CHEMISTRY OF MATERIALS

pubs.acs.org/cm

Article

A Machine Learning Approach for the Prediction of Formability and Thermodynamic Stability of Single and Double Perovskite Oxides

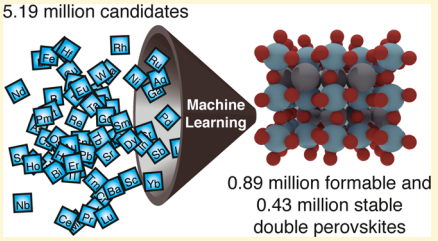
Anjana Talapatra,* Blas P. Uberuaga, Christopher R. Stanek, and Ghanshyam Pilania

Cite This: *Chem. Mater.* 2021, 33, 845−858

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Perovskite oxides continue to attract huge interest due to their fascinating and wide-ranging properties for diverse applications. The tunability of these properties may be further enhanced by increasing their compositional complexity via double perovskite-ordered configurations containing multiple cations. In this work, we focus on an exhaustive chemical space of single and double oxide perovskites and optimally explore this space to identify novel compositions that are likely to form stable compounds. Critically, we examine the relationship between formability, the practical ability to synthesize a compound, and stability, the thermodynamic preference to form the structure. Our formability and stability training data sets were enumerated from the available experimental literature and in-house density functional theory computations and contained 1505 and 3469 examples, respectively, representing state-of-the-art in the current open literature in perovskite and double perovskite compounds. Subsequently, cross-validated and highly accurate machine learning classification models are built using these training data sets and employed to screen for novel stable oxide perovskites. The study identifies (1) atomic features relevant to prediction of formability and stability in perovskite and double perovskite compounds, (2) the importance of including energy contributions due to local structural relaxations going beyond the high symmetry perovskite phase, and (3) 437,828 double perovskite compounds that are likely to be stable and 891,188 compounds that are likely to be formable. From the intersection of this large chemical space of formable and stable oxide perovskites, 414 compositions are identified as the most promising candidates for future experimental synthesis of novel oxide perovskites. The developed models may be generalized and have implications beyond perovskite discovery if applied to other families of compounds.

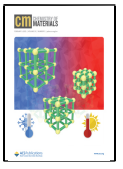


1. INTRODUCTION

In its simplest form, the aristotype perovskite represents any compound with an ABX_3 stoichiometry, with a cubic crystal structure composed of a three-dimensional framework of corner-sharing BX_6 octahedra. The A-site cations occupy the 12 coordinate sites formed by the BX_3 network, and each A cation is surrounded by 12 equidistant anions, as shown in Figure 1a. $CaTiO_3$ was the first perovskite to be discovered in 1839 in the Ural mountains of Russia by the geologist Gustav Rose.[1] In spite of being studied for almost 200 years, perovskites continue to attract wide interest due to their fascinating electrical and magnetic properties such as ferroelectricity,[2] piezoelectricity,[3] optical properties,[4] high-temperature superconductivity,[5] and magnetostrictive effects.[6] Their ubiquity stems from their ability to accommodate 90% of the metallic ions in the periodic table with a wide variety of different anions.[7] There is a close correlation between chemical and physical properties in these materials. It is this correlation that makes the perovskites as a group important, as the simple replacement of any of the atoms in these structures can be used to precisely tune the physical properties of interest. The tunability may be further enhanced when the A- and/or B-sites are occupied by two different types of cations giving rise to double perovskites with formulae $A_xA'_{2-x}B_2X_6$, $A_2B_yB'_{2-y}X_6$ and, in the most general case, $A_xA'_{2-x}B_yB'_{2-y}X_6$ for $0 \leq x, y \leq 2$, as shown in Figure 1b. Here A, A', B and B' are isovalent and X is generally a halogen, nitrogen or oxygen. Perovskites with an oxygen anion, that is, oxide perovskites, are of special interest since large concentrations of oxygen vacancies can exist in the ABO_3 perovskite oxides without structural collapse.[8,9]

In this paper, we focus on single and double oxide perovskites and explore this large chemical space to identify compositions that are likely to form stable perovskites. We use a multipronged approach, investigating the formability and thermodynamic stability of oxide perovskites. To do this, we employ a combination of first-principles calculations and literature data to build an initial training data set. Machine learning (ML) classification models are then built using these training data sets. The models are then used to explore an exhaustive chemical space of single and double perovskites to screen for formable and stable new oxide perovskites. With reference to the formability and stability training data sets of

Received: August 20, 2020
Revised: December 28, 2020
Published: January 15, 2021



https://dx.doi.org/10.1021/acs.chemmater.0c03402
Chem. Mater. 2021, 33, 845−858

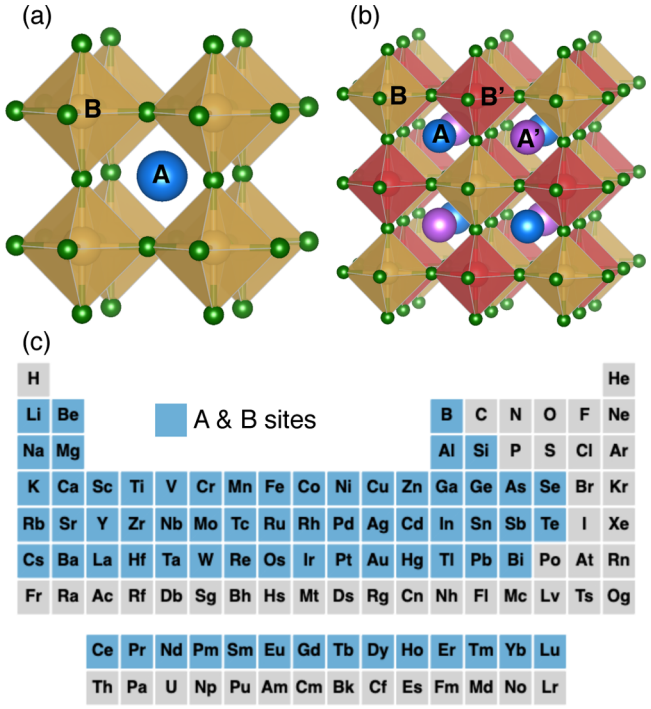


Figure 1. Crystal structure of (a) a single ABO_3 perovskite and (b) a double perovskite with two cations in each of the A- and B-sublattices: $AA'BB'O_6$. (c) Periodic table highlighting the 68 elements that are used to construct the comprehensive data set D_C of single and double oxide perovskites.

approximately 1500 and 3500 compounds, respectively, we identify about 125 times as many double perovskite compounds that are stable and more than 500 times as many compounds that are formable. From the intersection of this large chemical space of formable and stable oxide perovskites, 414 compositions are suggested as the most promising candidates for future experimental synthesis of novel oxide perovskites with very high confidence.

1.1. Ideal Perovskite Structure. The ideal cubic symmetry of the perovskite crystal is inflexible since there are no adjustable atomic position parameters in the unit, so that any compositional change must be concomitant with a change in lattice parameter. The cubic unit cell edge, a, is equal to twice the B−X bond length, that is, $2r_{B-X} = a$, and the width of the octahedral cage site, $\sqrt{2}\,a$, is equal to twice the A−X bond length, that is, $\sqrt{2}\,a = 2r_{A-X}$. Thus, the ideal structure forms when the ratio of the bond lengths is given by

$$\sqrt{2} = \frac{r_{A-X}}{r_{B-X}}$$

This relationship was first exploited by Goldschmidt in 1926,[10] who suggested that it could be used to predict the likelihood that a pair of ions would form a perovskite structure phase. When this was initially proposed, very few crystal structures had been determined, and so, ionic radii were used as a substitute for measured bond lengths.[1] Goldschmidt hypothesized that the number of anions surrounding a cation would necessarily be as large as possible, subject to the condition that all anions touch the cation (Goldschmidt's rule), thereby defining a tolerance factor t

$$t = \frac{r_A + r_X}{\sqrt{2}\,(r_B + r_X)}$$

where r_A, r_B, and r_X are the coordination-dependent Shannon's ionic radii[11] of the A-site cation, the octahedrally coordinated B cation, and the anion, respectively. Goldschmidt proposed that a perovskite structure phase would form if the value of the tolerance factor, t, was close to 1. This tolerance factor is found to be reasonably accurate in predicting the formability of cubic perovskites, especially for oxide perovskites, where ionic radii are precisely known. Empirical evidence suggests that, for compositions with t in the approximate range 0.9−1.0, a cubic perovskite structure is a reasonable possibility. If $t > 1$, that is, a composition with a large A cation and a small B cation, a hexagonal packing of the AX_3 layers is preferred, and hexagonal phases of the $BaNiO_3$ type[12] are more likely. In this work, such structures with hexagonal symmetry are not included in the purview of perovskites. In cases where t is between 0.71 and 0.9, the octahedral framework distorts to close down the octahedral coordination polyhedron, resulting in a lower symmetry crystal structure.[1] We shall discuss this in greater detail in the next section. The concept of the tolerance factor may be extended to perovskites with more complex compositions (*e.g.*, double perovskites) by using a weighted average value for ionic radii or bond lengths, albeit with limited utility.[13,14]

1.2. Crystallographic Variants of the ABX_3 Structure. As mentioned earlier, the ideal cubic symmetry of the perovskite crystal encompasses only one variable parameter—*viz.*, the lattice parameter. Consequently, a large number of perovskites stabilize in lower symmetry configurations via three different types of distortion: (i) distortions of the BX_6 octahedra, (ii) off-center displacements of the B cation within the octahedra, and (iii) the tilting of the BX_6 octahedra relative to one another (octahedral tilting).[15,16] Distortion of the BX_6 octahedra leads to elongated or flattened octahedra, which in extreme cases can lead to square planar or square pyramidal coordination. These distortions result from interactions between the cation electron orbitals and the surrounding anions and are exemplified by the Jahn−Teller effect which may be found in $KCuF_3$.[2] Octahedral distortion can also be caused by cation valence changes. The two different-sized cations then may adjust to the surroundings by a distortion of one or both of the cation-centred BX_6 octahedra, thereby giving rise to two different-sized octahedra. In $BaTiO_3$, B cation displacement is observed, which is conventionally associated with cations that are too undersized for the octahedral site, leading to a tolerance factor which is significantly less than 1.0. Depending upon the direction of cation displacement and the magnitude of the displacements, the structure may exhibit tetragonal, trigonal, or orthorhombic symmetry. Also, the displacements lead to permanent electric dipoles in the unit cell which can result in pyroelectric, ferroelectric, or antiferroelectric effects.[17] Octahedral tilting is frequently associated with A cations that are too small for the octahedral cage site, and so the BX_6 octahedra contort, effectively reducing the cavity dimensions, and thereby the structure accommodates values of t less than 1.0. Similar to B-site cation displacements, octahedral tilting also lowers the symmetry of the crystal and has a profound effect on the physical properties of these materials. Refer to Section 4.3 for a discussion on the effect of octahedral tilting on the stability and formability of perovskites.

https://dx.doi.org/10.1021/acs.chemmater.0c03402
Chem. Mater. 2021, 33, 845−858

These three distortions are not mutually exclusive, and they can occur independently or, sometimes, concomitantly. Moreover, the resulting changes may be cooperative, such that they affect all octahedra in a similar way, or non-cooperative, in which case the distortions may cancel out at a macroscopic level although they may still influence microscopic properties.[18] The degree of distortion is generally small and is readily influenced by the ambient conditions. Thus, changes in temperature, pressure, and crystal size or form may alter the magnitude of the distortion or the type of distortion present. The majority of perovskite phases manifest a series of symmetry changes as the temperature or pressure is changed, usually resulting in a cubic form at higher temperatures and pressure. Thus, the intrinsic flexibility and versatility of the perovskite structure allow for the formation of a wide number of derivatives, polymorphs, and superstructures, as well as the introduction of defects and structural disorder.[19]

2. PRIOR WORK

In the context of this work, we use two metrics to analyze elemental combinations within the ABO_3 perovskite structure with up to two A cations and two B cations to predict compositions likely to form stable oxide perovskites: (i) global formability and (ii) thermodynamic cubic stability. Global formability indicates if the material is experimentally known to form a stable perovskite or not. Consequently, since formability information is based off experimental information, it also includes perovskites that are synthesized in cubic as well as local distortion-induced lower symmetry phases. Thermodynamic cubic stability indicates whether it is energetically favorable for a particular composition to stabilize in the cubic perovskite structure and is determined by the position of the structure relative to the convex hull for the system. That is, predictions of formability start with experimentally known perovskites and, using trends in their structure, try to identify new compounds that would also be perovskites. On the other hand, predictions of stability start with the chemical composition of the compound and determine whether that chemistry would be stable in the perovskite structure. These two perspectives are complementary and have the same goal of identifying new perovskites but differ in the starting point to discover those materials. In the following sections, we describe these two metrics in greater detail and briefly explore prior work that has been done regarding the usage of these two metrics in the discovery of oxide perovskites.

2.1. Formability of Perovskites. Again, formability is the experimental realization of a perovskite with a given chemical composition. It accounts for any lowering of symmetry due to octahedral distortions or tilting. Because of the high interest in perovskites, a great deal of work has focused on developing empirical rules that relate basic structural properties with the potential formability of perovskites, with the goal of identifying new compounds that may form in the perovskite structure. A graphical representation of the way in which structures evolve as various physical parameters change has long been used to systematically study structure−property relationships in perovskites. In 1974, Muller and Roy proposed plotting a "structural map" which took the ionic radii of A and B as coordinates to study the distribution of different crystal structures in ABO_3 perovskites.[20] Such maps allow for the visual separation of the perovskites from other ABX_3-type structures and crystallographic variants within the perovskites themselves.

As mentioned earlier, the octahedron BX_6 is the basic motif of perovskite structures. The ratio of the radii of the cation B and anion X that form the octahedron gives rise to another dimensionless descriptor μ, called the octahedral factor,[21] and defined as

$$\mu = \frac{r_B}{r_X}$$

Naturally then, a structural map using the tolerance factor and the octahedral factor is most commonly used to predict perovskite formability. Lufaso and Woodward developed a software program SPuDs[22] to predict the crystal structures of perovskites, including those distorted by tilting of the octahedra based on the bond-valence (BV) model. In 2004, Li *et al.* plotted such a 2D (t, μ) map for 197 known ABO_3 systems and were able to successfully demarcate most known A−B combinations that led to perovskite structure phase formation from those systems that did not form perovskite phases.[23] Zhang and others then used two-dimensional structure maps based on BV-derived A−O and B−O bond distances to classify 376 experimentally known ABO_3-type perovskites and nonperovskites.[24] A few years ago, Filip and Giustino[14] extended the use of structure maps to also include $A_2BB'X_6$-type perovskites by introducing an additional octahedral mismatch parameter $\overline{\mu}_B = |r_B - r_{B'}|/2r_X$. Using three dimensional structure maps consisting of $(t, \mu, \overline{\mu}_B)$, they were able to predict the formability of perovskites with a fidelity of 80% and also generated a library of 94,232 hitherto-unknown perovskites and double perovskites that are expected to form.

Recently, research has focused on the use of data science and ML to screen for new perovskites. Artificial neural network classification models based on BV-based tolerance factors were used by Zhang *et al.*[25] to predict the formability of general ABO_3-type perovskites. Pilania *et al.*[26] successfully applied hyper-dimensional structure maps based on tolerance factors, ionic radii, and bond distances to the classification of ABO_3 perovskites and nonperovskites. Bartel *et al.*[13] used a compressed-sensing-based dimensionality reduction scheme to identify a one-dimensional tolerance factor τ, given by

$$\tau = \frac{r_X}{r_B} - n_A\left(n_A - \frac{r_A/r_B}{\ln(r_A/r_B)}\right)$$

where n_A is the oxidation state of A. Using this metric, they were able to demarcate the compounds as perovskite and nonperovskites with 92% fidelity for an experimental data set of 576 ABX_3 materials (oxides and halides). They also applied this model to the set of double perovskites and predicted 23,314 new double perovskites of the form $A_2BB'O_6$. Xu *et al.*[27] used crystal structure data made available via the Materials Project[28] database and developed a procedure to identify the perovskite formability of experimentally known ABX_3 and $A_2BB'X_6$ compounds. They also identified $A_2BB'X_6$ compounds with suspect formability information in current experimental data.

2.2. Thermodynamic Stability of Perovskites. As described in the preceding sections, conventionally, the formability of perovskites relies on geometric criteria derived using either ionic radii or bond distances and is a qualitative approach to identifying chemistries that will form perovskites. However, it is also fundamentally essential to develop a quantitative picture of the relationship between the perovskite

https://dx.doi.org/10.1021/acs.chemmater.0c03402
Chem. Mater. 2021, 33, 845−858

structure and its stability from a thermodynamic perspective. For example, it is necessary to know whether, if a perovskite is determined to be formable, does it follow that it is thermodynamically stable? Will it decompose into other phases under certain conditions? The thermodynamic stability of a compound may be accessed using an energy convex hull construction for the phase space of interest. The convex hull comprises phases that have an energy lower than any other phase or linear combination of phases at the composition of interest.[29] Thus, the convex hull is a formation energy surface that is a function of the chemical composition that passes through all lowest energy phases that are thermodynamically stable and do not decompose into other phases. To put it simply, a composition is thermodynamically stable if its free energy is the minimum of the free energy of all possible decomposition routes, including its elementary constituents and all possible binary, ternary, and more complex compounds.[30] The energy above the convex hull (ΔE_h) is defined as

$$\Delta E_h^{ABX_3} = E_f^{ABX_3} - E_{ch}$$

where $E_f^{ABX_3}$ is the formation energy of the ABX_3 compound and E_{ch} is the convex hull energy at the ABX_3 composition. (Note that, while we describe the energetics of the convex hull in terms of a single perovskite here, the same considerations apply for a double perovskite.) The formation energy $E_f^{ABX_3}$, which is the change of enthalpy during the formation of the ABX_3 compound from its constituent elements, may be calculated as

$$E_f^{ABX_3} = E^{ABX_3} - \mu_A - \mu_B - 3\mu_X$$

where E^{ABX_3} is the total energy of the ABX_3 compound and μ_A, μ_B, and μ_X are the chemical potentials of A, B, and X, respectively. All phases that lie on the convex hull are thermodynamically stable, while the ones above the convex hull are metastable or unstable. Ideally, this thermodynamic analysis should be based on Gibbs free energy comparisons; however, as an approximation, we have resorted to 0 K density functional theory (DFT)-computed enthalpies, and all the entropic contributions have been ignored.

Thus, the distance from the convex hull, ΔE_h, may be used as a direct measure of thermodynamic stability. To this end, Sun and Yin in 2017[31] used first-principles calculations to calculate the decomposition energies of 138 perovskite compounds and found that $(\mu + t)^\eta$, where η is the atomic packing fraction, demonstrates a remarkably linear correlation with thermodynamic stability. Zeng *et al.*[32] used first-principles calculations to demonstrate generalized trends in the formation energies of perovskite oxides using simple descriptors such as the number of valence electrons and oxidation states. Artini *et al.*[33] investigated the stability of inter-lanthanide perovskites and deduced a geometry-based stability criterion to derive the formability likelihood of interlanthanide perovskites. They also verified their approach by applying it to non-perovskite interlanthanide oxides. As with formability studies, due to the inherently arduous nature of energy hull construction and stability determination, recent years have seen an impetus in the use of ML approaches to tackle this problem. Li and Morgan developed ML models to predict the thermodynamic phase stability of ABO_3 perovskite oxides via convex hull analysis using a data set of more than 1900 DFT-calculated compounds.[34] Similarly, Liu *et al.*[35] built ML models using perovskite data from the Materials Project to screen thermodynamically stable ABO_3 perovskites.

In recent years, some research has focused on combining the ideas of perovskite formability and thermodynamic stability to predict new perovskites. Pilania *et al.*[36] employed a hierarchy of down-selection steps based on structure factors and thermodynamic stability in addition to constraints on band gap and band-edge positions to identify potential candidates residing in a target double perovskite chemical space for water-splitting applications. This screening resulted in four double perovskites, which were then extensively analyzed for thermodynamic stability, dynamical stability, electronic structure, and octahedral structural distortions. Balachandran *et al.*[37] applied ML methods to a database of 390 experimentally reported ABO_3 compounds to construct statistical models that predict possible new perovskite materials and cubic perovskites. They then compared the ML findings with DFT calculations and convex hull analyses using the Open Quantum Materials Database (OQMD).[38] They found that OQMD predicted 87 of the 235 ML-predicted perovskite compounds to be thermodynamically stable within 100 meV/atom of the convex hull, including 6 which were predicted to be cubic, and further presented these 87 as the most promising candidates for future experimental synthesis of novel perovskites.

In the light of all the work that has focused on formability and stability prediction of oxide perovskites, it is evident that almost all research—with few exceptions—has focused exclusively on ABO_3-type single perovskites. Double oxide perovskites, which offer an astronomically large composition space, remain comparatively unexplored. The current work looks to fill this gap by investigating the formability and stability in the chemical space of $A_2BB'O_6$, $AA'B_2O_6$, and $AA'BB'O_6$ double perovskites using ML classification techniques. The goal then is to identify hitherto unknown double perovskite oxides that may be formable and/or thermodynamically stable and offer unique chemistries which may be exploited to harness properties superior to those of existing known perovskites. A secondary goal is to compare predictions of formability against those of cubic stability to better understand the complementary aspects of these two views of structure prediction.

The rest of the article is organized as follows. In Section 3, we discuss the two training databases that are built to predict the formability and thermodynamic cubic stability of oxide perovskites, respectively, the features generated, and the technical details of the ML classification models that are then trained using these databases. In Section 4, we showcase the results obtained for the formability and stability classification and discuss some insights gleaned. Finally, we offer our conclusions and a brief outlook on the future of ML to predict new stable materials in Section 5.

3. INPUTS AND METHODS

3.1. Literature-Based Formability Database (D_F). In the course of this work, we built two databases, one for formability and another for stability predictions, respectively. The training data set for perovskite formability D_F was composed of 1505 experimentally known oxides of the form ABO_3 and $AA'BB'O_6$ collated from literature. Of these, 318 compounds are nonperovskites, while 1187 are perovskites. The data set includes both single and double perovskites and is a conglomeration of data from the Inorganic Crystal Structure Database (ICSD) and databases compiled by Zhang *et al.*,[24] Balachandran *et al.*,[37] and Vasala and Karppinen.[39] Figure 2

https://dx.doi.org/10.1021/acs.chemmater.0c03402
Chem. Mater. 2021, 33, 845−858

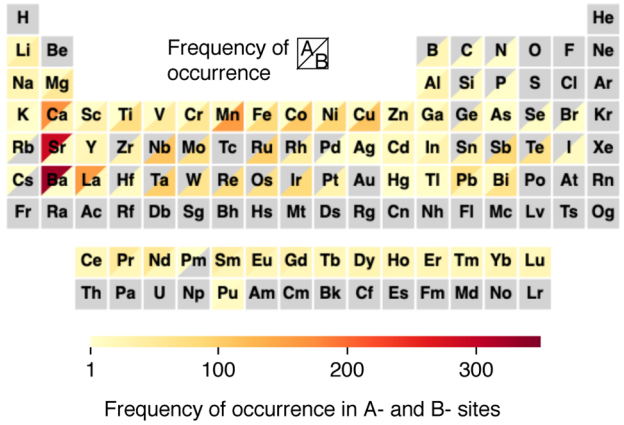


Figure 2. Periodic table showing the number of occurrences of each element in the perovskite formability database D_F. Each square indicates the respective element, with the upper left triangle indicating occurrences for the A-site cations and the lower right triangle indicating occurrences for the B-site cations. Gray elements indicate no occurrences.

shows the periodic table detailing the distribution of the elements occupying the A- and B-sites for all the compounds. It can be seen, for example, that there is a very large number of known perovskite oxides with elements Sr and Ba at the A-site.

3.2. DFT-Based Stability Database (D_S). For the second training data set, which is used to predict the thermodynamic stability of perovskites, we start by enumerating all possible ABO_3, $A_2BB'O_6$, $AA'B_2O_6$, and $AA'BB'O_6$ compounds. All elements up to Bi, with the exception of the noble gases, the halogens and C, N, S, and P are taken into account. This amounts to 68 elements, shown in Figure 1c, leading to ${}^{68}P_2 = 68!/(68-2)! = 4556$ different possible ABO_3 combinations. For the $A_2BB'O_6$, and $AA'B_2O_6$, accounting for equivalent A and A′ sites and B and B′ sites, we have ${}^{68}P_3 = 68!/2 \times (68-3)! = 150{,}843$ different combinations each. For the case of two A cations and two B cations, that is, the $AA'BB'O_6$ structures, there are a possible ${}^{68}P_4 = 68!/(68-4)! = 19{,}545{,}240$ combinations, which upon accounting for equivalent A and A′ sites and B and B′ sites reduces to 4,886,310 combinations. For the double perovskites, we have considered only equimolar cation compositions at the A- and/or B-site sublattices to keep the size of the data set tractable. These ~5.19 million combinations were then screened for charge neutrality, that is, the A, A′, B, B′ cations were assigned valences based on the most commonly exhibited oxidation states and those combinations for which the sum of the valences was 6 for the single perovskites and 12 for the double perovskites were kept for further consideration. This enumeration excluded double perovskite compositions where the A and B sublattices have a common cation chemistry. This resulted in a number of compounds with multiple valence state possibilities, all of which were included in the data set. This left us with a set of 946, 292 unique compounds, (some of which have multiple valence combinations) which is our foundational chemically compatible data set D_C.

From the list of compounds that belong to D_C, the training data set for thermodynamic stability, D_S, of perovskites was calculated. Compounds were selected for calculation to maintain substantial overlap with the formability database D_F and also with a view to ensuring maximum diversity in the included chemistries. For all the chemistries, we constrain the structure to cubic while considering the more commonly observed rocksalt type ordering[40,41] for the arrangement of cations on both the A- and B-site sublattices to keep the high-throughput calculations tractable. As mentioned in Section 1.2, the majority of oxide perovskites are known to be synthesizable in tetragonal, orthorhombic, and monoclinic symmetries instead of cubic. However, as we will explain later, we account for this by assigning a tolerance to the energy hull calculations for stability. In total, 3469 oxides with ABO_3 and $AA'BB'O_6$ stoichiometry were selected from D_C, and the structures were optimized and their total energies were calculated with DFT, as described below. Figure 3

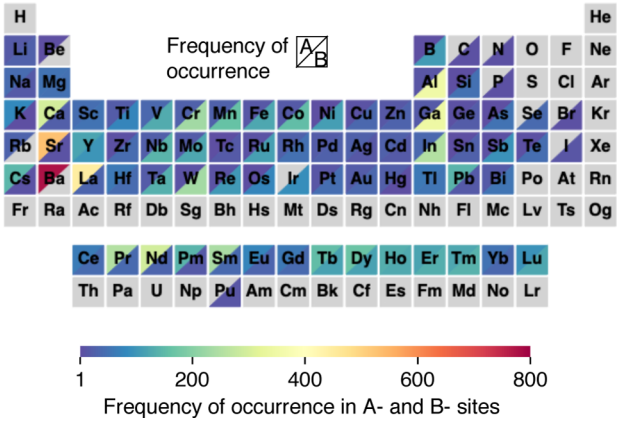


Figure 3. Periodic table showing the number of occurrences of each element in the calculated thermodynamic stability data set D_S. Each square indicates the respective element, with the upper left triangle indicating occurrences for the A-site cations and the lower right triangle indicating occurrences for the B-site cations. Gray elements indicate no occurrences.

shows the periodic table detailing the distribution of the elements occupying the A- and B-sites for all the compounds in the D_S dataset.All parameters and pseudopotential flavors were set to guarantee compatibility with the data available in the Materials Project[28] database. The Materials Project Application Programming Interface (API) was then used to extract the energy above the convex hull (ΔE_h) for each of these structures. For these energy above hull calculations, the Generalized Gradient Approximation (GGA) compatibility was ensured, that is, only GGA (and not GGA + U) calculations from Materials Project were used to compute the energy hull. Structures with ΔE_h below 50 meV were labeled as stable. Wu *et al.*,[42] in their high-throughput screening work on oxynitrides for water-splitting photocatalysts, obtained a DFT error value for convex hull analysis by examining the DFT-calculated energy above hull values of experimentally synthesized compounds contained in the ICSD. They found that greater than 80% of the compounds in the ICSD have a ΔE_h value of less than 36 meV/atom. To account for metastable cubic structures (structures that are not the ground state but still might be synthesizable) and/or potential relaxation to lower symmetry structures, we round up to a ΔE_h cutoff value of 50 meV/atom. Applying this 50 meV threshold criterion, of the 3469 oxides, 1501 were found to be thermodynamically stable, while the remaining 1955 were found to be thermodynamically unstable.

3.3. Feature Generation and Selection. For all the compounds, the databases include values of features associated with their A and B atoms in the case of single perovskites and A, A′, B, B′ atoms for the double perovskites. For the double perovskites, symmetric and antisymmetric compound features were used, an approach that has been previously explored by Pilania *et al.*[43,44] For a double perovskite $AA'BB'O_6$, for a given property P, the symmetric compound feature may be calculated as $P^{A+} = (P_A + P'_A)/2$ and the antisymmetric compound feature $P^{A-} = |P_A - P'_A|/2$ for the A-site, where P_A and P'_A are the elemental properties of A and A′; similar features were defined for the B-site.

Initially, a very large number of structural and chemical features were included to train the classification models. Recursive feature elimination (RFE)[45] as implemented in the open-source Scikit-learn[46] python package was used for feature selection. RFE works to select features by recursively removing those features which exhibit the smallest weight that are assigned by an extra trees classifier. The estimator is first trained on the initial set of features, and the importance of each feature is ascertained. The least important features are then pruned from the current set of features. This procedure is

https://dx.doi.org/10.1021/acs.chemmater.0c03402
Chem. Mater. 2021, 33, 845−858

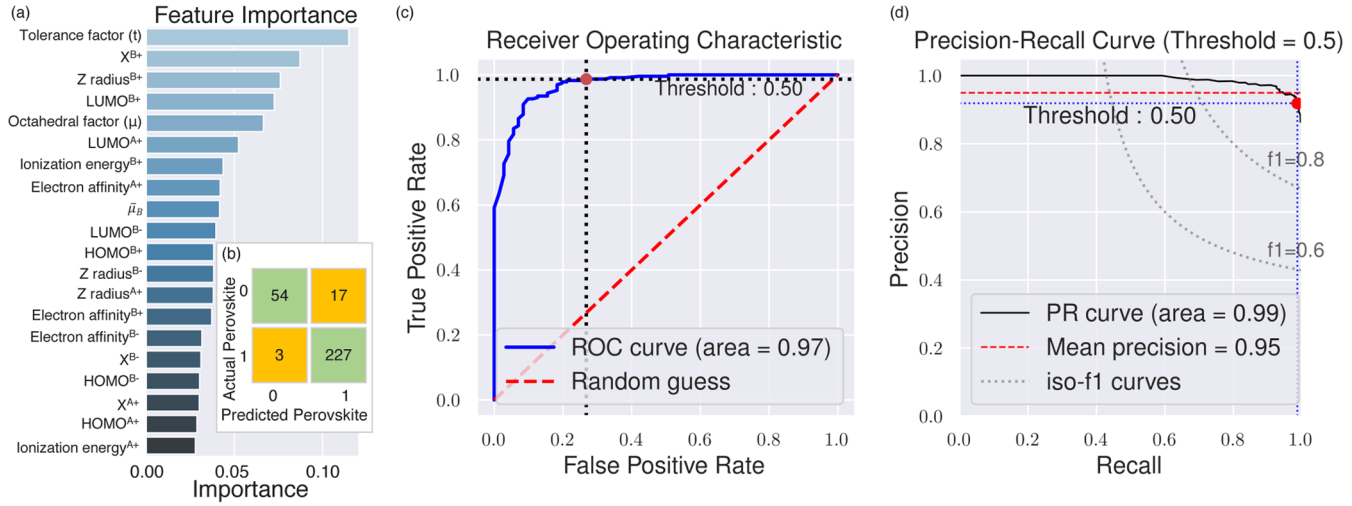


Figure 4. Random forest classification results for perovskite formability. (a) Feature importance plot for all the features with non-zero values, (b) confusion matrix, (c) receiver operating characteristic (ROC) curves, and (d) precision-recall curves of the cross-validated random forest classification on test data.

recursively repeated on the pruned set until the desired number of features to select is eventually reached. Prior to performing feature selection, all features were normalized to ensure zero mean and standard deviation of unity. The features thus retained clearly indicated that only a subset of the initially considered features are actually influential in describing the formability and stability of the oxide perovskite. These include atom-specific features—the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energies,[47] ionization energy, electronegativity,[48] Zunger's pseudopotential radius (sum of the radii for the s and p orbitals), and electron affinity—and geometric features of the compound—the tolerance factor (t), octahedral factor (μ), and mismatch factor ($\bar{\mu}_B$). Since we include compounds with two cations in the A-as well as the B-sites, we also introduce an additional mismatch factor for the A-site: $\bar{\mu}_A = |r_A - r_{A'}|/2r_X$. Symmetric and antisymmetric compound features for the A- and B-sites are used for the first six of these features resulting in 24 compound features. Including the four geometric features, we are left with a total of 28 features. Values of these 28 features for the D_F database and the D_S database are provided in the Supporting Information. The Supporting Information also includes the DFT-calculated compound features such as the total energy of the compound, the formation energy of the compound, and the energy above hull ΔE_h. In addition, the formability database D_F also includes information on whether the compound is known to be a perovskite or nonperovskite. The HOMO and LUMO Kohn–Sham levels of isolated atoms were computed using DFT computations without spin-polarized calculations in a large orthorhombic supercell (15 Å × 14.5 Å × 14 Å, to break the cubic symmetry) with respect to the vacuum level.[43,49]

3.4. First-Principles Calculation Details. The total energy calculations were performed within the DFT[50] framework, as implemented in the *Vienna ab initio* simulation package (VASP).[51,52] The GGA[53] is used in the form of the parameterization proposed by Perdew, Burke, and Ernzerhof (PBE).[54] Brillouin zone integrations were performed using a Monkhorst–Pack mesh with at least 5000 k points per reciprocal atom. Full relaxations were realized by using the Methfessel–Paxton smearing method[55] of order one and a final self-consistent static calculation with the tetrahedron smearing method with Blöchl corrections.[56] A cutoff energy of 533 eV was set for all of the calculations, and spin polarization was accounted for. The cell volumes and ionic positions of the structures were allowed to relax to their cubic ground states, and the relaxations were carried out in three stages: first stage by allowing changes in the volume (corresponding to the VASP ISIF = 7 tag), second stage by allowing only the ions to relax (corresponding to the VASP ISIF = 2 tag), and a final self-consistent static calculation run. All relaxations were carried out until changes in the total energy between relaxation steps were within 1×10^{-6} eV and atomic forces on each of the atoms were smaller than 0.01 eV/Å.

As noted earlier, the Materials Project database was used to extract the energy above hull per atom (ΔE_h) for all the compounds in the data sets. Materials Project consists of DFT calculations of experimentally observed compounds that are part of the ICSD database as well as hypothetical compounds that may or may not be experimentally observable. To construct the convex hull for each A–A′B–B′–O set of elements, the total energy data of all compounds that exist in the Materials Project database is taken into account and extracted using the database API. The calculated information for the compound in question is also added to this extracted data, and the distance from the hull is then determined. We note that in this work, we use the relatively inexpensive PBE exchange–correlation functional of DFT which is extensively used in the field and has a proven reliable track record for perovskite calculations. While more recent functionals such as the Heyd–Scuseria–Ernzerhof (HSE06)[57] hybrid functional provide a more accurate estimate, this is achieved at a much higher computational cost. A high-throughput study such as the present one necessitates a low-cost technique and recent studies have indicated that ground-state properties calculated using the PBE functional are sufficiently accurate.[58−60]

3.5. ML Classification Using Random Forests: Technical Details. Classification is a process of categorizing a given set of data into categories. In ML terminology, classification is a form of supervised learning—the prediction of a variable (or class) based on values of other variables or features. In this work, we used the random forest classifier[61] (RFC) for both formability and stability classification models. The random forest (also known as random decision trees) is a bagging-based ensemble learning method. Tree-based ensemble methods combine several decision trees to produce better predictive performance than utilizing a single decision tree. Bagging is a technique to grow ensemble decision trees, where to grow each tree, a random subset selection is made from the examples in the training set.[61] Bagging algorithms aim at reducing the complexity of models that overfit the training data in contrast to boosting, which is an approach to increase the complexity of models that underfit the training data.

Random Forest is an extension over bagging in that it takes several randomly selected subsets of the supplied feature sets to grow trees rather than using the entire feature set all at once. Thus, many random decision trees are constructed using randomly selected subsets of training data and subsets of the entire feature set. The prediction is improved due to better variance-bias trade-offs, and the technique is inherently robust against overfitting. In the case of a random forest,

https://dx.doi.org/10.1021/acs.chemmater.0c03402
Chem. Mater. 2021, 33, 845−858

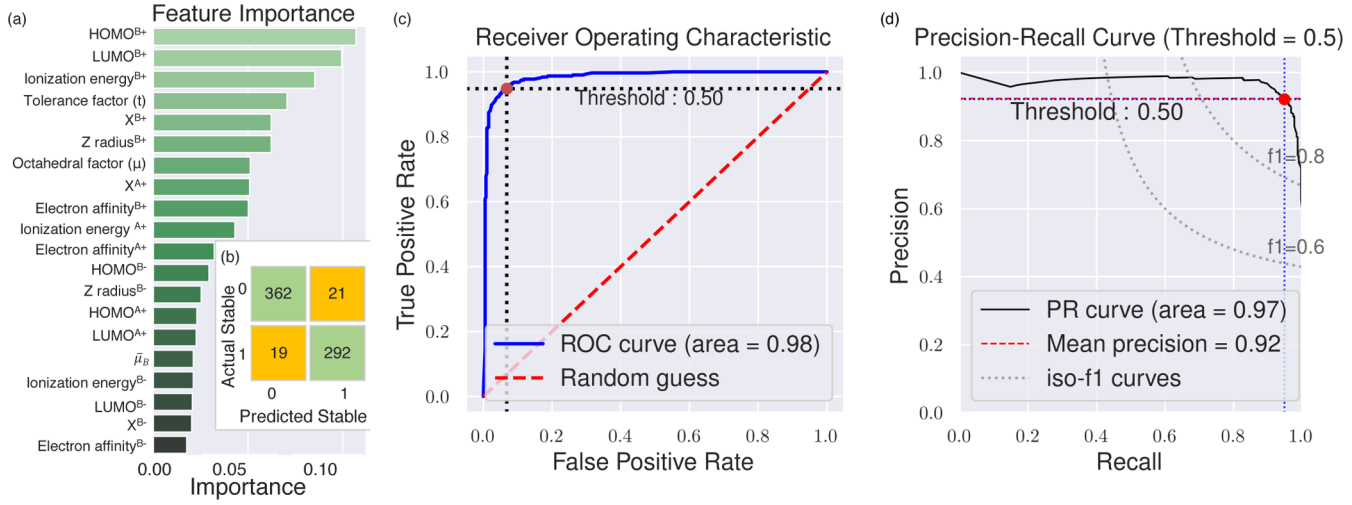


Figure 5. Random forest classification results for perovskite stability. (a) Feature importance plot for all the features with non-zero values, (b) confusion matrix, (c) receiver operating characteristic (ROC) curves, and (d) precision-recall curves of the cross-validated random forest classification on test data.

the hyperparameters include the number of decision trees in the forest and the tree depth considered by each tree when splitting a node. The predicted class probabilities of an input sample are computed as the mean predicted class probabilities of the trees in the forest. These class probabilities are used in this work to identify the best perovskite oxide candidates.

Here, we implement the random forest classifiers within the framework of the Scikit-learn package. We use a 80/20 training/test split for both the formability and stability classification models. In each case, the training and test data set selections were stratified over either the perovskite/non-perovskite candidates or stable/unstable chemistries. Thus, the models were trained on 80% of the data sets and then tested on the remaining 20%. The maximum tree depth was set at 22 for the formability classification and at 23 for the stability classification. The number of estimators or trees was 28 for both classification models. The hyper-parameters were determined using a 5-fold cross-validation with a view to maximizing accuracy while minimizing the standard deviation on unseen data. The cross-validation was also carried out using a 80% training subset. Mean and standard deviations for the model accuracy were calculated on randomly selected test sets over 100 independent runs. For the formability classification, the mean accuracy was 0.9401 with a standard deviation of 0.086, while, for the stability classification, a mean accuracy of 0.9409 with a standard deviation of 0.089 was achieved.

4. RESULTS AND DISCUSSION

The two RFC models for formability and stability prediction described above were built using the 28 features listed in Section 3.3 and the respective training data sets D_F and D_S. Both models were also tested using random features to gauge the robustness of the models and to identify the true importance values for the features. As expected, these random features were found to rank near the bottom in the relative feature importance lists. The models were then tested on the test data sets created using the 80/20 shuffle split mentioned earlier. Figures 4 and 5 show the detailed feature importances and performance metrics for the formability and cubic stability classification models, respectively. The relationship between the formability and stability of perovskites was then explored. Subsequent to the training and testing phases, the models were then applied to the comprehensive perovskite data set D_C separately to predict formable and/or stable oxide perovskites. In the following sections, we shall discuss the results obtained in detail.

4.1. Perovskite Formability Classification. In Figure 4a, we present the feature importances for the perovskite formability classification problem. Unsurprisingly, we find that the traditional geometric tolerance factor (t) and octahedral factor (μ) are prominently high on this list. In addition, a number of B-site symmetric features such as the Zunger pseudopotential radius (Z radius), the electronegativity (X), and LUMO are the most important in differentiating between compounds that form and do not form perovskites. The pseudopotential radius used was the sum of the pseudopotential radii of the s and p orbitals. Figure 4b shows the average test confusion matrix for 100 runs of the model. The off-diagonals (in red) indicate the False Positives (FP) and False Negatives (FN), while the diagonal elements (in green) indicate the True Positives (TP) and True Negatives (TN). Accuracy, precision, and recall metrics may be evaluated using this confusion matrix. Accuracy is the proportion of correct classifications which is given by the ratio of the sum of the TP and TN to the total number of data points. Precision is a ratio of the number of True Positives divided by the sum of the True Positives and False Positives. It describes how good a model is at predicting the positive class or in this case predicting the formable perovskite class. Recall is calculated as the ratio of the number of true positives divided by the sum of the true positives and the false negatives and is thus a measure of sensitivity. Hence, it is the appropriate metric to use when there is a high penalty associated with False Negative. The formability classification model yields accuracy, precision, and recall values of 0.9401, 0.9344, and 0.9913, respectively.

As mentioned earlier, Bartel *et al.*[13] applied their model to the set of double perovskites and predicted 23,314 $A_2BB'O_6$ perovskites. The formability predictions in this work were compared with their predictions. Of these 23,314 predicted compounds, when we consider only the oxides and neglect the compounds containing the elements that are not included in our study (actinides, C, P, Br, Cl, I, P, S, *etc.*), we are left with 11,048 predicted $A_2BB'O_6$ perovskites. Of these, 709 compounds are part of our perovskite formability training

https://dx.doi.org/10.1021/acs.chemmater.0c03402
Chem. Mater. 2021, 33, 845−858

data set. 608 of these compounds are experimentally known to be perovskites, while 101 are known to not be perovskites. Of the remaining 10,339 predicted $A_2BB'O_6$ compounds, 9450 (*i.e.*, 91%) are also predicted by this work to be perovskites. Thus, there is considerable overlap between the two sets of results, though it is not 100%.

Figure 4c shows the receiver operating characteristic (ROC) curves of the cross-validated random forest classification on the test data. It illustrates the performance of a binary classifier system as its discrimination threshold is varied. The ROC curve is a plot of true positive rate on the y axis and false positive rate on the x axis. The top left corner of the curve is thus the ideal point—where the true positive rate is 1 and the false positive rate is 0; hence, the greater the area under the curve (AUC), the better the performance of the classifier. Here, we highlight the value of the curve for a threshold value of 0.5, that is, compounds having a probability of >0.5 being a perovskite are classified as perovskites, while those with a probability <0.5 are classified as nonperovskites. The classifier performs very well with an AUC of 0.96. In Figure 4d, we show the precision and recall curves for the classifier. Similar to the ROC curve, a precision-recall (PR) curve is a plot of the precision rate on the y axis and the recall rate on the x axis for different threshold values. Thus, a model with perfect predictive capabilities is depicted as a point at a coordinate of (1,1). A skillful model is represented by a curve that tends toward a coordinate of (1,1). We also indicate the $f1$-score curves on the plot, where $f1$ is the harmonic mean of the precision and recall rates for a threshold value of 0.5. Again, the classifier performs very well, being very close to the ideal position on the plot for a threshold of 0.5. Just like the ROC AUC, the PR-AUC is similar in that it summarizes the curve with a range of threshold values as a single score, where a score of 1.0 indicates a perfect model. The PR-AUC for this model is 0.99, indicating a near-perfect model.

4.2. Perovskite Stability Classification. Similar results are shown for the perovskite stability classification in Figure 5. The feature importances are indicated in Figure 5a and are markedly different from those found from the formability classification. The tolerance factor t is the only geometric feature of importance, while again the B-site symmetric features of HOMO, ionization energy, LUMO, and pseudopotential radius are important indicators of stability. Figure 5b−d indicates the confusion matrix, ROC curves, and precision-recall curves, respectively, for the stability classification model. From the confusion matrix, that is Figure 5b, we see that the thermodynamic stability classification model yields accuracy, precision, and recall values of 0.941, 0.933, and 0.936, respectively. This classifier also performs admirably, with an AUC = 0.98 for a threshold value of 0.5, as seen in Figure 5c. The PR curve in Figure 5d also indicates that, for a threshold of 0.5, the classifier is satisfactorily close to the ideal (1,1) position on the plot with a PR-AUC of 0.97, again indicating a near-perfect model. Thus, all metrics indicate that the classifier is robust and is able to distinguish between stable and unstable perovskites exceptionally well. We attribute the remarkable performance of both the perovskite formability and thermodynamic stability classification models to the very large training data sets that were painstakingly developed, ensuring the inclusion of the maximum possible number of chemistries.

4.3. Comparison of Perovskite Formability and Cubic Stability. It is not known with certainty whether a formable perovskite is necessarily thermodynamically stable and vice versa. Are both formability and thermodynamic stability necessary to guarantee the viability of a composition as a perovskite candidate or is one a more robust metric compared to the other? The intersection of the compounds in the formability and stability data sets was extracted to form a new data set $D_{F\cap S}$ in which chemistries that are both experimentally known to form a perovskite or not and for which we have energies above hull from DFT are included. This data set is found to comprise 1153 compounds. Of these, 156 or approximately 10% are known to be nonperovskites, while the rest (997) are perovskites. Using an energy above a hull threshold of 50 meV/atom, 694 of the 1153 are labeled as stable, while the remaining 459 are unstable. It should be noted that, here, we are in principle comparing only chemistries. For the DFT calculations, we assume cubic symmetry; however, many of the known perovskites are found to have been experimentally synthesized at lower non-cubic symmetries. Recall that, for our stability data set, we fixed a 50 meV/atom stability threshold to account in part for stability at lower symmetries.

Figure 6a shows the energy above hull (ΔE_h) in meV plotted against the tolerance factor (t) for this data set $D_{F\cap S}$. Perovskite compounds are indicated in blue, while nonperovskites are indicated in orange. On close analysis (as shown in Figure 6c), it is found that, for a E_c threshold of 50 meV/atom, 56% of the compounds in the data set are perovskites and thermodynamically stable and 10% are not perovskites and not stable, leading us to conclude that, for 66% of the data set, there is agreement between the formability and stability metrics. However, 30% of the data set comprises compounds that are known to be perovskites but found to be unstable, while the remaining 4% of the data set are not perovskites but are calculated to be stable—resulting in disagreement between the metrics for 34% of the data set. This disagreement is expected and may be attributed to four reasons. First, our convex hull analysis may be incomplete due to missing phases, hence limiting the accuracy of our stability analysis. This would account for the 4% which are not perovskites but are stable. A cursory convex hull analysis using the OQMD[38] database indicated more overlap between the stable compounds and formable compounds in the $D_{F\cap S}$ database; however, since the pseudopotential flavors used in this work are not entirely compatible with the OQMD database, a detailed analysis was not feasible. Second, the formability data set includes oxides that are metastable and are not thermodynamic ground states. The 50 meV/atom threshold may not be sufficient to offset this metastability and hence would account for some of the formable perovskites that are not stable. Third, our thermodynamic stability database is restricted to cubic structures and does not account for all the perovskites that stabilize in lower symmetry structures due to the various distortions discussed in Section 1.2, accounting for the compositions in the data set that are perovskites, but not stable. Finally, the formability database comes directly from experimental observations, while the cubic stability database is based on DFT computations. Therefore, the accuracy of DFT (GGA-PBE) computations itself may be a source of disagreement. As shown in Figure 6a, increasing the E_c threshold to 200 meV/atom results in 85.4% agreement between the two metrics. A threshold of 400 meV/atom increases this to 91%, and a breakdown of the numbers is shown in Figure 6b. This indicates that a threshold for stability of E_c = 50 meV/atom fails to account for those compounds which stabilize at lower

https://dx.doi.org/10.1021/acs.chemmater.0c03402
Chem. Mater. 2021, 33, 845−858

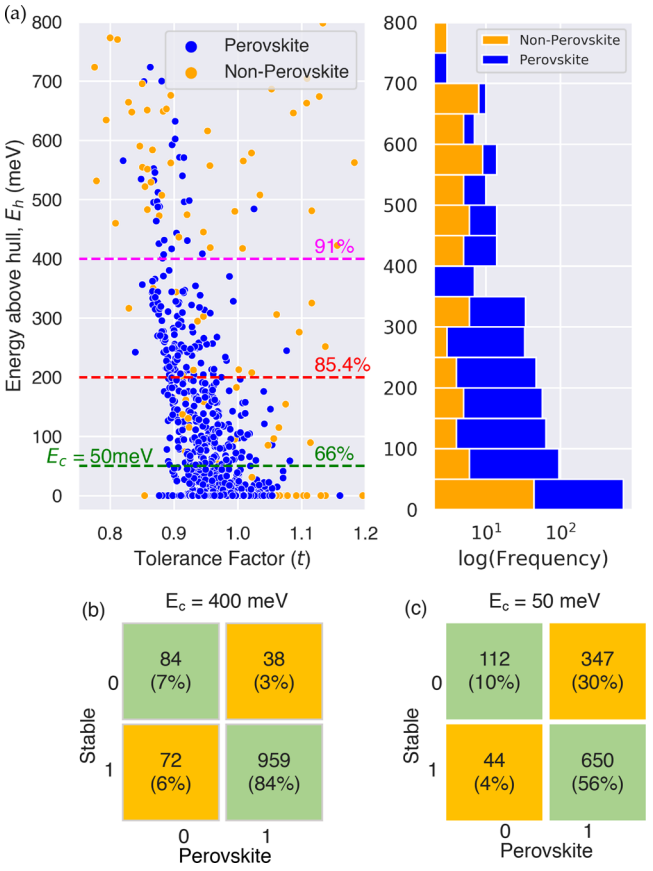
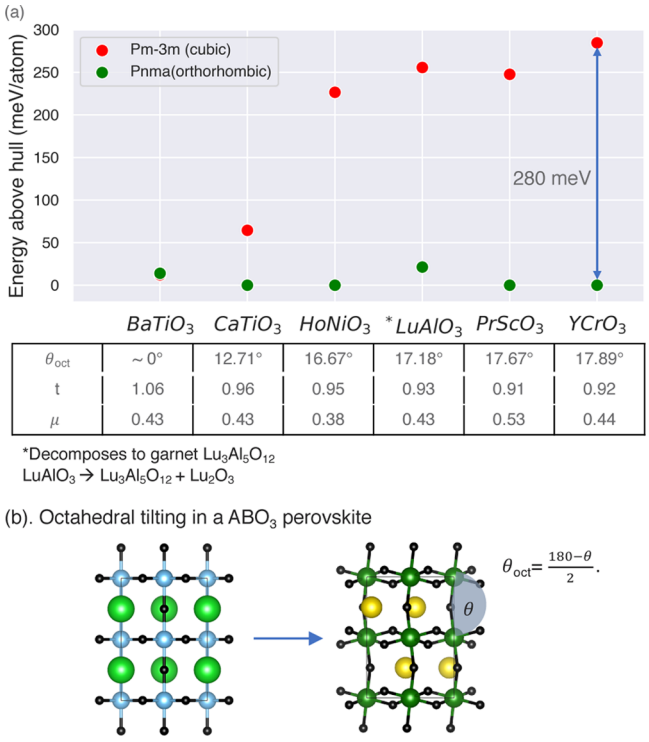


Figure 6. (a) Energy above hull (ΔE_h) in meV is plotted against the tolerance factor (t) for the data set $D_{F\cap S}$ which is the intersection of the perovskite formability and stability data sets (D_F and D_S). Experimentally known perovskite or nonperovskite nature of the compounds is also indicated by the color of the points. Horizontal lines at various threshold values of ΔE_h are shown along with percentage values which indicate the agreement between both data sets at those threshold values. On the right is a bar plot indicating the distribution of perovskites and nonperovskites along the energy above hull axis. (b,c) Tables indicating distribution of formability and stability data for a threshold criterion (b) E_c = 400 meV/atom and (c) E_c = 50 meV/atom.

symmetries (non-cubic) due to the various distortions discussed in Section 1.2.

To further elucidate this point, we compared the energy above hull (ΔE_h) of cubic and orthorhombic variants of 6 ABO_3 compounds, which is shown in Figure 7a. The orthorhombic variant *Pnma* was selected since it is the most commonly found stable symmetry in the case of single perovskites. For all the compounds shown, with the exception of $BaTiO_3$, we see that the *Pnma* variant is much more stable than the cubic $Pm\bar{3}m$ structure, with a difference in energy of upto 300 meV between the variants for $YCrO_3$. For $BaTiO_3$, the cubic and orthorhombic variants are very close in energy and above the convex hull—the actual stable symmetry is known to be the $Amm2$ orthorhombic variant, which is lower in energy than the $Pm\bar{3}m$ and the *Pnma* variants. Further, for $LuAlO_3$, we see that the *Pnma* variant is above the hull, and this is easily explained since $LuAlO_3$ is known to decompose into the garnet structure $Lu_3Al_5O_{12}$. However, $LuAlO_3$ highlights the necessity of the tolerance criterion E_c, as, while the perovskite structure is not the ground state for this compound, it can still be synthesized in a perovskite structure.



	$BaTiO_3$	$CaTiO_3$	$HoNiO_3$	*$LuAlO_3$	$PrScO_3$	$YCrO_3$
θ_{oct}	~0°	12.71°	16.67°	17.18°	17.67°	17.89°
t	1.06	0.96	0.95	0.93	0.91	0.92
μ	0.43	0.43	0.38	0.43	0.53	0.44

Figure 7. (a) Energy above hull (E_h) of cubic and orthorhombic variants of 6 ABO_3 compounds. The table lists the octahedral tilt angle θ_{oct}, tolerance factor (t), and octahedral factor (μ) for the six compounds under consideration. (b) Octahedral tilting in a ABO_3 perovskite.

If we limited our consideration to only those compounds that lie on the convex hull (even if we were able to account for all possible symmetry variants), metastable perovskite compounds such as $LuAlO_3$ would always be excluded from consideration.

As discussed in Section 1.2, octahedral tilting is associated with A cations that are too small for the octahedral cage site; thus, the BX_6 octahedra twist so as to effectively reduce the cavity dimensions, again allowing the structure to accommodate values of t less than 1. Figure 7b schematically depicts the octahedral tilting in a ABO_3 perovskite. The octahedral tilt angle θ_{oct} is defined as $\theta_{oct} = (180 - \theta)/2$, where θ is the B−O−B bond angle. The table at the bottom of Figure 7a lists the octahedral tilt angle θ_{oct} for the six compounds under consideration. On average, as the tilt angle increases, so does the energy difference between the $Pm\bar{3}m$ and the *Pnma* variants. The tolerance factor t is also seen to move further away from 1 with increasing θ_{oct} and correspondingly the energy difference between the variants.

Thus, cubic thermodynamic stability is a more conservative criterion for perovskite synthesizability than formability since the E_c = 50 meV/atom that we utilize only accounts for oxide perovskites that stabilize in or very close to cubic symmetry and does not allow for the stability of compounds in lower symmetry phases. The ranges of thermodynamic stability may be tuned to accept or reject lower symmetry variants by varying the energy above hull threshold criterion E_c. However, increasing E_c also increases the risk of identifying as stable, compositions that are inherently unstable, and thus it is necessary to be careful while choosing a suitable value for E_c.

In contrast, since it is more efficient to computationally predict or calculate the thermodynamic stability of a

https://dx.doi.org/10.1021/acs.chemmater.0c03402
Chem. Mater. 2021, 33, 845−858

compound than it is to experimentally classify a compound as a perovskite or non-perovskite, thermodynamic stability is a more accessible metric for synthesizability. For the purposes of this work, our interest lies in identifying compositions that are likely to stabilize as cubic perovskites, and thus, using an energy above hull criterion of E_c = 50 meV will ensure that one is able to screen for chemistries with a high probability of synthesizability in cubic or close-to-cubic symmetry.

4.4. Prediction of New Formable and/or Stable Perovskites. Subsequent to the training and testing phases, the formability and stability classification models were applied to the comprehensive perovskite data set D_C to predict new formable and stable oxide perovskites, respectively. This prediction data set D_C consists of 946,292 distinct compounds. Of these, 891,188 compounds were predicted to be formable, while 437,828 compounds were predicted to be stable. Greater that 99% of the predicted stable compounds are a subset of the predicted formable perovskites, further highlighting the conclusion that thermodynamic stability is a less liberal metric than formability. From the predictions, the candidates that are predicted to be both formable and stable were selected, and the distribution of the RFC predicted probabilities for those candidates is shown in Figure 8.

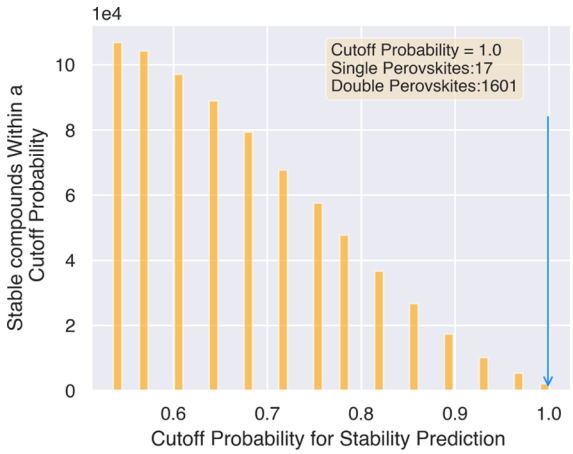


Figure 8. Distribution plot of predicted probabilities of candidate compounds that are likely to be both formable and stable.

We select as our final candidates all those compositions that satisfy a cutoff probability of 1.0; that is, all those compounds that are predicted to be stable with a probability of 100%. We thus shortlist 1618 oxide perovskite candidates of which 17 are single perovskites while 1601 are double perovskites. The periodic table in Figure 9 indicates the number of occurences of each element in the A-site and B-site positions of these 1618 compositions. The candidates are dominated by barium, strontium, sodium, and the lanthanide elements on the A-site. Aluminium, silicon, germanium, and arsenic are not predicted to be stable in the A-site at all. These elements appear in the prediction data set D_C on the A-site in very insignificant numbers as well, indicating that their charge states are not chemically compatible with other elements when they are located on the A-site. The case for cadmium, cesium, thallium, and gold on the B-site is similar as well, with very few compounds in D_C having these elements on the B-site. On the other hand, vanadium, chromium, manganese, and iron are predicted to be stable in the B-site in a very large number of compounds. A comprehensive list of these 1618 compounds

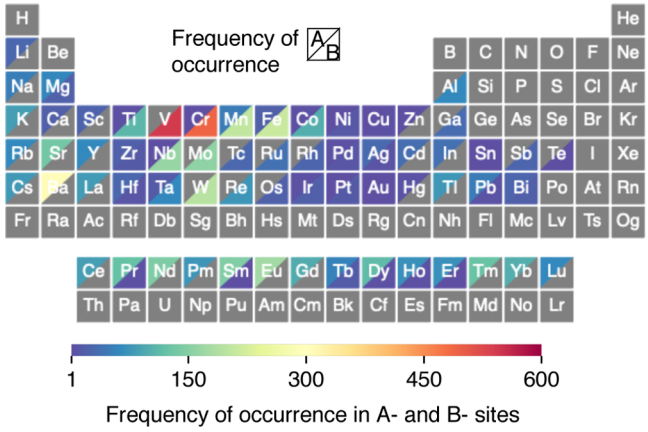


Figure 9. Periodic table showing the number of occurrences of each element in the A-site and B-site positions in the D_C data set that are predicted to be stable (with cutoff probability = 1.0) and formable. Gray elements indicate no occurrences.

may be found in the Supporting Information. Of these 1618 compositions, 414 novel compounds are predicted to also be formable with a cutoff probability of 1.0 as well, in addition to the perfect stability cutoff probability. Thus, these 414 compounds are presented as the most promising candidates for future experimental synthesis of novel oxide perovskites. These compounds are listed in Table 1. Also highlighted in the table in bold are the 21 perovskite oxides that present new A and B cation possibilities that have not been previously explored, that is, neither single nor double perovskites have been synthesized with these combinations of chemistries to the best of our knowledge.

Of the 414 listed compounds, 60 were randomly selected, and their ΔE_h values were calculated in an effort to computationally confirm our predictions. Of these 60 materials, 58 materials had ΔE_h < 25 meV/atom, while only 2 of the compounds were found to have ΔE_h > 50 meV/atom, thus computationally confirming our predictions. The ΔE_h values for these 60 compounds are also listed in Table 1.

Thus, a large number of hitherto unexplored and potentially stable and formable perovskite oxides with novel chemistries are predicted. As compared to the starting formability and stability databases of 1500 and 3500 compounds, respectively, we have identified 125 times as many double perovskite compounds that are stable and more than 500 times as many compounds that are formable. This dramatically increases the chemical space of perovskite compounds that can be considered for a wide range of applications.

5. CONCLUSIONS

To summarize, a systematic computational screening strategy is presented to identify novel single and double oxide perovskites by efficiently exploring a large fraction of the double perovskite chemical space. ML classification models based on random forests were built to predict formability and thermodynamic stability of single and double oxide perovskites. The ML models were trained to maximize accuracy while minimizing variance. To start with, a large number of elemental features were considered using domain knowledge, intuition, and prior work by researchers. These were finally whittled down to 28 features. The perovskite formability model was trained on a data set of experimentally known perovskites and nonperovskites culled from the literature (D_F).

https://dx.doi.org/10.1021/acs.chemmater.0c03402
Chem. Mater. 2021, 33, 845−858

Table 1. List of 414 Predicted Oxide Perovskite Compounds Satisfying Cutoff Probabilities = 1.0 from Both Stability and Formability Prediction Models[a]

predicted novel oxide perovskite candidates				
$PmVO_3$	**Ba_2BiHoO_6**	$GdErCr_2O_6$	$HoLuV_2O_6$	$EuLuCr_2O_6$
$PmNdTiMnO_6$	**$HoMnO_3$**	Ba_2OsDyO_6	$YTbV_2O_6$	$LaHoV_2O_6$
$ErTmFe_2O_6$	$PmTmTiMnO_6$	Ba_2TbTcO_6 (0)	Ba_2OsTmO_6	$DyTbV_2O_6$
$HoErMn_2O_6$	$ErLuV_2O_6$	$NdHoTiMnO_6$	Ba_2TbWO_6	Ba_2OsNdO_6 (2)
$YbTbV_2O_6$	$HoYbV_2O_6$	$LaErV_2O_6$	$PrNdTiMnO_6$	La_2VNiO_6 (0)
Ba_2BiErO_6	$LaDyV_2O_6$ (0)	$CeHoV_2O_6$	$ErTmV_2O_6$	$LaYbCrFeO_6$ (0)
Gd_2CrVO_6	Ba_2WPmO_6	$LaYbV_2O_6$ (0)	$YHoV_2O_6$	$ErYbV_2O_6$
$LaTbCrFeO_6$ (0)	Ba_2PrReO_6	Ba_2MoMgO_6	$PmGdCr_2O_6$	$HoDyV_2O_6$
$ErDyV_2O_6$	$LaDyCrFeO_6$	Eu_2CrCoO_6	Ba_2MoHoO_6	**$GdHoMn_2O_6$**
$HoTmV_2O_6$	$YErV_2O_6$	$PrSmTiMnO_6$	Sr_2TiNiO_6	Ba_2IrHoO_6
$GdHoV_2O_6$	$PmNdMn_2O_6$	$NdSmV_2O_6$	$PrTmTiMnO_6$	Ba_2RhEuO_6
Ba_2TaHoO_6	$NdGdCr_2O_6$	**$BaHfV_2O_6$**	$NdTbV_2O_6$	$SmDyMnMgO_6$
Ba_2RhPmO_6	**Ba_2SbHoO_6**	$PmGdV_2O_6$	$BaHfMn_2O_6$ (0)	$EuErCr_2O_6$
$LaYbMnMgO_6$ (0)	Ba_2RhGdO_6	Ba_2WHoO_6 (0)	$GdHoFe_2O_6$	$PmSmMn_2O_6$ (0)
$EuErV_2O_6$	$TbYbMnMgO_6$	Ba_2RhDyO_6	Ba_2MnHoO_6	$LaLuV_2O_6$
$PrSmFe_2O_6$	$NdYbV_2O_6$	$TbDyMnMgO_6$	Ba_2RhTbO_6 (0)	Ba_2MoPmO_6
$LaTbV_2O_6$ (0)	$PrTmV_2O_6$	$NdDyV_2O_6$	$EuDyMnMgO_6$	Nd_2TiMnO_6
Ba_2TaPmO_6	$SmGdV_2O_6$	$PrCeV_2O_6$	$NdYV_2O_6$	$NdSmMnMgO_6$
Sr_2IrEuO_6	Ba_2MoSmO_6	$GdLuMn_2O_6$	$PrYV_2O_6$	$NdLuV_2O_6$
$EuNdMnMgO_6$	Dy_2CrVO_6	Ba_2TaSmO_6	$GdTmMn_2O_6$	$PrDyV_2O_6$
$LaNdV_2O_6$	$SmCeCrVO_6$	Y_2CrVO_6	Ba_2TaNdO_6	$GdDyMn_2O_6$
$PrTbV_2O_6$	$NdYbCo_2O_6$	$LaNdTiFeO_6$ (0)	Tb_2CrVO_6	Ba_2CrYO_6 (0)
$PrGdV_2O_6$ (0)	$PrYbV_2O_6$	$NdSmFe_2O_6$	$TbDyTiMnO_6$	Ba_2ReScO_6
Ba_2CrDyO_6	$BaTbV_2O_6$	$PrSmMn_2O_6$	$NdYFe_2O_6$	$LaCaNiMnO_6$
Ba_2ReNdO_6	Ba_2CrTbO_6	$SmGdCr_2O_6$	$SmMgMn_2O_6$	$NdDyFe_2O_6$
$LaSrTiFeO_6$	Ba_2ReEuO_6	Ba_2MnDyO_6 (0)	$BaTbCr_2O_6$	$PmTbV_2O_6$
$NdErV_2O_6$	$PmTmCrVO_6$	Ba_2RePmO_6	Ba_2CrCaO_6 (0)	$GdTmCr_2O_6$
$PmLuV_2O_6$	$EuNdFe_2O_6$	$PmCeCrVO_6$	Ba_2RePrO_6	Ba_2MoTmO_6 (0)
$HoErFe_2O_6$	**$PmScV_2O_6$**	$SmErV_2O_6$	$PmScCrVO_6$	Ba_2ReGdO_6
Ba_2TaTmO_6 (0)	$HoErV_2O_6$	$PmCeV_2O_6$	$EuNdV_2O_6$	$BaCaTaAgO_6$ (21)
Ba_2RuGdO_6	Ba_2TcTmO_6	$GdTmFe_2O_6$	$PmTmV_2O_6$ (0)	$EuYFe_2O_6$
$LaNdFeTiO_6$	Ba_2RuHoO_6	**Ba_2SbTmO_6** (0)	$GdErMn_2O_6$	$LaPmV_2O_6$
$EuDyFe_2O_6$	$NdScCrVO_6$	Ba_2RuLaO_6	$PmSmV_2O_6$	**$EuHfV_2O_6$**
$PmDyFe_2O_6$	$NdTmV_2O_6$	$NdTmCrVO_6$	Ba_2RuSmO_6	$LaGdV_2O_6$
$SrHfV_2O_6$	$PmYFe_2O_6$	$NdCeV_2O_6$	$NdCeCrVO_6$	Ba_2RuErO_6
$PrPmMn_2O_6$	$GdDyFe_2O_6$	$PmYV_2O_6$ (0)	$NdScV_2O_6$	$PrSmCrVO_6$
Ba_2RuYO_6	$PrPmFe_2O_6$	$GdLuFe_2O_6$	$PmDyV_2O_6$ (0)	$EuSmV_2O_6$
$LaSrTiMnO_6$ (0)	Pm_2CrCoO_6	$PrPmV_2O_6$	$NdHfV_2O_6$	$PmYbV_2O_6$ (0)
$ZrDyCrVO_6$	$PrNdCrVO_6$ (0)	Sr_2IrTbO_6	$PmHoV_2O_6$	$NdHfCr_2O_6$
$PmYbCo_2O_6$	$YNdMnMgO_6$	$HoScCrVO_6$	Sr_2IrDyO_6	$PrHoV_2O_6$ (0)
$EuCaW_2O_6$	$BaSrTi_2O_6$	**$ScDyMnMgO_6$**	$PrHoCrVO_6$	Sr_2MoDyO_6
$PmNdFe_2O_6$	$LaYbMn_2O_6$ (0)	$DyHfMn_2O_6$	$DySmMnMgO_6$	$PrErCrVO_6$ (0)
Sr_2CrFeO_6	$PmNdV_2O_6$	$SmDyMn_2O_6$	$YbHfMn_2O_6$	$YEuMnMgO_6$
$PrMgCrVO_6$	La_2MnCrO_6	$SmNdMn_2O_6$	$CeSmMn_2O_6$	$EuHfMn_2O_6$
$DyEuMnMgO_6$	$BaFeMnTiO_6$	**Pb_2WFeO_6** (0)	$PrMgV_2O_6$	$LaYMn_2O_6$
$NdHfMn_2O_6$	$SmNdMnMgO_6$	**$LaSrMoMgO_6$** (0)	Ba_2MnEuO_6	$CeNdMn_2O_6$
$NdYbW_2O_6$	$BaSrMo_2O_6$	**$ScEuMnMgO_6$**	$PrTmCrVO_6$ (0)	Ba_2TaLuO_6
$GdLuV_2O_6$	$YLuV_2O_6$	$DyZrCr_2O_6$	$NdEuMnMgO_6$	$PrCeCrVO_6$
Ba_2TcYO_6 (0)	$PrEuMn_2O_6$	$DyLuV_2O_6$	$PrErV_2O_6$ (0)	$LaSrFeMnO_6$ (0)
$LaEuFeTiO_6$	Ba_2MoTbO_6	$GdTbV_2O_6$	$YbLuV_2O_6$	$PrNdFe_2O_6$ (0)
$LaSrMnTiO_6$ (0)	$LaCaTiCoO_6$	Ba_2MoYO_6	$LuTmNi_2O_6$	$TbLuV_2O_6$
$PrNdV_2O_6$	$NdErTiMnO_6$	$PmNdCrVO_6$	Ba_2MoYbO_6 (0)	$GdTmV_2O_6$
$CeTbV_2O_6$	$PrEuV_2O_6$ (0)	$PrKFeTiO_6$	$CeNdCrVO_6$	Ba_2IrErO_6 (0)
$GdDyV_2O_6$	$CeDyV_2O_6$	$PrEuFe_2O_6$	$PrKMnTiO_6$	$HoNdCrVO_6$
Ba_2NbInO_6 (2)	$GdYbV_2O_6$	$TbTmMn_2O_6$	$EuDyMn_2O_6$	$PmHoTiMnO_6$
$LaSrTiCuO_6$ (103)	Ba_2NbTbO_6	$GdYV_2O_6$	$DyTmMn_2O_6$	$EuYMn_2O_6$
$NdTmTiMnO_6$	$HoSmCrVO_6$	Ba_2NbTlO_6 (2)	$GdTbCr_2O_6$	$TmZrMn_2O_6$
$NdEuMn_2O_6$	**$PrKTiFeO_6$**	$PmHoCrVO_6$	Ba_2TaErO_6	$SmDyFe_2O_6$
$LaTmV_2O_6$ (0)	$SmEuMn_2O_6$	**$PrKTiMnO_6$**	$LaEuTiFeO_6$	Ba_2MoErO_6
$GdDyCr_2O_6$	$TbTmV_2O_6$	$EuZrMn_2O_6$	$HfNdCrFeO_6$	$PmEuCrVO_6$

https://dx.doi.org/10.1021/acs.chemmater.0c03402
Chem. Mater. 2021, 33, 845−858

Table 1. continued

predicted novel oxide perovskite candidates				
Ba_2WDyO_6	$GdYCr_2O_6$	$DyTmV_2O_6$	$NdDyMn_2O_6$	**$HfKWMnO_6$**
$PmSmCrVO_6$	Ba_2WYbO_6	$GdLuCr_2O_6$	$LuTmV_2O_6$ (0)	$NdYMn_2O_6$
$NdEuTiFeO_6$	$PmErCrVO_6$	Ba_2WTbO_6	$PrNdMn_2O_6$	$YTmV_2O_6$
$TbErMn_2O_6$	$PrPmCrFeO_6$	$HoEuCrVO_6$	Ba_2WTmO_6	$PmErV_2O_6$
$TmZrV_2O_6$	$DyErMn_2O_6$	$PrPmTiMnO_6$	$LaSrTiCoO_6$	Ba_2WErO_6
$PrPmCo_2O_6$	$YbTmV_2O_6$	$YYbV_2O_6$	**$LaKZrReO_6$** (188)	$CeSmCrVO_6$
Ba_2MoEuO_6	$DyTmFe_2O_6$	$SmYV_2O_6$	$DyYbV_2O_6$	$SmNdCrVO_6$
$HfEuCrVO_6$	Ba_2MnSmO_6 (0)	$LaGdCr_2O_6$	$SmZrCr_2O_6$	$YDyV_2O_6$
$YbYCrVO_6$	$LaCaMnNiO_6$	Ba_2IrYbO_6 (0)	$TbTmFe_2O_6$	$SmZrV_2O_6$
$DyZrV_2O_6$	$DyYCrVO_6$	$PrEuCrVO_6$	Ba_2OsEuO_6 (0)	$PmEuFe_2O_6$
$SmTmV_2O_6$	$EuYV_2O_6$	$TbYCrVO_6$	$NdSmCrVO_6$	Ba_2OsYbO_6
$PmEuV_2O_6$	$SmDyV_2O_6$	$EuTbV_2O_6$	$TbDyCrVO_6$	$ErNdCrVO_6$
Ba_2OsErO_6	$PrSmV_2O_6$	$SmTbV_2O_6$	$EuDyV_2O_6$	$YbDyCrVO_6$
$EuSmCrVO_6$	Ba_2OsHoO_6	$NdHoV_2O_6$	$SmCeV_2O_6$	$EuYbV_2O_6$
$LaLuCrVO_6$	$EuNdCrVO_6$	Pr_2TiMnO_6	$NdGdV_2O_6$	$SmYbV_2O_6$
$EuZrV_2O_6$	$LaTbCrVO_6$ (0)	$PrPmCrVO_6$	Ba_2TaPrO_6 (15)	$SmHoV_2O_6$
$SmLuV_2O_6$	$LaEuV_2O_6$	$LaYCrVO_6$	$PrPmFeCrO_6$ (0)	Ba_2IrGdO_6
$GdErV_2O_6$	$PmEuMn_2O_6$	$EuTmV_2O_6$	$LaDyCrVO_6$ (0)	$GdNdCrVO_6$
Ba_2CrGdO_6 (0)	$EuGdCr_2O_6$	$PmDyMn_2O_6$ (0)	$EuLuV_2O_6$	$LaYbCrVO_6$ (0)
$GdPmCrVO_6$	Ba_2MnGdO_6	$EuHoV_2O_6$	$PmYMn_2O_6$	$EuZrCr_2O_6$
$SmEuCrVO_6$	$LuKMnTiO_6$	Ba_2MoPrO_6	$GdErFe_2O_6$	$PmSmFe_2O_6$ (0)
$NdScCr_2O_6$	$NdEuFeTiO_6$	$GdPrCrVO_6$	Ba_2MnErO_6 (0)	$EuGdV_2O_6$
$HoTbV_2O_6$	$ErTbV_2O_6$	$PmErTiMnO_6$	$GdEuCrVO_6$	

[a]Highlighted in bold are the 21 perovskite oxides that present new A and B cation possibilities that have not been previously explored. ΔE_h values in meV/atom for 60 randomly selected candidates are also indicated in parentheses.

It does not directly encode any symmetry information and is fundamentally developed using geometric criteria based on the ionic radii of the constituent elements. However, a number of the compounds in this database have been experimentally synthesized in orthorhombic, rhombohedral, or tetragonal symmetries, and thus, the data set and therefore the model indirectly account for symmetry variants. During the literature search, it was also noticed that the number of known perovskites far exceeds the number of known nonperovskites with ABO_3 stoichiometry. In contrast, the thermodynamic stability data set (D_S) is solely built using DFT calculations and by assuming cubic symmetry. A threshold criterion of E_c = 50 meV is assumed while assigning stable and unstable labels to the compounds in this data set to account for both lower symmetry variants arising from crystal distortions and potential metastable compounds. The Materials Project database was used to determine the convex hull for all the compounds in the stability database. It is worth repeating that the utility of using convex hull estimations to predict stability is limited by the data available in the Materials Project database.

The thermodynamic cubic stability was found to be a more conservative metric than perovskite formability, perhaps owing to the E_c = 50 meV threshold. However, it is shown that, by varying this threshold value, the range of thermodynamic stability can be tuned to substantially overlap with the formability predictions and include lower symmetry phases as well. Increasing E_c inherently increases the risk of identifying as stable, compositions that are inherently unstable, and thus, a suitable value for E_c must be chosen carefully.

A very large and comprehensive data set of candidate single and double perovskites was assembled by considering 68 elements from the periodic table. For the double perovskites, only equimolar ordered compositions were considered with cubic rock-salt structure. Accounting for chemical compatibility, this resulted in 946,292 prospective distinct candidates. Of these, 891,188 were predicted to be formable, while 437,828 were predicted to be stable. Further, 414 compositions are presented as the most promising candidates for future experimental synthesis of novel oxide perovskites with very high confidence. Of these, 21 double oxide chemistries are presented with hitherto unexplored A and B cation chemistries. As indicated in Table 1 and Figure 9, a number of new previously overlooked chemistries are predicted to form stable perovskites.

This study highlights the potential for using machine-learning-guided high-throughput calculations to identify novel chemistries that may exhibit enhanced functionality. Thus, this study can serve as the first step in a more comprehensive research plan in which the functionality of promising compounds is then examined. However, the novelty of the new compounds alone demonstrates the utility of this approach and opens new avenues for materials design for advanced applications.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemmater.0c03402.

Three data sets: formability training data set for perovskite formability which comprises 1505 experimentally known oxides of the form ABO_3 and $AA'BB'O_6$ collated from the literature, DFT-computed stability data set of 3469 single and double oxide perovskites, and results of data set of 1618 single and double oxide perovskites which are predicted to be formable and thermodynamically stable along with values of features used in ML models (ZIP)

https://dx.doi.org/10.1021/acs.chemmater.0c03402
Chem. Mater. 2021, 33, 845−858

AUTHOR INFORMATION

Corresponding Author

Anjana Talapatra − *Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, United States;* orcid.org/0000-0002-6446-2437; Email: atalapatra@lanl.gov

Authors

Blas P. Uberuaga − *Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, United States;* orcid.org/0000-0001-6934-6219

Christopher R. Stanek − *Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, United States*

Ghanshyam Pilania − *Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, United States;* orcid.org/0000-0003-4460-1572

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemmater.0c03402

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Research presented in this paper was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory under project numbers 20190656PRD4 and 20190043DR. Computational support for this work was provided by LANL's high-performance computing clusters. This work was supported by the U.S. Department of Energy through the Los Alamos National Laboratory. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (contract no. 89233218CNA000001).

REFERENCES

(1) Tilley, R. J. The ABX3 Perovskite Structure. *Perovskites*; John Wiley & Sons, Ltd.: Chichester, U.K., 2016; Vol. *1*−*41*.

(2) Towler, M. D.; Dovesi, R.; Saunders, V. R. Magnetic interactions and the cooperative Jahn-Teller effect in $KCuF_3$. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1995**, *52*, 10150.

(3) Uchino, K. Glory of piezoelectric perovskites. *Sci. Technol. Adv. Mater.* **2015**, *16*, 046001.

(4) DiDomenico, M., Jr.; Wemple, S. H. Optical properties of perovskite oxides in their paraelectric and ferroelectric phases. *Phys. Rev.* **1968**, *166*, 565.

(5) Galasso, F. Perovskite Type Compounds and High Tc Superconductors. *JOM* **1987**, *39*, 8−10.

(6) Visser, D. W.; Ramirez, A. P.; Subramanian, M. A. Thermal conductivity of manganite perovskites: Colossal magnetoresistance as a lattice-dynamics transition. *Phys. Rev. Lett.* **1997**, *78*, 3947.

(7) Dulian, P. *Perovskite Materials: Synthesis, Characterisation, Properties, and Applications*; Pan, L., Guang, Z., Eds.; BoD−Books on Demand, 2016; p 1.

(8) Bhalla, A. S.; Guo, R.; Roy, R. The perovskite structure—a review of its role in ceramic science and technology. *Mater. Res. Innovations* **2000**, *4*, 3−26.

(9) Ji, Q.; Bi, L.; Zhang, J.; Cao, H.; Zhao, X. S. The role of oxygen vacancies of ABO_3 perovskite oxides in the oxygen reduction reaction. *Energy Environ. Sci.* **2020**, *13*, 1408−1428.

(10) Goldschmidt, V. M. Die gesetze der krystallochemie. *Naturwissenschaften* **1926**, *14*, 477−485.

(11) Shannon, R. D.; Prewitt, C. T. Revised values of effective ionic radii. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1970**, *26*, 1046−1048.

(12) Sawaguchi, E.; Akishige, Y.; Yamamoto, T.; Nakahara, J. Phase transition in hexagonal type $BaTiO_3$. *Ferroelectrics* **1989**, *95*, 29−36.

(13) Bartel, C. J.; Sutton, C.; Goldsmith, B. R.; Ouyang, R.; Musgrave, C. B.; Ghiringhelli, L. M.; Scheffler, M. New tolerance factor to predict the stability of perovskite oxides and halides. *Sci. Adv.* **2019**, *5*, No. eaav0693.

(14) Filip, M. R.; Giustino, F. The geometric blueprint of perovskites. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 5397−5402.

(15) Woodward, P. M. Octahedral tilting in perovskites. I. Geometrical considerations. *Acta Crystallogr., Sect. B: Struct. Sci.* **1997**, *53*, 32−43.

(16) Lee, J.-H.; Bristowe, N. C.; Lee, J. H.; Lee, S.-H.; Bristowe, P. D.; Cheetham, A. K.; Jang, H. M. Resolving the physical origin of octahedral tilting in halide perovskites. *Chem. Mater.* **2016**, *28*, 4259−4266.

(17) Johnsson, M.; Lemmens, P. *Handbook of Magnetism and Advanced Magnetic Materials*; American Cancer Society, 2007; Chapter 4, pp 1−9.

(18) Tilley, R. J. *Perovskites: Structure-Property Relationships*; John Wiley & Sons, 2016; pp 1−41.

(19) Fop, S.; McCombie, K. S.; Wildman, E. J.; Skakle, J. M. S.; Mclaughlin, A. C. Hexagonal perovskite derivatives: a new direction in the design of oxide ion conducting materials. *Chem. Commun.* **2019**, *55*, 2127−2137.

(20) Muller, O.; Roy, R. The major ternary structural families. *Cryst. Chem. Non-Metal. Mater.* **1974**, *37*, 496.

(21) Li, C.; Lu, X.; Ding, W.; Feng, L.; Gao, Y.; Guo, Z. Formability of ABX_3 (X= F, Cl, Br, I) Halide Perovskites. *Acta Crystallogr., Sect. B: Struct. Sci.* **2008**, *64*, 702−707.

(22) Lufaso, M. W.; Woodward, P. M. Prediction of the crystal structures of perovskites using the software program SPuDS. *Acta Crystallogr., Sect. B: Struct. Sci.* **2001**, *57*, 725−738.

(23) Li, C.; Soh, K. C. K.; Wu, P. Formability of ABO_3 perovskites. *J. Alloys Compd.* **2004**, *372*, 40−48.

(24) Zhang, H.; Li, N.; Li, K.; Xue, D. Structural stability and formability of ABO_3-type perovskite compounds. *Acta Crystallogr., Sect. B: Struct. Sci.* **2007**, *63*, 812−818.

(25) Zhang, Y. M.; Ubic, R.; Xue, D. F.; Yang, S. Predicting the structural stability and formability of ABO_3-type perovskite compounds using artificial neural networks. *Materials Focus* **2012**, *1*, 57−64.

(26) Pilania, G.; Balachandran, P.; Gubernatis, J. E.; Lookman, T. Classification of ABO_3 perovskite solids: a machine learning study. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2015**, *71*, 507−513.

(27) Xu, Q.; Li, Z.; Liu, M.; Yin, W.-J. Rationalizing perovskite data for machine learning and materials design. *The journal of physical chemistry letters* **2018**, *9*, 6948−6954.

(28) Jain, A.; et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013**, *1*, 011002.

(29) Emery, A. A.; Wolverton, C. High-Throughput DFT calculations of formation energy, stability and oxygen vacancy formation energy of ABO_3 perovskites. *Sci. Data* **2017**, *4*, 170153.

(30) Schmidt, J.; Shi, J.; Borlido, P.; Chen, L.; Botti, S.; Marques, M. A. L. Predicting the thermodynamic stability of solids combining density functional theory and machine learning. *Chem. Mater.* **2017**, *29*, 5090−5103.

(31) Sun, Q.; Yin, W.-J. Thermodynamic stability trend of cubic perovskites. *J. Am. Chem. Soc.* **2017**, *139*, 14905−14908.

(32) Zeng, Z.; Calle-Vallejo, F.; Mogensen, M. B.; Rossmeisl, J. Generalized trends in the formation energies of perovskite oxides. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7526−7533.

(33) Artini, C.; Pani, M.; Lausi, A.; Costa, G. A. Stability of interlanthanide perovskites ABO3 (A ≡ La−Pr; B ≡ Y, Ho−Lu). *J. Phys. Chem. Solids* **2016**, *91*, 93−100.

(34) Li, W.; Jacobs, R.; Morgan, D. Predicting the thermodynamic stability of perovskite oxides using machine learning models. *Comput. Mater. Sci.* **2018**, *150*, 454−463.
(35) Liu, H.; Cheng, J.; Dong, H.; Feng, J.; Pang, B.; Tian, Z.; Ma, S.; Xia, F.; Zhang, C.; Dong, L. Screening stable and metastable ABO_3 perovskites using machine learning and the materials project. *Comput. Mater. Sci.* **2020**, *177*, 109614.
(36) Pilania, G.; Mannodi-Kanakkithodi, A. First-principles identification of novel double perovskites for water-splitting applications. *J. Mater. Sci.* **2017**, *52*, 8518−8525.
(37) Balachandran, P. V.; Emery, A. A.; Gubernatis, J. E.; Lookman, T.; Wolverton, C.; Zunger, A. Predictions of new ABO_3 perovskite compounds by combining machine learning and density functional theory. *Phys. Rev. Mater.* **2018**, *2*, 043802.
(38) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *Jom* **2013**, *65*, 1501−1509.
(39) Vasala, S.; Karppinen, M. $A_2B'B''O_6$ perovskites: a review. *Prog. Solid State Chem.* **2015**, *43*, 1−36.
(40) Zheng, K.; Świerczek, K.; Zając, W.; Klimkowicz, A. Rocksalt ordered-type double perovskite anode materials for solid oxide fuel cells. *Solid State Ionics* **2014**, *257*, 9−16.
(41) Samanta, K.; Saha-Dasgupta, T. Rocksalt versus layered ordering in double perovskites: A case study with La_2CuSnO_6 and La_2CuIrO_6. *Phys. Rev. B* **2017**, *95*, 235102.
(42) Wu, Y.; Lazic, P.; Hautier, G.; Persson, K.; Ceder, G. First principles high throughput screening of oxynitrides for water-splitting photocatalysts. *Energy Environ. Sci.* **2013**, *6*, 157−168.
(43) Pilania, G.; Mannodi-Kanakkithodi, A.; Uberuaga, B.; Ramprasad, R.; Gubernatis, J.; Lookman, T. Machine learning bandgaps of double perovskites. *Sci. Rep.* **2016**, *6*, 19375.
(44) Pilania, G.; Balachandran, P. V.; Gubernatis, J. E.; Lookman, T. Data-Based Methods for Materials Design and Discovery: Basic Ideas and General Methods. *Synth. Lect. Mater. Optic.* **2020**, *1*, 1−188.
(45) Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389−422.
(46) Pedregosa, F.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(47) Beurich, H.; Madach, T.; Richter, F.; Vahrenkamp, H. Experiments on the HOMO-LUMO Nature of Metal-Metal Bonds. *Angew Chem. Int. Ed. Engl.* **1979**, *18*, 690−691.
(48) Zunger, A. A pseudopotential viewpoint of the electronic and structural properties of crystals. *Struct. Bonding Cryst.* **1981**, *1*, 73−135.
(49) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big data of materials science: critical role of the descriptor. *Phys. Rev. Lett.* **2015**, *114*, 105503.
(50) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev. [Sect.] A* **1965**, *140*, A1133.
(51) 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.* **1996**, *54*, 11169.
(52) 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**, *6*, 15−50.
(53) Perdew, J. P.; Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1992**, *45*, 13244.
(54) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(55) Methfessel, M.; Paxton, A. T. High-precision sampling for Brillouin-zone integration in metals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1989**, *40*, 3616.
(56) Blöchl, P. E.; Jepsen, O.; Andersen, O. K. Improved tetrahedron method for Brillouin-zone integrations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *49*, 16223.
(57) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **2003**, *118*, 8207−8215.
(58) Armiento, R.; Kozinsky, B.; Fornari, M.; Ceder, G. Screening for high-performance piezoelectrics using high-throughput density functional theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2011**, *84*, 014103.
(59) He, L.; Liu, F.; Hautier, G.; Oliveira, M. J.; Marques, M. A.; Vila, F. D.; Rehr, J.; Rignanese, G.-M.; Zhou, A. Accuracy of generalized gradient approximation functionals for density-functional perturbation theory calculations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 064305.
(60) Zhang, G.-X.; Reilly, A. M.; Tkatchenko, A.; Scheffler, M. Performance of various density-functional approximations for cohesive properties of 64 bulk solids. *New J. Phys.* **2018**, *20*, 063020.
(61) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.

https://dx.doi.org/10.1021/acs.chemmater.0c03402
Chem. Mater. 2021, 33, 845−858