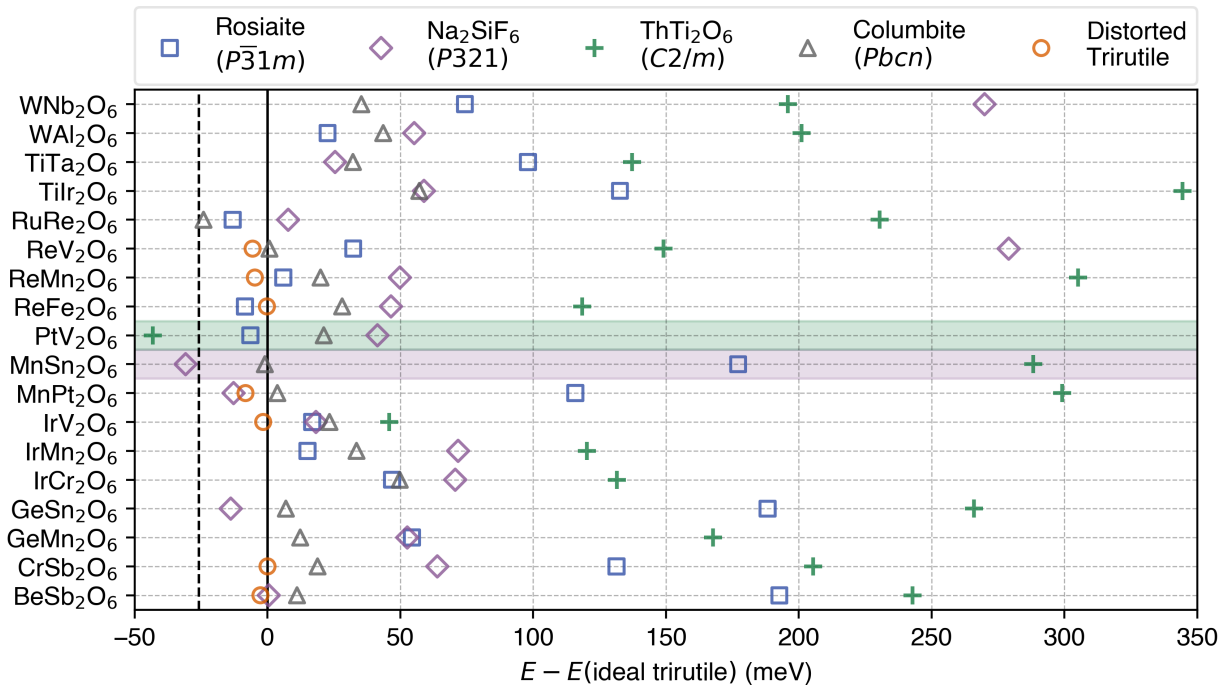
Figure 7: DFT-based energy calculations for the most prevalent alternative 1-2-6 oxide structures plotted relative to the ideal trirutile case. The vertical dashed line indicates the threshold above which the ideal trirutile structure is within kT for $T = 300\,\mathrm{K}$. Two of the 18 candidate compounds (highlighted rows) are predicted to adopt alternate structures at 300 K.
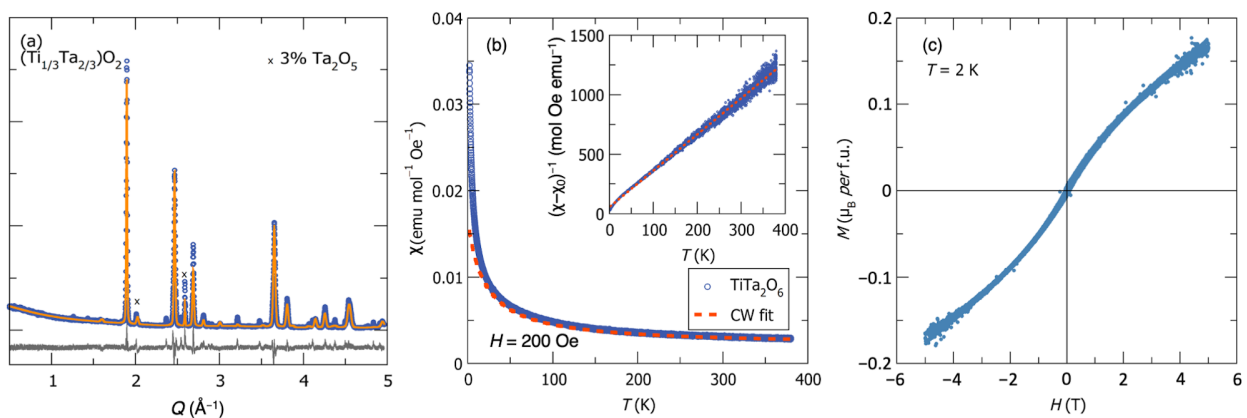

Figure 8: (a) Laboratory Cu-Kα diffraction patterns on (b) TiTa$_2$O$_6$ show that the structure corresponds to a disordered $(A_{1/3}, B_{2/3})O_2$ rutile structure. There is a slight Ta$_2$O$_5$ impurity in (Ti$_{1/3}$,Ta$_{2/3}$)O$_2$. (b) Temperature–dependent magnetization of TiTa$_2$O$_6$ taken between T = 1.8 and 380 K under a constant field of H = 200 Oe is shown. There is no ordering transition in this temperature range, suggesting a paramagnetic regime. (c) Field–dependent magnetization of TiTa$_2$O$_6$ taken between $H = -5$ and 5 T at a constant temperature of T = 2 K shows paramagnetic behavior.

17

With a candidate list of new trirutiles, we attempted synthesis via traditional solid state methods. We found that compounds with noble metals (Pt and Ir) were difficult to react because the low stability of the 4+ valence state limited the synthesis conditions to low temperatures (typically $< 700^\circ$C), where reactions proceeded slowly or not at all. We did not, for this study, attempt any of the Re-based candidates. Although we attempted synthesis of CrSb$_2$O$_6$, X-ray fluorescence measurements revealed a 1:1 Cr:Sb ratio suggesting that we were only able to make CrSbO$_4$. TiTa$_2$O$_6$ did react at high temperatures (1100°C) to form a ternary compound. X-ray diffraction data can be fit to a $(A_{1/3}, B_{2/3})$O$_2$ rutile model [figure 8a]. While we were unable to refine occupancies on the transition metal site reliably due to limits of the data resolution, X-ray fluorescence measurements on (Ti$_{1/3}$,Ta$_{2/3}$)O$_2$ indicate the product composition (32.8% Ti, 67.2% Ta) is close to the nominal composition.

Figure 8b displays the raw and fitted field-cooled temperature-dependent magnetization data of TiTa$_2$O$_6$ between 1.8 and 380 K under an applied field of $H = 200$ Oe. This data reveals that TiTa$_2$O$_6$ remains paramagnetic down to at least 1.8 K. A Curie-Weiss fit between 25 K and 300 K gave a small but significant χ_0 correction of 0.002 emu mol^{-1} Oe^{-1}. From fitting the linear region of the inverse susceptibility with the inverse Curie-Weiss law, a Curie constant of 0.33 emu mol^{-1} Oe^{-1} and a Θ_{CW} of about -22 K were calculated. While this Θ_{CW} is negative, it is very close 0 K, as expected in a paramagnet. The field-dependent magnetization taken at $T = 2$ K in Figure 8c does not saturate at high field and is reminiscent of an ideal paramagnet that can be described with a Brillouin function.

An effective moment of Ti can be calculated using the approximation $\mu_{eff} = \sqrt{8C}$. Using the Curie constant obtained from the Curie-Weiss fit, $\mu_{eff} = 1.62\mu_B$. An analysis of spin-only and spin-orbit moments can reveal the oxidation state of the magnetic ion. The spin-only moment for Ti^{2+}, calculated as $\mu_{eff} = g\sqrt{S(S+1)}$, $g \approx 2$ and $S = 1$, so the effective moment is 2.8 μ_B. The spin-orbit moment is calculated as $\mu_{eff} = g^*\sqrt{J(J+1)}$, where $g^* = 1 + \frac{J(J+1)+S(S+1)-L(L-1)}{2J(J+1)}$ and $J = |L - S|$ if the shell is less than half full, and

$J = L + S$ if the shell is more than half full. For Ti^{2+}, $L = 3$ and $S = 1$, so $J = 2$ and $g = 2/3$. This gives an effective spin-orbit moment of 1.63 μ_B. If a similar analysis is performed for Ti^{3+}, $S = 1/2$ and $L = 2$, giving a spin-only moment of 1.73 μ_B and a spin-orbit moment of 1.55 μ_B. From this, we conclude that the oxidation state of Ti in $TiTa_2O_6$ is iclose to $+2$.

The analysis of XPS spectra (presented in the supplementary information in figure S3) in the 2p region for Ti and the 4d region for Ta confirms the 1:2 ratio of the cations. However the $2p_{3/2}$ and $2p_{1/2}$ spin orbit doublets for Ti are suggestive of an oxidation state that is indistinguishable from the reported values for Ti^{4+}. We infer that there may be some surface oxidation of the Ti that is being captured in the XPS, but is distinct from the bulk d electron count represented in the magnetic measurement.

We have so far been unable to find an annealing scheme that would order the A and B cations to form the trirutile structure. The erroneous prediction of a stable ordered trirutile phase for these compositions is likely due to the failure to account for disorder in either the machine learning model or the DFT. A simple energy threshold relative to the binaries as calculated by DFT is not a perfect metric for synthesizability,[40] and more sophisticated means for determining stability could allow us to optimize the efficiency of a computational pipeline to overcome synthesis challenges. Because of the difficulty of modeling and understanding disordered materials, any disordered materials were thrown out in our original data cleaning process for the machine learning training data, and disordered supercells were not calculated in the DFT screening. Despite the disorder, the formation of a ternary rutile phase indicates the model has a consistent ability to predict the local coordination environment in this composition space.

Conclusion

Machine learning is a powerful tool for making connections in complex systems with many variables. While the AB_2X_6 composition space of oxides and fluorides is small compared to many databases used in machine learning, it has a multitude of competing crystal structures. The most common crystal structure in this region, the trirutile family, makes up only 12% of the members, competing with vacancy-ordered double perovskites, layered compounds, columbites, and many others. We find that the trirutile structure is preferred overall when A and B atoms are small and less electronegative, and that the oxide trirutiles are typically a combination of a $3d$ or $3s$ atom with a $5d$ or $5p$ atom. The fluoride trirutiles are typically Li with a $3d$ atom or Na with a $4d$ atom. We discover that similar geometric and bonding features that are important for the trirutiles appear also to be important for other families, such as the perovskites. In combination with DFT screening for stability, we predict 18 novel trirutile oxides. Of those, we are able to synthesize ternary oxides of 2 and show that they form in a disordered rutile structure, highlighting the success of the model in predicting local coordination environments. However, the energetics governing disorder are challenging to understand and predict and require special consideration. Additionally, the goal of accounting for synthesizability in inorganic materials, for example by including Ellingham diagrams or other solid state chemistry benchmarks in a machine learning model is crucial to accelerate the timeline for the design and realization of novel functional materials.

Supporting Information Available

Detailed description of python packages and machine learning methods used, including feature libraries and data processing tools. A correlation matrix for the top 10 features is shown in Figure S1, Figure S2 displays the types of computed structural distortions of the AB_2 trirutile, and Figure S3 displays XPS spectra for $TiTa_2O_6$.

Acknowledgements

References

(1) Tehrani, A. M.; Oliynyk, A. O.; Parry, M.; Rizvi, Z.; Couper, S.; Lin, F.; Miyagi, L.; Sparks, T. D.; Brgoch, J. Machine Learning Directed Search for Ultraincompressible, Superhard Materials. *J. Am. Chem. Soc.* **2018**, *140*, 9844–9853, DOI: `10.1021/jacs.8b02717`.

(2) Avery, P.; Wang, X.; Oses, C.; Gossett, E.; Proserpio, D. M.; Toher, C.; Curtarolo, S.; Zurek, E. Predicting superhard materials via a machine learning informed evolutionary structure search. *npj Computational Materials* **2019**, *5*, 1–11.

(3) Lee, J.; Seko, A.; Shitara, K.; Nakayama, K.; Tanaka, I. Prediction Model of Band Gap for Inorganic Compounds by Combination of Density Functional Theory Calculations and Machine Learning Techniques. *Phys. Rev. B* **2016**, *93*, 115104, DOI: `10.1103/PhysRevB.93.115104`.

(4) Woods-Robinson, R.; Broberg, D.; Faghaninia, A.; Jain, A.; Dwaraknath, S. S.; Persson, K. A. Assessing High-Throughput Descriptors for Prediction of Transparent Conductors. *Chem. Mater.* **2018**, *30*, 8375–8389, DOI: `10.1021/acs.chemmater.8b03529`.

(5) Isayev, O.; Oses, C.; Toher, C.; Gossett, E.; Curtarolo, S.; Tropsha, A. Universal fragment descriptors for predicting properties of inorganic crystals. *Nature communications* **2017**, *8*, 1–12.

(6) Oliynyk, A. O.; Antono, E.; Sparks, T. D.; Ghadbeigi, L.; Gaultois, M. W.; Meredig, B.; Mar, A. High-Throughput Machine-Learning-Driven Synthesis of Full-Heusler Compounds. *Chem. Mater.* **2016**, *28*, 7324–7331, DOI: `10.1021/acs.chemmater.6b02724`.

(7) Pauling, L. The Principles Determining the Structure of Complex Ionic Crystals. *J. Am. Chem. Soc.* **1929**, *51*, 1010–1026, DOI: `10.1021/ja01379a006`.

(8) Slater, J. C. Atomic Radii in Crystals. *J. Chem. Phys.* **1964**, *41*, 3199, DOI: 10.1063/1.1725697.

(9) Shannon, R. D. Revised Effective Ionic Radii and Systematic Studies of Interatomic Distances in Halides and Chalcogenides. *Acta Cryst. A* **1976**, *32*, 751–767, DOI: 10.1107/S0567739476001551.

(10) Phillips, J. C.; Van Vechten, J. A. Dielectric Classification of Crystal Structures, Ionization Potentials, and Band Structures. *Phys. Rev. Lett.* **1969**, *22*, 705–708, DOI: 10.1103/PhysRevLett.22.705.

(11) Zunger, A. Systematization of the Stable Crystal Structure of all AB-Type Binary Compounds: A Pseudopotential Orbital-Radii Approach. *Phys. Rev. B* **1980**, *22*, 5839, DOI: 10.1103/PhysRevB.22.5839.

(12) Shoemaker, D. P.; Shoemaker, C. B. Concerning the Relative Numbers of Atomic Coordination Types in Tetrahedrally Close Packed Metal Structures. *Acta Cryst. B* **1986**, *42*, 3–11, DOI: 10.1107/S0108768186098671.

(13) Fischer, C. C.; Tibbetts, K. J.; Morgan, D.; Ceder, G. Predicting Crystal Structure by Merging Data Mining with Quantum Mechanics. *Nat. Mater.* **2006**, *5*, 641–646, DOI: 10.1038/nmat1691.

(14) Hautier, G.; Fischer, C. C.; Jain, A.; Mueller, T.; Ceder, G. Finding Nature's Missing Ternary Oxide Compounds Using Machine Learning and Density Functional Theory. *Chem. Mater.* **2010**, *22*, 3762–3767, DOI: 10.1021/cm100795d.

(15) Oliynyk, A. O.; Adutwum, L. A.; Harynuk, J. A.; Mar, A. Classifying Crystal Structures of Binary Compounds AB through Cluster Resolution Feature Selection and Support Vector Machine Analysis. *Chem. Mater.* **2016**, *28*, 6672–6681, DOI: 10.1021/acs.chemmater.6b02905.

(16) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big data of materials science: critical role of the descriptor. *Physical review letters* **2015**, *114*, 105503.

(17) Oganov, A. R.; Glass, C. W. Evolutionary Crystal Structure Prediction as a Tool in Materials Design. *J. Phys: Condensed Matter* **2008**, *20*, 064210, DOI: 10.1088/0953-8984/20/6/064210.

(18) Wang, Y.; Lv, J.; Zhu, L.; Ma, Y. Crystal Structure Prediction via Particle-Swarm Optimization. *Phy. Rev. B* **2010**, *82*, 094116.

(19) Wang, Y.; Lv, J.; Zhu, L.; Ma, Y. CALYPSO: A method for crystal structure prediction. *Computer Physics Communications* **2012**, *183*, 2063–2070.

(20) Kunnmann, W.; Placa, S. L.; Corliss, L. M.; Hastings, J. M.; Banks, E. Magnetic Structures of the Ordered Trirutiles Cr_2WO_6, Cr_2TeO_6 and Fe_2TeO_6. *J. Phys. Chem. Solids* **1968**, *29*, 1359–1364, DOI: 10.1016/0022-3697(68)90187-X.

(21) Law, J. M.; Koo, H.-J.; Whangbo, M.-H.; Brücher, E.; Pomjakushin, V.; Kremer, R. K. Strongly Correlated One-Dimensional Magnetic Behavior of $NiTa_2O_6$. *Phys. Rev. B* **2014**, *89*, 014423, DOI: 10.1103/PhysRevB.89.014423.

(22) Kato, M.; Hatazaki, A.; Yoshimura, K.; Kosuge, K. One-Dimensional Magnetic Behavior of $CuSb_2O_6$. *Physica B Condens. Matter* **2000**, *281–282*, 663–664, DOI: 10.1016/S0921-4526(99)00944-8.

(23) Ehrenberg, H.; Wltschek, G.; Rodriguez-Carvajal, J.; Vogt, T. Magnetic Structures of the Tri-Rutiles $NiTa_2O_6$ and $NiSb_2O_6$. *J. Magn. Magn. Mater.* **1998**, *184*, 111–115, DOI: 10.1016/S0304-8853(97)01122-0.

(24) Morin, F. J. Oxides Which Show a Metal-to-Insulator Transition at the Neel Temperature. *Phys. Rev. Lett.* **1959**, *3*, 34–36, DOI: 10.1103/PhysRevLett.3.34.

(25) Hiroi, Z. Structural Instability of the Rutile Compounds and Its Relevance to the Metal–Insulator Transition of VO_2. *Progr. Solid State Chem.* **2015**, *43*, 47–69, DOI: `10.1016/j.progsolidstchem.2015.02.001`.

(26) Kasinathan, D.; Koepernik, K.; Rosner, H. Quasi-One-Dimensional Magnetism Driven by Unusual Orbital Ordering in $CuSb_2O_6$. *Phys. Rev. Lett.* **2008**, *100*, 237202, DOI: `10.1103/PhysRevLett.100.237202`.

(27) Allen, F. H.; Bergerhoff, G.; Brown, I. D. *Crystallographic Databases*; International Union of Crystallography, 1987.

(28) Wang, A. Y.-T.; Murdock, R. J.; Kauwe, S. K.; Oliynyk, A. O.; Gurlo, A.; Brgoch, J.; Persson, K. A.; Sparks, T. D. Machine Learning for Materials Scientists: An introductory guide towards best practices. *Chemistry of Materials* **2020**,

(29) Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N. E. R.; Bajaj, S.; Wang, Q.; Montoya, J. H.; Chen, J.; Bystrom, K.; Dylla, M.; Chard, K.; Asta, M.; Persson, K.; Snyder, G. J.; Foster, I.; Jain, A. Matminer: An Open Source Toolkit for Materials Data Mining. *Comput. Mater. Sci.* **2018**, *152*, 60–69, DOI: `10.1016/j.commatsci.2018.05.018`.

(30) Dunn, A.; Wang, Q.; Ganose, A.; Dopp, D.; Jain, A. Benchmarking Materials Property Prediction Methods: The Matbench Test Set and Automatminer Reference Algorithm. *npj Comp. Mater.* **2020**, *6*, 159.

(31) Le, T. T.; Fu, W.; Moore, J. H. Scaling Tree-Based Automated Machine Learning to Biomedical Big Data with a Feature Set Selector. *J. Bioinform.* **2020**, *36*, 250–256, DOI: `10.1007/978-3-319-31204-0_9`.

(32) Kresse, G.; Furthmuller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **1996**, *54*, 11169–11186, DOI: `10.1103/PhysRevB.54.11169`.

(33) Blochl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50*, 17953–17979, DOI: `10.1103/PhysRevB.50.17953`.

(34) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B* **1999**, *59*, 1758–1775, DOI: `10.1103/PhysRevB.59.1758`.

(35) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868, DOI: `10.1103/PhysRevLett.77.3865`.

(36) Togo, A.; Tanaka, I. First Principles Phonon Calculations in Materials Science. *Scr. Mater.* **2015**, *108*, 1–5, DOI: `10.1016/j.scriptamat.2015.07.021`.

(37) Muller, O.; Roy, R. *The Major Ternary Structural Families*; Springer Verlag: Berlin, Heidelberg, and New York, 1974.

(38) Esmaeilzadeh, S.; Grins, J. Meta-stable Phases in the Mn–Ta–O System. *Solid State Sci.* **2002**, *4*, 117–123, DOI: `10.1016/S1293-2558(01)01216-X`.

(39) Sala, F.; Trifiró, F. Relationship Between Structure and Activity of Antimony Mixed Oxides in 1-Butene Oxidation. *J. Catal.* **1976**, *41*, 1–13, DOI: `10.1016/0021-9517(76)90194-9`.

(40) Sun, W.; Dacek, S. T.; Ong, S. P.; Hautier, G.; Jain, A.; Richards, W. D.; Gamst, A. C.; Persson, K. A.; Ceder, G. The thermodynamic scale of inorganic crystalline metastability. *Science advances* **2016**, *2*, e1600225.

For Table of Contents
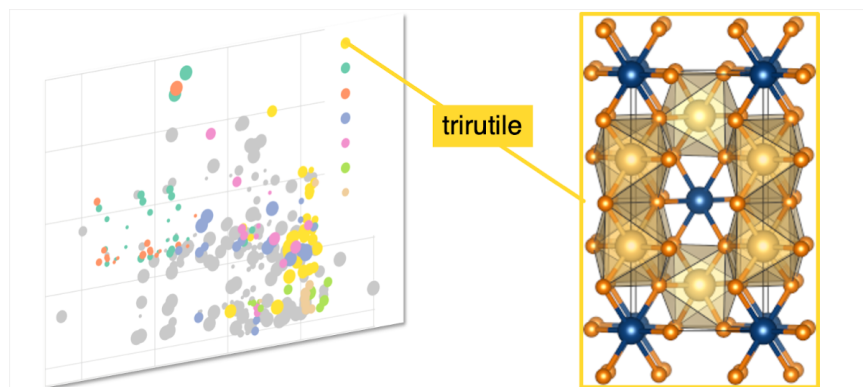
The composition space of AB_2X_6 oxide and fluoride compounds has been examined using a combination of machine learning from known structures and density functional theory-based electronic structure calculations with the specific goal of finding new compounds in the trirutile family.

Supporting Information for:

AB_2X_6 Compounds and the Stabilization of Trirutile

Oxides

Emily C. Schueller,[†] Yuzki M. Oey,[†] Kyle D. Miller,[‡] Kira E. Wyckoff,[†]

Ruining Zhang,[¶] William Zhang,[¶] Stephen D. Wilson,[†] James M. Rondinelli,[‡] and

Ram Seshadri[∗,†,§]

[†]Materials Department and Materials Research Laboratory

University of California, Santa Barbara, Santa Barbara, CA, 93106

[‡]Department of Materials Science and Engineering, Northwestern University, Evanston,

Illinois 60208, USA

[¶]Materials Research Laboratory

University of California, Santa Barbara, Santa Barbara, CA, 93106

[§]Department of Chemistry and Biochemistry

University of California, Santa Barbara, Santa Barbara, CA, 93106

E-mail: seshadri@mrl.ucsb.edu

The machine learning model relied heavily on automatminer pipeline attributes. Here is specific information about the pipeline steps.

Featurizing the matrix:

```
"autofeaturizer": AutoFeaturizer(preset="all")
```

```
List of feature libraries:
```

```
 AtomicOrbitals(),
 ElementProperty.from\_preset(''matminer") (these are from pymatgen),
          ElementProperty.from\_preset(''magpie"),
          ElementProperty.from\_preset(''deml"),
          ElementFraction(),
          Stoichiometry(),
          TMetalFraction(),
          BandCenter(),
          ValenceOrbital(),
          YangSolidSolution(),
          CationProperty.from\_preset(preset\_name='deml'),
          OxidationStates.from\_preset(preset\_name='deml'),
          ElectronAffinity(),
          ElectronegativityDiff(),
          IonProperty(),
          Miedema()
```

```
pymatgen stats: ["minimum", "maximum", "range", "mean", "std\_dev"]
```

```
magpie stats: ["minimum", "maximum", "range", "mean", "avg\_dev", "mode"]
```

```
deml stats: ["minimum", "maximum", "range", "mean", "std\_dev"]
```

Cleaning the data:

```
``cleaner": DataCleaner()
```

If feature value missing (N/A) for >3% of samples (or <3% of features available for sample), remove. Remove duplicates.

```
``reducer": FeatureReducer(reducers=('corr', 'tree'),
  tree\_importance\_percentile=0.99)
```

If features >95% correlated, remove. Also, use a random forest to remove features that together sum to <1% of importance by weight. Together, these lead to a reduction of 629 intial features to 46.

Machine learning genetic algorithm parameters:

```
``learner'': TPOTAdaptor(max\_time\_mins=720, max\_eval\_time\_mins=20,
  scoring=``f1'')
```

he genetic algorithm runs for 12 hours and is scored on an f_1 metric, as explained in the methods.

The genetic algorithm continues feature selection, chooses recursive feature elimination (RFE) to remove half of the features with importances determined via extra trees classifier and gini importance, resulting in 23 features, with parameters shown below.

```
RFE(estimator=ExtraTreesClassifier(bootstrap=False, class\_weight = None,
criterion = 'gini', max\_depth = None, max\_features = 0.35,
max\_leaf\_nodes = None, min\_impurity\_decrease = 0.0,
min\_impurity\_split = None, min\_samples\_leaf = 1, min\_samples\_split = 2,
min\_weight\_fraction\_leaf = 0.0, n\_estimators = 100, n\_jobs = None,
oob\_score = False, random\_state = None, verbose = 0, warm\_start = False),
n\_features\_to\_select = None, step = 0.35, verbose = 0)),
```

These are the RFE hyperparameters the genetic algorithm iterated over:

```
  'sklearn.feature\_selection.RFE': (
      'step': np.arange(0.05, 1.01, 0.05),
      'estimator': (
          'sklearn.ensemble.ExtraTreesClassifier': (
              'n\_estimators': [100],
              'criterion': ['gini', 'entropy'],
              'max\_features': tree\_max\_features
          )
```

Preprocessing steps follow, such as scaling and normalizing the features. The genetic algorithm finds polynomial features to order 2 is best, with the specific parameters as follows.

```
('polynomialfeatures',
 PolynomialFeatures(degree=2, include\_bias=False, interaction\_only=False)),
```

Finally, the genetic algorithm chooses the best machine learning model. For a classification problem, TPOT chooses between models like random forest, gradient boosting classifier, linear SVC, and others. For this dataset, the best model was found to be an extra trees classifier with the parameters shown below:

```
('extratreesclassifier',
 ExtraTreesClassifier(bootstrap=False, class\_weight=None, criterion='entropy',
           max\_depth=None, max\_features=0.95,
           max\_leaf\_nodes=None, min\_impurity\_decrease=0.0,
           min\_impurity\_split=None, min\_samples\_leaf=7,
           min\_samples\_split=14, min\_weight\_fraction\_leaf=0.0,
           n\_estimators=100, n\_jobs=None, oob\_score=False,
           random\_state=None, verbose=0, warm\_start=False))
```

These are the hyperparameters TPOT iterates over and the acceptable ranges in the nested CV:

```
tree\_estimators = [20, 100, 200, 500, 1000]
```

```
tree\_max\_features = np.arange(0.05, 1.01, 0.1)


tree\_learning\_rates = [1e-2, 1e-1, 0.5,  1.]


tree\_max\_depths = range(1, 11, 2)


tree\_min\_samples\_split = range(2, 21, 3)


tree\_min\_samples\_leaf = range(1, 21, 3)


tree\_ensemble\_subsample = np.arange(0.05, 1.01, 0.1)
```


We fit the same features and train/test split to other models using a grid search cross validation with a 5-fold stratified split to benchmark our model. Other models included a random forest, neural net, and support vector classifier (for the last two the features were scaled), but the original extra trees classifier had the best f_1 score on both cross validation and test.

The legend for Fig. 4 (in the main text) is shown below.

1. Maximum atomic radius * band center: The band center is related to the mean electronegativity of the compound: when it's closer to 0, the A and B atoms are less electronegative.

2. Range atomic radius * band center: The range of the atomic radius is the difference between the largest and smallest atom size in the compound.

3. Maximum covalent radius * band center

4. Maximum atomic radius

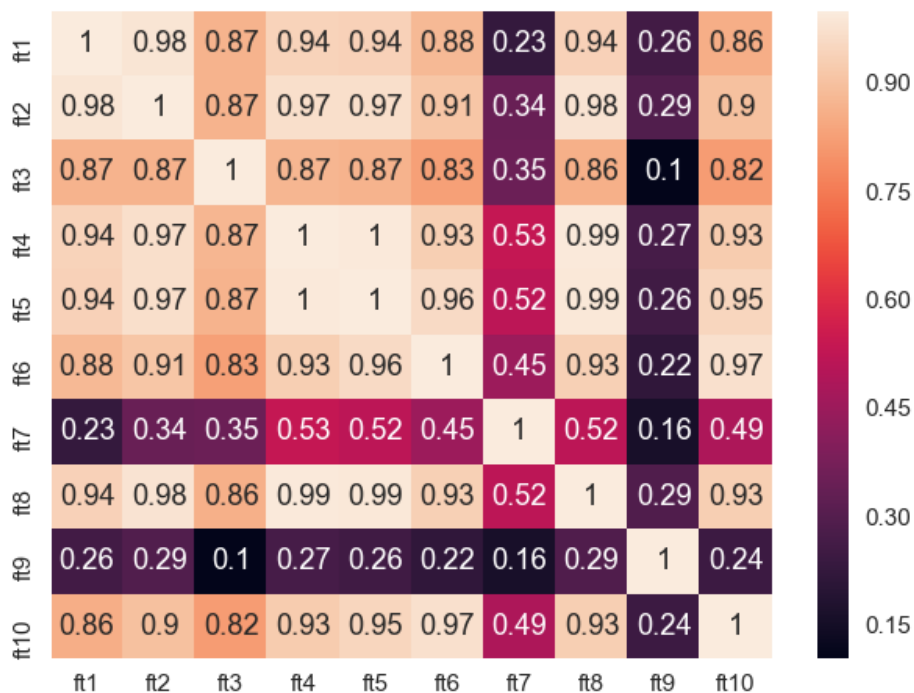5. (Maximum atomic radius)^2

Figure S.1: A correlation matrix for the 10 most important features of the extra trees classifier shows that 7 of the top 10 features are highly correlated with the most important feature. This indicates a combination of geometry and bonding is the most important factor in understanding which compositions form in the trirutile structure.

6. Maximum covalent radius * maximum ground state volume per atom: The ground state volume per atom is the space that the atom takes up in its elemental ground state crystal structure from density functional theory; strongly correlated with other atomic and covalent radii for the atom.

7. (Band center)^2

8. Range atomic radius

9. Average deviation electronegativity * average deviation number of valence electrons: The average deviation of electronegativity is controlled mostly by the A and B atoms since O and F are strongly electronegative. This is related to band structure. The average deviation in number of valence electrons, as explained in more detail in the main text, helps to differentiate 4d vs 5d atoms.

10. Maximum covalent radius * mean ground state volume per atom

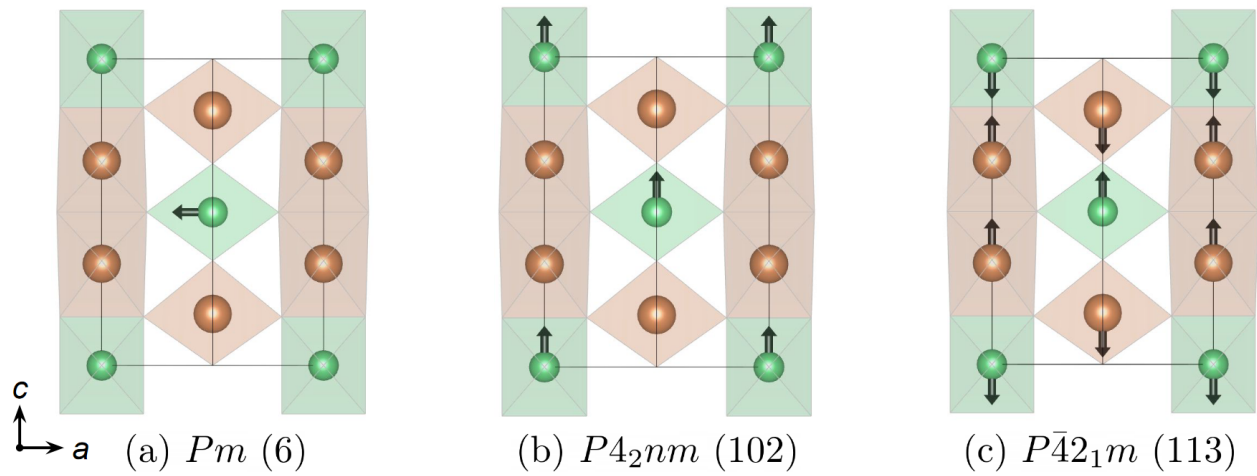(a) Pm (6) (b) $P4_2nm$ (102) (c) $P\bar{4}2_1m$ (113)

Figure S.2: Three types of lattice distortions are predicted among the 18 trirutile structures, which reduce the symmetry of the ideal trirutile phase. We observe distortion (a) in $BeSb_2O_6$, distortion (b) in $CrSb_2O_6$, and distortion (c) in ReV_2O_6, $ReMn_2O_6$, IrV_2O_6. These distortions produce noncentrosymmetric space groups and indicate the possibility of ferroelectricity and/or piezoactivity.
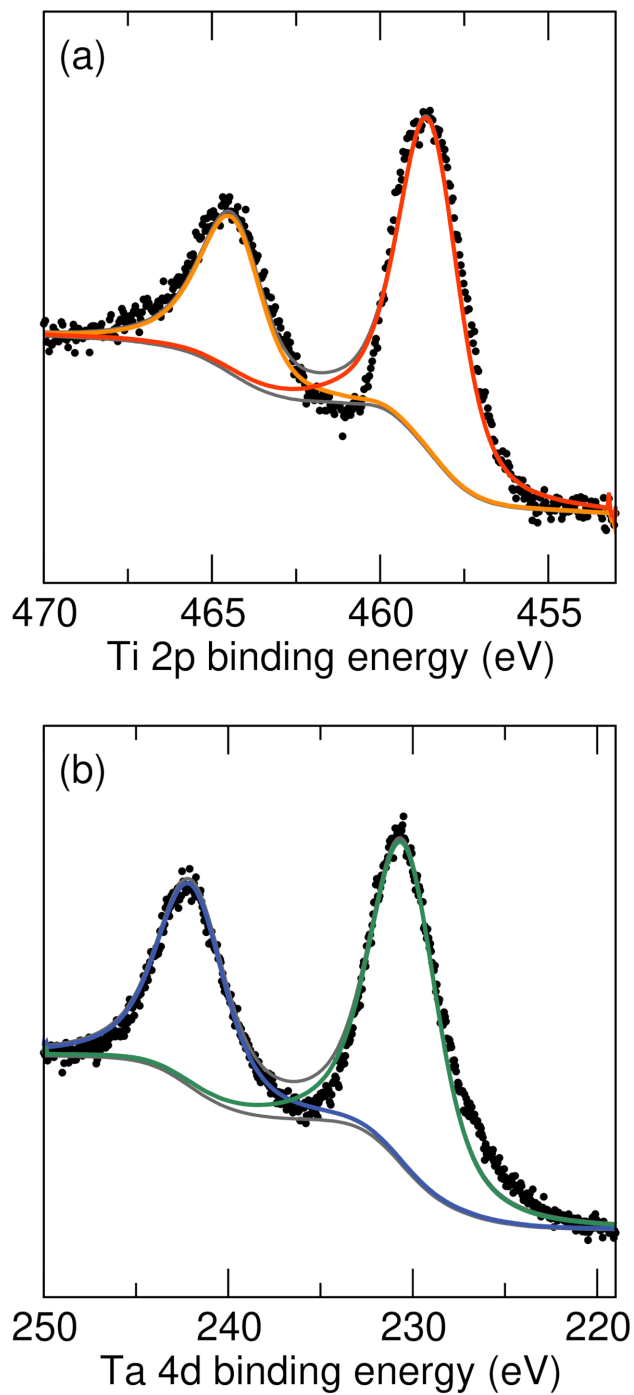
Figure S.3: X-ray photoelectron spectra displaying the region of (a) Ti 2p spin-orbit doublets $2p_{3/2}$ and $2p_{1/2}$ and (b) Ta 4d spin orbit doublets $4d_{5/2}$ and $4d_{3/2}$.