SCIENTIFIC DATA

OPEN

Data Descriptor: High-throughput DFT calculations of formation energy, stability and oxygen vacancy formation energy of ABO_3 perovskites

数据描述符：ABO3钙钛矿的地层能量，稳定性和氧空位形成能的高通量DFT计算

Received: 26 May 2017
Accepted: 22 August 2017
Published: 17 October 2017

Antoine A. Emery[1] & Chris Wolverton[1]

ABO_3 perovskites are oxide materials that are used for a variety of applications such as solid oxide fuel cells, piezo-, ferro-electricity and water splitting. Due to their remarkable stability with respect to cation substitution, new compounds for such applications potentially await discovery. In this work, we present an exhaustive dataset of formation energies of 5,329 cubic and distorted perovskites that were calculated using first-principles density functional theory. In addition to formation energies, several additional properties such as oxidation states, band gap, oxygen vacancy formation energy, and thermodynamic stability with respect to all phases in the Open Quantum Materials Database are also made publicly available. This large dataset for this ubiquitous crystal structure type contains 395 perovskites that are predicted to be thermodynamically stable, of which many have not yet been experimentally reported, and therefore represent theoretical predictions. The dataset thus opens avenues for future use, including materials discovery in many research-active areas.

ABO3钙钛矿是氧化物材料，用于各种应用，如固体氧化物燃料电池，压电，铁电和水分解。由于它们在阳离子取代方面的显着稳定性，用于此类应用的新化合物可能等待发现。在这项工作中，我们提出了使用第一原理密度泛函理论计算的5,329立方和扭曲钙钛矿的地层能量的详尽数据集。除了形成能之外，还公开了几种额外的性质，例如氧化态，带隙，氧空位形成能和关于开放量子材料数据库中所有相的热力学稳定性。这种普遍存在的晶体结构类型的大型数据集包含395个钙钛矿，预计其具有热力学稳定性，其中许多尚未经过实验报道，因此代表了理论预测。因此，该数据集为未来的使用开辟了道路，包括许多研究活跃领域的材料发现。

Design Type(s)	database creation objective
Measurement Type(s)	physicochemical characterization
Technology Type(s)	computational modeling technique
Factor Type(s)	compound by chemical composition
Sample Characteristic(s)	

[1]Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208, USA. Correspondence and requests for materials should be addressed to C.W. (email: c-wolverton@northwestern.edu).

背景与摘要
由于ABO3钙钛矿对氧空位具有较大的耐受性，因此广泛用于各种应用，如固体氧化物燃料电池，压电，铁电和热化学水分解1-3。此外，它们对其组成元素的显着结构稳定性表明潜在的新化合物仍有待发现。由于可能的ABO3化合物的数量很大，我们使用高通量密度泛函理论（HT-DFT）以穷举的方式计算5,329种组合物的热力学稳定性。除了化合物的稳定性外，我们还计算了氧空位形成能，因为它是涉及还原这些化合物1-3的许多应用的相关量。由于它们的多功能性，钙钛矿已经在几个高吞吐量的工作中进行了研究4-7。但是，据我们所知，这些数据集并非公开和/或容易获得。
通过在ABO3钙钛矿晶体结构的A和B位置（732 = 5,329）上取代元素周期表中的73种金属和半金属（参见图1），产生所有5,329种化合物。理想的ABO3立方钙钛矿晶体结构由与氧原子八面体六重配位的B阳离子和由氧原子配位12倍的A阳离子组成。除了理想的立方结构外，许多钙钛矿从这种立方结构中经历局部变形，这些扭曲的钙钛矿可以具有多种对称性，包括菱形，四方和正交变形3（见图2）。在这项工作中，所有5,329种成分都是在理想的立方结构中计算出来的，其中一部分（见“方法”部分中的结构扭曲段落）是在前面提到的3种扭曲中计算出来的。
对于使用凸壳结构的A-B-O三元相图中存在的所有可能的相线性组合，评估所有ABO3化合物的T = 0K，P = 0巴基态稳定性。用于稳定性计算的所有阶段均来自开放量子材料数据库（OQMD）8和（截至2017年7月）包括来自无机晶体结构数据库（ICSD）9,10和约430,000个假设化合物的约40,000个阶段关于共同的装饰

Background & Summary

Due to their large tolerance to oxygen vacancy, ABO_3 perovskites are widely used for a variety of applications such as solid oxide fuel cells, piezo-, ferro-electricity and thermochemical water splitting[1–3]. Furthermore, their remarkable structural stability with respect to their constituent elements suggests that potential new compounds remain to be discovered. As the number of possible ABO_3 compounds is large, we use high-throughput density functional theory (HT-DFT) to compute the thermodynamical stability of 5,329 compositions in an exhaustive manner. In addition to the compounds stability, we also calculate the oxygen vacancy formation energies as it is a relevant quantity for many applications involving reduction of these compounds[1–3]. Due to their versatility, perovskites have been studied in several high-throughput works[4–7]. However, to our knowledge, those datasets are not publicly and/or easily available.

All the 5,329 compounds are created by substituting 73 metals and semi-metals of the periodic table of the elements (see Fig. 1) on the A and B sites ($73^2 = 5{,}329$) of the ABO_3 perovskite crystal structure. The ideal ABO_3 cubic perovskite crystal structure is composed of a B cation that is octahedrally 6-fold coordinated with oxygen atoms and an A cation that is 12-fold coordinated by oxygen atoms. Aside from the ideal cubic structure, many perovskites undergo a local distortion from this cubic structure, and these distorted perovskites can have a variety of symmetries, including the rhombohedral, tetragonal and orthorhombic distortions[3] (see Fig. 2). In this work, all 5,329 compositions are calculated in the ideal cubic structure and a subset of those (see structural distortions paragraph in the Methods section) are calculated in the 3 aforementioned distortions.

The T = 0 K, P = 0 bar ground state stability of all ABO_3 compounds was assessed with respect to all possible linear combinations of phases present in the A-B-O ternary phase diagram using a convex hull construction. All the phases that are used for the stability calculation are from the Open Quantum Materials Database (OQMD)[8] and (as of July 2017) include ~40,000 phases from the Inorganic Crystal Structure Database (ICSD)[9,10] and ~430,000 hypothetical compounds based on decoration of common

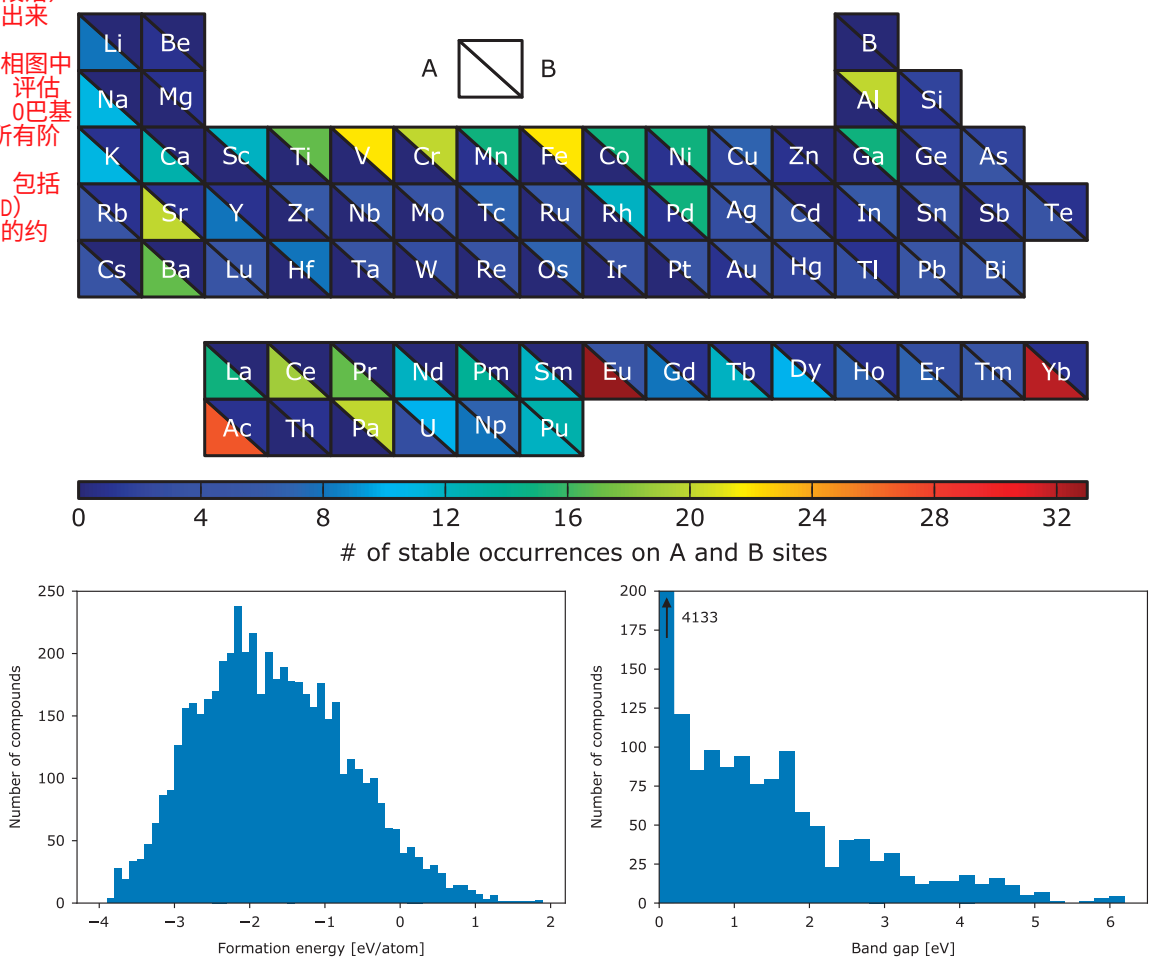

Figure 1. List of elements considered for the A and B sites. Elements are color-coded as a function of the number of stable perovskites with the respective elements on the A and B sites. (Bottom) Histogram representation of formation energies and band gap of compounds calculated in this work.

结构原型。通过使用A2B2O5结构计算氧空位形成能，所述A2B2O5结构对应于去除氧原子的2个钙钛矿单元电池。此外，还报告了DFT计算中容易获得的其他属性，包括松弛结构，带隙和总磁矩。图2显示了用于获取所有数量的工作流程。
本数据集用于一项研究，旨在利用稳定性和氧空位形成能作为筛选参数，确定用于热化学水分解应用的合适钙钛矿11。该数据更一般地用于指导预测的新化合物的实验合成，进一步筛选多种应用（除了水分解）或训练机器学习（ML）模型。虽然材料数据集上的机器学习是一个活跃的研究领域12-14，但是各个研究小组使用的数据集通常彼此差异很大，无法比较各种ML模型。拥有可供各种研究小组用于培训机器学习模型的大型，一致的材料数据集将允许对该领域中使用的各种方法进行更透明的比较。
方法
密度泛函理论计算
使用Vienna Ab initio模拟软件包（VASP）15,16通过密度泛函理论计算所有化合物。自始至终使用投影仪增强波方法（PAW）17和GGA-PBE18作为交换相关功能的近似。 3d过渡金属和

structural prototypes. The oxygen vacancy formation energy was calculated by using an $A_2B_2O_5$ structure, which corresponds to 2 perovskite unit cells with an oxygen atom removed. Additionally, other properties readily available from DFT calculations are reported, including the relaxed structure, band gap, and total magnetic moment. Figure 2 shows the workflow used to obtain all the quantities.

The present dataset was used in a study aiming at identifying suitable perovskites for thermochemical water-splitting applications using both the stability and oxygen vacancy formation energy as screening parameters[11]. This data is valuable more generally in guiding experimental synthesis of predicted new compounds, further screening for a large variety of applications (other than water splitting) or to train machine learning (ML) models. While machine learning on materials dataset is an area of active research[12–14], the datasets used by various research groups are often vastly different from one another with no way to compare various ML models. Having a large, consistent materials dataset that can be used by a variety of research groups to train machine learning models will allow a more transparent comparison of various methods being used in the field.

Methods

Density functional theory calculations

All compounds were calculated by density functional theory using the Vienna Ab initio simulation package (VASP)[15,16]. The projector-augmented wave method (PAW)[17] and GGA-PBE[18], as an approximation to the exchange-correlation functional, were used throughout. 3d transition metals and

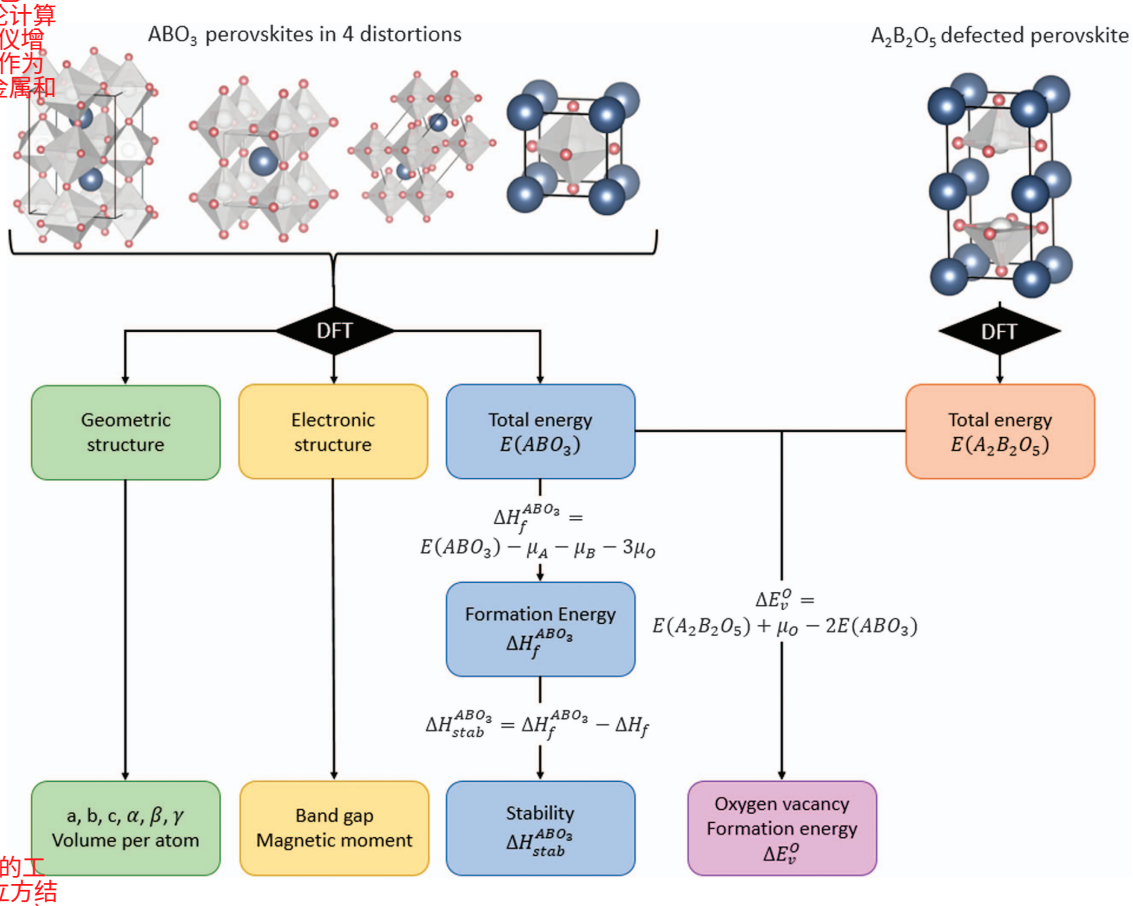

图2.计算当前数据集中所有属性的工作流程。 （左上）我们从所有立方结构开始，并使用密度泛函理论（DFT）计算它们的总能量。如果立方钙钛矿（在等式2中定义）的稳定性小于每个原子0.5eV（即，立方相在基态凸壳的每个原子的0.5eV内），我们还计算3个额外的扭曲（正交，四角形），菱形）。从计算中可以容易地获得几何特性（晶格参数，角度和每个原子的体积）和电子特性（带隙和磁矩）。使用元素化学势计算形成能，并且相对于OQMD中存在的所有其他A-B-O相计算热力学稳定性。 （右上）使用DFT计算具有缺失氧原子的2 × 1 × 1超级单元的受损钙钛矿，并且将它们的总能量与原始立方细胞的能量一起用于计算氧空位形成能。

Figure 2. Workflow to calculate all the properties in the current dataset. (top left) We start with all the cubic structures and compute all their total energies using density functional theory (DFT). If the stability of the cubic perovskite (defined in Equation 2) is less than 0.5 eV per atom (i.e., the cubic phase is within 0.5 eV per atom of the ground state convex hull), we also compute 3 additional distortions (orthorhombic, tetragonal, rhombohedral). The geometric properties (lattice parameters, angles, and volume per atom) and electronic properties (band gap and magnetic moment) are readily available from the calculations. Formation energies are calculated using elemental chemical potentials and thermodynamic stability is calculated with respect to all the other A-B-O phases present in the OQMD. (top right) Defected perovskites, a 2 × 1 × 1 supercell with a missing oxygen atom, are calculated using DFT and their total energies, in conjunction with those of pristine cubic cells, are used to compute the oxygen vacancy formation energies.

大多数act系元素使用DFT + U19,20形式计算（U值表见表1，U表示过渡金属的d电子和act系元素的f电子），并用自旋的铁磁排列进行自旋极化。计算是在开放量子材料数据库8的框架内进行的，该数据库包含来自实验数据库的470,000个不同阶段，如无机晶体结构数据库（ICSD）9,10和基于共同的假设化合物
结构原型。
地层能量的计算
根据等式1计算化合物HABO f 3的形成能：
HABO3
f¼EABOðÞ3 - μA - μB - 3μOð1Þ
其中E（ABO3）是钙钛矿的总能量，μA，μB和μO分别是A，B和氧的化学势。对于大多数元素，化学势等于其基态的DFT总能量。对于某些元素，其中T = 0 K DFT基态不是足够的参考状态，我们应用化学势的校正：双原子气体（O），室温液体（Hg），结构相变在0和298 K之间的几个元素（Na，Ti，Sn）和具有DFT + U校正的元素（见表格1）。对于这些元素，化学势适合实验形成能量。所用实验形成能量的细节和所得到的校正在参考文献中给出。 8。

相位稳定性的计算
使用能量凸壳结构21,22评估热力学稳定性。根据定义，凸包由具有比各个组成的任何其他相或线性相组合低的能量的相组成。稳定性或凸包船体距离定义为HABO3
stab¼HABOf 3 - Hfð2Þ
HABO3
f是钙钛矿的形成能，在等式1中定义，ΔHf是在ABO3组成的凸壳能量。对于给定相位P，如果P不稳定，则凸包距离为正
并且表示P与该组合物处的凸包之间的能量差。如果P是稳定的，则凸包的距离是负的，并且表示P和没有化合物P计算的（假设的）凸包之间的能量差。但是，在这项工作中，我们希望略微放松这个限制以解释几乎稳定的化合物和与DFT相关的可能的不确定性/错误。
因此，在本文中，我们将稳定性低于0.025 eV /原子（室温下约为kT）的化合物标记为稳定。
结构扭曲
在计算了所有立方钙钛矿后，我们研究了扭曲对钙钛矿稳定性的影响。为此，我们随机选择了三分之一的成分（1,776）并计算了它们在菱形，四角形和斜方形畸变中的稳定性。我们看到，扭曲通常会降低理想立方结构的能量，但没有发现扭曲化合物的能量低于立方相超过每个原子0.5 eV的情况。因此，我们仅计算了具有低于每个原子0.5eV的立方稳定性的组合物的变形。这导致2,162（1,776 + 386）种组合物，其中计算出四种扭曲（立方菱形，四方和斜方晶）。

most actinides were calculated using the DFT+U[19,20] formalism (see Table 1 for U-values, U is applied on d-electrons for transition metals and f-electrons for actinides) and were spin-polarized with a ferromagnetic alignment of spin. Calculations were performed within the framework of the Open Quantum Materials Database[8] which contains 470,000 different phases from experimental databases such as the Inorganic Crystal Structure Database (ICSD)[9,10] and hypothetical compounds based on common structural prototypes.

Calculation of formation energy

The formation energy of a compound, $H_f^{ABO_3}$, is calculated according to Equation 1:

$$H_f^{ABO_3} = E(ABO_3) - \mu_A - \mu_B - 3\mu_O \quad (1)$$

where $E(ABO_3)$ is the total energy of the perovskites and μ_A, μ_B and μ_O are the chemical potentials of A, B and oxygen, respectively. For most elements, chemical potentials are equal to the DFT total energies of their ground states. For some elements where the T = 0 K DFT ground state is not an adequate reference state, we apply corrections to the chemical potentials: diatomic gases (O), room temperature liquid (Hg), several elements with structural phase transformations between 0 and 298 K (Na, Ti, Sn) and elements with DFT+U correction (see Table 1). For these elements, chemical potentials are fitted to experimental formation energies. Details of the experimental formation energies used and the resulting corrections are given in ref. 8.

Calculation of phase stability

Thermodynamic stability was assessed using an energy convex hull construction[21,22]. By definition, the convex hull consists of phases that have an energy lower than any other phase or linear combination of phases at the respective compositions. The stability, or convex hull distance is defined as

$$H_{stab}^{ABO_3} = H_f^{ABO_3} - H_f \quad (2)$$

swhere $H_f^{ABO_3}$ is the formation energy of the perovskite, defined in Equation 1 and ΔH_f is the convex hull energy at the ABO_3 composition. For a given phase, P, the convex hull distance is positive if P is unstable and represents the energy difference between P and the convex hull at that composition. The convex hull distance is negative if P is stable and represents the energy difference between P and the (hypothetical) convex hull calculated without compound P. However, in this work, we wish to slightly loosen this restriction to account for nearly-stable compounds and possible uncertainties/errors associated with DFT. Hence, in this paper, we label compounds with a stability below 0.025 eV per atom (approximately kT at room temperature) as stable.

Structural distortions

After calculating all cubic perovskites, we investigated the effect of distortions on the stability of perovskites. To do so, we randomly selected one-third of the compositions (1,776) and computed their stability in the rhombohedral, tetragonal and orthorhombic distortion. We saw that distortions generally lower the energy of the ideal cubic structure but found no case where the distorted compound was lower in energy than the cubic phase by more than 0.5 eV per atom. Thus, we only calculated the distortion of compositions having a cubic stability lower than 0.5 eV per atom. This resulted in 2,162 (1,776+386) compositions where the four distortions (cubic rhombohedral, tetragonal and orthorhombic) where calculated.

Element	U-value [eV]
V	3.1
Cr	3.5
Mn	3.8
Fe	4
Co	3.3
Ni	6.4
Cu	4
Th	4
Pa	4
U	4
Np	4.0
Pu	4.0

Table 1. U-values used for the calculations of compounds containing the listed elements.

氧空位形成能的计算
使用立方未变形的A2B2O5 9原子超晶胞计算氧空位形成能由EO v =EðA2B2O5Þgivenμ0 - 2EðABO3Þð3Þ给出其中E（A2B2O5），E（ABO3）分别是缺陷和原始细胞的DFT总能量，μ0是氧的化学势。 9原子超级电池的选择使得能够高通量计算所有化合物的氧空位形成能。
氧化态和离子大小
使用键价法计算元素的氧化态23。如果化合物包含至少一个不具有BV参数的元素，则无法确定该化合物的化合价。利用氧化数和配位数（分别为A，B和O原子的12,6和2），使用Shannon radii24,25作为每个离子的离子大小的量度。
稳定钙钛矿的实验数据
为了将我们的稳定性预测与实验数据进行比较，我们从实验已知的钙钛矿的文献中收集了数据。汇总并策划了四篇综述论文26-29，以鉴定已经通过实验合成的223种ABO3钙钛矿组合物。策展是由仔细检查源纸，确保化合物真正合成并准确表征为钙钛矿。另外，除去在高压（> 1GPa）下合成的化合物。我们注意到，对于大多数这些化合物，在综合评价论文中没有报道元素的最稳定的变形或氧化状态。
代码可用性
高通量计算和热力学分析的自动化使用（https://github.com/wolverton-research-group/qmpy）上提供的qmpy python软件包完成，并在MIT许可下发布。 qmpy还用于管理高吞吐量工作流程并可视化计算的输出。 VASP是用于在此工作中生成数据的DFT代码，是http://www.vasp.at提供的专有软件。
数据记录
可以在figshare上找到5,329个ABO3钙钛矿的清单（数据引用1）。所有计算以及用于稳定性计算的所有470,000种化合物均可从www.oqmd.org下载或直接咨询。Kirklin等人8还包含有关计算的详细信息参数。
文件格式
数据存储在CSV电子表格中。每行包含不同的组合物，每列是该组合物的性质（在表2中描述）。未收敛到最终解的计算由表中针对该组合的连字符（'-'）表示。这种缺乏收敛可能由于各种原因而发生，例如，与最终松弛几何形状相距很远的初始几何形状，由难以实现电子自洽的元素组成的化学等等。这些类型的计算问题是在高通量方法中很常见，其中必须使用一致的设置在合理的时间内计算所有化合物。
数据的图形表示
图1的顶部显示了稳定钙钛矿的数量与A-和B-位点上发生的元素的函数关系。在73种元素中，只有硼不会出现在任何稳定的钙钛矿中。镧系元素和碱土金属通常位于A位，用于稳定的钙钛矿，而过渡金属，特别是第一排，在B位上是常见的。这些元素偏好与文献中实验报道的内容一致[1,26]。
图1的下半部分显示了在这项工作中计算的所有化合物的形成能和带隙分布。我们看到大多数化合物具有负的形成能量（只有276个是正的）。负形成能和稳定性之间的差异是显而易见的，因为大多数材料具有负的形成能但是不稳定。至于带隙，其中大部分为0，这在使用GGA-PBE时是预期的（参见技术验证部分的带隙段落）。
技术验证
开放量子材料数据库
开放量子材料数据库使用密度泛函理论（DFT）来计算每种化合物的总能量。 DFT由于其准确性和可重复性而被广泛用于固态物理学30-32。此外，先前的研究表明，与实验测量的那些相比，使用DFT计算的地层能量具有与来自两个不同来源的实验值之间的比较相似的精确度8。

Calculation of oxygen vacancy formation energy

Oxygen vacancy formation energies are calculated using a cubic undistorted $A_2B_2O_5$ 9-atom supercell and is given by

$$E_v^O = E(A_2B_2O_5) + \mu_O - 2E(ABO_3) \quad (3)$$

Where $E(A_2B_2O_5)$, $E(ABO_3)$ are the DFT total energies of the defect and pristine cell, respectively and μ_O is the chemical potential of oxygen. The choice of 9-atom supercell enables the high-throughput calculation of oxygen vacancy formation energy of all compounds.

Oxidation states and ionic size

Oxidation states for elements were computed using a bond valence method[23]. If a compound contains at least one element that does not have a BV parameter, the valence of that compound cannot be determined. With the oxidation numbers and coordination numbers (12, 6 and 2 for the A, B and O atoms, respectively) at hand, Shannon radii[24,25] were used as a measure of the ionic size of every ion.

Experimental data on stable perovskites

To compare our stability predictions with experimental data, we collected data from the literature on experimentally known perovskites. Four review papers[26–29] were aggregated and curated to identify 223 ABO_3 perovskite compositions that have been experimentally synthesized. The curation was made by careful examination of source papers to ensure that the compounds were really synthesized and accurately characterized as a perovskite. In addition, compounds that were synthesized at high pressure (>1 GPa) were removed. We note that, for most of those compounds, the most stable distortion or oxidation states of the elements are not reported in the aggregated review papers.

Code availability

The automation of the high-throughput calculations and thermodynamic analysis were done using the qmpy python package available at (https://github.com/wolverton-research-group/qmpy) and is released under the MIT license. qmpy was also used to manage the high-throughput workflow and visualize the output from calculations. VASP, the DFT code used to generate the data in this work, is a proprietary software available at http://www.vasp.at.

Data Records

The list of 5,329 ABO_3 perovskites can be found on figshare (Data Citation 1). All the calculations, along with all the 470,000 compounds used for the stability calculations are available for download or for direct consultation at www.oqmd.org. Kirklin *et al.*[8] also contains detailed information about the calculation parameters.

File format

The data is stored in a CSV spreadsheet. Each row contains a different composition and each column is a property of that composition (described in Table 2). A calculation that did not converge to a final solution is indicated by a hyphen ('–') in the table for that composition. This lack of convergence can happen for variety of reasons, for instance an initial geometry that is very far from the final relaxed geometry, a chemistry consisting of elements for which electronic self-consistency is difficult to achieve, etc. These kinds of computational issues are commonplace in high-throughput methods where *consistent settings* have to be used for the calculation of *all* compounds in a reasonable amount of time.

Graphical representation of the data

The top part of Fig. 1 shows the number of stable perovskites as a function of the elements occurring on the A- and B-sites. Out of 73 elements, only boron does not appear in any stable perovskites. Lanthanides and alkaline earths are frequently on the A-site for stable perovskites whereas transition metals, specially the first row, are common on the B-site. These elemental preferences are consistent with what is experimentally reported in the literature[1,26].

The bottom part of Fig. 1 shows the formation energy and band gap distribution for all the compounds calculated in this work. We see that most compounds have negative formation energies (only 276 are positive). The difference between negative formation energy and stability is apparent as the majority of materials have a negative formation energy but are not stable. As for the band gaps, the majority of them are 0 which is expected when using GGA-PBE (see band gap paragraph in the technical validation section).

Technical Validation

The open quantum materials database

The Open Quantum Materials Database uses density functional theory (DFT) to compute the total energies of every compound. DFT is widely used in solid states physics due to its accuracy and reproducibility[30–32]. In addition, previous studies have shown that formation energies calculated using DFT, when compared against those measured experimentally, have a similar accuracy as a comparison between experimental values from two different sources[8].

格子参数
对于预测稳定且在ICSD中具有条目的所有化合物，我们将DFT松弛结构的晶格参数与实验结构的晶格参数进行比较（图3）。113种化合物的所有晶格参数的平均误差（ME），平均绝对误差（MAE），平均相对误差（MRE）和平均绝对相对误差（MARE）分别为0.011Å，0.048Å，0.20和0.82%。晶格参数的大小和过高估计与DFT-PBE33文献中的其他晶格参数研究一致。
超空泡氧气空位形成能量的大小
我们比较了使用9原子超级单元计算的氧空位形成能与使用较大超晶胞尺寸（79原子）34计算的氧空位形成能。我们看到我们的高通量方法与文献数据之间存在良好的一致性（见图4）。此外，Curnan和Kitchin35已经表明，氧空位形成能量趋势在很大程度上不受LaBO3和SrBO3（B = Sc-Cu）的超晶胞尺寸的影响。
带隙
用GGA-PBE计算带隙，对于一些3d-过渡金属和act系元素具有U值（参见表1）。 GGA-PBE倾向于低估半导体的带隙36,37，这意味着在这项工作中提出的带隙值必须作为下限，并且有助于识别绝缘体。可以进行不同的，更昂贵的计算，例如混合功能或准粒子计算（GOWO，GWO和GW），以更准确地计算带隙值36

Lattice parameter

For all the compounds that are predicted to be stable and have an entry in the ICSD, we compared the lattice parameters of the DFT relaxed structure with the lattice parameter of the experimental structure (Fig. 3). The mean error (ME), mean absolute error (MAE), mean relative error (MRE) and mean absolute relative error (MARE) across all lattice parameter for the 113 compounds are 0.011 Å, 0.048 Å, 0.20 and 0.82%, respectively. The magnitude and overestimation of the lattice parameters are consistent with other lattice parameters studies in the literature for DFT-PBE[33].

Supercell size for oxygen vacancy formation energy

We compared our oxygen vacancy formation energies calculated using 9-atom supercells with oxygen vacancy formation energy calculated using larger supercell sizes (79-atom)[34]. We see good agreement between our high-throughput approach and data from the literature (see Fig. 4). In addition, Curnan and Kitchin[35] have showed that oxygen vacancy formation energy trend is largely unaffected by the supercell size for $LaBO_3$ and $SrBO_3$ (B = Sc-Cu).

Band gap

The band gaps were calculated with GGA-PBE, with U-values for some 3d-transition metals and actinides (see Table 1). GGA-PBE tends to underestimate the band gaps of semiconductors[36,37] meaning that band gap values presented in this work have to be taken as lower bound and are useful to identify insulators. Different, much more expensive, calculations, such as hybrid functionals or quasiparticle calculations (G_0W_0, GW_0 and GW), can be done to compute band gap values more accurately[36].

Name	Type	Unit	Description
Chemical formula	string	None	Chemical composition of the compound. The first and second elements correspond to the A- and B-site, respectively. The third element is always oxygen
A	string	None	Chemical element on the A-site
B	string	None	Chemical element on the B-site
In literature	boolean	None	Report of experimental synthesis of compound in the literature. True indicates that the compound is present in one of the four review papers.
Valence A	number or string	None	Valence of atom A as estimated by bond valence (BV) theory. If a compound is not balanced, it is denoted by 'not balanced'. If the compound contains a least one element without a BV parameter, it is denoted by 'element not in BV'
Valence B	number or string	None	Valence of atom B as estimated by bond valence (BV) theory. If a compound is not balanced, it is denoted by 'not balanced'. If the compound contains a least one element without a BV parameter, it is denoted by 'element not in BV'
Radius A	number	Å	Shannon ionic radius of atom A. When possible, the oxidation state and coordination number (12) of the A atom was used to estimate its radius.
Radius B	number	Å	Shannon ionic radius of atom B. When possible, the oxidation state and coordination number (6) of the B atom was used to estimate its radius.
Lowest distortion	string	None	Distortion with the lowest energy (among cubic, rhombohedral, tetragonal and orthorhombic corresponding to space group 221, 167, 99 and 62, respectively)
Formation energy	number	eV per atom	Formation energy as calculated by equation (1) of the distortion with the lowest energy
Stability	number	eV per atom	Stability (hull distance) as calculated by equation (2) of the distortion with the lowest energy. A compound is considered stable if it is within 0.025 eV per atom of the convex hull
Magnetic moment	number	μ_B	Resulting magnetic moment of the relaxed structure. If the composition does not contain any magnetic element, the magnetic moment is set to a hyphen ('–').
Volume per atom	number	$Å^3$ per atom	Volume per atom of the relaxed structure
Band gap	number	eV	PBE band gap obtained from the relaxed structure
a	number	Å	Lattice parameter a of the relaxed structure
b	number	Å	Lattice parameter b of the relaxed structure
c	number	Å	Lattice parameter c of the relaxed structure
alpha	number	°	α angle of the relaxed structure. $\alpha = 90$ for the cubic, tetragonal and orthorhombic distortion.
beta	number	°	β angle of the relaxed structure. $\beta = 90$ for the cubic, tetragonal and orthorhombic distortion.
gamma	number	°	γ angle of the relaxed structure. $\gamma = 90$ for the cubic, tetragonal and orthorhombic distortion.
Vacancy energy	number	eV per O atom	Oxygen vacancy formation energy as calculated by equation (3)

Table 2. Description of column keys for the CSV spreadsheet containing the dataset (Data Citation 1).

Magnetism

Several perovskites are experimentally observed to have complex magnetic structures, e.g., antiferromagnetic order[38]. However, in order to save computational times as the magnetic ordering of an unknown structure cannot be known *a priori*, only ferromagnetic configurations are calculated in the present study. Stevanović *et al.*[39] showed that the error associated with different magnetic ordering is of the order of 0.01 to 0.02 eV per atom by calculating ternary compounds with up to ten different spin orderings. To take this potential error into consideration, compounds with a hull distance below 0.025 eV per atom are labelled as stable in this study.

Comparison with experimentally observed perovskites

Of the 5,329 different compositions that were calculated, 395 are predicted to be thermodynamically stable by density functional theory. Out of those, 165 are reported in the literature. As a result, 230 *new* compounds are predicted to be DFT stable but not yet experimentally reported. This set of compounds represents a wide range of predictions amenable for materials synthesis.

The stability values of the 223 compounds that we found in 4 review papers[26–29] are plotted in Fig. 5

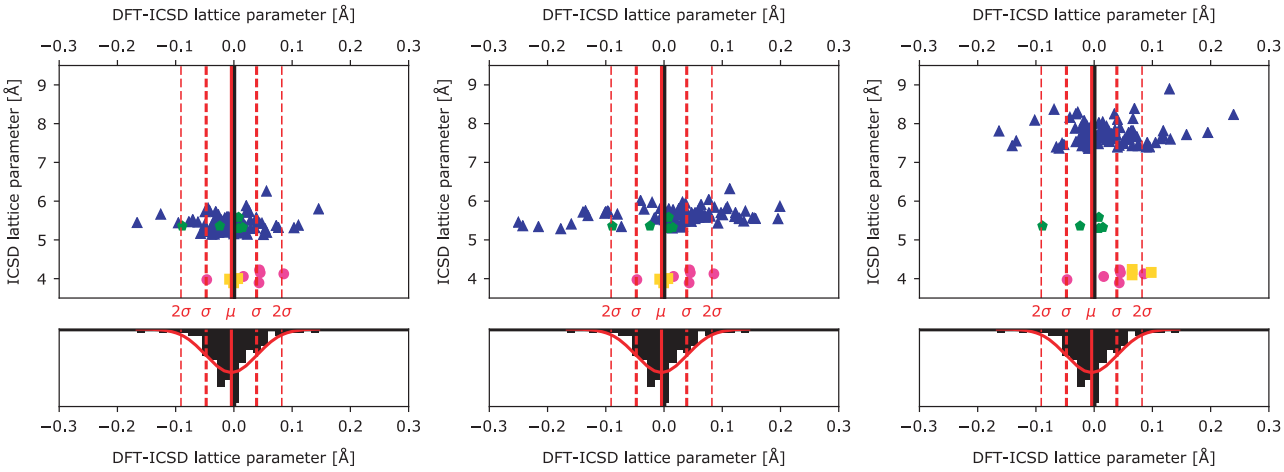

Figure 3. Comparison between DFT and ICSD lattice parameters for 113 compounds. (left) lattice parameter a, (center) lattice parameter b, and (right) lattice parameter c. In the top panels, the horizontal axes measure the difference between the computed and experimental lattice parameters while the vertical axes are the experimental lattice parameters. The lower plots correspond to a histogram of the difference in lattice parameters from DFT and experiment. The solid and dashed red lines indicate the average error, first and second standard deviations between DFT and experiment, respectively.

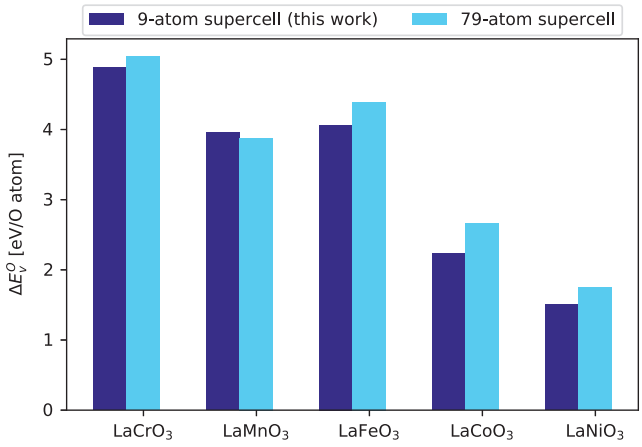

Figure 4. Comparison of oxygen vacancy formation energy between the present work (9-atom supercell) and larger supercells (79-atom). Data for the 79-atom supercell are taken from Deml *et al.*[34]

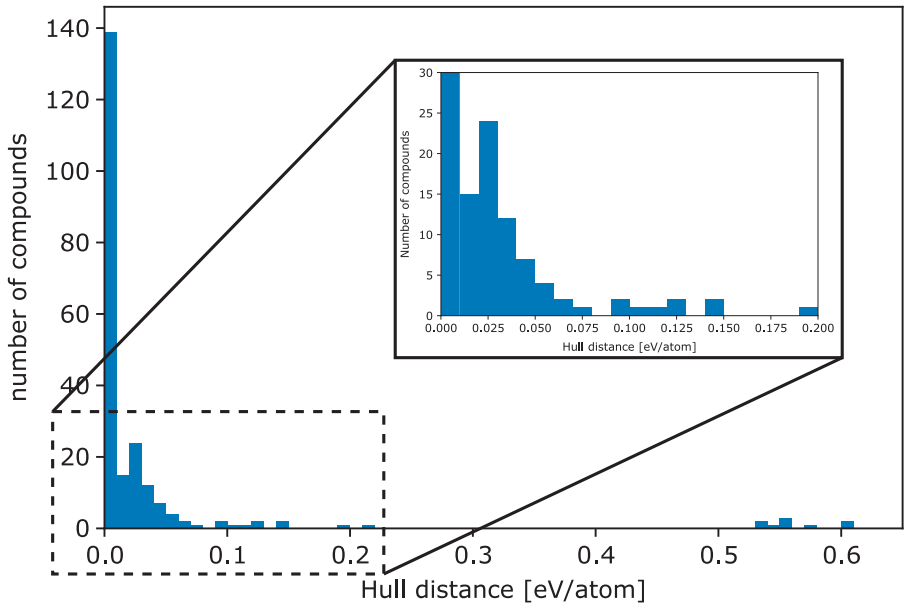

Figure 5. Histogram of the DFT stability of 223 ABO3 perovskite compounds reported in the literature. The inset shows the rapid decay of stable compounds as a function of stability.

(for clarity, all stable compounds are merged into the '0 hull distance bin'). The plot shows that a large number of these experimentally reported compounds are stable according to our DFT T = 0 K calculations. However, the remainder of the phases are above the convex hull, and hence metastable (or unstable). The results of Fig. 5 shows the measure of metastability in terms of convex hull distance: there is rapid decay of the number of synthesized compounds as the convex hull distance increases, reaching almost 0 at a hull distance of 0.1 eV per atom. This 0.1 eV/atom metric for metastability is consistent with the results from another recent high-throughput study of metastabiity by Sun *et al.*[40]

Nine compounds reported in the literature are seen with a stability above 0.5 eV per atom. All these compounds contain rare earth elements, which are difficult to treat accurately with DFT because of the complexities associated with f-electron systems. In our high-throughput study, f-electrons are not included in the valence electrons of the pseudopotentials used, and therefore the DFT calculations of rare-earth-containing perovskites could have physical errors associated with the approximations made in the DFT calculations. For a more detailed discussion about f-electrons and frozen-core potentials, we refer the reader to Kirklin *et al.*[8] Errors can also come from erroneous experimental characterization and/or classification.

Usage Notes

We suggest using the data as it is in the spreadsheet. If one chooses to access the data from OQMD via qmpy, we note that the OQMD is a constantly-growing database. Indeed, as a result of compounds being constantly calculated and added to the database, the stability of the already-present compounds can change: adding new stable compounds may change the predicted stability of a perovskite.

References

1. Bhalla, A. S., Guo, R., Roy, R., Ruyan, A. S. B. & Rustum, G. The perovskite structure—a review of its role in ceramic science and technology. *Mater. Res. Innov.* **4,** 3–26 (2000).
2. Chroneos, A., Vovk, R. V., Goulatis, I. & Goulatis, L. I. Oxygen transport in perovskite and related oxides: A brief review. *J. Alloys Compd* **494,** 190–195 (2010).
3. Ishihara, T.. *Perovskite Oxide for Solid Oxide Fuel Cells* (Springer US, 2009). Available at http://www.springerlink.com/index/10.1007/978-0-387-77708-5.
4. Castelli, I. E. *et al.* New cubic perovskites for one- and two-photon water splitting using the computational materials repository. *Energy Environ. Sci* **5,** 9034–9043 (2012).
5. Landis, D. D. *et al.* The Computational Materials Repository. *Comput. Sci. Eng.* **14,** 51–57 (2012).
6. Kim, C., Pilania, G. & Ramprasad, R. Machine Learning Assisted Predictions of Intrinsic Dielectric Breakdown Strength of ABX3 Perovskites. *J. Phys. Chem. C* **120,** 14575–14580 (2016).
7. Castelli, I. E. & Jacobsen, K. W. Designing rules and probabilistic weighting for fast materials discovery in the Perovskite structure. *Model. Simul. Mater. Sci. Eng.* **22,** 55007 (2014).
8. Kirklin, S. *et al.* The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater* **1,** 15010 (2015).
9. Bergerhoff, G., Hundt, R., Sievers, R. & Brown, I. D. The inorganic crystal structure data base. *J. Chem. Inf. Model.* **23,** 66–69 (1983).

10. Belsky, A., Hellenbrandt, M., Karen, V. L. & Luksch, P. New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Crystallogr. Sect. B Struct. Sci* **58,** 364–369 (2002).
11. Emery, A. A., Saal, J. E., Kirklin, S., Hegde, V. I. & Wolverton, C. High-Throughput Computational Screening of Perovskites for Thermochemical Water Splitting Applications. *Chem. Mater.* **28,** 5621–5634 (2016).
12. Lookman, T., Alexander, F. J. & Bishop, A. R. Perspective: Codesign for materials science: An optimal learning approach. *APL Mater* **4,** 053501 (2016).
13. Agrawal, A. & Choudhary, A. Perspective: Materials informatics and big data: Realization of the 'fourth paradigm' of science in materials science. *APL Mater* **4,** 53208 (2016).
14. Rajan, K. Materials Informatics: The Materials ' Gene ' and Big Data. *Annu. Rev. Mater. Res* **45,** 153–169 (2015).
15. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6,** 15–50 (1996).
16. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54,** 11169–11186 (1996).
17. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59,** 1758–1775 (1999).
18. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77,** 3865–3868 (1996).
19. Dudarev, S. L., Botton, G. A., Savrasov, S. Y., Humphreys, C. J. & Sutton, A. P. Electron-energy-loss spectra and the structural stability of nickel oxide: An LSDA+U study. *Phys. Rev. B* **57,** 1505–1509 (1998).
20. Wang, L., Maxisch, T. & Ceder, G. Oxidation energies of transition metal oxides within the GGA+U framework. *Phys. Rev. B* **73,** 195107 (2006).
21. Barber, C. B., Dobkin, D. P. & Huhdanpaa, H. The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw* **22,** 469–483 (1996).
22. Akbarzadeh, A. R., Ozolins, V. & Wolverton, C. First-Principles Determination of Multicomponent Hydride Phase Diagrams: Application to the Li-Mg-N-H System. *Adv. Mater.* **19,** 3233–3239 (2007).
23. O'Keefe, M. & Brese, N. E. Atom sizes and bond lengths in molecules and crystals. *J. Am. Chem. Soc.* **113,** 3226–3229 (1991).
24. Shannon, R. D. & Prewitt, C. T. Effective ionic radii in oxides and fluorides. *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem* **25,** 925–946 (1969).
25. Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr. Sect. A* **32,** 751–767 (1976).
26. Roth, R. S. Classification of perovskite and other ABO3-type compounds. *J. Res. Natl. Bur. Stand. (1934)* **58,** 75–88 (1957).
27. Giaquinta, D. M. & zur Loye, H.-C. Structural Predictions in the ABO3 Phase Diagram. *Chem. Mater.* **6,** 365–372 (1994).
28. Li, C., Soh, K. C. K. & Wu, P. Formability of ABO3 perovskites. *J. Alloys Compd* **372,** 40–48 (2004).
29. Zhang, H., Li, N., Li, K. & Xue, D. Structural stability and formability of ABO3 -type perovskite compounds. *Acta Crystallogr. Sect. B Struct. Sci* **63,** 812–818 (2007).
30. Lejaeghere, K. *et al.* Reproducibility in density functional theory calculations of solids. *Science (80 − .)* **351,** aad3000 (2016).
31. Curtarolo, S., Morgan, D. & Ceder, G. Accuracy of ab initio methods in predicting the crystal structures of metals: A review of 80 binary alloys. *Calphad Comput. Coupling Phase Diagrams Thermochem* **29,** 163–211 (2005).
32. Hautier, G., Ong, S. P., Jain, A., Moore, C. J. & Ceder, G. Accuracy of density functional theory in predicting formation energies of ternary oxides from binary oxides and its implication on phase stability. *Phys. Rev. B* **85,** 155208 (2012).
33. Haas, P., Tran, F. & Blaha, P. Calculation of the lattice constant of solids with semilocal functionals. *Phys. Rev. B—Condens. Matter Mater. Phys* **79,** 1–10 (2009).
34. Deml, A. M. *et al.* Tunable Oxygen Vacancy Formation Energetics in the Complex Perovskite Oxide Srx La1- x Mny Al1- y O3. *Chem. Mater.* **26,** 6595–6602 (2014).
35. Curnan, M. T. & Kitchin, J. R. Effects of Concentration, Crystal Structure, Magnetism, and Electronic Structure Method on First-Principles Oxygen Vacancy Formation Energy Trends in Perovskites. *J. Phys. Chem. C* **118,** 28776–28790 (2014).
36. Shishkin, M. & Kresse, G. Self-consistent GW calculations for semiconductors and insulators. *Phys. Rev. B* **75,** 235102 (2007).
37. Perdew, J. P. Density functional theory and the band gap problem. *Int. J. Quantum Chem.* **30,** 451–451 (1986).
38. Lee, Y.-L., Kleis, J., Rossmeisl, J. & Morgan, D. Ab initio energetics of LaBO3 (001) (B = Mn, Fe, Co, and Ni) for solid oxide fuel cell cathodes. *Phys. Rev. B* **80,** 224101 (2009).
39. Stevanović, V., Lany, S., Zhang, X. & Zunger, A. Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies. *Phys. Rev. B* **85,** 115104 (2012).
40. Sun, W. *et al.* The Thermodynamic Scale of Inorganic Crystalline Metastability. *Sci. Adv* **2,** e1600225 (2016).

Data Citation

1. Emery, A. A. & Wolverton, C. Figshare http://dx.doi.org/10.6084/m9.figshare.5334142 (2017).

Acknowledgements

This work was also funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Grant DE-FG02-07ER46433. This research used computational resources provided by the Quest high-performance computing facility at Northwestern University which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology as well as the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Author Contributions

A.A.E. ran the calculations and analyzed the data. C.W. conceived the overall research plan, provided advice for the work, and helped analyze the data.

Additional Information

Competing interests: The authors declare no competing financial interests.

How to cite this article: Emery, A. A. & Wolverton, C. High-throughput DFT calculations of formation energy, stability and oxygen vacancy formation energy of ABO_3 perovskites. *Sci. Data* 4:170153 doi: 10.1038/sdata.2017.153 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.