Title	Prediction of perovskite-related structures in ACuO₃ (A = Ca, Sr, Ba, Sc, Y, La) using density functional theory and Bayesian optimization
Author(s)	Seko, Atsuto; Ishiwata, Shintaro
Citation	Physical Review B (2020), 101(13)
Issue Date	2020-04-01
URL	http://hdl.handle.net/2433/259827
Right	© 2020 American Physical Society;
Type	Journal Article
Textversion	publisher

Prediction of perovskite-related structures in ACuO$_{3-x}$ (A = Ca, Sr, Ba, Sc, Y, La) using density functional theory and Bayesian optimization

Atsuto Seko [1,2,3,*] and Shintaro Ishiwata [4]

[1]*Department of Materials Science and Engineering, Kyoto University, Kyoto 606-8501, Japan*
[2]*Center for Elements Strategy Initiative for Structure Materials, Kyoto University, Kyoto 606-8501, Japan*
[3]*Center for Materials Research by Information Integration, National Institute for Materials Science, Tsukuba 305-0047, Japan*
[4]*Division of Materials Physics, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan*

Oxygen vacancy ordering in perovskite-type transition-metal oxides plays an important role in the emergence of exotic electronic properties, as typified by superconducting cuprates. In this study, we predict the stability of oxygen-deficient perovskite structures in ACuO$_{3-x}$ (A = Ca, Sr, Ba, Sc, Y, La) by density functional theory calculation. We introduce a combination of the cluster expansion method, Gaussian process, and Bayesian optimization to find stable oxygen-deficient structures among a considerable number of candidates. Our calculations not only reproduce the reported structures but suggest the presence of several unknown oxygen-deficient perovskite structures, some of which are stabilized at high pressures. This work demonstrates the great applicability of the present computational procedure for the elucidation of the structural stability of strongly correlated oxides with a large tolerance for oxygen deficiency.

I. INTRODUCTION

Perovskite-type transition-metal oxides, namely, ABO$_3$, have been extensively studied as strongly correlated systems showing a large variety of electronic properties, typified by high-T_c cuprates and ferromagnetic manganites. One of the major methods of controlling the physical properties of ABO$_3$ is the chemical substitution for A-site cations, which dominates the bandwidth and valence of B-site cations [1–3]. Their potential to provide novel electronic properties can be further enhanced by introducing oxygen deficiency, which plays an important role not only in the valence of B-site cations but also in the topology of the B-O lattice. For instance, the typical superconducting cuprates YBa$_2$Cu$_3$O$_{7-d}$ [4] and Sr$_{1-x}$La$_x$CuO$_2$ [5] are well-ordered oxygen-deficient perovskites with CuO$_2$ planes. In addition, SrCuO$_{2.5}$ [6] and LaCuO$_{3-x}$, with $x =$ 0.4 and 0.5 [7], have been reported as potential superconducting cuprates with a quasi-one-dimensional Cu-O motif.

Considering the recent development of computational structure prediction methods and computational resources, it is worth performing a comprehensive search for oxygen-deficient perovskite-type structures and related ones to discover novel cuprates that have been overlooked so far. Recently, computational studies on perovskite-type transition-metal oxides, ABO$_{3-x}$, have been reported, but their structural variations are quite limited [8]. In this study, we systematically investigate the stability of oxygen-deficient perovskite structures in ACuO$_{3-x}$ ($0 \leqslant x \leqslant 1$) with divalent and trivalent cations (A = Ca, Sr, Ba, Sc, Y, La) by density functional

theory (DFT) calculation. This study is based on a large set of oxygen-vacancy configurations derived from the cubic perovskite lattice to search for a stable structure with a large unit cell volume. Enumeration of the nonequivalent configurations is shown in Sec. II.

A common way to find stable structures in such a configurational system is the cluster expansion (CE) method [9–11]. The CE method has successfully been used to describe alloy thermodynamics and to efficiently find stable alloy configurations among a considerable number of candidates. Although the CE method has also been successful at predicting atom-deficient nonstoichiometric structures [12–18], a simple use of the CE method fails to accurately represent the configurational energy in oxygen-deficient ACuO$_{3-x}$ with large geometry relaxation from the ideal cubic perovskite lattice, as shown in Sec. III. Therefore, we introduce Bayesian optimization [19], which was recently applied to the exploration of materials and structure prediction [20–28]. As shown in Sec. IV, Bayesian optimization with the Gaussian process (GP) [29] and structural representations of the CE method is very useful for predicting oxygen-deficient structures. Although a nonlinear neural network CE approach that was recently proposed can also estimate the configurational energy efficiently [30], a probabilistic approach such as GP plays an essential role in predicting oxygen-deficient structures using Bayesian optimization.

II. NONEQUIVALENT OXYGEN-DEFICIENT PEROVSKITE STRUCTURES

A pool of oxygen-deficient structures with the cubic perovskite lattice is first generated using a derivative structure algorithm proposed by Hart and Forcade [31,32] implemented
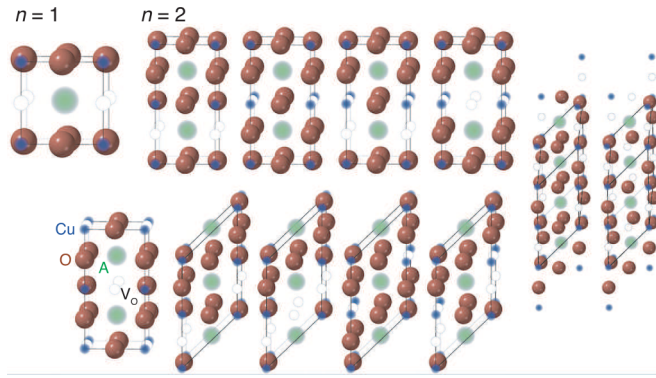
*seko@cms.mtl.kyoto-u.ac.jp

FIG. 1. A complete set of nonequivalent oxygen-deficient structures of cubic perovskite for supercell sizes of $n = 1$ and $n = 2$ in ACuO$_{3-x}$ ($0 \leqslant x \leqslant 1$).

in the CLUPAN code [33,34]. The derivative structure algorithm consists of two steps of enumeration. In the first step, a complete set of nonequivalent supercell shapes of the cubic perovskite is enumerated for a given supercell size n. The supercell shape is identified by the three-dimensional square Hermite normal form (HNF) $\boldsymbol{H}$, with the determinant $n = |\boldsymbol{H}|$, modifying the axis matrix of the conventional unit cell $\boldsymbol{A}$ as

$$\boldsymbol{A}' = \boldsymbol{A}\boldsymbol{H}. \tag{1}$$

Therefore, the enumeration of nonequivalent supercell shapes corresponds to that of nonequivalent HNFs. For instance, a complete set of nonequivalent HNFs for up to $n = 8$ is listed in Appendix A (Table III). In the second step, nonequivalent oxygen-deficient structures are enumerated for each HNF using the finitely presented group of the supercell lattice constructed by the HNF. Figure 1 illustrates a complete set of nonequivalent oxygen-deficient structures with supercell sizes of $n = 1$ and $n = 2$. Moreover, Table I lists the number of nonequivalent oxygen-deficient structures in ACuO$_{3-x}$ ($0 \leqslant x \leqslant 1$). The number of nonequivalent structures increases exponentially as the supercell size increases.

TABLE I. Number of nonequivalent oxygen-deficient structures of cubic perovskite in ACuO$_{3-x}$ ($0 \leqslant x \leqslant 1$). Symbols N_{sites}, N_{config}, and $N_{\text{config,acc}}$ denote the number of oxygen lattice sites, the number of nonequivalent configurations for a given n, and the cumulative number of nonequivalent configurations with up to a given n, respectively.

n	N_{sites}	N_{config}	$N_{\text{config,acc}}$
1	3	1	1
2	6	11	12
3	9	42	54
4	12	444	498
5	15	1245	1743
6	18	16 343	18 086
7	21	49 403	67 489
8	24	706 622	774 111

III. CLUSTER EXPANSION METHOD

A. Formalism

In general, it is impossible to estimate the configurational energy for a complete set of nonequivalent structures by DFT calculation owing to the considerable number of nonequivalent structures. Therefore, it has been a practical way to employ a machine learning model of the configurational energy estimated from a set of DFT calculations for sampled structures. The CE method [9–11] has been a widely used approach and gives a reliable model of configurational energy; hence, it has enabled us to accurately predict the ground-state structures and finite-temperature thermodynamics in many multicomponent systems.

In a binary system, the configurational energy E is described in a linear form using the pseudospin configurational variable σ_i for the respective lattice site i as

$$E = V_0 + \sum_i V_i \sigma_i + \sum_{i,j} V_{ij} \sigma_i \sigma_j + \sum_{i,j,k} V_{ijk} \sigma_i \sigma_j \sigma_k + \cdots$$

$$= \sum_\alpha V_\alpha \cdot \varphi_\alpha, \tag{2}$$

where φ_α denotes the correlation function of cluster α. Coefficients V_α are called the effective cluster interactions (ECIs), which are estimated from a set of DFT calculations for sampled configurations using general linear regression methods such as conventional linear regression and the least absolute shrinkage and selection operator (LASSO) [35,36].

B. Application to oxygen-deficient perovskite

We prepared a training data set that contains the formation energies of all nonequivalent structures with up to $n = 4$ estimated by DFT calculation performed using the plane-wave-basis projector augmented-wave method [37] within the Perdew-Burke-Ernzerhof generalized gradient approximation exchange-correlation functional [38], as implemented in the VASP code [39–41]. The cutoff energy was set to 400 eV. The total energies converged to less than 10^{-3} meV/supercell. The atomic positions and lattice constants of the nonequivalent structure were optimized until the residual forces were less than 10^{-2} eV/Å to consider the relaxation effect of the ideal cubic perovskite lattice. We assumed the ferromagnetic configuration for copper atoms in the DFT calculation for simplicity, although ACuO$_{3-x}$ is expected to show an antiferromagnetic configuration as shown experimentally in Ca$_{1-x}$Sr$_x$CuO$_2$ [42]. We adopted the $+U$ approach to improve the description of the ground state of correlated systems [43]. The value of U for copper was set to 4 eV, taken from the literature [44].

Figure 2 shows the formation energies of nonequivalent oxygen-deficient structures with up to $n = 4$ in ACuO$_{3-x}$ systems. They are measured from the energy of the cubic perovskite ACuO$_3$ and the lowest energy among the structures in ACuO$_2$. Since the DFT calculation fails to finish successfully in some structures, we exclude the structures from the data set. The sizes of the resultant training data sets are 485, 473, 480, 457, 468, and 480 for BaCuO$_{3-x}$, CaCuO$_{3-x}$, SrCuO$_{3-x}$, ScCuO$_{3-x}$, YCuO$_{3-x}$, and LaCuO$_{3-x}$, respectively.
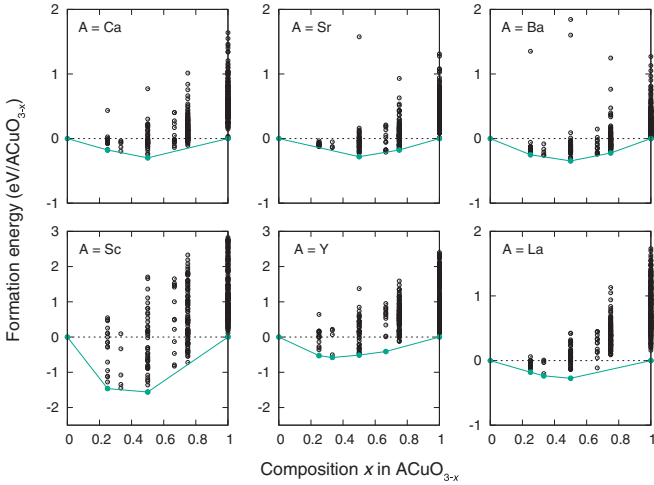
FIG. 2. Formation energies of nonequivalent oxygen-deficient perovskite structures with up to $n = 4$. The green line shows the convex hull of formation energy.
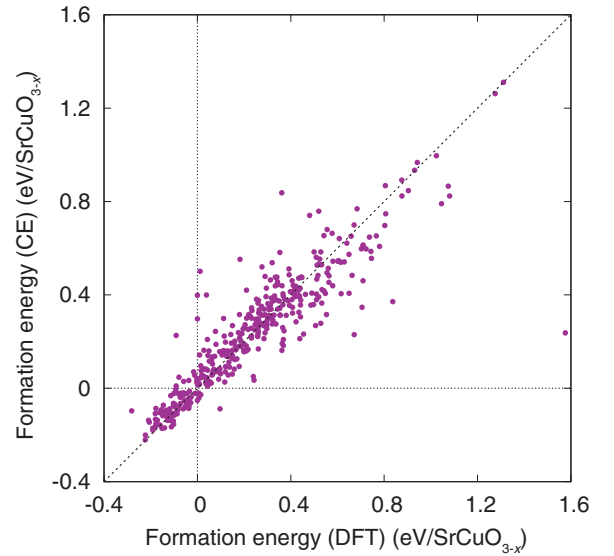

FIG. 3. Distribution of energies of nonequivalent oxygen-deficient structures with up to $n = 4$ predicted using a CE model with 36 clusters up to four bodies (model 3 in Table II) in $SrCuO_{3-x}$. The Pearson correlation coefficient between the DFT and CE energies is $R = 0.90$.

Then, we construct various CE models for each system using the training data set. We employ spin values of $+1$ and -1 for the oxygen atom and vacancy, respectively, which are required to define the correlation function. Table II shows the fitting errors of the four CE models, which indicate that all of the CE models have significant errors. They are much more significant than the order of the difference of the formation energies between structures. Figure 3 also shows the distribution of the formation energies of nonequivalent oxygen-deficient structures estimated using a CE model for $SrCuO_{3-x}$. The formation energies deviate from those estimated by DFT calculations. Therefore, it is impossible to find the true

TABLE II. Fitting errors for four CE models (in units of eV/$ACuO_{3-x}$). The fitting error normalized by the absolute value of the lowest formation energy is also shown in parentheses. N_c denotes the number of regression coefficients included in a CE model. Models 1, 2, 3, and 4 are composed of 11 pairs, 19 pairs, 11 pairs and 23 multiplets of up to four bodies, and 11 pairs and 39 multiplets of up to six bodies, respectively. All of them also contain the empty and point clusters corresponding to the bias term and the composition x, respectively.

	Model 1	Model 2	Model 3	Model 4
$CaCuO_{3-x}$	0.171	0.167	0.157	0.152
	(0.572)	(0.559)	(0.525)	(0.508)
$SrCuO_{3-x}$	0.135	0.133	0.127	0.125
	(0.480)	(0.473)	(0.451)	(0.445)
$BaCuO_{3-x}$	0.170	0.169	0.159	0.155
	(0.491)	(0.488)	(0.459)	(0.448)
$ScCuO_{3-x}$	0.669	0.657	0.614	0.606
	(0.429)	(0.421)	(0.393)	(0.388)
$YCuO_{3-x}$	0.399	0.386	0.355	0.350
	(0.687)	(0.664)	(0.611)	(0.602)
$LaCuO_{3-x}$	0.283	0.274	0.255	0.250
	(1.033)	(1.000)	(0.932)	(0.912)
N_c	13	21	36	52

ground states using only the energy predicted using the CE model.

The significant error in the CE model arises from (1) the fact that different oxidation states for copper are not distinguished in the CE model and (2) the fact that the CE model describes the energy of the oxygen-deficient structure obtained by the local geometry optimization using the correlation functions defined for the ideal lattice of cubic perovskite. Although the CE models should be improved by distinguishing different oxidation states for copper (e.g., Ref. [18]), we adopt another straightforward approach to search for the stable oxygen-deficient perovskite structures.

IV. BAYESIAN OPTIMIZATION

A. Gaussian process regression

To find stable oxygen-deficient perovskite structures, we employ Bayesian optimization [19] based on the probabilistic GP [29] specified by its mean function and covariance function instead of a linear CE model. Representing a structure by vector $\boldsymbol{d}$, a radial basis function covariance or kernel between structures $\boldsymbol{d}_i$ and $\boldsymbol{d}_j$ for noise-free observation is given by

$$k(\boldsymbol{d}_i, \boldsymbol{d}_j) = \sigma_f^2 \exp\left(-\frac{|\boldsymbol{d}_i - \boldsymbol{d}_j|^2}{2l^2}\right), \qquad (3)$$

where l and σ_f^2 are tuning parameters controlling the length scales for structure $\boldsymbol{d}$ and the observation, respectively. Here, we represent oxygen-deficient structures by the correlation functions of point clusters and 11 pair clusters used in the CE models shown in Fig. 6(c) below because many-body clusters improve the CE models slightly, as can be seen in the fitting errors for the data set of structures with up to $n = 4$. As described above, the discrepancy between the DFT calculation and the CE model is too large to predict the stable

oxygen-deficient perovskite structures using the CE model itself. On the other hand, the CE models have a strong correlation with the DFT calculation, as shown in Fig. 3. This implies that a set of correlation functions is useful for representing a structure in GP-based Bayesian optimization.

Describing the formation energies of structures in the training data set as y, the mean function μ at structure d_* and the variance function σ_*^2 are given as

$$\mu(d_*) = k_*^{\top} K^{-1} y \tag{4}$$

and

$$\sigma_*^2 = k(d_*, d_*) - k_*^{\top} K^{-1} k_*, \tag{5}$$

respectively, where $k_* = [k(d_*, d_1), \ldots, k(d_*, d_N)]^{\top}$ is the vector of kernel functions between structure d_* and N structures in the training data set. A symmetric kernel matrix K is composed of kernel functions for all pair arrangements of the training data, expressed as

$$K = \begin{pmatrix} k(d_1, d_1) & k(d_1, d_2) & \cdots & k(d_1, d_N) \\ k(d_2, d_1) & k(d_2, d_2) & \cdots & k(d_2, d_N) \\ \vdots & \vdots & \ddots & \vdots \\ k(d_N, d_1) & k(d_N, d_2) & \cdots & k(d_N, d_N) \end{pmatrix}. \tag{6}$$

Note that the variance of a probabilistic model has also been used for the structure selection to efficiently determine the ECIs in the framework of the CE [33,45,46].

B. Procedure of Bayesian optimization

Our procedure of Bayesian optimization is as follows. First, a GP model is developed from an initial set of formation energies. The initial set is composed of the formation energies of all nonequivalent structures with up to $n = 4$ and 100 nonequivalent structures that are randomly selected from the pool of 773 613 nonequivalent structures with up to $n = 8$. The model is then iteratively updated by (i) sampling the structure for which the formation energy is expected to be the lowest among the remaining candidate structures for each composition and (ii) updating the GP model using the training data set updated by the structures observed in step (i). These steps are repeated until the convex hull of the formation energy converges. We exclude some structures from the data set, i.e., structures for which DFT calculations fail to converge.

There are several well-known procedures to sample the structure that is expected to show the lowest formation energy in step (i). Here, we adopt a simple procedure of the probability of improvement (PI) strategy [19]. In the minimization problem of formation energy, the PI involves sampling the structure at which the probability that the formation energy is lower than y_{best} is maximized, where y_{best} denotes the current best formation energy among the training data. This means that structure $d_{i'}$ is selected by maximizing the probability as

$$d_{i'} := \underset{d_*}{\mathrm{argmax}} \, \Phi\left(\frac{y_{\text{best}} - \mu(d_*)}{\sigma_*}\right), \tag{7}$$

where $\Phi[y - \mu(d_*)/\sigma_*]$ denotes the cumulative distribution function of the normal distribution $N(\mu, \sigma^2)$. Since the structure with the maximum probability shows the highest Z score,
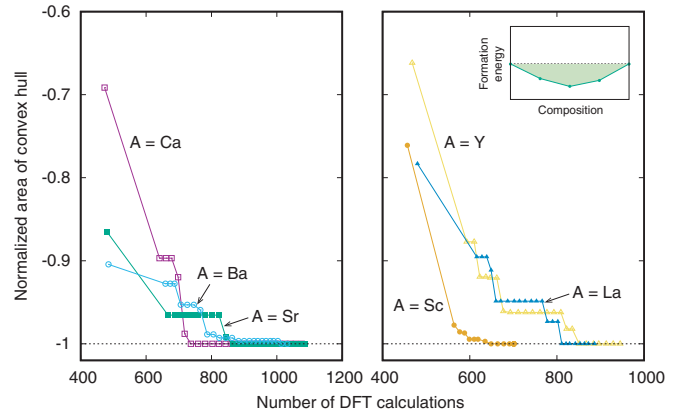

FIG. 4. Convergence behavior of the convex hull of formation energy in Bayesian optimization for $A\mathrm{CuO}_{3-x}$ systems. The vertical axis indicates the area of the convex hull below the line indicating zero formation energy, schematically illustrated as the shaded area at the top of the right panel. The area is normalized by the area of the converged convex hull.

defined as

$$Z(d_*) = \frac{y_{\text{best}} - \mu(d_*)}{\sigma_*}, \tag{8}$$

we select the structure with the highest Z score in practice.

C. Oxygen-deficient perovskite structures at 0 GPa

Figure 4 shows the convergence behavior of the convex hull of formation energy in Bayesian optimization. We measure the convergence degree of the convex hull using the area of the convex hull below the line indicating zero formation energy. As can be seen in Fig. 4, the convex hull converges well in all systems. The numbers of DFT calculations required for the convergence are 1076, 1085, 1032, 701, 945, and 886 in CaCuO_{3-x}, SrCuO_{3-x}, BaCuO_{3-x}, ScCuO_{3-x}, YCuO_{3-x}, and LaCuO_{3-x}, respectively.

Figure 5 shows the formation energies of the oxygen-deficient structures with up to $n = 8$ selected by Bayesian optimization, including the structures in the initial training data set of Bayesian optimization. The dashed line in Fig. 5 indicates the convex hull of formation energy representing the set of oxygen-deficient perovskite structures that are more stable than the other oxygen-deficient perovskite structures. We have found three, five, eight, five, four, and three oxygen-deficient perovskite structures on the convex hull in CaCuO_{3-x}, SrCuO_{3-x}, BaCuO_{3-x}, ScCuO_{3-x}, YCuO_{3-x}, and LaCuO_{3-x}, respectively. Four of the oxygen-deficient perovskite structures on the convex hull, i.e., CaCuO_2, SrCuO_2, $\mathrm{SrCuO}_{2.5}$, and $\mathrm{LaCuO}_{2.5}$, are identical to the experimental structures in the literature [6,7,47,48]. They are denoted by p-1 and p-2 in Fig. 5.

D. Stability relative to experimental structures

Although the convex hull of oxygen-deficient perovskite structures has many vertices or oxygen-deficient perovskite structures, experimental structures that are not classified into the perovskite family have also been known in $A\mathrm{CuO}_{3-x}$
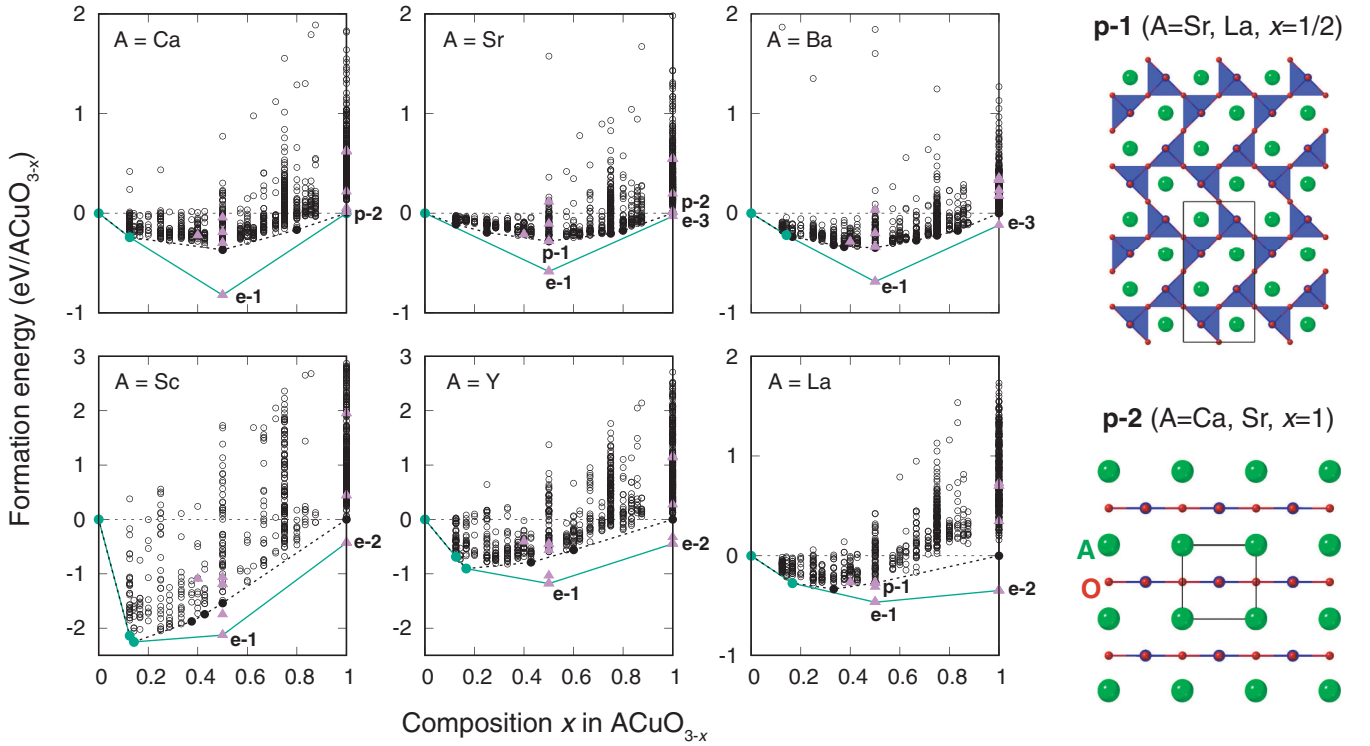
FIG. 5. Formation energy distribution of oxygen-deficient perovskite structures with up to $n = 8$ selected by Bayesian optimization. The dashed black line and black solid circles show the convex hull of formation energy and its vertices, respectively, obtained only from oxygen-deficient perovskite structures. The purple solid triangles show the formation energies of experimental prototype structures. The green line shows the convex hull obtained from the combination of oxygen-deficient perovskite structures and the experimental prototype structures. The oxygen-deficient perovskite structures reported in the literature the convex hulls are indicated by p-1 ($Ca_2Mn_2O_5$ type) and p-2 ($LaNiO_2$ type). Their crystal structures are also shown. Moreover, the experimental structures on the convex hulls are indicated as e-1 ($Sc_2Cu_2O_5$ type), e-2 (delafossite $FeCuO_2$ type), and e-3 ($SrCuO_2$ type).

systems. In the six $ACuO_{3-x}$ systems, seven structure prototypes with no partial occupancy of $Sc_2Cu_2O_5$, $Y_2Cu_2O_5$, $Nd_2Cu_2O_5$, $SrCuO_2$, $BaCuO_2$, delafossite $FeCuO_2$, and $CaCuO_2$ types are known in addition to the oxygen-deficient perovskite $Ca_2Mn_2O_5$ type (p-1) and $LaNiO_2$ type (p-2). Therefore, we estimate the formation energy for the seven prototypes in each system by DFT calculation.

Figure 5 shows the formation energy distribution of the experimental prototype structures. The convex hull obtained from the oxygen-deficient perovskite structures and the experimental prototype structures is also shown in Fig. 5. In all compositions of $ACuO_{2.5}$, the $Sc_2Cu_2O_5$-type structure (e-1) is stable, although the $Y_2Cu_2O_5$-type structure is competitive with the $Sc_2Cu_2O_5$-type structure in $LaCuO_{2.5}$. Also, the $SrCuO_2$ type (e-3) is stable in $SrCuO_2$ and $BaCuO_2$, whereas the delafossite $FeCuO_2$ type (e-2) is stable in $ScCuO_2$, $YCuO_2$, and $LaCuO_2$. The stable structure of $CaCuO_2$ is of the $LaNiO_2$ type (p-2), which is consistent with the experimental structure. On the other hand, most of the oxygen-deficient perovskite structures are not vertices of the convex hull, which indicates that they are not thermodynamically stable at 0 GPa.

E. Clustering of oxygen-deficient perovskite structures

To understand the common features of the oxygen-deficient perovskite structures on the convex hull, we carry out a machine learning clustering of the oxygen-deficient perovskite structures. Here, we introduce the Warren-Cowley short-range order (WC-SRO) parameter to represent the oxygen-deficient perovskite structures. The WC-SRO parameters can be derived from the correlation functions used in both the CE method and Bayesian optimization. In $ACuO_{3-x}$, the WC-SRO parameter of pair α for structure $\boldsymbol{d}$ is defined as

$$SRO_\alpha(\boldsymbol{d}) = \frac{\varphi_\alpha(\boldsymbol{d}) - q^2}{1 - q^2}, \tag{9}$$

where $\varphi_\alpha(\boldsymbol{d})$ denotes the correlation function of pair α and $q = (3 - 2x)/3$.

We employ a hierarchical clustering approach using Pearson correlation coefficients between the oxygen-deficient perovskite structures. For the distance metric to define the similarity between the structures, we adopt a complete-linkage clustering, which uses the maximum correlation-based distances between all observations of the two sets. Figure 6(a) shows Pearson correlation coefficients between the oxygen-deficient perovskite structures on the convex hull, represented by WC-SRO parameters of the 11 pairs shown in Fig. 6(b). As can be seen in Fig. 6(a), the hierarchical clustering classifies the oxygen-deficient perovskite structures into four large groups.

Figure 6(c) shows the WC-SRO parameters averaged over the oxygen-deficient perovskite structures belonging to each
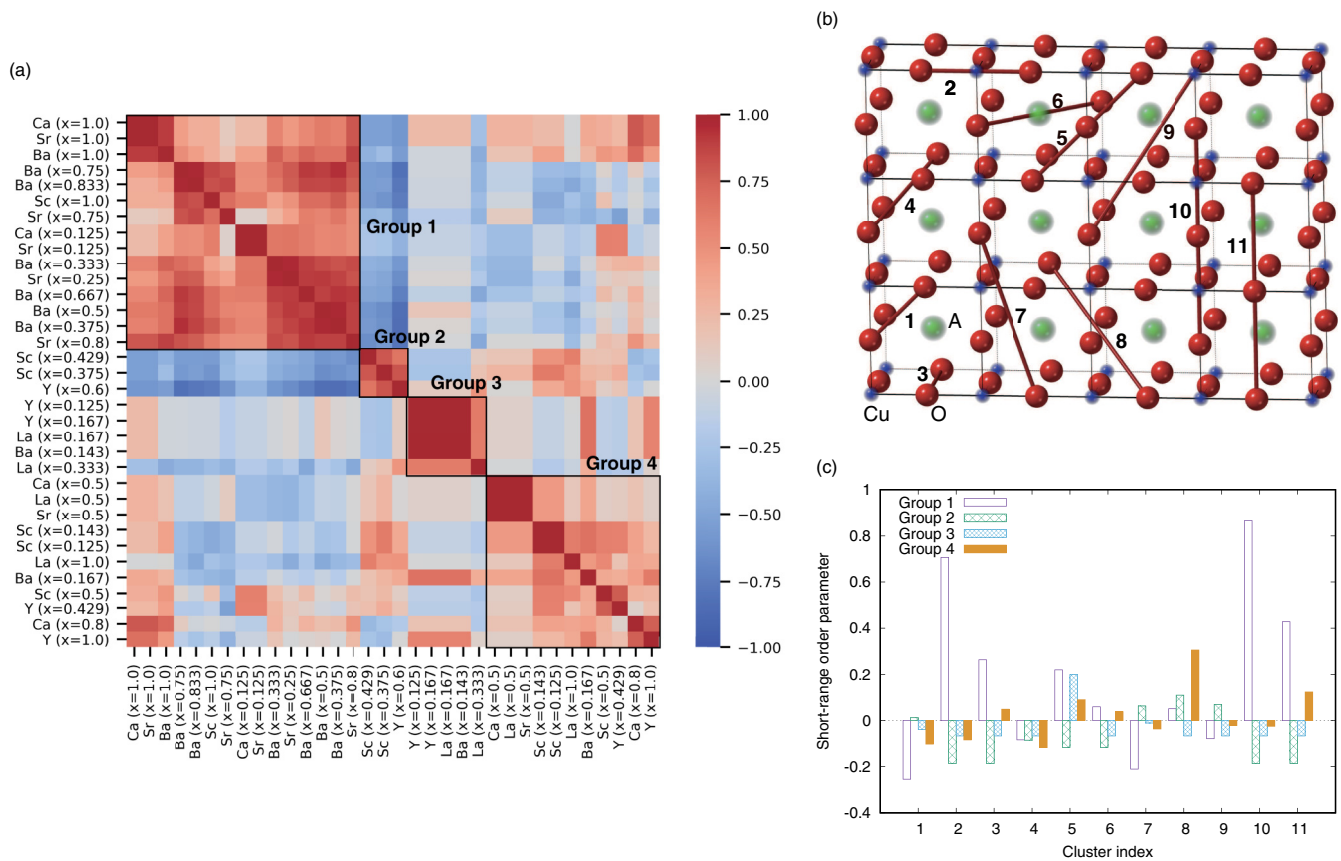
FIG. 6. (a) Pearson correlation coefficients between oxygen-deficient perovskite structures on the convex hull shown by the solid circles in Fig. 5. (b) Pair clusters used in the CE method and Bayesian optimization. (c) Dependence of WC-SRO parameters on the group classified by a hierarchical clustering. A bar shows the WC-SRO parameter averaged over the oxygen-deficient perovskite structures belonging to each group. The cluster indexes correspond to those of pairs shown in (b).

group. Group 1 shows positive large values of the WC-SRO parameters for the second, tenth, and eleventh pairs; hence, most of those pairs are composed of the same species in the structures of group 1. This indicates that the structures in group 1 consist mainly of planar fourfold copper units. An example is $SrCuO_2$ with the p-2 structure illustrated in Fig. 5, where all copper atoms are planar fourfold units. Also, group 4 is characterized by a positive value of the WC-SRO parameter of the nearest pair along the $\langle 111 \rangle$ direction (the eighth pair). This means that $\langle 111 \rangle$ nearest pairs of vacancies are components of the structures in group 4. An example is $SrCuO_{2.5}$ with the p-1 structure illustrated in Fig. 5.

Note that we evaluate WC-SRO parameters of the oxygen-deficient perovskite structures with the ideal cubic perovskite lattice. Some of the structures optimized by the DFT calculation significantly differ from their ideal structures.

F. Stability at 10 GPa

As shown in Sec. IV C, most of the oxygen-deficient perovskite structures are not stable at 0 GPa. On the other hand, oxygen-deficient perovskite compounds such as $CaCuO_2$, $SrCuO_2$, and $SrCuO_{2.5}$ have been synthesized under high-pressure conditions [6,47,48]. Therefore, we investigate the

pressure dependence of the phase stability of the oxygen-deficient perovskite structures. We restrict the DFT calculation to the experimental prototype structures and the oxygen-deficient perovskite structures satisfying the condition that their energies measured from the convex hull of the oxygen-deficient perovskite structures are less than 0.1 eV/$ACuO_{3-x}$.

Figure 7(a) shows the formation enthalpy distribution of the experimental structures and the oxygen-deficient perovskite structures at 10 GPa. Contrary to the phase stability at 0 GPa, all of the experimental oxygen-deficient perovskite structures, namely, $CaCuO_2$, $SrCuO_2$, $SrCuO_{2.5}$, $LaCuO_{2.5}$, and $LaCuO_{2.6}$, are stable at 10 GPa. In addition, the convex hull indicates that some of the oxygen-deficient perovskite structures are stable at 10 GPa in $SrCuO_{3-x}$, $BaCuO_{3-x}$, and $LaCuO_{3-x}$. Figure 7(b) shows the crystal structures of the oxygen-deficient perovskites with the composition $x \geqslant 0.4$ on or almost on the convex hull, where only $LaCuO_{2.6}$ (p-7) is experimentally reported [7]. The crystal structures with the other compositions are also shown in Appendix B. In $CaCuO_{3-x}$, $ScCuO_{3-x}$, and $YCuO_{3-x}$, the structures on the convex hull at 10 GPa are almost the same as those at 0 GPa. However, some oxygen-deficient perovskite structures such as $CaCuO_{2.5}$ and $YCuO_{2.571}$ are close to the convex hull. They are expected to be stable at higher pressures.
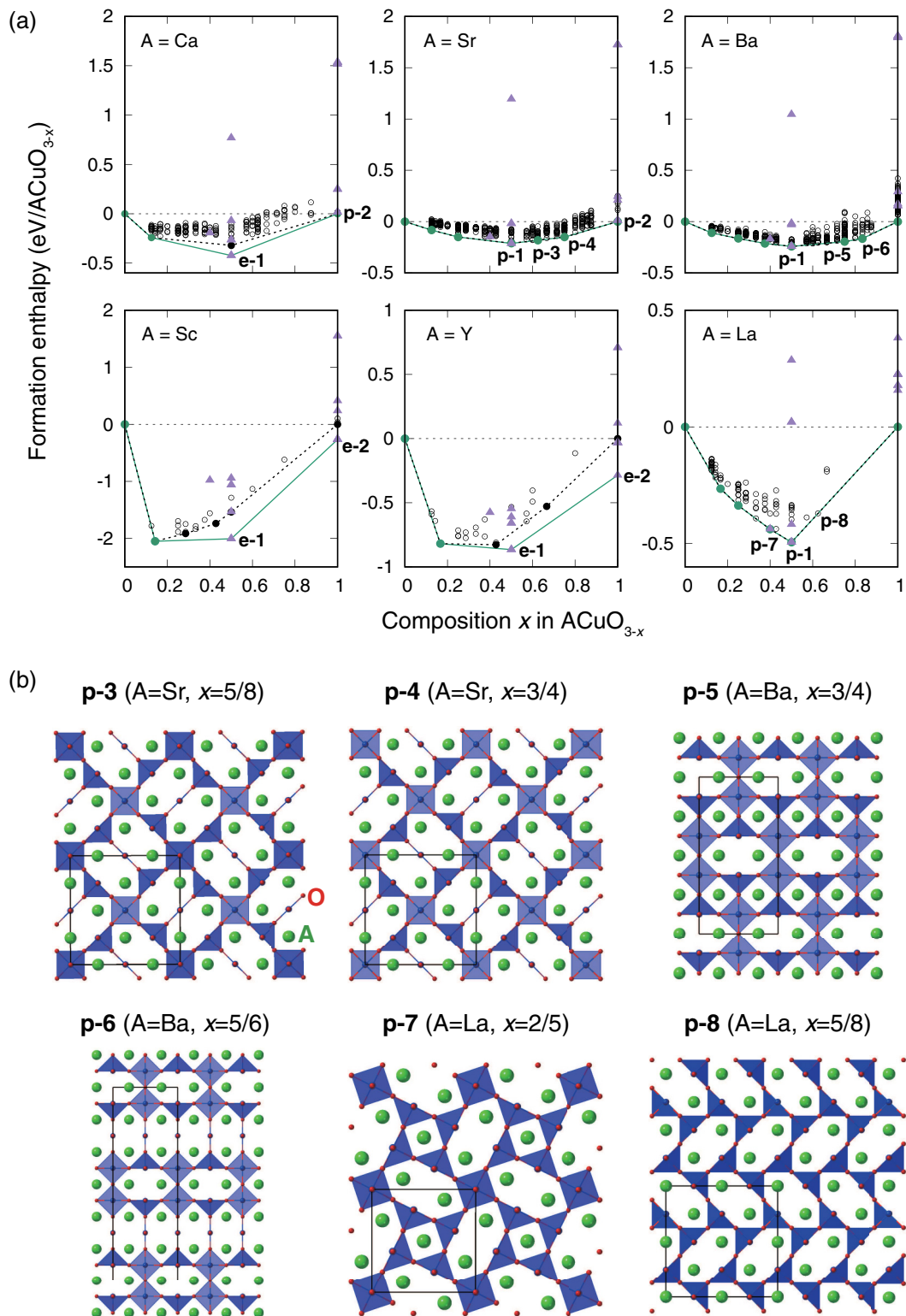
FIG. 7. (a) Formation enthalpy distribution at 10 GPa of the oxygen-deficient perovskite structures satisfying the condition that their energies measured from the convex hull of the oxygen-deficient perovskite structures are less than 0.1 eV/$ACuO_{3-x}$. The dashed black line and solid circles show the convex hull of the oxygen-deficient perovskite structures and its vertices, respectively. The purple closed triangles show the formation enthalpy values of experimental prototype structures. The green line shows the convex hull obtained from the combination of the oxygen-deficient perovskite structures and the experimental prototype structures. (b) Crystal structures of the oxygen-deficient perovskites with composition $x \geqslant 0.4$ on the convex hull or almost on the convex hull. Only $LaCuO_{2.6}$ (p-7) is experimentally reported [7]. It has a crystal structure type identical to that of $SrMnO_{2.6}$ [49].

As can be seen in Fig. 7, the oxygen configuration in the predicted structure of $SrCuO_{2.375}$ (p-3) is similar to that in the experimental structure of $La_{0.5}Ca_{0.5}CuO_{2.5}$ [50]. The difference between them is the Cu-O polyhedron located at the center of the unit cell; that is, they are planar fourfold and octahedral sixfold structures, respectively. The predicted structure of $SrCuO_{2.25}$ (p-4) is also similar to the structures of $SrCuO_{2.375}$ (p-3) and $La_{0.5}Ca_{0.5}CuO_{2.5}$. At the same time, since $La_{0.5}Ca_{0.5}CuO_{2.25}$ [50] and $La_{0.25}Sr_{0.75}CuO_{2.25}$ [51] are reported as existing compounds, it seems to be natural to compare $SrCuO_{2.25}$ (p-4) with them. However, their crystal structures remain unclarified. The results in Ref. [51] reveal only that $La_{0.25}Sr_{0.75}CuO_{2.25}$ has a space group of $P42_12$ or $P\bar{4}2_1m$. Nevertheless, they are subgroups of the space group $P4/mbm$ of $SrCuO_{2.25}$ (p-4) with index 2. Hence, the oxygen configuration in $SrCuO_{2.25}$ (p-4) may be identical to those in $La_{0.5}Ca_{0.5}CuO_{2.25}$ and $La_{0.25}Sr_{0.75}CuO_{2.25}$.

In addition, we compare the oxygen-deficient structures of $BaCuO_{2.333}$ with that of $YBa_2Cu_3O_{7-d}$ (YBCO) because the present structure data set contains the oxygen configuration of YBCO. As shown in Appendix B (Fig. 9), although the YBCO-like structure cannot be found as an oxygen-deficient structure with the lowest enthalpy at $x = 2/3$ for $BaCuO_{3-x}$, the formation enthalpy of the YBCO-like structure is only 52 meV/$BaCuO_{3-x}$ larger than that of the lowest formation enthalpy. This implies that the formation of the YBCO-type structure consisting of the alternate stacking of Cu-O chains and Cu-O double layers is dominated not by the cation ordering but by the oxygen vacancy ordering.

G. Pressure dependence of stability

Finally, we demonstrate the pressure dependence of the phase stability between the oxygen-deficient perovskite structures and the experimental structures. For each structure, the pressure-enthalpy curve is obtained from the energy-volume curve fitted to the Vinet equation of state [52] using the DFT energies of ten structures constructed by expansions and contractions of the equilibrium structure. Figure 8 shows the pressure-enthalpy curves of the oxygen-deficient perovskite structure (p-1) and $Sc_2Cu_2O_5$-type structure (e-1) in $SrCuO_{2.5}$. As can be seen in Fig. 8, the oxygen-deficient perovskite structure becomes stable as the pressure increases. It is found that the transition pressure from the $Sc_2Cu_2O_5$-type structure to the oxygen-deficient perovskite p-1 structure is 5.6 GPa in $SrCuO_{2.5}$. Similarly, $LaCuO_{2.5}$ involves the phase transition from the $Sc_2Cu_2O_5$-type structure to the $Nd_2Cu_2O_5$-type structure at 1.9 GPa and the phase transition from the $Nd_2Cu_2O_5$-type structure to the oxygen-deficient perovskite p-1 structure at 4.4 GPa.

V. CONCLUSION

The stability of oxygen-deficient perovskite structures in $ACuO_{3-x}$ (A = Ca, Sr, Ba, Sc, Y, La) has been investigated systematically using a combination of the DFT calculation,
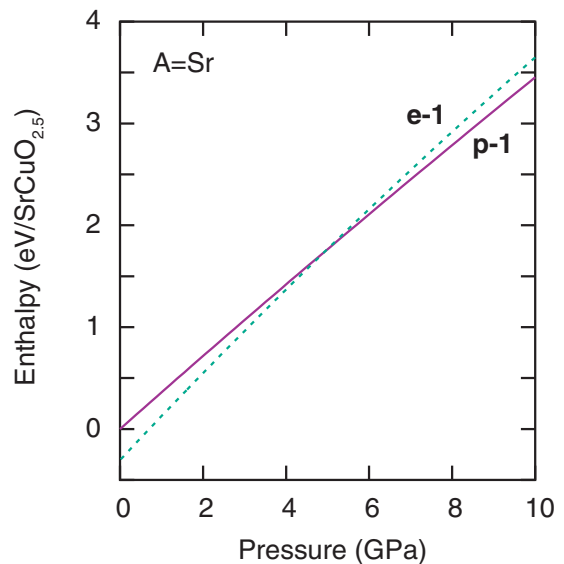
FIG. 8. Pressure-enthalpy curves of the oxygen-deficient perovskite structure (p-1) and $Sc_2Cu_2O_5$-type structure (e-1) in $SrCuO_{2.5}$.

the CE method, GP, and Bayesian optimization. We have successfully reproduced the reported oxygen-deficient perovskite structures and found a series of stable oxygen-deficient structures that are not reported in the literature. Our results indicate that high-pressure synthesis is advantageous for obtaining the oxygen-deficient perovskite structures. By adopting the machine learning clustering for the obtained structures, we found that they can be classified into four groups and one of the largest groups shows a common structural feature characterized by the square planar coordination for copper.

Regarding the computational aspects of structure prediction, the CE method fails to derive an accurate model for the formation energy of the oxygen-deficient structure owing to its substantial geometry relaxation from the ideal cubic perovskite lattice. On the other hand, the combination of the CE method, GP model, and Bayesian optimization is very useful for predicting the oxygen-deficient perovskite structures. This procedure is also applicable in a straightforward manner to a wide range of atom-deficient structures involving extensive geometry relaxation from their ideal lattice.

APPENDIX A: HERMITE NORMAL FORMS

Table III shows a complete set of nonequivalent HNFs with up to $n = 8$ for the cubic perovskite lattice.

TABLE III. Complete set of nonequivalent HNFs for the cubic perovskite lattice.

Determinant of HNF	Nonequivalent HNF

$n = 1$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$n = 2$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 2 \end{pmatrix}$$

$n = 3$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 3 \end{pmatrix}$$

$n = 4$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$

$n = 5$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 2 & 5 \end{pmatrix}$$

$n = 6$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 2 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 3 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 2 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 3 & 6 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 3 & 6 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

$n = 7$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 7 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 7 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 7 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 2 & 7 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 3 & 7 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 3 & 7 \end{pmatrix}$$

$n = 8$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 2 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 3 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 4 & 8 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 2 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 3 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 4 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 4 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 4 & 8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 2 & 4 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 2 & 0 & 4 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$
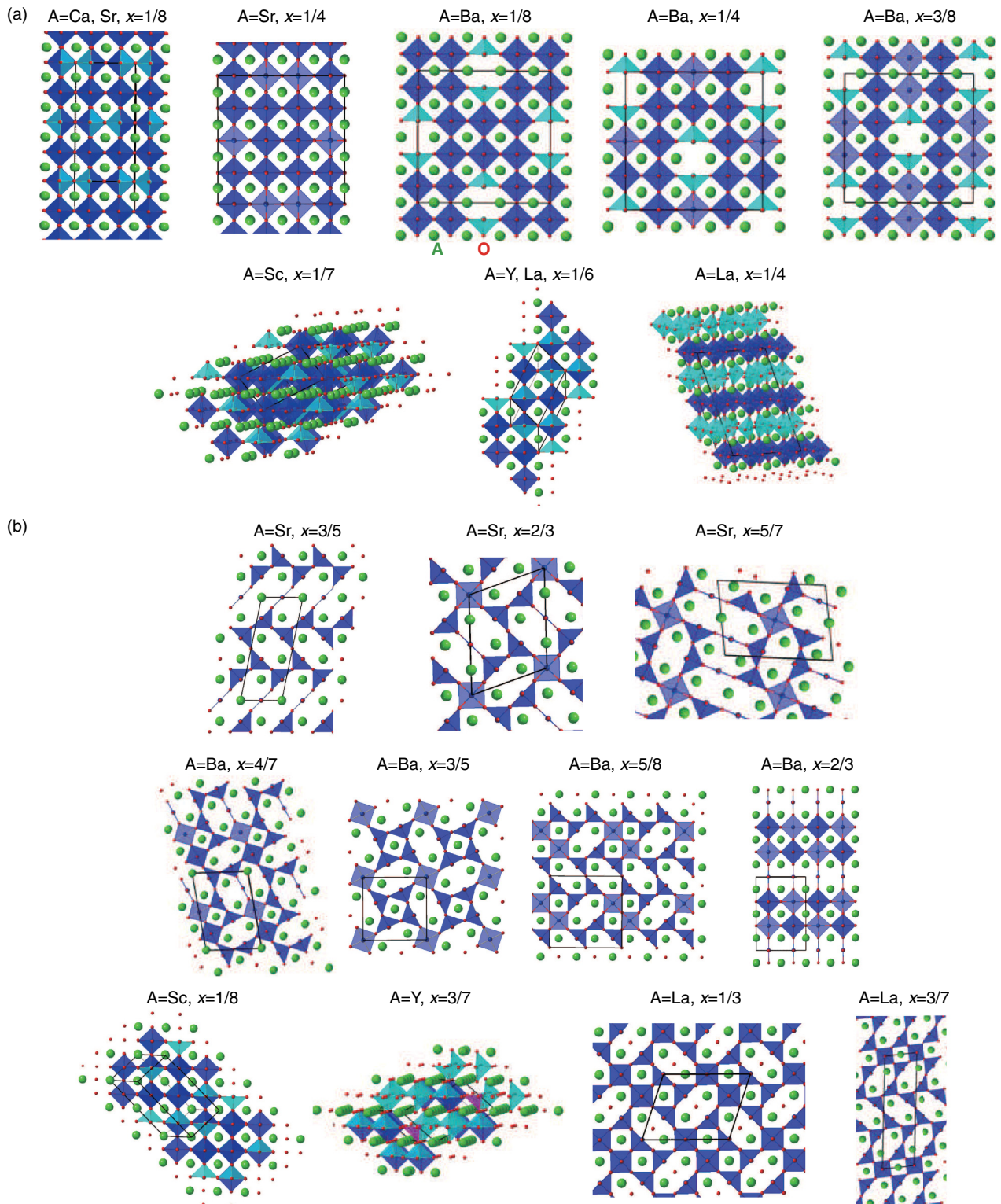
FIG. 9. Crystal structures of the oxygen-deficient perovskites (a) on the convex hull and (b) almost on the convex hull at 10 GPa except those shown in Figs. 5 and 7. The dark blue, sky blue, light blue, and pink polyhedrons show sixfold, fivefold, planer fourfold, and tetrahedral fourfold Cu-O polyhedrons.

APPENDIX B: OXYGEN-DEFICIENT PEROVSKITE STRUCTURES ON CONVEX HULL

Figure 9 shows the crystal structures of oxygen-deficient perovskites on or almost on the convex hull at 10 GPa.

[1] M. T. Anderson, J. T. Vaughey, and K. R. Poeppelmeier, Chem. Mater. **5**, 151 (1993).

[2] S. Stølen, E. Bakken, and C. E. Mohn, Phys. Chem. Chem. Phys. **8**, 429 (2006).

[3] R. J. Tilley, *Perovskites: Structure-Property Relationships* (Wiley, Hoboken, NJ, 2016).

[4] R. J. Cava, B. Batlogg, R. B. van Dover, D. W. Murphy, S. Sunshine, T. Siegrist, J. P. Remeika, E. A. Rietman, S. Zahurak, and G. P. Espinosa, Phys. Rev. Lett. **58**, 1676 (1987).

[5] S. Ishiwata, D. Kotajima, N. Takeshita, C. Terakura, S. Seki, and Y. Tokura, J. Phys. Soc. Jpn. **82**, 063705 (2013).

[6] B.-H. Chen, D. Walker, B. A. Scott, and D. B. Mitzi, J. Solid State Chem. **121**, 498 (1996).

[7] J. Bringley, B. Scott, S. La Placa, R. Boehme, T. Shaw, M. McElfresh, S. Trail, and D. Cox, Nature (London) **347**, 263 (1990).

[8] A. A. Emery and C. Wolverton, Sci. Data **4**, 170153 (2017).

[9] J. M. Sanchez, F. Ducastelle, and D. Gratias, Physica A (Amsterdam, Neth.) **128**, 334 (1984).

[10] D. de Fontaine, Solid State Phys. **47**, 33 (1994).

[11] F. Ducastelle, *Order and Phase Stability in Alloys* (Elsevier, Amsterdam, 1994).

[12] A. Van der Ven, M. K. Aydinol, G. Ceder, G. Kresse, and J. Hafner, Phys. Rev. B **58**, 2975 (1998).

[13] G. L. W. Hart and A. Zunger, Phys. Rev. Lett. **87**, 275508 (2001).

[14] M. E. Arroyo y de Dompablo, A. Van der Ven, and G. Ceder, Phys. Rev. B **66**, 064112 (2002).

[15] A. Seko, A. Togo, F. Oba, and I. Tanaka, Phys. Rev. Lett. **100**, 045702 (2008).

[16] A. Predith, G. Ceder, C. Wolverton, K. Persson, and T. Mueller, Phys. Rev. B **77**, 144104 (2008).

[17] H. Das, A. Urban, W. Huang, and G. Ceder, Chem. Mater. **29**, 7840 (2017).

[18] F. Zhou, T. Maxisch, and G. Ceder, Phys. Rev. Lett. **97**, 155704 (2006).

[19] D. Jones, J. Global Optim. **21**, 345 (2001).

[20] A. Seko, T. Maekawa, K. Tsuda, and I. Tanaka, Phys. Rev. B **89**, 054303 (2014).

[21] A. Seko, A. Togo, H. Hayashi, K. Tsuda, L. Chaput, and I. Tanaka, Phys. Rev. Lett. **115**, 205901 (2015).

[22] K. Toyoura, D. Hirano, A. Seko, M. Shiga, A. Kuwabara, M. Karasuyama, K. Shitara, and I. Takeuchi, Phys. Rev. B **93**, 054112 (2016).

[23] D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue, and T. Lookman, Nat. Commun. **7**, 11241 (2016).

[24] S. Kiyohara, H. Oda, T. Miyata, and T. Mizoguchi, Sci. Adv. **2**, e1600746 (2016).

[25] T. Yamashita, N. Sato, H. Kino, T. Miyake, K. Tsuda, and T. Oguchi, Phys. Rev. Mater. **2**, 013803 (2018).

[26] K. Kanamori, K. Toyoura, J. Honda, K. Hattori, A. Seko, M. Karasuyama, K. Shitara, M. Shiga, A. Kuwabara, and I. Takeuchi, Phys. Rev. B **97**, 125124 (2018).

[27] M. Todorović, M. U. Gutmann, J. Corander, and P. Rinke, npj Comput. Mater. **5**, 35 (2019).

[28] M. S. Jørgensen, U. F. Larsen, K. W. Jacobsen, and B. Hammer, J. Phys. Chem. A **122**, 1504 (2018).

[29] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2006).

[30] A. R. Natarajan and A. Van der Ven, npj Comput. Mater. **4**, 1 (2018).

[31] G. L. W. Hart and R. W. Forcade, Phys. Rev. B **77**, 224115 (2008).

[32] G. L. W. Hart and R. W. Forcade, Phys. Rev. B **80**, 014120 (2009).

[33] A. Seko, Y. Koyama, and I. Tanaka, Phys. Rev. B **80**, 165122 (2009).

[34] A. Seko, CLUPAN, http://sourceforge.net/projects/clupan.

[35] L. J. Nelson, G. L. W. Hart, F. Zhou, and V. Ozoliņš, Phys. Rev. B **87**, 035125 (2013).

[36] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. (Springer, New York, 2009).

[37] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[38] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[39] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).

[40] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).

[41] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[42] D. Vaknin, E. Caignol, P. K. Davies, J. E. Fischer, D. C. Johnston, and D. P. Goshorn, Phys. Rev. B **39**, 9122 (1989).

[43] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton, Phys. Rev. B **57**, 1505 (1998).

[44] L. Wang, T. Maxisch, and G. Ceder, Phys. Rev. B **73**, 195107 (2006).

[45] A. van de Walle and G. Ceder, J. Phase Equilib. **23**, 348 (2002).

[46] T. Mueller and G. Ceder, Phys. Rev. B **82**, 184107 (2010).

[47] J. Karpinski, H. Schwer, I. Mangelschots, K. Conder, A. Morawski, T. Lada, and A. Paszewin, Physica C (Amsterdam, Neth.) **234**, 10 (1994).

[48] M. Takano, Y. Takeda, H. Okada, M. Miyamoto, and T. Kusaka, Physica C (Amsterdam, Neth.) **159**, 375 (1989).

[49] L. Suescun, O. Chmaissem, J. Mais, B. Dabrowski, and J. D. Jorgensen, J. Solid State Chem. **180**, 1698 (2007).

[50] A. M. Guloy, B. A. Scott, and R. A. Figat, J. Solid State Chem. **113**, 54 (1994).

[51] W. Fu, Q. Xu, A. Verheijen, J. van Ruitenbeek, H. Zandbergen, and L. de Jongh, Solid State Commun. **73**, 291 (1990).

[52] P. Vinet, J. Ferrante, J. H. Rose, and J. R. Smith, J. Geophys. Res. **92**, 9319 (1987).