Explainable active learning in investigating structure–stability of $SmFe_{12-\alpha-\beta}X_\alpha Y_\beta$ structures X, Y = {Mo, Zn, Co, Cu, Ti, Al, Ga}

Duong-Nguyen Nguyen,* Hiori Kino, Takashi Miyake, and Hieu-Chi Dam*

In this article, we propose a query-and-learn active learning approach combined with first-principles calculations to rapidly search for potentially stable crystal structure via elemental substitution, to clarify their stabilization mechanism, and integrate this approach to $SmFe_{12}$-based compounds with $ThMn_{12}$ structure, which exhibits prominent magnetic properties. The proposed method aims to (1) accurately estimate formation energies with limited first-principles calculation data, (2) visually monitor the progress of the structure search process, (3) extract correlations between structures and formation energies, and (4) recommend the most beneficial candidates of $SmFe_{12}$-substituted structures for the subsequent first-principles calculations. The structures of $SmFe_{12-\alpha-\beta}X_\alpha Y_\beta$ before optimization are prepared by substituting X, Y elements—Mo, Zn, Co, Cu, Ti, Al, Ga—in the region of $\alpha + \beta < 4$ into iron sites of the original $SmFe_{12}$ structures. Using the optimized structures and formation energies obtained from the first-principles calculations after each active learning cycle, we construct an embedded two-dimensional space to rationally visualize the set of all the calculated and not-yet-calculated structures for monitoring the progress of the search. Our machine learning model with an embedding representation attained a prediction error for the formation energy of 1.25×10^{-2} (eV/atom) and required only one-sixth of the training data compared to other learning methods. Moreover, the time required to recall most potentially stable structures was nearly four times faster than the random search. The formation energy landscape visualized using the embedding representation revealed that the substitutions of Al and Ga have the highest potential to stabilize the $SmFe_{12}$ structure. In particular, $SmFe_9[Al/Ga]_2Ti$ showed the highest stability among the investigated structures. Finally, by quantitatively measuring the change in the structures before and after optimization using OFM descriptors, the correlations between the coordination number of substitution sites and the resulting formation energy are revealed. The negative-formation-energy-family $SmFe_{12-\alpha-\beta}[Al/Ga]_\alpha Y_\beta$ structures show a common trend of increasing coordination number at substituted sites, whereas structures with positive formation energy show a corresponding decreasing trend.

Impact statement

Seeking the next generation of high-performance magnets is a crucial demand for replacing the widely accepted Nd-Fe-B magnets developed in the middle 80s. The iron-rich compounds with the original tetragonal $ThMn_{12}$ structure appear as the most potential candidates except for the hard synthesizing it in nature due to its high energy of formation. Stabilization for this material system is expected by substituting new elements, but the vast number of possible structures makes the exploration difficult even for theoretical calculations. This article proposes an integration of first-principles calculations and explainable active learning to efficiently explore the crystal structure space of this material system. In particular, the explored crystal structure space can be rationally visualized, on which the relationship between substitution elements, substitution sites, and crystal structure stabilization can be intuitively interpreted.

Duong-Nguyen Nguyen, School of Knowledge Science, Japan Advanced Institute of Science and Technology, Nomi, Ishikawa, Japan; nguyennd@jaist.ac.jp
Hiori Kino, Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, Tsukuba, Ibaraki, Japan
Takashi Miyake, Research Center for Computational Design of Advanced Functional Materials, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, Japan
Hieu-Chi Dam, School of Knowledge Science, Japan Advanced Institute of Science and Technology, Nomi, Ishikawa, Japan; International Center for Synchrotron Radiation Innovation Smart, Tohoku University, Sendai, Miyagi, Japan; dam@jaist.ac.jp
*Corresponding author

Introduction

Finding functional and useful crystals or molecular structures is highly challenging, and numerous methods have been proposed.[1–6] Even when the crystal structures are limited to those for which prototypes are known, the enormous number of possible substitutions of elements leads to considerable difficulty in characterizing the function of physical properties associated with these structures. Within the vast pool of possible crystals or molecular structures, researchers frequently confined their explorations to a looping search set, formulating hypotheses about structures of interest and then verifying these hypotheses by validating physical properties using experiments or theoretical calculations. Naively, this approach is closely associated with the trial-and-error problem-solving method with solution-oriented and problem-specific features.

Solution-oriented approaches seeking to find optimal solutions invest relatively little effort to reveal why and how the solution is found (e.g., optimal structures in materials discovery). Besides, problem-specific approaches make little effort to generalize the solution to different problem scopes (e.g., extending the structure search space). Therefore, these approaches inevitably involve limitations when deployed in a screening space with wider bounds than the known material structures.

In this study, we propose a query-and-learn architecture based on active learning to assist researchers in actively monitoring the material structure discovery process. The query-and-learn method aims to (1) accurately estimate physical properties from the most limited first-principles data, (2) accelerate the search for outstanding structures, (3) interpret the structure search process, and (4) generalize findings by extracting the structure–property correlations. The problem regarding the formation mechanism of SmFe$_{12}$-based compounds with the ThMn$_{12}$ structure is used to demonstrate the effectiveness of the proposed method.

The original structure of iron-rich SmFe$_{12}$ compounds were first discovered in the late 1980s.[9–11] It was expected that they would show high saturation magnetization, magnetocrystalline anisotropy, and Curie temperature.[12] However, SmFe$_{12}$ and other families of RFe$_{12}$ with R denoting a rare earth element have not been widely adopted to produce the excellent magnets that can be obtained owing to the practical difficulty of stabilizing the material. Numerous studies have substituted elements such as Co, Ti, V, Cr, Mo, W, or Ga to obtain a stable ThMn$_{12}$-type phase.[13–19] Unrestricted from ternary compounds, recently researchers have emphasized searching for the most potentially stable SmFe$_{12}$-based quaternary compounds using the bi-element substitution method.[20–27] Because the stabilizing elements are assumed to be substituted at the Fe sites, a large supercell of SmFe$_{12}$ should be considered as a host structure to investigate substitution structures with the possibility of diverse elemental substitutions. Therefore, a more efficient methodology to investigate the structure space, where the number of candidates increases combinatorially, is urgently required.

Figure 1 summarizes key components in the query-and-learn active learning design in discovering formable SmFe$_{12}$-based compounds in the ThMn$_{12}$ structure. At the beginning of the query step, a pool of not-yet-calculated structures is created by applying substitution operators on the prototype of SmFe$_{12}$. The system queries the most informative candidates to estimate their properties before updating them to the training data of machine learning predictors. Canonically, the informativeness of queried structures is assumed to show the most significant impact to improve the accuracy of the prediction model. However, the predictive ability term is usually challenging to clarify explicitly because predictive evaluations often lack information on the relative position among new queried-training–testing data. For example, authors in References 28 and 29 reported exploration strategies by assuming out-of-distribution structures as superior structures. Therefore, querying then accurately predicting structures in the out-of-distribution region are on the top demand rather than the task of predicting properties for all not-yet-calculated structures in the pool. Furthermore, the methods by which the estimator inferred the predicted value and the learned function changed by adding queried data are often blind to researchers' monitoring the discovery process. In the learn step of the query-and-learn design, we extend the prediction model's interpretability by introducing metric learning in transforming the original structure representation vector into a low-dimensional space, which preserves the smoothness of the function of formation energy. Consequently, information in the structure search progress can be actively monitored including prediction accuracy; features of the learned model, regions of outstanding structures, or inter-correlations between query structures with training structures. Studies of active learning designs used in materials science are shown in References 28 and 30–32, besides other machine learning-assisted material designs shown in References 33–35.

The contributions of this work are summarized as follows:

- We investigate systematically the formation energy and magnetization of 3307 SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ with X, Y as Mo, Zn, Co, Cu, Ti, Al, and Ga, limited by $\alpha + \beta < 4$ using the VASP calculation procedure from OQMD.[36]
- We confirm that SmFe$_9$[Al/Ga]$_2$Ti structures have the highest stability and SmFe$_9$Co$_3$ structures have optimal magnetization value.
- We confirm that the SmFe$_{12-\alpha-\beta}$[Al/Ga]$_\alpha$Y$_\beta$ structures show on average negative formation energies and an increase in the coordination number at substituted sites (Al/Ga), whereas other families showed opposite trends.
- We propose an active learning design with embedding representation of orbital-field matrix that achieves an optimal prediction accuracy and recalls outstanding structures using limited training data.
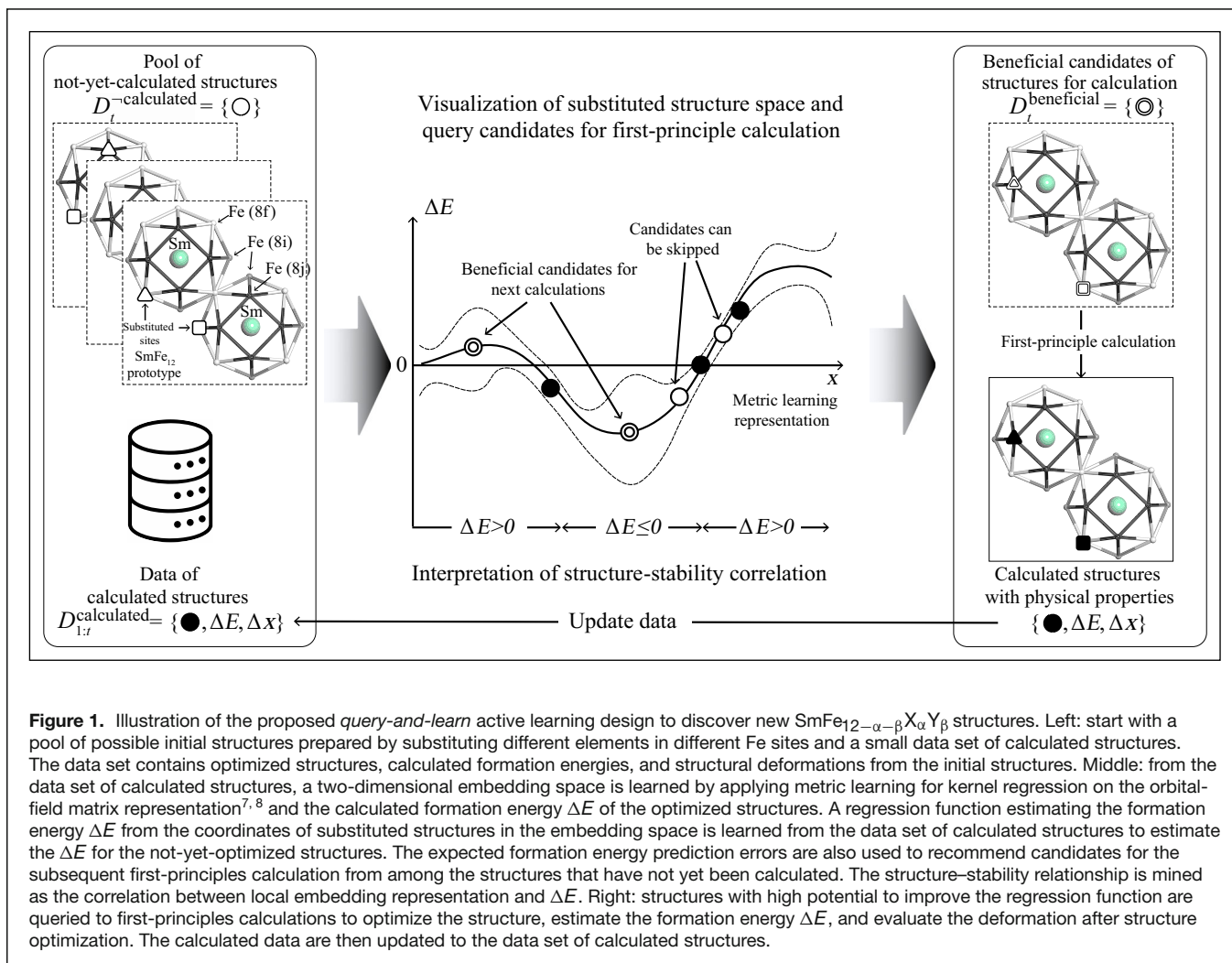
Figure 1. Illustration of the proposed *query-and-learn* active learning design to discover new SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ structures. Left: start with a pool of possible initial structures prepared by substituting different elements in different Fe sites and a small data set of calculated structures. The data set contains optimized structures, calculated formation energies, and structural deformations from the initial structures. Middle: from the data set of calculated structures, a two-dimensional embedding space is learned by applying metric learning for kernel regression on the orbital-field matrix representation[7, 8] and the calculated formation energy ΔE of the optimized structures. A regression function estimating the formation energy ΔE from the coordinates of substituted structures in the embedding space is learned from the data set of calculated structures to estimate the ΔE for the not-yet-optimized structures. The expected formation energy prediction errors are also used to recommend candidates for the subsequent first-principles calculation from among the structures that have not yet been calculated. The structure–stability relationship is mined as the correlation between local embedding representation and ΔE. Right: structures with high potential to improve the regression function are queried to first-principles calculations to optimize the structure, estimate the formation energy ΔE, and evaluate the deformation after structure optimization. The calculated data are then updated to the data set of calculated structures.

- We extract a relationship of bi-elements substitution to the stability, that is, SmFe$_{12-\alpha-\beta}$[Al/Ga/Ti]$_\alpha$Y$_\beta$ is potentially stable, and SmFe$_{12-\alpha-\beta}$[Mo]$_\alpha$Y$_\beta$ is potentially unstable, which can be interpreted using the embedding representation.

In the following sections, we will explain the proposed approach in detail, and use it for finding potentially stable SmFe$_{12}$-based compounds. The exploration space for discovering potentially stable SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ structures is set with X and Y as Mo, Zn, Co, Cu, Ti, Al, and Ga, limited by $\alpha + \beta < 4$, where α and β are integers. We will demonstrate the efficiency of the proposed approach, and show how to extract information associated with structural stability. Details of the data preparation are shown in the "First-principles calculation" section. The "Active learning design" section presents the components of the active learning architecture in detail. Last, the "Experiment and discussion" section shows the performance of active learning designs and the results of interpreting correlations extracted from the embedding space regarding the formation energy.

First-principles calculation

Creation of SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ structures

We focus on SmFe$_{12}$-based crystalline magnetic materials under the formula SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ with X and Y as the substituted elements from Mo, Zn, Co, Cu, Ti, Al, and Ga; α and β are integer numbers of X and Y compositions, respectively. A hypothetical-not-yet-calculated structure is created by substituting α iron sites with the element X and β iron sites with the element Y. There are numerous possible hypothetical structures; hence, we limit our investigation to $\alpha + \beta < 4$. Owing to the symmetrical properties of the iron sites in the host SmFe$_{12}$ structure, new substituted structures were compared with one another to remove duplications. We followed the comparison procedure proposed by qmpy, a Python application programming interface of OQMD.[36] The internal coordinates of the structures were compared by examining all rotations allowed by each lattice and searching for rotations and translations to map the atoms of the same species into one another within a given level of tolerance. Here, any two structures with a percent deviation in lattice parameters and angles smaller than

0.1 were considered identical. Furthermore, we applied our designed orbital-field matrix (OFM)[7,8] to eliminate duplication. Notably, two structures were considered the same when the L_2 norm of the OFM difference was less than 10^{-3}.

To initialize the active learning model data set, we substituted one atom from Mo, Zn, Co, Cu, Ti, Al, and Ga to one iron site of the SmFe$_{12}$ host structure. Consequently, there were 283 structures under the formula SmFe$_{12-\alpha}$X$_\alpha$ with $\alpha \in \{1, 2, 3\}$. By substituting two elements, we created 3024 structures using the formula SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ with $\alpha + \beta < 4$. We used this data set as an initial of not-yet-calculated data set $\mathcal{D}_1^{\neg\text{calculated}}$; a detailed description is provided in the "Data set notation" section. To rephrase, this data set is considered a screening space/exploration space for the exploration process; we retain all these unknown structures as distinct from the initial space. Subsequently, all structures were subjected to structural optimization through first-principles calculations to obtain the optimal structures.

Assessment of formation energy of structures

The first-principles calculations using density functional theory (DFT)[37,38] are among the most practical calculation methods used in materials science. DFT calculations precisely estimate the total energy of the materials, which can be used to determine the formation energy of the substituted structure. The formation energy of a given structure s is defined as follows:

$$\Delta E[s] = \frac{1}{N}(E[s] - \Sigma_i^N E[s_i]), \qquad 1$$

where $\Delta E[s]$, $E[s]$, and $E[s_i]$ are the formation energy, total energy of structure s per formula unit, and simple substance s_i per atom, respectively. Finally, N is the total number of atoms in the formula unit of s. The simple substances were chosen as (1) Im-$3m$ with Fe and Mo, (2) R-$3m$ with Sm and Al, (3) Fm-$3m$ with Cu and Co, (4) $P6/mmm$ with Ti, and (5) $P63/mmc$ with Zn. Details of the substances chosen are provided in the Supplementary Information. A structure whose formation energy lies below or lower than zero, that is, $\Delta E \leq 0$, is a potentially formable material in nature, whereas a structure associated with $\Delta E > 0$ could be considered unstable. For the competing phases, the stability of the structure should be discussed using the hull distance. In this study, we use the formation energy defined in Equation 1 as an index for simplicity. The relationship between the experimental material and the hull distance at $T = 0$ K has been summarized in References 39 and 40. The stability of the magnets at finite temperature can be found in Reference 41. We discuss in detail the reliability of this calculation in the Supplementary Information.

In this study, we follow the computational settings of OQMD[36,42] to estimate the total energy of all structures. The calculations were performed using the Vienna *ab initio* simulation package (VASP)[43,44] by utilizing the projector-augmented wave method potentials[45,46] and the Perdew–Burke–Ernzerhof[47] exchange–correlation functional. Pseudopotentials used in this work were collected from POTCAR library version 5.4 of VASP.[45,48–51] With the 4f element of Sm, Sm^{3+} potentials were applied where five electrons in f shell were treated as core electrons. Details of potential for other elements is shown in the Supplemental Information, with notation as shown in Reference 49.

All calculations were spin-polarized with the ferromagnetic alignment of the spins. For a given structure, we performed three optimization steps following the coarse relax, fine relax, and standard procedures provided by OQMD. The k-points per reciprocal lattice for these calculation series were selected as 4000, 6000, and 8000 for coarse relax, fine relax, and standard, respectively. Optimal lattice parameters from the last step were used as the initial setting for the next step. We set 520 eV as the cutoff energy in the standard calculation step. The total energies of the final converged calculations were used to estimate the formation energy, ΔE.

In addition, the total magnetic moment of these materials $\mu[s]$ was recalculated because we used an open-core approximation to treat the 4f electrons of Sm, as follows:

$$\mu[s] = \Sigma_i m[s_i] + \Sigma_k J_{4f}g_{J_{4f}}[s_k], \qquad 2$$

where $m[s_i]$ is the magnetic moment of atom i, $J_{4f}g_{J_{4f}}[s_k]$ is the correction term with $g_{J_{4f}}$ as the Lande factor, and J_{4f} is the angular momentum of lanthanide s_k. Index i represents all atoms, and index k represents all lanthanide atoms in the structure. The contribution of the 4f electrons of Sm to the magnetization is $J_{g_J} = 0.714$. In this paper, this value is finally converted to magnetization per formula unit, M (T/f.u.).

Active learning design

There are three essential components in the proposed active learning approach, including (1) a pool $\mathcal{D}$ of not-yet-calculated structures (non-optimized) and first-principles calculated (optimized) structures, (2) an estimator E to predict the target formation energy, and (3) an acquisition function α to estimate the structures that should be queried in order of priority to enhance the prediction ability of E.

Data set notation

For a given query time t, we denote $\mathcal{D}_{[1:t]-1}^{\text{calculated}}$ as the data set comprising all the structures queried and optimized by first-principle calculation at the start of the query time t. We also denote $\mathcal{D}_t^{\neg\text{calculated}}$ as the data set with the remainder of not-yet-calculated structures at the start of the query time t. From $\mathcal{D}_t^{\neg\text{calculated}}$, we evaluate data sets $\mathcal{D}_t^{\text{beneficial}}$ such that by adding the calculated results of $\mathcal{D}_t^{\text{beneficial}}$ to $\mathcal{D}_{[1:t]-1}^{\text{calculated}}$ we can improve the prediction ability of the estimator E. $\mathcal{D}_t^{\text{beneficial}}$ is queried by the acquisition functions described in the "Acquisition function" section ($\mathcal{D}_t^{\text{beneficial}} \subset \mathcal{D}_t^{\neg\text{calculated}}$). To evaluate the ability of the active learning system to search potentially stable structures, we also collect $\mathcal{D}_t^{\text{outstanding}}$ from $\mathcal{D}_t^{\neg\text{calculated}}$ as a set of structures that are expected to be stable. Within the scope of finding the most potentially stable substituted SmFe$_{12}$ families, if the calculated or predicted formation energy ΔE is

smaller than $-0.1\,(\mathrm{eV/atom})$, the structure is considered potentially stable. At the time t of querying process, a predetermined number of structures with the lowest ΔE_{pred} predicted by $\mathtt{E}$ are then added to $\mathcal{D}_t^{\mathrm{outstanding}}$ for verification using first-principles calculations. First-principles calculations are then carried out for all the structures in $\mathcal{D}_t^{\mathrm{beneficial}}$, and the obtained optimized structures are added to $\mathcal{D}_{[1:t]-1}^{\mathrm{calculated}}$ to get $\mathcal{D}_{[1:t]}^{\mathrm{calculated}}$. We then use $\mathcal{D}_{[1:t]}^{\mathrm{calculated}}$ as the training data for learning the estimator $\mathtt{E}$. All the optimized structures confirmed using first-principles calculation with a formation energy lower below the specified limit are considered as potentially stable structures, and they are added to data set $\mathcal{D}_{\mathrm{confirmed}}^{\mathrm{outstanding}}$, which comprises all the potentially stable structures that are confirmed up to this point. The set of all the structures estimated using the estimator $\mathtt{E}$ as potentially stable structures is denoted by $\mathcal{D}_{\mathrm{estimated}}^{\mathrm{outstanding}}$. The pseudo-code summarization of the entire query-and-learn process is shown in the Supplemental Information.

In $\mathcal{D}_{[1:t]}^{\mathrm{calculated}}$, we represent calculated structures accumulating up to t with representation vectors as $\mathbf{x}_{[1:t]}$ and formation energy as $\mathbf{y}_{[1:t]}$. The formation energy of SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ structures is described in the "First-principles calculation" section. For simplicity, we denote $\mathbf{x}$ as a representation vector of not-yet-calculated structures, normal subscript denotes data point index, bracket subscript $[1:t]$ represent for collected data up to t, and superscript represents the index of feature. In this study, we applied the OFM[7,8] as a descriptor to represent all structures. In OFM representation, the most outer-shell electron configuration is set as a representation of each composition site. Details of OFM atomic representation is used in Element.electronic_structure in pymatgen[52] and the summary in Table I in the Supplemental Information. All elements in the OFM appear in the form of (u^i, u^j), which counts the average coordination number of neighbors u^j surrounding the center u^i. By representing each atom using outer-shell electron configuration, each individual matrix element is associated with one specific coordination number of a pair of elements in a given structure. Practical interpretation samples are shown in References 53 and 54. In this work, after removing features with zero in all structures, we finally required an 88-dimensional orbital-field vector to represent all SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ structures.

Gaussian process estimator

The Gaussian process estimator assumes that the joint distribution of the observed values $\mathbf{y}_{[1:t]}$ and predicted values $\hat{\mathbf{y}}$ follow the Gaussian prior distribution, expressed as follows:

$$\begin{bmatrix} \mathbf{y}_{[1:t]} \\ \hat{\mathbf{y}} \end{bmatrix} = \mathcal{N}\left(0, \begin{bmatrix} \kappa(\mathbf{x}_{[1:t]}, \mathbf{x}_{[1:t]})\kappa(\mathbf{x}_{[1:t]}, \mathbf{x}) \\ \kappa(\mathbf{x}, \mathbf{x}_{[1:t]})\kappa(\mathbf{x}, \mathbf{x}) \end{bmatrix}\right). \qquad 3$$

With these assumptions, the predicted values for the unknown state points follow the conditional distribution calculated by updating the prior probability distribution after observing the sampled state points. Thus, $\hat{\mathbf{y}} \approx \mathcal{N}(\boldsymbol{\mu}(\mathbf{x}), \boldsymbol{\sigma}(\mathbf{x}))$ with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}$ are estimated as

$$\boldsymbol{\mu}(\mathbf{x}) = \kappa(\mathbf{x}, \mathbf{x}_{[1:t]})\kappa(\mathbf{x}_{[1:t]}, \mathbf{x}_{[1:t]})^{-1}\mathbf{y}_{[1:t]}, \qquad 4$$

$$\boldsymbol{\sigma}(\mathbf{x}) = \kappa(\mathbf{x}_{[1:t]}, \mathbf{x}) \\ - \kappa(\mathbf{x}, \mathbf{x}_{[1:t]})\kappa(\mathbf{x}_{[1:t]}, \mathbf{x}_{[1:t]})^{-1}\kappa(\mathbf{x}_{[1:t]}, \mathbf{x}). \qquad 5$$

The mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}$ are the main components used to construct the acquisition functions, which are introduced in the "Acquisition function" section. The most conventional kernel, known as the Gaussian kernel κ_{ij} is defined as the kernel between $\mathbf{x}_i$ and $\mathbf{x}_j$ as follows:

$$\kappa_{ij} := \kappa(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{\epsilon\sqrt{2\pi}} e^{-\left[\frac{d(\mathbf{x}_i, \mathbf{x}_j)}{\epsilon}\right]^2}, \qquad 6$$

where ϵ is a hyperparameter that is tunable to learn the best form of the kernel and d is conventionally defined as the Euclidean distance.

Metric learning

Human intuition regarding the Euclidean distance among data points from three-dimensional spaces often does not apply to higher-dimensional cases. In high-dimensional spaces (e.g., the 88-dimensional orbital-field vector in this work), if an enormous number of examples are distributed uniformly in a high-dimensional hypercube, most examples are closer to the face of the hypercube than to their nearest neighbor. If we approximate a hypersphere by a hypercube, in high dimensions, almost all the volume of the hypercube is outside the hypersphere.[55] Moreover, with increasing dimensionality, the distance to the nearest neighbor approaches the distance to the farthest neighbor,[56] which implies that the learned weight of the Gaussian process could be meaningless in distinguishing between neighbors and distant data points. In the following, we observe that estimators working on high-dimensional spaces show more difficulty in converging to obtain suitable prediction accuracy; in other words, it is more difficult to estimate both distant and neighbor data points.

To overcome the curse of high dimensionality as well as perform tracking to see how the learned function is created, we propose the use of a metric learning algorithm for kernel regression—MLKR,[57] which optimizes the smoothness of dependence between a representation vector and a target property. First, the Mahalanobis distance $d(\mathbf{x}_i, \mathbf{x}_j)$ is defined as a linear transformation of conventional Euclidean distance as follows:

$$d(\mathbf{x}_i, \mathbf{x}_j) = ||\mathsf{A}(\mathbf{x}_i - \mathbf{x}_j)||^2, \qquad 7$$

where A is a linear transformation matrix. The MLKR method attempts to optimize the loss function $\mathcal{L}$ defined by the training error as

$$\mathcal{L} = \Sigma_i (y_i - \hat{y}_i)^2. \qquad 8$$

With the defined kernel in Equation 6, we can iteratively find the optimal A by $\Delta\mathsf{A}$, defined as

$$\Delta A = \lambda \frac{\partial \mathcal{L}}{\partial A} = 4\lambda A \Sigma_i (y_i - \hat{y}_i) \Sigma_j (y_j - \hat{y}_j) \kappa_{ij} \mathbf{x}_{ij} \mathbf{x}_{ij}^\top, \qquad 9$$

with $\mathbf{x}_{ij} := \mathbf{x}_i - \mathbf{x}_j$. The matrix A is gradually optimized to find the best embedding space. Therefore, we obtain a new embedding representation $\mathbf{u} := A\mathbf{x}$ by linear transformation of the original vector $\mathbf{x}$. From the definition of $\mathcal{L}$, the function of the target property learned on $\mathbf{u}$ is optimized to smoothly traverse through data points. Moreover, $\mathbf{u}$ with its low dimension, 2D in our setting, is expected to be of benefit for both prediction estimators and human intuition regarding the Euclidean distance compared with the conventional $\mathbf{x}$ for 88 dimensions in the OFM representation.

Embedding function interpretation

Maximizing the prediction ability of the machine learning estimator using the most limited training data is the first priority of the active learning method. The process of querying new labeled data are equivalent with correcting the form of learned function with respect to target property. For example, in binary classification, asserting data points with maximal variance of predicted class labels is equivalent to locating the boundary that separates two observed classes. Therefore, as an alternative advantage, following the correction process leads to better insight regarding the phenomena of interest. In this work, we introduce a method to localize information on the learned function, monitoring its change to improve the querying data process in interpreting the phenomena of interest.

With the target property as a continuous variable, we consider the learned formation energy function $\mathbf{y} = f(\mathbf{u})$, which is called interpretable if it is possible to allocate on the representation space $\mathbf{u}$, where the function meets a predefined condition g. In detail, given a condition g, the probability distribution spanning on the embedding space $\mathbf{u}$ is defined as follows:

$$p(\mathbf{u}|g) = \frac{1}{nh} \Sigma_{i=1}^n \mathbb{1}[g(\mathbf{u}_i)] e^{-\frac{|\mathbf{u}_i - \mathbf{u}|}{h}}, \qquad 10$$

with $p(\mathbf{u}|g)$ as the probability density at $\mathbf{u}$ under g, $\mathbf{u}_i$ as the location of an observed data point i (i.e., $\mathbf{u}_i = A\mathbf{x}_i$); h as a tuning kernel width. The indicator $\mathbb{1}[\cdot]$ returns 1 if the condition $[\cdot]$ is true, and 0 otherwise. In the present work, we consider two forms of relevant conditions.

$$g_y(\mathbf{u}_i) \overset{\text{def}}{=} \hat{\mathbf{y}}(\mathbf{u}_i) < 0, \qquad 11$$

$$g_{\mathbf{x}^j}(\mathbf{u}_i) \overset{\text{def}}{=} x_i^j > 0, \qquad 12$$

where $g_y(\mathbf{u}_i)$ and $g_{\mathbf{x}^j}(\mathbf{u}_i)$ intuitively represent a region of interest with potentially stable materials and regions spanned by structures incorporating the nonzero OFM element $\mathbf{x}^j$. Then, we measure the Bhattacharyya coefficient[58] between a pair of $(g_y, g_{\mathbf{x}^j})$ as

$$BC(g_y, g_{\mathbf{x}^j}) = \int \sqrt{p(\mathbf{u}|g_y)p(\mathbf{u}|g_{\mathbf{x}^j})} \, d\mathbf{u}, \qquad 13$$

with the integral taken over the space spanned by $\mathbf{u}$. The Bhattacharyya coefficient $BC(g_y, g_{\mathbf{x}^j})$ measures the probability of joint occurrence between two conditions g_y and $g_{\mathbf{x}^j}$. Higher BC values indicate a higher possibility to obtain correlation between conditions g_y and $g_{\mathbf{x}^j}$ and vice versa; this makes it easier to understand the meaning of the BC coefficient in identifying overlapping distributions. In the discussion of the results provided in the "Results and discussion" section, we characterize any distribution $p(\mathbf{u}|g)$ using a single-level contour representation.

Acquisition function

The acquisition function $\Gamma(\mathbf{x})$ quantifies the reward of structures in each $\mathcal{D}_t^{\neg\text{calculated}}$ that contributes to the prediction accuracy of the estimation models, as well as the exploration process. Structures $\mathbf{x}^*$ are queried to $\mathcal{D}_t^{\text{beneficial}}$ to calculate their formation energy if their acquisition function values reach an optimal value.

$$\mathbf{x}^* = \arg\max \Gamma(\mathbf{x}). \qquad 14$$

The majority form of Γ is designed to determine the optimum of a fixed expensive-to-compute function. In this work, we examine the two most canonical functions as follows:

$$\Gamma_{\text{exr}}(\mathbf{x}) = \sigma(\mathbf{x}), \qquad 15$$

$$\Gamma_{\text{exp}}(\mathbf{x}) = -\mu(\mathbf{x}), \qquad 16$$

where $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are the mean and variance of estimated values of not-yet-calculated structure $\mathbf{x}$, respectively. In representation space upon which the estimator is located, $\mathbf{x}$ is either an OFM vector or embedding vector $\mathbf{u} = A\mathbf{x}$ learned by the metric learning method. The first acquisition function, Γ_{exr}, based on the exploration strategy, assumes not-yet-calculated structures with higher variance to enhance the prediction ability of the estimator (i.e., are beneficial to the machine learning model). This acquisition function does not support directly finding superior structures because the information of the absolute value of the target property has not been included. The second acquisition function, Γ_{exp}, based on the exploitation strategy, selects not-yet-calculated structures with the lowest predicted target values as potential candidates to enhance the prediction ability of the estimator. Numerous acquisition functions[31,59–61] have been introduced to balance the exploration and exploitation assumptions. Finally, we also examine an acquisition strategy Γ_{uni} that randomly selects from the pool of not-yet-calculated structures.

Experiment and discussion

Experimental setup

We designed an experiment to simulate the process of exploring SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ structures with X, Y as Mo, Zn, Co, Cu, Ti, Al, and Ga using the proposed query-and-learn method. We collected ternary compounds—SmFe$_{12-\alpha}$X$_\alpha$ structures ($\alpha < 4$) to use as the initial training data and quaternary compounds SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ and $\alpha + \beta < 4$ as the initial pool of

not-yet-calculated data. Consequently, at the initial time of the exploration process, all not-yet-calculated structures were created using the bi-element substitution method rather than the single element substitution method as training structures. We summarize the initial training structures in **Figure** 2, which shows the primary state of the training data $\mathcal{D}^{\text{calculated}}$ with SmFe$_{12-\alpha}$X$_\alpha$ structures ($\alpha < 4$). In this figure, the structures are all referenced to SmFe$_{12}$ values of formation energy (0.07 eV/atom) and magnetization (2.011 T/f.u.). Substituting Ti, Al, Co, and Ga regularly creates substituted structures with formation energies lower than the reference value of SmFe$_{12}$. Among them, Ti and Al show a higher rate in creating negative formation energy structures than others. With Mo, Zn, and Cu, several substituted structures are more stable than the host SmFe$_{12}$, whereas the others are not. A part of our calculations were found to be consistent with other first-principles calculation methods such as the Quantum MAterials Simulator (QMAS),[62–64] OpenMX,[20] or experimental results.[13,65,66] Details of comparisons are shown in the Supplementary Information section. Figure 2 in the Supplemental Information shows summarization of all SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ structures in the region of $\alpha + \beta < 4$. All structures were described using 88-dimensional OFM vectors after eliminating duplicated columns.

For a time query t, two batches of structures were selected, denoted by $\mathcal{D}_t^{\text{beneficial}}$ and $\mathcal{D}_t^{\text{outstanding}}$. A detailed description of all batches is provided in the "Data set notation" section. We set 40 as the number of selected structures for each $\mathcal{D}_t^{\text{beneficial}}$ and $\mathcal{D}_t^{\text{outstanding}}$. Besides this, to evaluate performance of each strategy, we added 20 random structures to $\mathcal{D}_t^{\text{beneficial}}$. Finally, there were 30 query times to collect all structures in the screening space.

Results and discussion

Query-and-learn in monitoring the SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ structures discovery process

We now present the proposed query-and-learn method designed to monitor the materials discovery process. The relative positions of not-yet-calculated, calculated and queried structures, the form of the formation energy function, and generalizing knowledge of the structure–stability mechanism of SmFe$_{12}$ family are discussed.

Figure 3 shows the learned embedding function regarding the formation energy of SmFe$_{12}$ structures. In this figure, we show the results of a random querying strategy with the initial query $t = 1$ on the upper panel and the last query $t = 30$ on the lower panel. We demonstrate the results of different strategies in querying structures in the Supplemental Information. The calculated structures are denoted using face and edge color, which indicate the portion of each substituted element. For each query time t, non-calculated structures are shown as gray dots. White rhombus markers indicate structures that were queried at t in $\mathcal{D}_t^{\text{beneficial}}$ and white triangle markers indicate estimation regarding the most potentially stable structures in $\mathcal{D}_t^{\text{outstanding}}$.
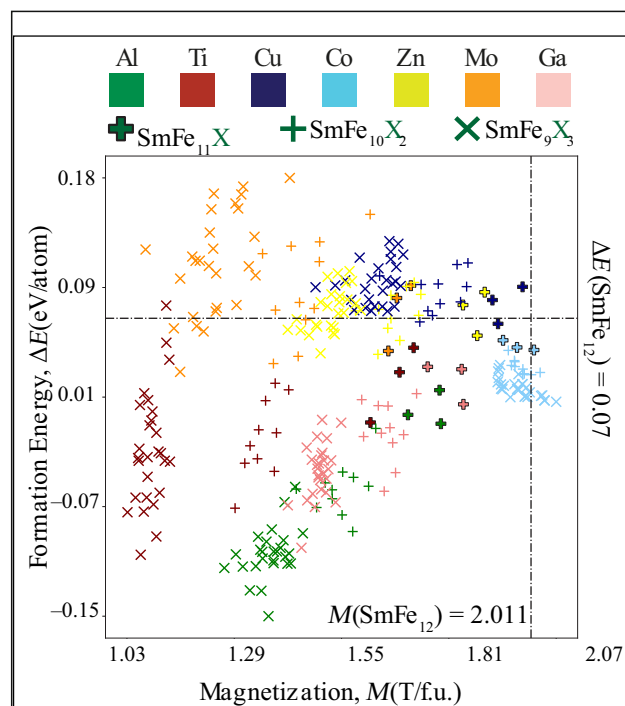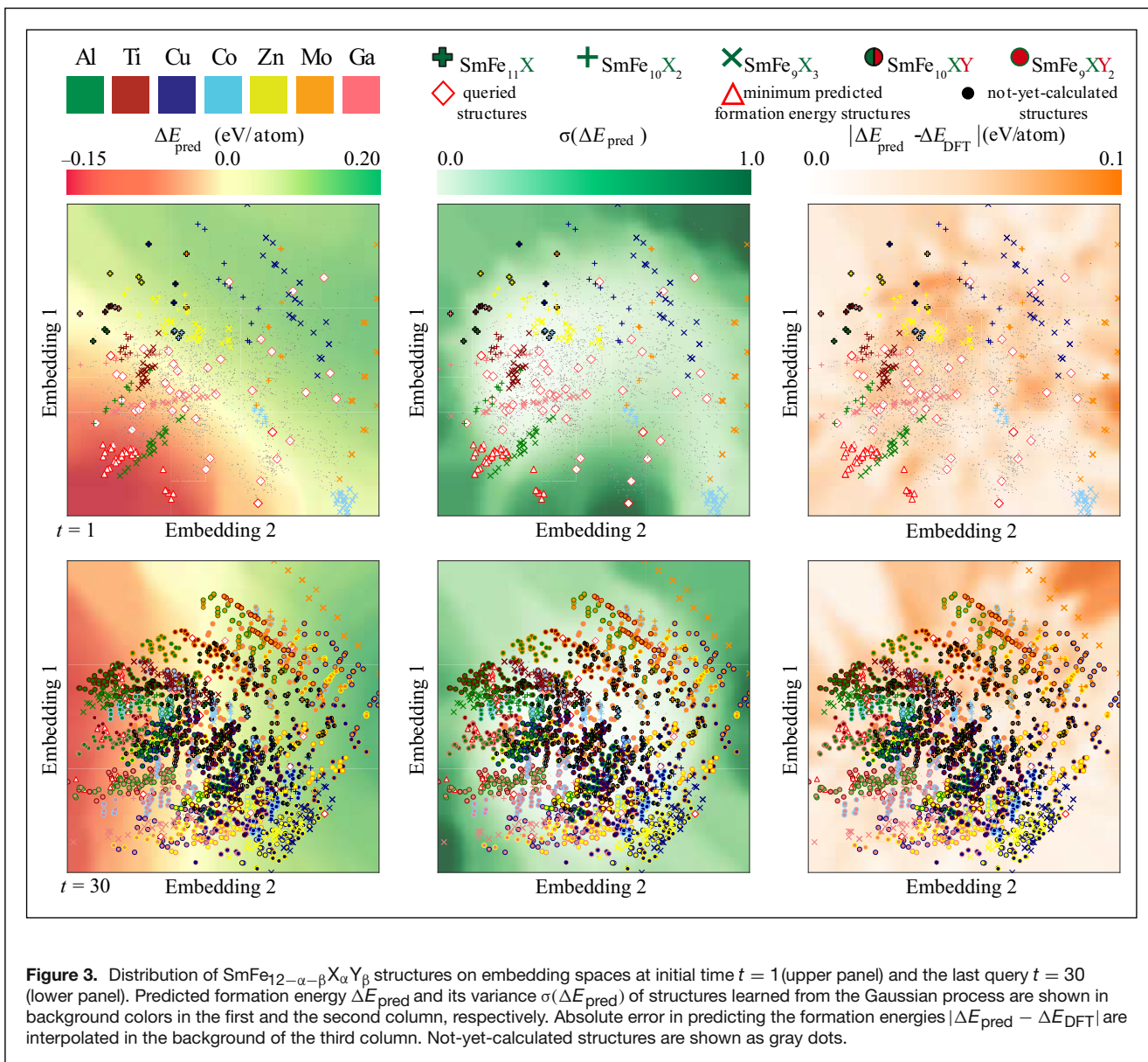

Figure 2. Formation energy and magnetization of SmFe$_{12-\alpha}$X$_\alpha$ structures ($\alpha < 4$), used as an initial training set of active learning systems. The substances are denoted as Ga (pink), Mo (orange), Zn (yellow), Co (light blue), Cu (navy), Ti (maroon), and Al (green).

For each query time, we show in the left and middle column of Figure 3 the predicted formation energy $\hat{y}$ and the estimated variance $\sigma(\hat{y})$ deriving from f, respectively. Moreover, we show in the right of Figure 3 the absolute error $|y - \hat{y}|$ in prediction between the ground truth first-principles method y and the calculated formation energy Gaussian process regression $\hat{y}$. In each t, we evaluated the error in predicting formation energy for all not-yet-calculated structures in $\mathcal{D}_t^{\neg\text{calculated}}$.

Values of all these attributes $\hat{y}, \sigma(\hat{y})$ and $|y - \hat{y}|$ shown in background color with nearest-neighbor interpolation in the embedding space. From this figure, the learned function of $\hat{y}$ appears as a smooth function traversing throughout all structures between negative to positive formation energy regions. Although queried structures were randomized well and distributed throughout the entire structure space using Γ_{uni}, our predicted potentially stable structures (white triangles) were also accurately allocated in the most negative formation energy region. Moreover, in $t = 1$, not-yet-calculated structures using the bi-element substituted method are uniformly dispersed throughout calculated structures with the single substituted element method.

Next, we investigated the learned formation energy function on embedding space via extremum interpretation. **Figure** 4 shows the formation energy landscape generated by embedding representation in the first and the last query time. Aiming

Figure 3. Distribution of SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ structures on embedding spaces at initial time $t = 1$ (upper panel) and the last query $t = 30$ (lower panel). Predicted formation energy ΔE_{pred} and its variance $\sigma(\Delta E_{\text{pred}})$ of structures learned from the Gaussian process are shown in background colors in the first and the second column, respectively. Absolute error in predicting the formation energies $|\Delta E_{\text{pred}} - \Delta E_{\text{DFT}}|$ are interpolated in the background of the third column. Not-yet-calculated structures are shown as gray dots.

to stabilize the SmFe$_{12}$ structure, the local minima of the formation energy function is defined as our region of interest, which contains the most negative formation energy structures in $\mathcal{D}_{\text{estimated}}^{\text{outstanding}}$. This region is defined as the distribution spanned by $p(\mathbf{u}|g_y)$ in the "Embedding function interpretation" section. In Figure 4, $p(\mathbf{u}|g_y)$ distributions are shown in red contours. In the following discussion, we refer to these distributions as the target contours for simplicity. By contrast, distributions of structures with nonzero OFM features defined as $p(\mathbf{u}|g_{\mathbf{x}^i})$ are shown in the embedding space via other contour lines. Intuitively, higher overlapping contours show a higher correlation between these properties. In the middle and the right of Figure 4, we show the projected OFM features $p(\mathbf{u}|g_{(\text{p}^1,M)})$ and $p(\mathbf{u}|g_{(\text{d}^5,M)})$, with M as s^1, s^2, p^1, d^2, d^5, d^6 and d^7, respectively. OFM features (d^5, M) show the average coordination number of sites owning M atomic representation surrounding Mo. Similarly,

(p^1, M) shows the average coordination number of atoms with M representation surrounding Al or Ga because these two elements share p^1 in their most outer-shell electron configuration. In the last query time, $t = 30$ or equivalently after collecting all calculated structures, one might recognize that the region of negative formation energy mostly overlaps with all (p^1, M) contours—regions spanning Al- and Ga-substituted structures. Among them, the (p^1, d^2) contour shows the distribution of SmFe$_{12-\alpha-\beta}$[Al/Ga]$_\alpha$Ti$_\beta$ structures within the most negative formation energy region. In the end of labeling all structures, Figure 2 in the Supplemental Information shows SmFe$_9$[Al/Ga]$_2$Ti structures as the most negative formation energy. By contrast, structures with Mo-substituted elements show distancing from the potentially stable regions. Notably, these correlations between the substituted element and corresponding stability could be found at the beginning of the querying process.
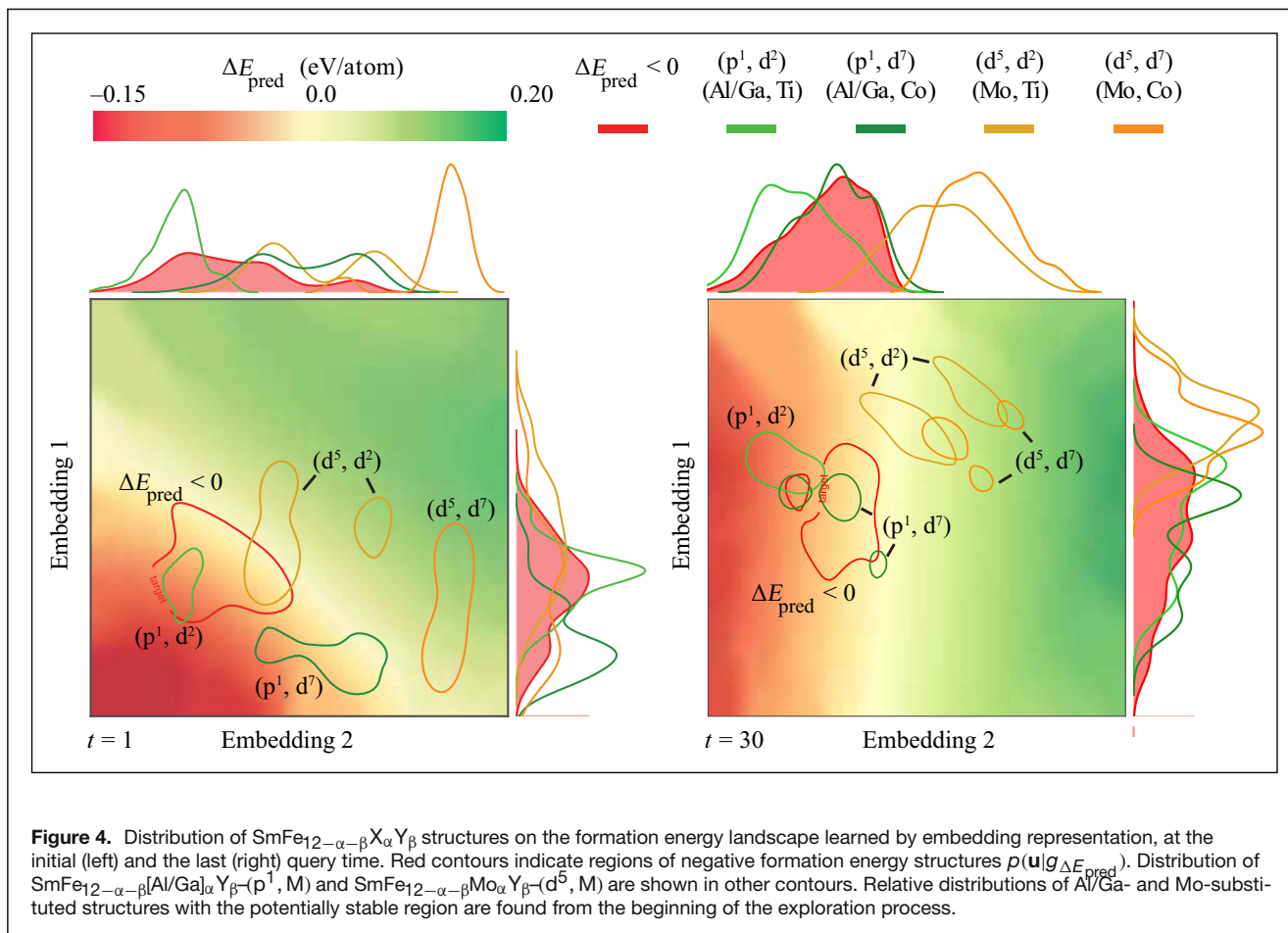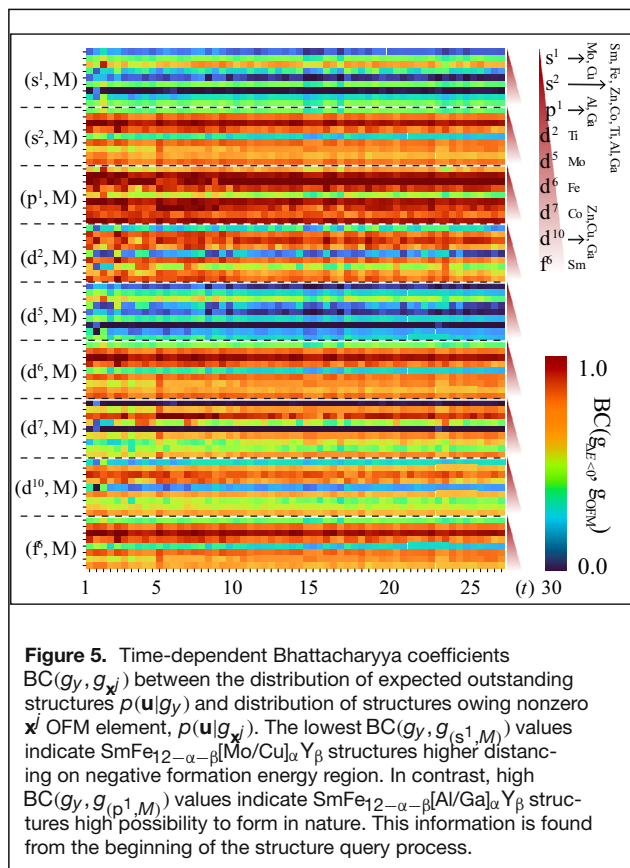
Figure 4. Distribution of SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ structures on the formation energy landscape learned by embedding representation, at the initial (left) and the last (right) query time. Red contours indicate regions of negative formation energy structures $p(\mathbf{u}|g_{\Delta E_{pred}})$. Distribution of SmFe$_{12-\alpha-\beta}$[Al/Ga]$_\alpha$Y$_\beta$–(p^1, M) and SmFe$_{12-\alpha-\beta}Mo_\alpha$Y$_\beta$–(d^5, M) are shown in other contours. Relative distributions of Al/Ga- and Mo-substituted structures with the potentially stable region are found from the beginning of the exploration process.

Figure 5 shows the dependence of normalized BC($g_y, g_{\mathbf{x}^j}$) on query time t in the active learning process for all OFM features. OFM features show a matrix with blocks of similar center atom representation; each block is presented keeping a similar order of neighbor representation. From Figure 5, structures with (s^1, M) and (d^5, M) features (i.e., Cu- and Mo-substituted structures) showed the lowest BC scores for all t. In other words, these structures were not located within the region of negative formation energy. By contrast, BC($g_y, g_{(p^1,M)}$) always remained at the highest score, or as we showed previously in the learned embedding space, these SmFe$_{12-\alpha-\beta}$[Al/Ga]$_\alpha$ B$_\beta$ structures were mostly associated with negative formation energy. We show another example in interpreting the substituted effect using BC($g_y, g_{(d^2,M)}$) or Ti-substituted structures. Structures excluding (d^2, s^1) and (d^2, d^5), that is, except SmFe$_{12-\alpha-\beta}$[Mo/Cu]$_\alpha$Ti$_\beta$, showed high possibility of negative formation energy. All these correlations were established by analyzing all queried data shown in the Supplementary information. Interestingly, these correlations could be performed very early, even at the beginning of the exploration process. In summary, the BC score on a learned embedding space is potentially useful in understanding the form of the formation

energy function and determining where interesting information is located without labeling all data.

Prediction ability of active learning designs

We examine the prediction accuracies of different active learning designs. For any query time t, we measured the mean absolute error (MAE) between the predicted and observed formation energies of structures in $\mathcal{D}_t^{\neg calculated}$. Because different structure querying strategies update their training data differently, not-yet-calculated structures in $\mathcal{D}_t^{\neg calculated}$ also differed among experiments. Therefore, MAE measured on $\mathcal{D}_t^{\neg calculated}$ can be approximated as the natural prediction loss of our designed system. **Figure** 6a shows the MAE results of active learning designs drawn from possible combinations of representation methods, estimators, and querying strategies. In this figure, with three acquisition functions, including uniform, exploration, and exploitation functions, experiments using the OFM representation are denoted in cyan, green, and blue, respectively. By contrast, active learning designs based on embedding representations are shown in yellow, orange, and red, respectively, with these three acquisition functions. Finally, we independently evaluate each of the six active

Figure 5. Time-dependent Bhattacharyya coefficients $BC(g_Y, g_{\mathbf{x}^i})$ between the distribution of expected outstanding structures $p(\mathbf{u}|g_Y)$ and distribution of structures owing nonzero $\mathbf{x}^i$ OFM element, $p(\mathbf{u}|g_{\mathbf{x}^i})$. The lowest $BC(g_Y, g_{(s^1, M)})$ values indicate SmFe$_{12-\alpha-\beta}$[Mo/Cu]$_\alpha$Y$_\beta$ structures higher distancing on negative formation energy region. In contrast, high $BC(g_Y, g_{(p^1, M)})$ values indicate SmFe$_{12-\alpha-\beta}$[Al/Ga]$_\alpha$Y$_\beta$ structures high possibility to form in nature. This information is found from the beginning of the structure query process.

Figure 6. Learning rate of active learners on orbital-field matrix representation space (cyan, green, and blue) and embedded metric learning space (yellow, orange, and red) with different acquisition functions α. Exploration, exploitation, and uniform acquisition strategies are denoted by Γ_{exr}, Γ_{exp}, and Γ_{uni}, respectively. (a) Dependence of the mean absolute error in predicting structures in $\mathcal{D}_t^{-calculated}$ on querying time t. (b) Recall rate in querying the most potentially stable structure $\mathcal{D}_t^{outstanding}$ depending on querying time t.

learning designs ten times with different initial random structures in order to evaluate the prediction accuracies.

The difference between active learning designs primarily depended on the nature of the representation method. At $t = 1$, all active learning systems obtained MAE at 2.5×10^{-2} (eV/atom). Overall, MAEs gradually decreased with increasing t for all the active learning systems. However, the performance of designs with high-dimensional OFM representation showed gradual linear improvement by adding new queried structures. This could be explained as new queried data points that are added using this strategy help the estimator forecast their neighbor only, rather than correcting the estimator learning on the entire dataspace. The MAE curve with the highest fluctuation belonged to a system using exploitation querying strategies. In other words, adding excessively biased data (e.g., low energy structures), as in the exploitation strategy, into the prediction model misguided the model to estimate other structures and directly reduced its prediction ability. By contrast, the lowest-bounded MAE curve always belonged to a design that utilized a uniform sampling strategy operating on the embedding representation. By querying up to $t = 5$, one-sixth of all not-yet-calculated structures, the design using the uniform querying strategy on embedding space quickly reached the optimal MAE at 1.25×10^{-2} (eV/atom) and then remained at this performance level for the remainder of the experiment. The model outperforms others because the Mahalanobis metric
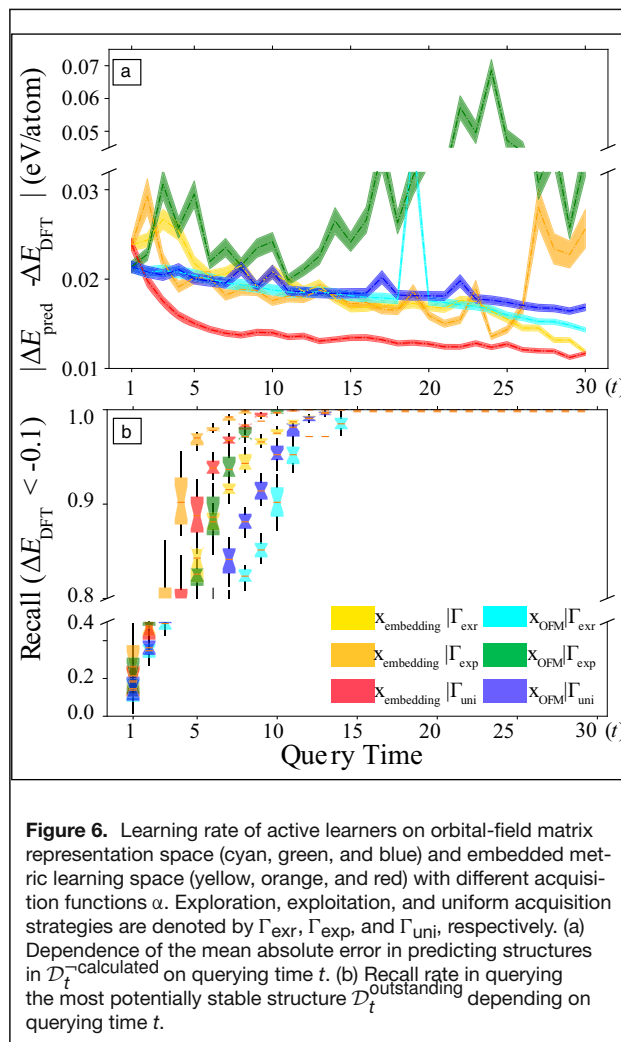
learned using MLKR preserving both Euclidean distance and following the direction of the target function[57] helps to correct the form of the estimator locally and globally. Thus, given several queried data points uniformly sampling on the embedding space, we could improve these two aspects simultaneously.

Next, we evaluated active learning designs in recalling the most potentially stable structures. We heuristically defined -0.1 (eV/atom) as the upper-limit formation energy for the set of most potentially stable structures. Consequently, the ground truth of the $\mathcal{D}_{confirmed}^{outstanding}$ set contained 74 structures incorporated with formation energy lower than -0.1 (eV/atom) or equivalent with 2.2% total not-yet-calculated candidates. Figure 6b shows the recall rate results for all active learning designs. The colors and patterns denoted for different active learning designs are synchronized with the MAE results, as shown in Figure 6a. This figure shows that all active learning designs recalled all $\mathcal{D}_{confirmed}^{outstanding}$ structures without querying all unlabeled structures. The worst recall performance of the active learning design by an exploration querying strategy and OFM representation required 14/30 query steps to recall all these potentially stable structures. By contrast, methods

with the best recall performance required 8/30 query steps. In the naivest case, when we randomly selected a structure from an unlabeled structure data set and avoided using any structures to update all active learning components, we needed to query all not-yet-calculated structures to recall all $\mathcal{D}_{\text{confirmed}}^{\text{outstanding}}$ structures. Equivalently, the rate of recall of the top 2.2% of structures with the lowest formation energy was enhanced between 2.1 and 3.7 times compared with the basic random selection method. We also report the results of using active learning with different initialization training data in Supplemental Information Materials.

Structure–stability relationship

In this section, we discuss the structure–stability relationship of this SmFe$_{12}$ family in detail. We investigate how different substituted elements distorted the host structure by measuring displacement of the OFM elements before and after performing a structure optimization step based on calculation from first-principles. The displacement $\Delta(\cdot)$ was measured as $\Delta\mathbf{x} = \mathbf{x}_{\text{opt}} - \mathbf{x}_{\text{org}}$ with $\mathbf{x}_{\text{opt}}$; $\mathbf{x}_{\text{org}}$ shows the value of an OFM element of calculated and initial structure, respectively. In **Figure** 7, we show correlations between formation energy and displacement OFM elements (d^6, M) in the upper panel and (p^1, M) in the lower panel, where M refers to s^2, s^1,... Because d^6 represents the Fe element and p^1 represents the Al/Ga elements in the OFM, we focused on analyzing the change in the coordination number of Fe site and Al/Ga sites, respectively. Correlations between formation energy and other OFM elements are shown in the Supplementary Information. Here, a violin plot with blue (yellow) show the displacements of OFM elements with the mean negative (positive) of the full displacement, respectively. By contrast, the formation energy of the corresponding structures is shown in red (green) for the negative (positive) mean of the formation energy.

In the upper panel with (d^6, M), structures owning (d^6, p^1) and (d^6, d^2) , that is, Al/Ga and Ti-substituted structures, respectively, show on average negative formation energies, indicating a trend of potentially stable structures. Further, the distribution of structures owning (d^6, d^2) show on average a reduction in coordination number, (d^6, p^1) structures appear with a distribution of a positive mean value. As an interpretation, in SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$-substituted structures, only Al/Ga-substituted sites come close to Fe sites on average (i.e., increasing coordination number). In the lower panel with (p^1, M), we confirmed again that in all Al/Ga-substituted families, there is a tendency of increasing coordination number of neighbors surrounding all p^1-like OFM element (yellow violin distribution). Moreover, almost all structures with (p^1, M) exhibited a mean negative formation energy. By contrast, as shown in the Supplementary information section, structures with other OFM elements all showed decreasing trends of the average coordination number and mean positive formation energy except (d^2, M)-Ti element. The lowest mean value of formation energy belonged to (p^1, d^2) structures (i.e., the SmFe$_{12-\alpha-\beta}$[Al/Ga]$_\alpha$Ti$_\beta$ family group).

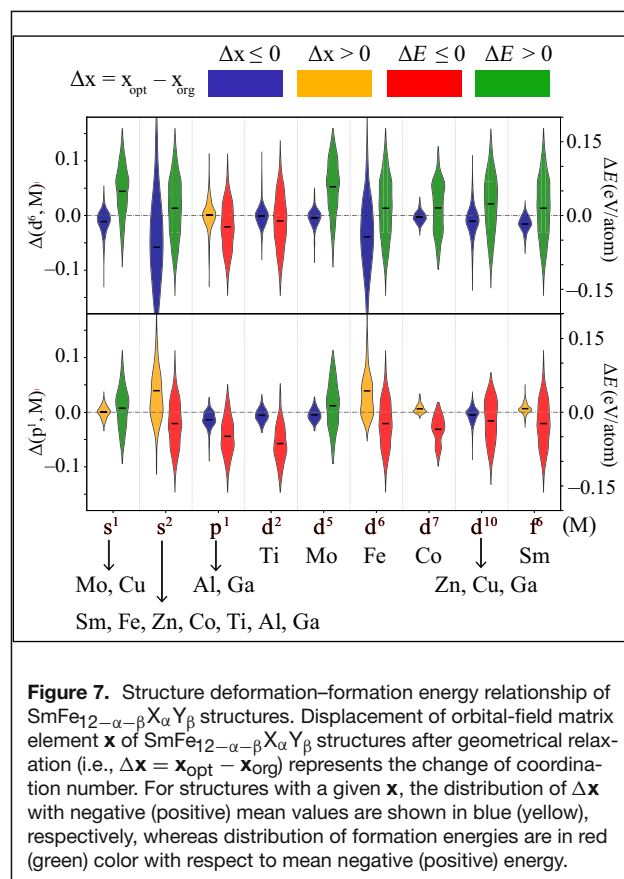**Figure 7.** Structure deformation–formation energy relationship of SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ structures. Displacement of orbital-field matrix element $\mathbf{x}$ of SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ structures after geometrical relaxation (i.e., $\Delta\mathbf{x} = \mathbf{x}_{\text{opt}} - \mathbf{x}_{\text{org}}$) represents the change of coordination number. For structures with a given $\mathbf{x}$, the distribution of $\Delta\mathbf{x}$ with negative (positive) mean values are shown in blue (yellow), respectively, whereas distribution of formation energies are in red (green) color with respect to mean negative (positive) energy.

Ideal structures in the SmFe$_{12}$ family should meet one more qualification about maximizing the magnetization of the substituted one. In the Supplemental Information, the most potential structures are mixed between Al, Co, and Cu-substituted structures that show optimal stability and magnetization. In **Figure** 8, we show the non-optimized original structure SmFe$_{12}$ compared to other Al, Co, and Cu-substituted structures after the optimization process. Three structures, SmFe$_{10}$Al$_2$, SmFe$_{10}$CoAl, and SmFe$_{10}$CuCo are shown with formation energy lower than SmFe$_{12}$ and sorted in increasing value of formation energy, respectively. Overall, these structures are shown with smaller sizes than the original structure SmFe$_{12}$ and the decreasing distance at the Fe-8f site to neighbors reflects an increasing coordination number at this Fe site. In detail, structures with two Al-substituted elements, SmFe$_{10}$Al$_2$ structure show the highest shrinkage level to the lattice parameter on the x- and y-axis while slightly expanding the lattice parameter on the z-axis. Substituting one Al and one Co site, SmFe$_{10}$CoAl structure obtains a smaller volume compared to the original but slightly larger than SmFe$_{10}$Al$_2$. The largest volume among these three substituted structures belongs to SmFe$_{10}$CuCo. In other words, Cu- and Co-substituted sites cannot distort other Fe and Sm sites. This evidence highlights the difference between the increased coordination number of Al-substituted structures and others.

Conclusion

In this study, we have introduced a query-and-learn active learning approach in exploring SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ structures with X, Y as Mo, Zn, Co, Cu, Ti, Al, Ga, and $\alpha + \beta < 4$. Our proposed method was developed to accelerate the rate of discovery of potentially stable structures and generalize our understanding of the stability mechanism of this family. 3307 SmFe$_{12-\alpha-\beta}$X$_\alpha$Y$_\beta$ structures with formation energy and magnetism calculated using first-principles calculations were used to form the exploration space. MAE of active learning designs showed the lowest values at 1.25×10^{-2} (eV/atom)—3.7% of the range calculated from first-principles by utilizing the embedded descriptor originating from the OFM. Moreover, the design reached this irreducible error approximately six times faster than the alternatives

compared. In the experiment aiming to find the most potentially stable structures, all active learning designs presented a successful recall rate 2.1–3.7 times faster than the random search strategy. Finally, we interpreted the formation energy landscape learned by embedding representation via smooth correlations between distributions of the local extreme and different coordination number information. We discovered that structures with substitution of non-transition-metal elements of like Al and Ga, associated with Ti, in particular SmFe$_9$[Al/Ga]$_2$Ti, had the highest possibility of stabilizing the SmFe$_{12}$ structure. Moreover, the mean negative formation energy SmFe$_{12-\alpha-\beta}$[Al/Ga]$_\alpha$Y$_\beta$ structures exhibited an increasing trend of neighbor atoms surrounding Al/Ga-substituted sites on average, whereas other families showed opposite trends.
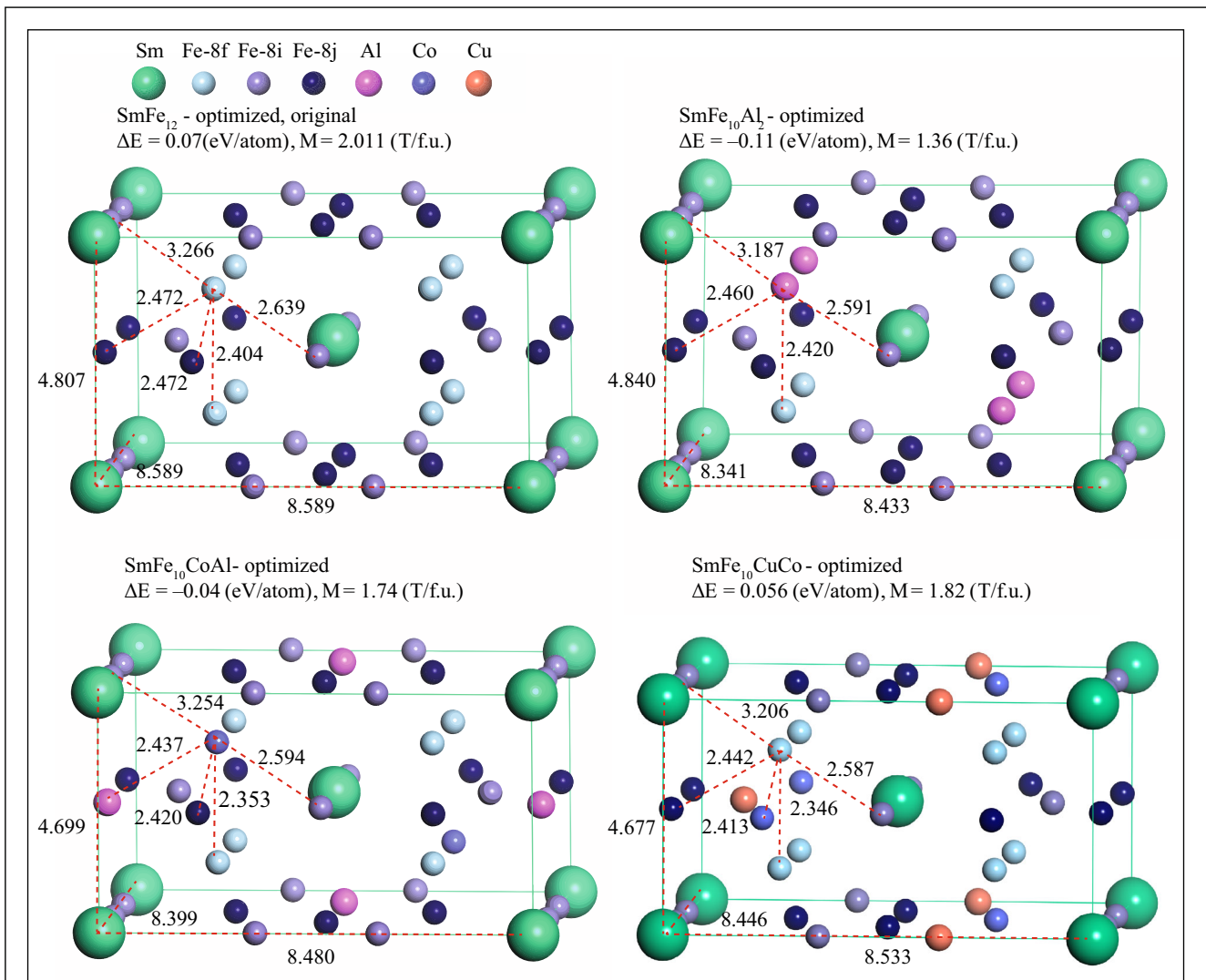

Figure 8. The original structure of SmFe$_{12}$ and the typical Al, Co, and Cu-substituted structures after the optimization process. SmFe$_{10}$Al$_2$, SmFe$_{10}$CoAl, and SmFe$_{10}$CuCo have lower formation energy and smaller unit cell than that of the original SmFe$_{12}$. The distance from the Fe-8f site to other adjacent sites (in Å) are highlighted to show the distortion caused by the substitution.

Author contributions
Duong-Nguyen Nguyen and Hieu-Chi Dam designed the study. Duong-Nguyen Nguyen performed the experiments. Duong-Nguyen Nguyen, Hiori Kino, and Takashi Miyake performed VASP calculation in methodology. Duong-Nguyen Nguyen wrote the manuscript. All authors analyzed the results and revised the manuscript.

Data availability
Data set of $SmFe_{12-\alpha-\beta}X_\alpha Y_\beta$ structures with X, Y as Mo, Zn, Co, Cu, Ti, Al, Ga, and $\alpha + \beta < 4$ including VASP calculations, OFM descriptor are published in Zenodo at https://doi.org/10.5281/zenodo.5763325. The calculated data from the three steps VASP are published upon reasonable requests.

Code availability
Code for active learning framework and analysis is available at https://github.com/nguyennd9192/query-and-learn.git.

Consent for publication
The participant has consented to the submission of the case report to the journal.

Conflict of interest
On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval
Hereby, I, D-N.N consciously assure that for the manuscript "Explainable active learning in investigating structure-stability of $SmFe_{12-\alpha-\beta}X_\alpha Y_\beta$ structures X, Y = {Mo, Zn, Co, Cu, Ti, Al, Ga}" is the authors' own original work, which has not been previously published elsewhere. The paper reflects the authors' own research and analysis in a truthful and complete manner. Besides, the paper properly credits the meaningful contributions of co-authors and co-researchers and appropriately placed in the context of prior and existing research.

Informed consent
Informed consent was obtained from all individual participants included in the study.

Supplementary information
The online version of this article (https://doi.org/10.1557/s43577-022-00372-9) contains supplementary material, which is available to authorized users.

References
1. S. Goedecker, *J. Chem. Phys.* **120**, 9911 (2004). https://doi.org/10.1063/1.1724816
2. A.R. Oganov, C.W. Glass, *J.. Chem. Phys.* **124**, 244704 (2006). https://doi.org/10.1063/1.2210932
3. C.J. Pickard, R.J. Needs, *Phys. Rev. Lett.* **97**, 045504 (2006). https://doi.org/10.1103/PhysRevLett.97.045504
4. Y. Wang, J. Lv, L. Zhu, Y. Ma, *Phys. Rev. B* **82**, 094116 (2010). https://doi.org/10.1103/PhysRevB.82.094116
5. M. Takagi, T. Taketsugu, H. Kino, Y. Tateyama, K. Terakura, S. Maeda, *Phys. Rev. B* **95**, 184110 (2017). https://doi.org/10.1103/PhysRevB.95.184110
6. T. Yamashita, N. Sato, H. Kino, T. Miyake, K. Tsuda, T. Oguchi, *Phys. Rev. Mater.* **2**, 013803 (2018). https://doi.org/10.1103/PhysRevMaterials.2.013803
7. T. Lam Pham, H. Kino, K. Terakura, T. Miyake, K. Tsuda, I. Takigawa, H. Chi Dam, *Sci. Technol. Adv. Mater.* **18**, 756 (2017)
8. T.-L. Pham, N.-D. Nguyen, V.-D. Nguyen, H. Kino, T. Miyake, H.-C. Dam, *J. Chem. Phys.* **148**, 204106 (2018)
9. K.H.J. Buschow, D.B. de Mooij, M. Brouha, H.H. Smit, R.C. Thiel, *IEEE Trans. Magn.* **24**, 1611 (1988)
10. A. Müller, *J. Appl. Phys.* **64**, 249 (1988). https://doi.org/10.1063/1.341473
11. K. Ohashi, Y. Tawara, R. Osugi, M. Shimao, *J. Appl. Phys.* **64**, 5714 (1988). https://doi.org/10.1063/1.342235
12. Y.K. Takahashi, H. Sepehri-Amin, T. Ohkubo, *Sci. Technol. Adv. Mater.* **22**, 449 (2021). https://doi.org/10.1080/14686996.2021.1913038
13. R. Coehoorn, *Phys. Rev. B* **41**, 11790 (1990). https://doi.org/10.1103/PhysRevB.41.11790
14. K. Buschow, *J. Magn. Magn. Mater.* **100**, 79 (1991). https://doi.org/10.1016/0304-8853(91)90813-P
15. Y. Wang, G. Hadjipanayis, *J. Magn. Magn. Mater.* **87**, 375 (1990). https://doi.org/10.1016/0304-8853(90)90774-K
16. T. Fukazawa, H. Akai, Y. Harashima, T. Miyake, *J. Magn. Magn. Mater.* **469**, 296 (2019). https://doi.org/10.1016/j.jmmm.2018.08.071
17. T. Fukazawa, Y. Harashima, Z. Hou, T. Miyake, *Phys. Rev. Mater.* **3**, 053807 (2019). https://doi.org/10.1103/PhysRevMaterials.3.053807
18. A. Schönhöbel, R. Madugundo, O.Y. Vekilova, O. Eriksson, H. Herper, J. Barandiarán, G. Hadjipanayis, *J. Alloys Compd.* **786**, 969 (2019). https://doi.org/10.1016/j.jallcom.2019.01.332
19. T. Miyake, Y. Harashima, T. Fukazawa, H. Akai, *Sci. Technol. Adv. Mater.* **22**, 543 (2021). https://doi.org/10.1080/14686996.2021.1935314
20. M. Matsumoto, T. Hawai, K. Ono, *Appl. Phys. Rev.* **13**, 064028 (2020). https://doi.org/10.1103/PhysRevApplied.13.064028
21. P. Tozman, Y. Takahashi, H. Sepehri-Amin, D. Ogawa, S. Hirosawa, K. Hono, *Acta Mater.* **178**, 114 (2019). https://doi.org/10.1016/j.actamat.2019.08.003
22. A. Gabay, G. Hadjipanayis, *Scr. Mater.* **154**, 284 (2018). https://doi.org/10.1016/j.scriptamat.2017.10.033

23. M. Hagiwara, N. Sanada, S. Sakurada, *J. Magn. Magn. Mater.* **465**, 554 (2018). https://doi.org/10.1016/j.jmmm.2018.06.042

24. A.M. Gabay, A. Martín-Cid, J.M. Barandiaran, D. Salazar, G.C. Hadjipanayis, *AIP Adv.* **6**, 056015 (2016). https://doi.org/10.1063/1.4944066

25. A. Gabay, G. Hadjipanayis, *J. Magn. Magn. Mater.* **422**, 43 (2017). https://doi.org/10.1016/j.jmmm.2016.08.064

26. N. Sakuma, S. Suzuki, T. Kuno, K. Urushibata, K. Kobayashi, M. Yano, A. Kato, A. Manabe, *AIP Adv.* **6**, 056023 (2016). https://doi.org/10.1063/1.494452

27. S. Suzuki, T. Kuno, K. Urushibata, K. Kobayashi, N. Sakuma, K. Washio, M. Yano, A. Kato, A. Manabe, *J. Magn. Magn. Mater.* **401**, 259 (2016). https://doi.org/10.1016/j.jmmm.2015.10.042

28. E.V. Podryabinkin, E.V. Tikhonov, A.V. Shapeev, A.R. Oganov, *Phys. Rev. B* **99**, 064114 (2019). https://doi.org/10.1103/PhysRevB.99.064114

29. K. Terayama, M. Sumita, R. Tamura, D.T. Payne, M.K. Chahal, S. Ishihara, K. Tsuda, *Chem. Sci.* **11**, 5959 (2020). https://doi.org/10.1039/D0SC00982B

30. K. Terayama, R. Tamura, Y. Nose, H. Hiramatsu, H. Hosono, Y. Okuno, K. Tsuda, *Phys. Rev. Mater.* **3**, 033802 (2019). https://doi.org/10.1103/PhysRevMaterials.3.033802

31. C. Dai, S.C. Glotzer, *J. Phys. Chem. B* **124**, 1275 (2020). https://doi.org/10.1021/acs.jpcb.9b09202

32. E.E. Marinero, A. Strachan, J.C. Verduzco, *Integr. Mater. Manuf. Innov.* **10**, 299 (2021). https://doi.org/10.1007/s40192-021-00214-7

33. D. Xue, P.V. Balachandran, R. Yuan, T. Hu, X. Qian, E.R. Dougherty, T. Lookman, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13301 (2016). https://doi.org/10.1073/pnas.1607412113

34. P.V. Balachandran, J. Young, T. Lookman, J.M. Rondinelli, *Nat. Commun.* **8**, 14282 (2017). https://doi.org/10.1038/ncomms14282

35. M. Spellings, S.C. Glotzer, *AIChE J.* **64**, 2198 (2018). https://doi.org/10.1002/aic.16157

36. J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, *JOM* **65**, 1501 (2013)

37. W. Kohn, L.J. Sham, *Phys. Rev.* **140**, A1133 (1965)

38. P. Hohenberg, W. Kohn, *Phys. Rev.* **136**, B864 (1964)

39. W. Sun, S.T. Dacek, S.P. Ong, G. Hautier, A. Jain, W.D. Richards, A.C. Gamst, K.A. Persson, G. Ceder, *Sci. Adv.* **2**, e1600225 (2016). https://doi.org/10.1126/sciadv.1600225

40. Y. Wu, P. Lazic, G. Hautier, K. Persson, G. Ceder, *Energy Environ. Sci.* **6**, 157 (2013). https://doi.org/10.1039/C2EE23482C

41. G. Xing, T. Ishikawa, Y. Miura, T. Miyake, T. Tadano, *J. Alloys Compd.* **874**, 159754 (2021). https://doi.org/10.1016/j.jallcom.2021.159754

42. S. Kirklin, J.E. Saal, B. Meredig, A. Thompson, J.W. Doak, M. Aykol, S. Rühl, C. Wolverton, *NPJ Comput. Mater.* **1**, 15010 (2015)

43. G. Kresse, J. Hafner, *Phys. Rev. B* **47**, 558 (1993)

44. G. Kresse, J. Hafner, *Phys. Rev. B* **49**, 14251 (1994)

45. P.E. Blöchl, *Phys. Rev. B* **50**, 17953 (1994)

46. G. Kresse, D. Joubert, *Phys. Rev. B* **59**, 1758 (1999)

47. J.P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996)

48. G. Kresse, D. Joubert, *Phys. Rev. B* **59**, 1758 (1999). https://doi.org/10.1103/PhysRevB.59.1758

49. I.G. Kresse, D. Joubert, P.E. Blöchl, VASP. https://www.vasp.at/wiki/index.php/Available_PAW_potentials. Accessed 31 Aug 2021

50. G. Kresse, D. Joubert, *Phys. Rev. B* **59**, 1758 (1999). https://doi.org/10.1103/PhysRevB.59.1758

51. P.E. Blöchl, *Phys. Rev. B* **50**, 17953 (1994). https://doi.org/10.1103/PhysRevB.50.17953

52. S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder, *Comput. Mater. Sci.* **68**, 314 (2013). https://doi.org/10.1016/j.commatsci.2012.10.028

53. D.-N. Nguyen, D.-A. Dao, T. Miyake, H.-C. Dam, *J. Chem. Phys.* **153**, 114111 (2020). https://doi.org/10.1063/5.0015977

54. T.-L. Pham, D.-N. Nguyen, M.-Q. Ha, H. Kino, T. Miyake, H.-C. Dam, *IUCrJ* **7**, 1036 (2020). https://doi.org/10.1107/S2052252520010088

55. P. Domingos, *Commun. ACM* **55**, 5 (2012). https://doi.org/10.1145/2347736.2347755

56. K. Beyer, J. Goldstein, R. Ramakrishnan, U. Shaft, in *International Conference on Database Theory, ICDT 1999,* Lecture Notes in Computer Science Book Series, vol. 1540 (1999)

57. K. Weinberger, G. Tesauro, *J. Mach. Learn. Res. Proc.* **2**, 612 (2007)

58. A.K. Bhattacharyya, *Bull. Calcutta Math. Soc.* **35**, 99 (1943)

59. H.J. Kushner, *J. Basic Eng.* **86**, 97 (1964). https://doi.org/10.1115/1.3653121

60. R. Urtasun, T. Darrell, A. Kapoor, K. Grauman, *Int. J. Comput. Vis.* **88**, 169 (2010). https://doi.org/10.1007/s11263-009-0268-3

61. P.I. Frazier, W.B. Powell, S. Dayanik, *SIAM J. Control Optim.* **47**, 2410 (2008). https://doi.org/10.1137/070693424

62. I. Dirba, Y. Harashima, H. Sepehri-Amin, T. Ohkubo, T. Miyake, S. Hirosawa, K. Hono, *J. Alloys Compd.* **813**, 152224 (2020). https://doi.org/10.1016/j.jallcom.2019.152224

63. Y. Harashima, K. Terakura, H. Kino, S. Ishibashi, T. Miyake, *J. Appl. Phys.* **120**, 203904 (2016). https://doi.org/10.1063/1.4968798

64. Y. Harashima, K. Terakura, H. Kino, S. Ishibashi, T. Miyake, in *Proceedings of Computational Science Workshop 2014 (CSW2014)* (2014). https://doi.org/10.7566/JPSCP.5.011021

65. Y. Hirayama, Y. Takahashi, S. Hirosawa, K. Hono, *Scr. Mater.* **138**, 62 (2017). https://doi.org/10.1016/j.scriptamat.2017.05.029

66. S. Hirosawa, *IEEE Trans. Magn.* **55**(2), 1 (2019). https://doi.org/10.1109/TMAG.2018.2863737 □