CHEMISTRY OF MATERIALS

pubs.acs.org/cm

Article

Revealing Hidden Patterns through Chemical Intuition and Interpretable Machine Learning: A Case Study of Binary Rare-Earth Intermetallics *RX*

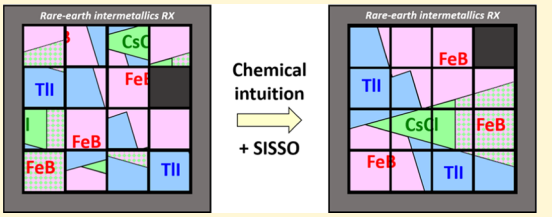
Volodymyr Gvozdetskyi,* Balaranjan Selvaratnam, Anton O. Oliynyk, and Arthur Mar*

Cite This: *Chem. Mater.* 2023, 35, 879–890

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Machine learning algorithms have been applied successfully in many areas of materials chemistry but often suffer from an inability to extract chemical insight. To demonstrate that an approach combining machine learning and chemical intuition can be effective in generating interpretable models, the structures of binary equiatomic rare-earth intermetallics *RX*, whose relationships have long defied understanding, were investigated as a case study. A structure map was developed based on only two parameters, which are derived from simple elemental descriptors (atomic number, period and group numbers, radii, and electronegativity) of the *R* and *X* components. This map reveals the previously hidden patterns of structural regularities of *RX* intermetallics. It explains the preference for CsCl-, TlI-, or FeB-type structures among these compounds, predicts the structures of missing members, rationalizes the occurrence of metastable phases and polymorphic transformations, and offers strategies for structure stabilization through the addition of a third component.



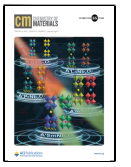
1. INTRODUCTION

In an age when the corpus of scientific data, gathered experimentally or computationally, is overwhelmingly large, high-throughput machine learning methods have become viable and indispensable to accelerate discovery. In materials chemistry, many machine learning models have been proposed to predict new candidates with desired structures or properties, often with high accuracy.[1−5] Skeptics rightfully criticize some of these models as being opaque ("black box"), devoid of physical meaning. Thus, there is a strong impetus to develop interpretable machine learning models. In the context of machine learning, interpretability means that users can see transparently how the model arrived at its predictions, in terms of domain-specific knowledge.[6−10] When machine learning is applied to materials chemistry, an interpretable model is one that is understandable in terms of chemical concepts. Such interpretable models are valuable because users can decide for themselves if the predictions are reasonable on the basis of sound scientific ideas, and equally important they can gain useful insight and practical guidance to prepare new materials. Some machine learning algorithms, such as linear regression and decision trees, are more easily interpretable than others, such as neural networks. Contrary to expectations, improving interpretability does not have to be at the expense of diminishing the accuracy or completeness of the model.

A natural way to build interpretability in any chemical model is to relate the structures and properties of substances to familiar elemental descriptors such as electronegativity, atomic size, and electron configuration. This desire to couch explanations in terms of chemical concepts is so visceral and innate that the collective body of wisdom and experience is often called "chemical intuition." Examples of such empirical models include simple principles to rationalize the structures of ionic solids,[11,12] tolerance factors to evaluate the degree of distortion in perovskite structures,[13,14] or electron counting rules to predict the electrical properties of half-Heusler compounds.[15] All of these models remain highly appealing, even if they may not be completely accurate, because they present a causal relationship between elemental descriptors and physical phenomena that can be grasped by the human mind, something that machine learning models often fail or are not designed to do.

Predicting the crystal structure of a compound is a fundamental classification problem in chemistry and a prerequisite to evaluating its physical properties accurately. The number of experimentally determined structures of unique inorganic phases reported to date is on the order of 10^5.[16−20] These are augmented by a huge number of computationally generated phases, most of which have yet to be experimentally verified;[21−25] it remains difficult to judge whether such hypothetical phases are truly stable (or metastable), given that some discrepancies have been noted in their formation energies vis-à-vis actual structures observed.[26,27]

Received: August 8, 2022
Revised: January 13, 2023
Published: January 30, 2023



There has been a long tradition in solid-state chemistry to classify crystal structures, starting from the earliest empirical models proposed by Pauling[11,12] and the structure maps developed by Pettifor, who emphasized that individual atomic properties (such as radii, electronegativity, and number of valence electrons) alone are insufficient to discriminate these structures.[28−31] Rather, he assumed that the structures could be classified by a single atomic parameter (which he called a "chemical scale," χ), presaging the virtue of dimensionality reduction commonly performed in machine learning methods. Compared to other approaches,[32] these structure maps were advantageous in facilitating chemical understanding; for example, closely related structure types are found to be clustered together. However, there is no pretense that such structure maps can offer detailed explanations for how crystal structures form. It is interesting that, even during his time, Pettifor recognized that DFT calculations could be helpful in providing energetic reasons for the relative stability of competing crystal structures. Later developments in structure maps occurred as more compounds of specific compositions were rapidly discovered and when modifications to atomic properties (such as various electronegativity scales)[33] were introduced (Figure S1). Recent statistical analyses of crystal structures suggest that caution has to be exercised when models are created to classify them.[34−36] Some compositions are over-represented simply because they have been well studied or are more easily synthesized (e.g., oxides, sulfides, and selenides with the composition AB_2 are 4 times more common than other compositions), and over 90% of the most frequently occurring structure types are based on close-packed atomic arrangements. That is, creating a fair model to classify crystal structures relies on the data being representative. Despite concerns that materials data may not fully satisfy the assumption of being independent and identically distributed, carefully developed machine learning models have demonstrated their utility in predicting new materials, especially when supported by experimental validation.[37,38]

Among early efforts to classify structures of binary equiatomic phases (having a 1:1 composition), Villars created structure maps based on three parameters involving pseudopotential radii, Martynov−Batsanov electronegativity, and number of valence electrons.[39] More recently, machine learning studies have been conducted to tackle this problem for selected subsets of these binary phases.[40,41] For example, we have combined partial least-squares discriminant analysis and support vector machine methods to achieve a 93% accuracy in classifying structures of binary equiatomic phases and to experimentally confirm the predicted CsCl-type structure of the new compound RhCd.[42] The most serious violations in the Villars structure map occur among binary rare-earth intermetallic compounds *RX* (*R* = rare-earth metal; *X* = metal or metalloid), whose three most common structure types (CsCl, TlI, and FeB) fall into overlapping clusters and could not be resolved. It is interesting that the machine learning model also suffers from this ambiguity. The reasons for the preference of these *RX* structures, which are highly similar, remain unknown.

Rare-earth intermetallics have many important applications, especially those that take advantage of their magnetic properties (e.g., ferromagnets and magnetocalorics).[43,44] Because f-electron systems present technical challenges in first-principles calculations (such as electron correlation and mixed valency), machine learning is an attractive alternative to predicting the structures, stability, and properties of rare-earth intermetallics. We hypothesized that the structural preference of binary *RX* intermetallics may be understood if the hidden patterns are revealed by combining chemical intuition, namely that factors such as size and electronegativity must play a role, with machine learning methods, which will help choose the best functional form of these factors. By ensuring that these models are chemically interpretable, we developed rules to make sense of the structural patterns and make predictions about the stability of these *RX* compounds, including experimental verification. Finally, we evaluated how such interpretable models perform relative to more "black box" models.

2. RESULTS AND DISCUSSION

2.1. Structural Regularities among *RX* Compounds.

As an overview, it is useful to survey the structures of all *RX* compounds (where *X* is any other element), not just limited to intermetallics. About 95% of them (~400 compounds) fall into six major structure groups.[16] The most common are NaCl-type ($Fm\bar{3}m$), CsCl-type ($Pm\bar{3}m$), TlI-type (*Cmcm*), and FeB-type (*Pnma*) structures, constituting 88%. The remaining 7% are represented by ZrCl-type ($R\bar{3}m$) and a small recondite group of aluminides *R*Al comprising DyAl-type (*Pbcm*), CeAl-type (*Cmcm*), and EuAl-type (*Pmmn*) structures. If one attempts to identify shared features, the CsCl-, TlI-, and FeB-type structures could be derived from 4^4 nets built from alternating *R* and *X* atoms as found in the NaCl-type structure but distorted to varying degrees and stacked in varying orientations (Figure 1), whereas the ZrCl-type and *R*Al structures show completely unrelated atomic arrangements (Figure S2).

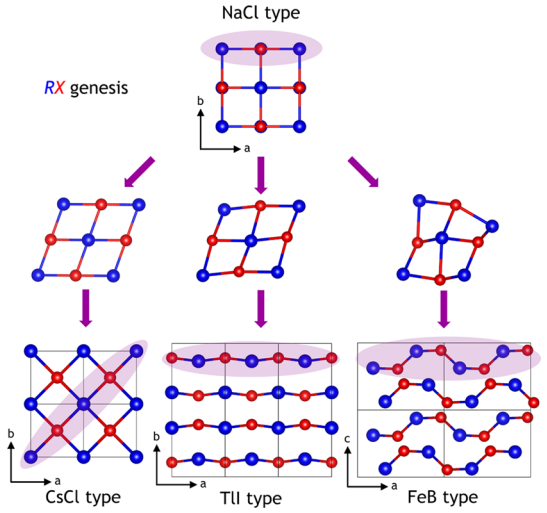


Figure 1. In rare-earth intermetallics *RX*, the square nets in NaCl-type structures are distorted and stacked in different ways resulting in CsCl-, TlI-, and FeB-type structures.

The simplest way to search for regularities among these structures is to see if they can be grouped according to the locations of the elements *R* and *X* in the periodic table. A diagram based on the atomic number of *R* and the period and group numbers of *X* already reveals some patterns (Figure 2). The halides (ZrCl type) are clearly segregated from the chalcogenides and pnictides (NaCl type). It is perhaps too facile to invoke ionic bonding character as the sole factor which dictates the formation of the ZrCl-type structure for the halides: first, this structure entails significant metal−metal

https://doi.org/10.1021/acs.chemmater.2c02425
Chem. Mater. 2023, 35, 879−890

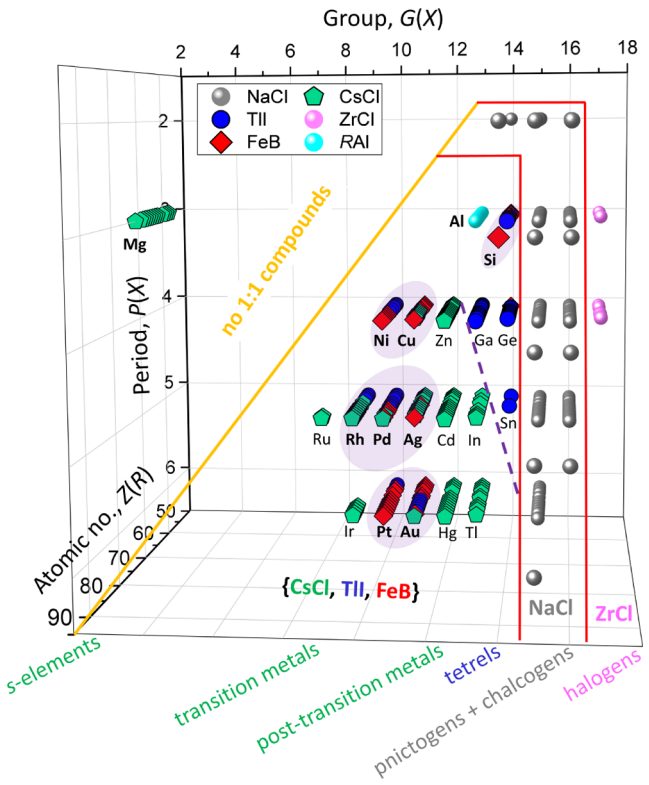


Figure 2. Structures of binary compounds *RX* plotted according to the positions of *R* [atomic number $Z(R)$] and *X* [period $P(X)$ and group $G(X)$] in the periodic table.

bonding, and, second, these compounds are not true binaries but rather hydrogen-stabilized ternaries.[45] Similarly, the binary chalcogenides and pnictides show significant covalent bonding character notwithstanding the widespread impression that the NaCl-type structure is a common one for ionic solids.[46] Nevertheless, for the purpose of classifying structures, it is sufficient to note that *RX* halides, chalcogenides, and pnictides are distinct from the rest, which are all intermetallics. The aluminides form a special category with unique structures noted above.[47] The remaining intermetallic compounds adopt CsCl-, TlI-, and FeB-type structures with no easily discernible patterns. The gallides, germanides, and stannides (TlI-type) seem to be delineated from the neighboring cluster of CsCl-type compounds (dashed purple line), close to the metal–nonmetal boundary line in the periodic table. Many of the remaining intermetallics adopt structures that depend on a haphazard way with *R* and *X* (circled clusters such as *R*Pd). Finally, equiatomic *RX* compounds do not form with the s-block and early d-block elements (diagonal orange line), except for *R*Mg. Instead, phases occur with other compositions (e.g., RX_2 and R_3X_2) or solid solution form, adopting W- ($Im\bar{3}m$), Mg- ($P6_3/mmc$), or Cu-type ($Fm\bar{3}m$) structures.

2.2. "Why Is It Called 'Puzzles?' That's the Puzzle".[48]

The jumble of CsCl-, TlI-, and FeB-type structures found for the compounds *RX* (*X* = Si, Ni, Cu, Rh, Pd, Ag, Pt, and Au) has been noted in the literature before but has remained an unsolved mystery for 60 years.[49−52] Because the position of an element within the periodic table serves as a proxy for many properties, it is not an outlandish idea to see where these structures would fall in a two-dimensional plot based on the atomic numbers of *R* and *X* (Figure 3a). Although it is mixed up, this structure map tantalizingly suggests systematic regularities but within very limited ranges and eliciting many curious questions. For example, why are *R*Pd phases metastable for a limited range of *R* (Dy, Ho, and Er)?[53]

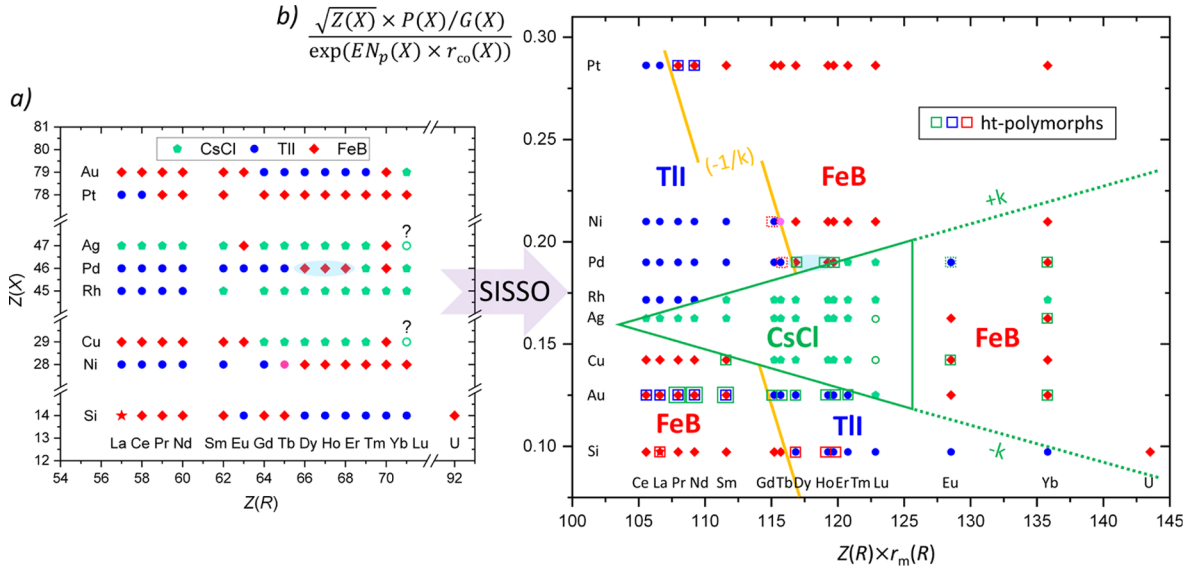


Figure 3. (a) Distribution of CsCl-, TlI-, and FeB-type structures for equiatomic rare-earth intermetallics *RX*, plotted according to the atomic numbers of *R* and *X*. The *R*Pd (*R* = Dy, Ho, and Er) phases (within the blue loop) are metastable. The existence of LuCu and LuAg (open green circles) has been unclear in the literature. TbNi (magenta circle) adopts a unique structure containing fragments of the TlI- and FeB-type structures. LaSi (red star) also adopts a unique structure closely resembling the FeB-type structure. (b) A structure map plotted using SISSO-determined parameters derived from simple elemental descriptors reveals hidden patterns among *RX* intermetallics. This map explains occurrence rules and regularities among high-temperature polymorphs, structural ambiguities at boundary conditions [TbNi and *R*Pd (*R* = Dy, Ho, and Er)], and structure stabilization by a third component. It predicts the structures of missing members such as LuCu and LuAg. Some of the high-temperature polymorphs (GdNi, TbPd, and EuPd), which are assumed to require specific experimental conditions to form, are enclosed in dashed squares; see text for further discussion.

https://doi.org/10.1021/acs.chemmater.2c02425
Chem. Mater. 2023, 35, 879−890

Why does TbNi adopt a unique structure?[54] What are the structures predicted for LuCu and LuAg? Given the relationship of the CsCl-, TlI-, and FeB-type structures (Figure 1), it is not surprising that some of these *RX* intermetallics can undergo transformation to other polymorphs by changing the temperature or doping with a third element. These transformations would most likely occur at the boundaries separating the structure types. The problem resembles a sliding puzzle whose hidden pattern could be revealed if only the tiles were ordered properly.

Can this puzzle be solved? Challenge accepted! We take a leap of faith (otherwise known as an "assumption") that an appropriate combination of elemental descriptors is capable of classifying these structures. Initially, the *RX* entries were sorted in terms of other descriptors besides atomic number (e.g., radii, electronegativity, and number of valence electrons), but all these attempts failed to reveal any patterns if the parameters are considered individually. Through trial and error, the *RX* entries were sorted in other sequences until a reasonable separation of structure types was achieved. The best result was obtained when the product of atomic number and metallic radius (Å) of the rare-earth component *R* is plotted along the *x*-axis

$$x = Z(R) \times r_m(R)$$

and the metal or metalloid component *X* is sorted in the sequence Si, Au, Cu, Ag, Rh, Pd, Ni, and Pt along the *y*-axis (Figure 3b). However, it is not obvious how this sequence of *X* components depends on elemental descriptors if any such meaningful relationship exists at all.

For the purpose of determining this relationship, machine learning played a crucial role. The sure independence screening and sparsifying operator (SISSO) method was applied,[55] in which a small number of simple descriptors [atomic number *Z*, period number *P*, group number *G*, metallic radius r_m (Å), covalent radius r_{co} (Å), and Pauling electronegativity EN_P] were considered. After ~4×10^7 combinations of these descriptors were screened, the optimum expression for the *y*-parameter was determined by SISSO to be

$$y = \frac{(\sqrt{Z(X)} \times P(X))/G(X)}{\exp(EN_P(X) \times r_{co}(X))}$$

The resulting structure map is based only on these two parameters and achieves the goal of classifying these *RX* intermetallics, with only two misclassifications. The *x*-parameter, with units of Å, is relatively simple, indicating that size [$r_m(R)$] and properties related to atomic number [$Z(R)$] are sufficient to capture the essential features of the *R* component; this parameter thus serves its purpose, and no attempt was made to search for more complicated expressions. The unitless *y*-parameter is more complex, with the addition of an electronegativity term (EN_P) in the denominator serving to convey information about the degree of electron transfer in the bonds between *R* and *X* atoms. This map provides chemical insight that resolves many of the previously unsettled questions, which we can now address, including confirmation from new experimental results.

2.3. Occurrence Rules for *RX* Intermetallics and High-Temperature Polymorphs. The CsCl-, TlI-, and FeB-type structures are now clearly delineated in the revised map (Figure 3b). Within the central part of this structure map, the triangular CsCl-type cluster is bounded by two green lines with slopes of $\pm k$ (where $k = 1.87 \times 10^{-3}$) and intersecting at $x = 103.5$ and $y = 0.159$. Starting from the middle row of *R*Ag compounds, the stability field of the CsCl-type structure increases proportionally to $\pm k$ until the *x*-parameter reaches 125. Beyond this point, the FeB-type structure becomes more stable for $x = 125 - 145$. Note, however, if the sloping lines are extended, the formation of high-temperature polymorphs for EuCu, YbAu, YbAg, and YbPd with CsCl-type structures can be rationalized.

Beyond the $\pm k$ limits, the *RX* intermetallics crystallize with TlI- or FeB-type structures in regions separated by orange lines with a slope of $-1/k$. The FeB-type structure is formed when the *x*- and *y*-parameters are both small or large, whereas the TlI-type structure is formed when one is large and the other is small. Where high-temperature polymorphs have been reported,[16] they tend to occur close to the boundary lines, exhibiting structures belonging to the closest neighboring fields. For example, among the TlI-type members, the aurides *R*Au (*R* = Gd−Tm) transform to CsCl-type and the silicides *R*Si (*R* = Dy−Er) transform to FeB-type structures at high temperatures. Conversely, among the FeB-type members, the aurides *R*Au (*R* = Pr, Nd, and Sm) transform successively to TlI- and CsCl-type structures at high temperatures.

2.4. Other Descriptors and Generalization to All *RX* Intermetallics. A few technical points are worth mentioning. In its regular classification mode, the SISSO algorithm does not lead to the easily interpretable structure map (Figure 3b). Instead, it tries to separate the *RX* entries into three clusters for the CsCl-, TlI-, and FeB-type structures, unfortunately not without overlap (Figure S3). Moreover, the elemental descriptors for *R* and *X* are intermingled in the derived parameters used to plot the *x* and *y* axes, making it impractical to apply this diagram to predict how chemical substitution affects structural transformations. If we insist that the derived parameters depend separately on *R* and *X*, there are still many combinations of descriptors that are possible, and it is interesting to compare how well they separate structure types (Figure S4). The optimized parameters depend on the metallic radius of *R* and the covalent radius of *X*. However, if the same type of radius scale is used for both *R* and *X*, or if they are interchanged, the structure types are not well separated. Out of all the combinations of descriptors screened by SISSO, two other derived parameters, $y = \frac{(Z(X)/\sqrt{r_{co}(X)})/G(X)}{\exp(EN_P(X) \times r_{co}(X))}$ (Å$^{-1/2}$) and $y = \frac{(Z(X)/\sqrt{r_m(X)})/G(X)}{\exp(EN_P(X) \times r_{co}(X))}$ (Å$^{-1/2}$), resulted in structure maps that perform nearly as well at classification as the one presented above. This is not surprising given that the expressions that appear in the numerator ($\sqrt{Z(X)} \times P(X)$, $Z(X)/\sqrt{r_{co}(X)}$, or $Z(X)/\sqrt{r_m(X)}$) are similar. Indeed, they can be shown to be mathematically equivalent because $r \propto Z/P^2$. The principle of parsimony favors the simpler parameter that depends only on an element's position in the periodic table (through the atomic number *Z* and the period number *P*) instead of its radius, which is a property that is not as easy to define or determine. Moreover, these alternative structure maps incorrectly classify USi as belonging to the TlI-type structure, and they lack the esthetic appeal of a symmetrical triangular shape for the CsCl-type cluster bounded by lines with a slope of $\pm k$.

The high accuracy of the structure map, which has only two misclassifications, was achieved with a relatively small set of descriptors. The results suggest that the model is optimum,

https://doi.org/10.1021/acs.chemmater.2c02425
Chem. Mater. 2023, 35, 879−890

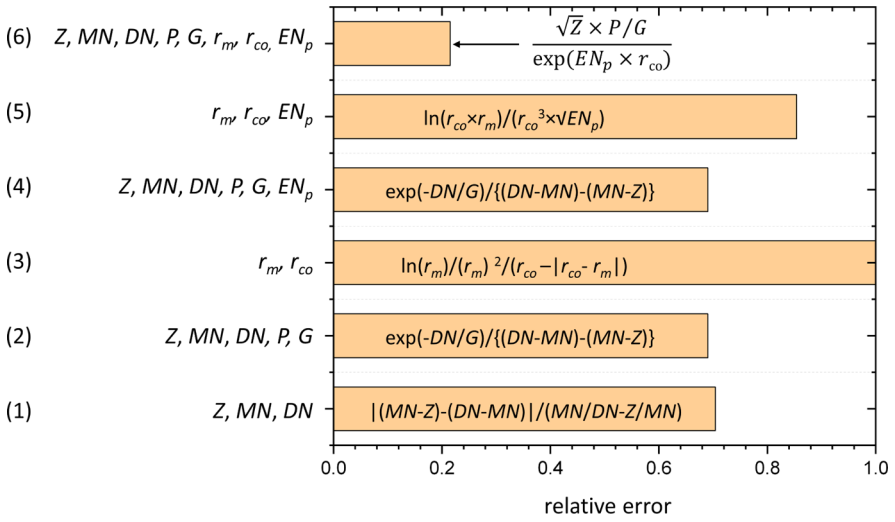


Figure 4. Comparison of relative errors in SISSO-generated models (1 to 6) based on a selection of atomic descriptors (left; abbreviations explained in the main text) and the optimized combination of these descriptors (mathematical expressions shown within the plot).

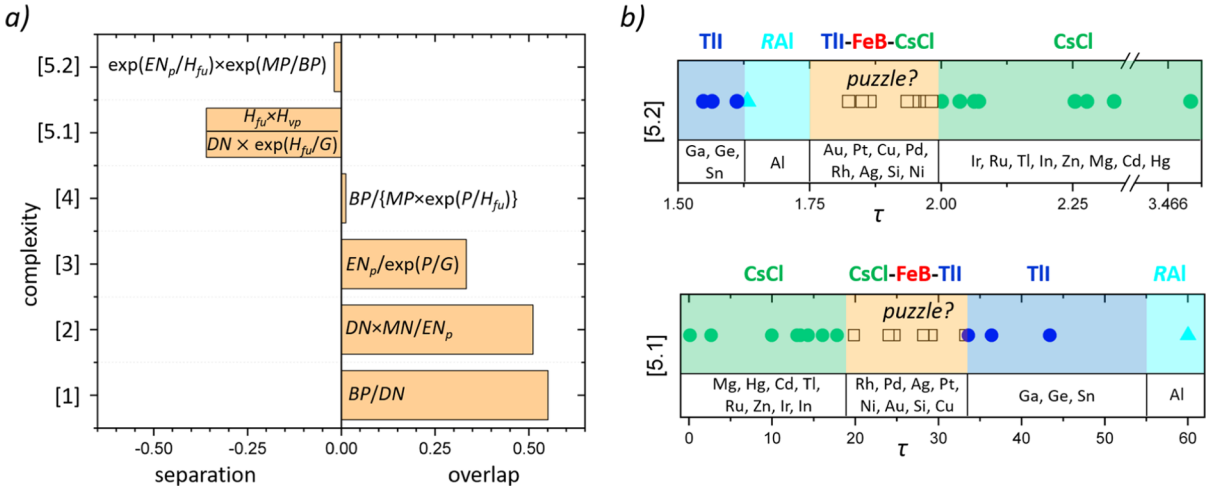


Figure 5. (a) Comparison of separation versus overlap of structure types in different SISSO-generated models (1 to 5.2) based on a parameter τ with different combinations of descriptors. (b) Distribution of structure types of *RX* intermetallics plotted over a τ-scale based on models 5.1 and 5.2.

and we can be reasonably confident that these descriptors have captured the essential factors that influence the crystal structures of *RX* intermetallics. The six descriptors considered here fall into three categories: (i) position in the periodic table (atomic number *Z*, period number *P*, and group number *G*); (ii) radii (metallic radius r_m and covalent radius r_{co}); and (iii) electronegativity scale (Pauling electronegativity EN_P), all for the element *X*. Besides examining different combinations of these descriptors, it is natural to wonder if selecting a subset of these or related descriptors is sufficient to classify the structures of *RX* intermetallics equally well. Several classification models were generated using the best combination of these types of descriptors, as optimized by the SISSO algorithm, and their performance was compared, as gauged by the relative errors (Figure 4). Models 1 and 2, which are based solely on the position of an element in the periodic table (atomic number *Z*, Mendeleev number *MN*, diagonal number *DN*, period number *P*, and group number *G*), do not perform well. Model 3, which is based solely on radii (metallic radius r_m and covalent radius r_{co}), performs even worse, consistent with expectations that the structures of such complex compounds cannot be governed by geometrical factors alone. Models 4 and 5, which combine element position descriptors or radii descriptors with Pauling electronegativity EN_P, do not give significant improvement. Unsurprisingly, none of these five models gives clear separation of the CsCl-, TlI-, and FeB-type structures comparable to the structure map of Figure 3b. Model 6, which combines all three types of descriptors, gives excellent performance with the lowest error. The optimized parameter, $y = \frac{(\sqrt{Z} \times P)/G}{\exp(EN_P \times r_{co})}$, is exactly the same as the one derived earlier and excludes *MN* and *DN*, even though they were considered as descriptors. The SISSO algorithm proved to be a powerful means to reveal hidden relationships, based on the most critical features.

The initial focus of this study was on *RX* intermetallics (*X* = Si, Au, Cu, Ag, Rh, Pd, Ni, and Pt) whose structures were the most jumbled (we call this the "puzzle" part of the picture) and appeared to violate the more regular patterns followed by other *RX* intermetallics (*X* = Mg, Ru, Ir, Zn, Cd, Hg, Al, Ga, In, Tl, Ge, and Sn) (Figure 2). The structure map of Figure 3b performs well in classifying the structure types within this

https://doi.org/10.1021/acs.chemmater.2c02425
Chem. Mater. 2023, 35, 879−890

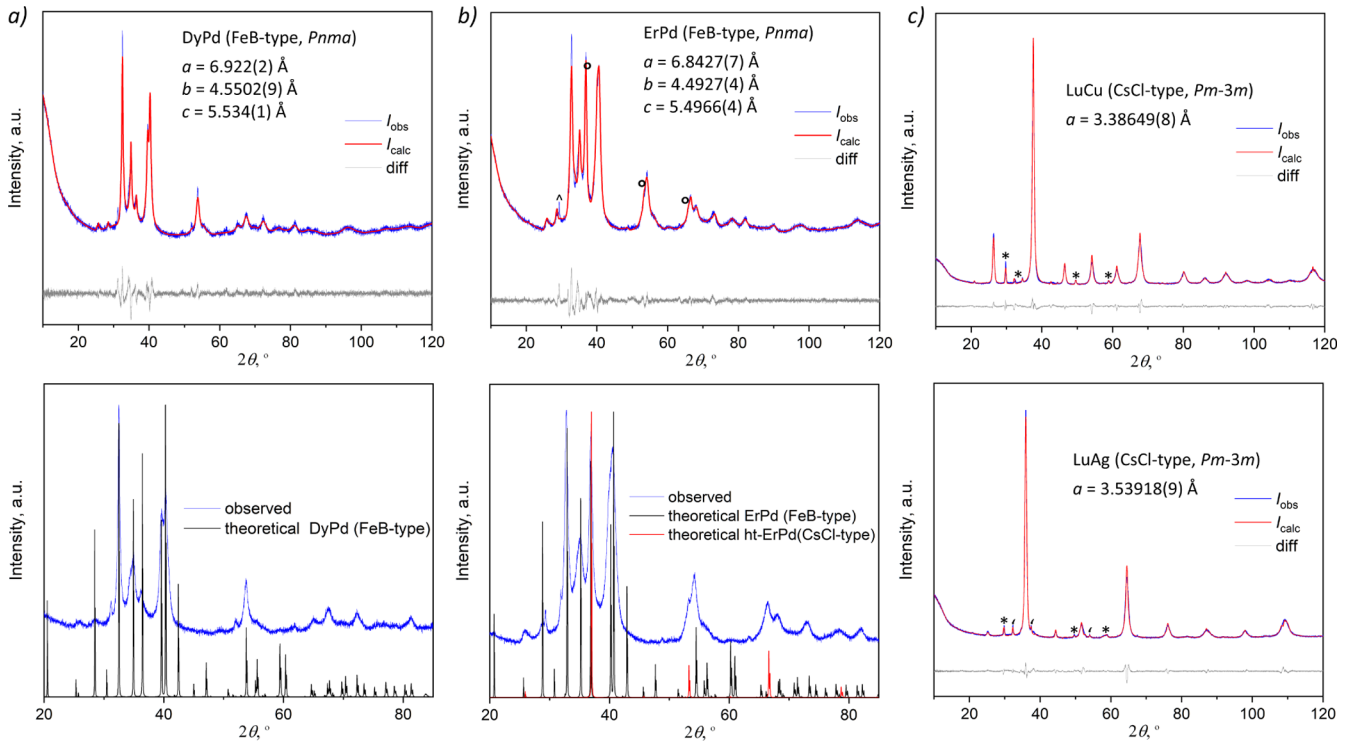


Figure 6. Powder XRD patterns with Le Bail fittings for (a) DyPd, (b) ErPd, and (c) LuCu and LuAg. In (a and b), the simulated patterns are also provided on the bottom panels. Traces of Er_2O_3 (^) and ~10 wt % of ht-ErPd (°) are present in the ErPd sample. Traces of Lu_2O_3 (*) are present in the LuCu and LuAg samples, and traces of Ag_2O (') are in the LuAg sample.

subset, but it would not be expected to be extrapolatable to the whole set of *RX* intermetallics because the model was not trained on them. It would be desirable to discover a coarser parameter τ that would be able to classify all *RX* intermetallics more broadly, while the finer parameter $y = \frac{(\sqrt{Z} \times P) / G}{\exp(EN_P \times r_{co})}$ (along with the x parameter) would be reserved to distinguish members within the puzzle region. The set of descriptors to be considered was enlarged to include the melting point *MP* (K), boiling point *BP* (K), heat of atomization H_{at} (kJ/mol), heat of vaporization H_{vp} (kJ/mol), heat of fusion H_{fu} (kJ/mol), specific heat H_{sp} (kJ/mol), atomic mass *W* (g/mol), and density ρ (tonne/m^3). Models were then developed based on SISSO-optimized combinations of these descriptors in varying levels of complexity in terms of mathematical operations allowed, and their ability to separate the structure types was compared (Figure 5a). The models with minimal complexity (1 to 4) led to a significant overlap of structure types. However, out of 2×10^8 combinations of descriptors tested, two models (5.1 and 5.2) gave stark segregation of CsCl, TlI, puzzle (CsCl−FeB−TlI), and the *R*Al structure types (Figure 5b). Between these two, model 5.1 is preferred because it sorts the structure types sequentially in one dimension according to $\tau = \frac{H_{fu} \times H_{vp}}{DN \times \exp(H_{fu} / G)}$ (kJ^2/mol^2) and gives the best magnitude of separation. The location of the puzzle region between the CsCl and TlI regions reflects the idea illustrated in the genesis of *RX* structure types (Figure 1), where the three structural variants (CsCl, FeB, and TlI) are possible and polymorphism is frequently encountered. We could even hypothesize that FeB is an intermediate structure ("twisted" nets) that evolves along the transformation from CsCl (flat nets) to TlI (buckled nets).

2.5. Structural Instabilities at Boundary Conditions.

Some of the *RX* intermetallics are located nearly or exactly on boundary lines, where structural instability may naturally be expected. Two examples can be highlighted: TbNi and *R*Pd (*R* = Dy, Ho, and Er). TbNi lies precisely on the orange line separating the TlI- and FeB-type structures, which explains why it crystallizes in its own unique structure type built up from fragments of these structures (Figure S5).[54] High-temperature polymorphs with the CsCl-type structure have been well established for *R*Pd (*R* = Dy, Ho, and Er).[53] Curiously, room-temperature polymorphs are also known but they were described as being metastable, poorly crystalline, and difficult to prepare as pure phases (always accompanied by mixtures of other phases).[53] They were tentatively assigned as having the FeB-type structure, but neither experimental powder X-ray diffraction (XRD) patterns nor cell parameters were provided. To revisit this problem, DyPd and ErPd were newly synthesized by arc melting, and they were indeed confirmed by their powder XRD patterns to exhibit poor crystallinity, making Rietveld refinements difficult to perform (Figure 6a,b). Instead, through Le Bail fitting, the lattice parameters could be refined based on an orthorhombic cell [a = 6.922(2) Å, b = 4.5502(9) Å, c = 5.534(1) Å for DyPd; a = 6.8427(7) Å, b = 4.4927(4) Å, c = 5.4966(4) Å for ErPd] and were found to be similar to those of the Pt analogues adopting FeB-type structures (a = 6.969 Å, b = 4.481 Å, c = 5.544 Å for DyPt; a = 6.906 Å, b = 4.451 Å, and c = 5.509 Å for ErPt).[56,57] The difference plots indicate that the fitting to the FeB-type model is not perfect. A more detailed investigation, perhaps through pair distribution function analysis with synchrotron radiation, may reveal further information about the local structure but is beyond the scope of the present study. It has been noted in the literature that the diffuse lines in the powder

https://doi.org/10.1021/acs.chemmater.2c02425
Chem. Mater. 2023, 35, 879−890

XRD pattern for DyPd remain unchanged even after prolonged annealing (4 months!), but the crystallinity of the FeB-type phase can be dramatically improved by substituting Pd with as little as 10% Pt in the solid solution $DyPd_{1-x}Pt_x$.[56] A similar structural stabilization of the FeB-type phase can be achieved in $DyPd_{1-x}Ni_x$, but this requires a higher Ni content ($x > 0.5$).[56] These observations can be understood by noting that DyPt lies much deeper than DyNi in the FeB-type stability field in the structure map, so that only a relatively smaller degree of Pt substitution is required to displace the position of DyPd away from the boundary line.

2.6. Structure Stabilization by a Third Component. The structure map provides a unified explanation for numerous examples in the literature of stabilizing CsCl-, TlI-, or FeB-type structures by adding a third component to a parent binary intermetallic *RX*. These include several series of *R*Ni phases: $Gd_{1-x}R_xNi$ (*R* = Dy−Tm, Lu; TlI → FeB), $Sm_{1-x}R_xNi$ (*R* = Dy−Er; TlI → FeB), and $PrPt_{1-x}Ni_x$ (FeB → TlI), and others.[16,56,58] The extensive focus on solid solutions derived from *R*Ni stems from an interest in their magnetic properties. For example, GdNi shows an extraordinary magnetoelastic behavior[59] and is well established to adopt the TlI-type structure.[52,60] Given its proximity to the TlI−FeB boundary line in the structure map, it could exhibit polymorphism. There is an old report of GdNi adopting the FeB-type structure,[61] which may be assumed to be a high-temperature polymorph (dashed red square in Figure 3b), but this result could not be reproduced by others.[52,59,60] Later, in the course of investigating the ternary Gd−Ni−Mn phase diagram at 803 K, an FeB-type phase for GdNi was reported, with a maximum solid solubility of 4 at % Mn.[62] Although no equiatomic GdMn phase exists, it would be hypothetically located at the top of the structure map (above the *R*Pt row), well within the FeB-type stability field. This provides a plausible explanation, in which partial replacement of Ni with Mn shifts the position of GdNi past the orange boundary line so that the FeB-type structure is stabilized. The earlier claim for an FeB-type phase of GdNi may have been an instance of impurity stabilization, imparted by the experimental conditions used in the original study.[52,60] We propose that examining this other polymorph in more detail could have important implications for improving the magnetoelastic properties of GdNi. Similarly, high-temperature polymorphs of TbPd and EuPd have been reported but not reproduced;[63,64] perhaps they may also be impurity-stabilized.

Other examples of *RX* intermetallics (*X* = Si, Au, Cu, Ag, Rh, Pd, Ni, and Pt) are known in which introducing a third component for *X* induces a structural transformation, and they can all be explained by shifts along the appropriate directions in the structure map: $TbSi_{1-x}Ge_x$ (FeB → TlI, upward),[65] $SmCu_{1-x}Mg_x$ (FeB → CsCl, upward),[66] $LaSi_{1-x}Ga_x$ (FeB → CsCl, upward),[67] and $ErPd_{1-x}In_x$ (FeB → CsCl, downward).[68] A particularly interesting case involves multiple phase transitions in $ErSi_{1-x}Al_x$ by substituting with Al (TlI → CsCl → FeB, upward),[52] but a re-entrant transition also occurs at $x = 0.5$ by increasing the temperature (FeB → TlI, neighboring cluster).[69]

2.7. Missing *RX* Intermetallics. When the distribution of structures of *RX* intermetallics was initially contemplated (Figure 3a), we could not find any information about LuCu and LuAg in experimental crystal structure databases. These cases are critical because their structures are not so obvious to predict by inspection of the neighboring data points in the plot. If these compounds are assumed to truly exist, then LuCu probably has a FeB-type structure, whereas LuAg is equally likely to have a CsCl- or FeB-type structure. Once the hidden patterns are revealed in the true structure map (Figure 3b), it becomes clear that both LuCu and LuAg are predicted to adopt a CsCl-type structure. We successfully synthesized LuCu and LuAg and confirmed the CsCl-type structure from their powder XRD patterns (Figure 6c). Because the samples are highly ductile, they were difficult to grind to fine powders. The intensities are slightly affected by preferred orientation (arising from thin plate shapes of microcrystallites), but from the fixed peak positions, the cubic cell parameters could be refined [a = 3.38649(8) Å for LuCu; a = 3.53918(9) Å for LuAg]. Subsequently, in the course of an extensive literature search in preparation of this manuscript, we found an old report that makes reference to work by A. E. Dwight claiming the existence of LuCu (a = 3.390 Å) and LuAg (a = 3.541 Å).[70] Because this work was never published, these compounds were not included in structural databases used to construct the machine learning model.

Among other missing *RX* intermetallics, EuNi, EuRh, EuPt, and nearly all the uranium-containing members U*X* (Figure 3b) may be difficult to synthesize, or perhaps they do not exist, so it would be worthwhile to pursue further experiments. Lastly, there are two outliers, EuPd (TlI type)[53] and YbRh (CsCl type),[71] which would have been expected to be FeB type and may require reinvestigation. LaSi (red star) crystallizes in its own structure type, closely resembling the FeB type (which is adopted by a high-temperature polymorph); this peculiarity may be connected with the unusual use of $LaBr_3$ in its synthesis.[72]

2.8. DFT Calculations. As computational power increases, it is now possible to apply first-principles calculations to help classify the crystal structures of many compounds, including *RX* intermetallics. Within 120 possible binary *R*−*X* systems, DFT calculations have been reported in the Open Quantum Materials Database (OQMD) for 83 of them, in which at least 2 out of 3 structure types (CsCl, TlI, and FeB) were considered.[21] Formation energies for these compounds in these structure types were compared. In 38 out of the 83 systems (or 46%), there is no uncertainty, with the alternative structures lying much higher in energy ($\Delta E \gg 0.05$ eV/atom) than the most stable one. It is very easy to identify the CsCl-type structure when it is most stable but nearly impossible to distinguish between the TlI- and FeB-type structures because their energy differences are generally minimal. This might suggest that polymorphic transformations between the TlI- and FeB-type structures are common. Further, in 45 out of the 83 systems (or 54%), the differences in DFT formation energies for CsCl-, TlI-, and FeB-type structures are rather small ($\Delta E <$ 0.05 eV/atom), so that polymorphism should be virtually guaranteed for all of them (Table S1). In reality, only 24 out of these 45 systems (about half) have shown experimental evidence for polymorphism. Out of the remaining 21 systems, there are 11 (Eu−Au, Eu−Ag, Eu−Pd, Lu−Ni, Tm−Pd, Gd−Pd, Gd−Ni, Nd−Cu, Nd−Rh, Pr−Cu, and La−Cu) in which our structure map predicts that polymorphs are likely because they are located close to boundary lines (Figure S6), and 10 (Lu−Pt, Yb−Pt, Tm−Ni, Er−Cu, Ho−Cu, Dy−Cu, Dy−Pt, Pr−Ag, La−Ag, and Ce−Si) in which polymorphs are *not* expected. If there are discrepancies between DFT and experiment, both deserve a closer evaluation. We have already discussed the cases of EuPd and GdNi earlier, representing

https://doi.org/10.1021/acs.chemmater.2c02425
Chem. Mater. 2023, 35, 879−890

Table 1. Classification Performance of Machine Learning Models[a]

	accuracy		precision		recall	
model	mean	max	mean	max	mean	max
SVC	0.86 (0.94)	0.86 (0.94)	0.88 (0.95)	0.88 (0.95)	0.86 (0.94)	0.86 (0.94)
DT	0.58 (0.72)	0.91 (0.81)	0.62 (0.74)	0.93 (0.81)	0.58 (0.72)	0.91 (0.81)
RF	0.89 (1.00)	0.96 (1.00)	0.92 (1.00)	0.96 (1.00)	0.87 (1.00)	0.96 (1.00)

[a]Values given are for the test set, and in parentheses are the training set values.

true positives, and it will now be incumbent to test the predictions for true negatives.

The poor correlation of DFT in computational databases with actual experimental results can be explained by factors such as the use of frozen core potentials for lanthanides, the exclusion of spin−orbit coupling, and the occurrence of magnetic ordering in the systems.[73] In other computational materials databases, such as the Materials Project, the same challenges of choosing appropriate functionals are anticipated.[23] It could well be that f-electrons do play a role in bonding, influencing the crystal structures. An evaluation of these systems will require treatment of electron correlation using appropriate functionals and a DFT + U approach, which uses an additional Hubbard-like term to describe the strongly correlated 4f electrons, while other valence electrons are treated normally with DFT approximations.[74,75] This also presents more challenges such as the selection of the Hubbard U parameter for DFT + U and the percentage of Fock exchange in the case of hybrid functionals.[75] Moreover, to validate the observation of high-temperature polymorphs, nonstandard conditions (beyond 0 K) need to be considered. This treatment is beyond the scope of the present work.

2.9. Other Machine Learning Approaches. Three other commonly used machine learning algorithms—support vector classifier (SVC), decision tree (DT), and random forest (RF)[76]—were applied to evaluate how well they classify the structures of *RX* intermetallics. In SVC, data with *K*-dimensional features are separated into classes by the best hyperplane (having dimensions of $K-1$). In DT, a tree-like flowchart is grown by finding the best features to minimize the classification error, with threshold splitting points at each node. The lower tree depths are easily visualized and interpreted in terms of these features. In RF, an ensemble of decision trees is built by training each tree with a different subset of training data and features. Then, the ensemble's collective knowledge is used to assign a class by majority vote.

The models were trained using exactly the same set of elemental descriptors (atomic number, period, group, radii, and electronegativity) as in the SISSO approach. The data set of *RX* intermetallics (consisting of 29 CsCl-, 33 TlI-, and 44 FeB-type entries) was split into training (80%) and test (20%) sets using stratified splitting on these structure types. The features were scaled by subtracting their mean values and dividing by their standard deviations (for SVC and RF). For each model, the hyperparameters were optimized (Table S2) using a cross-validated grid search across predefined parameter space. The randomized nature inherent in the selection of feature subsets (for DT and RF) and data (for RF only) introduces run-to-run variations, so the training and evaluation steps were performed 100 times, and those models with high test accuracies were selected for the comparison below. All models showed very good performance, with RF ranking first and DT ranking last (Table 1). The probabilities predicted by the RF model for all samples are plotted (Figure 7). Recall that

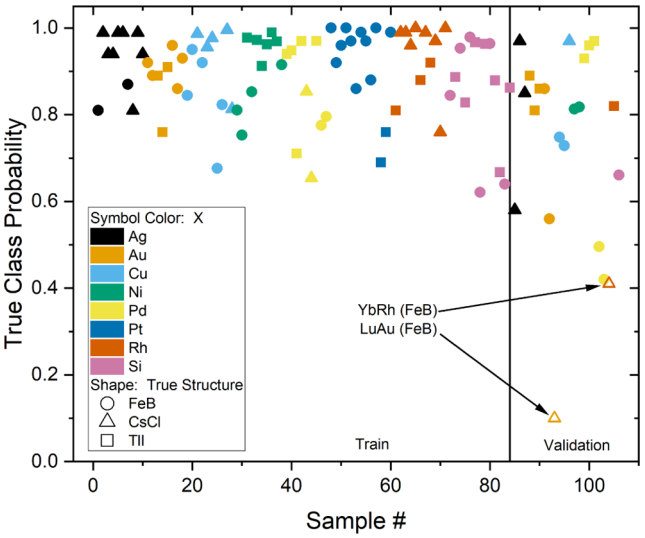


Figure 7. True class probabilities for *RX* intermetallics predicted by the RF model. The colors indicate the *X* element. The vertical line separates the training (left) from the test set (right). There are two false negatives (open symbols), YbRh and LuAu, incorrectly predicted as FeB-type but actually adopting the CsCl-type structure.

LuCu and LuAg were missing from the data set but both were subsequently confirmed by experimental synthesis to adopt the CsCl-type structure. All three models correctly predict the CsCl-type structure for both compounds.

The DT model does not perform well compared to SVC and RF, but it has the merit of being easily interpreted (Figure 8). This model was trained using features without any scaling applied, and the minimum number of samples at the tree leaves was set to 3. Some generalizations can be inferred from the decision tree. The CsCl-type structure is favored by large but lighter *X* atoms [$r_{co}(X) > 1.33$ Å, $Z(X) \leq 63$]. Nearly half of the TlI-type phases consist of later *R* elements [Tb or later, $Z(R) \geq 65$] in combination with *X* atoms of intermediate size [1.16 Å < $r_{co}(X) \leq 1.33$ Å]. Despite being more accurate, the SVC and RF models offer little "out-of-the-box" insight. As a way to extract some insight from these models, the permutation importance of chemical features can be evaluated (Figure S7).[77] First, the accuracy is calculated with correct feature values associated with the samples. Then, the data set is permuted by shuffling the feature values so that they are no longer correctly associated with the samples. If the accuracy is only minimally affected, then it can be concluded that a feature is not important to the model. Among SVC, DT, and RF models, there is consensus that the most important features are the atomic number of *R* and the covalent radius of *X*, whereas the least important is the period number of *X*. However, there is no simple, explicit relationship between these features and the structure type adopted, unlike the SISSO-generated structure map.

https://doi.org/10.1021/acs.chemmater.2c02425
Chem. Mater. 2023, 35, 879−890

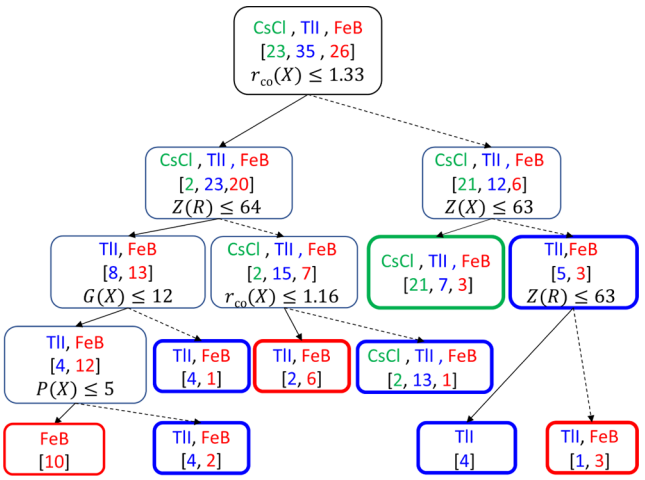


Figure 8. Decision tree to predict structures of *RX* intermetallics, trained without feature scaling. The number of entries of each structure type in the training data set is shown in brackets. Solid arrows lead to the subset of entries satisfying the condition and dashed arrows to those not satisfying. The final output may contain a mixture of structure types but the classification is based on the majority representative.

In comparison to these conventional machine learning models, the structure map constructed with the aid of chemical intuition (Figure 3b) offers practical advantages. All of the models are capable of addressing the question, "What?", but the structure map also addresses the questions, "Why and how?" An experimentalist would undoubtedly find comfort in its greater ease of interpretability to guide the design of new materials by answering chemically meaningful questions to understand when the CsCl-, TlI-, or FeB-type structures are favored, why high-temperature polymorphs can easily occur for some *RX* compounds but not others, why TbNi and *R*Pd (*R* = Dy, Ho, and Er) experience instability, and how introducing a third component can stabilize a different structure. This map thus encourages new hypotheses to be formulated. If machine learning models are to be embraced more widely by chemists and materials scientists, continued effort should be made to ensure that they are interpretable. In the end, "all models are wrong, but some are useful."[78]

3. CONCLUSIONS

By combining chemical intuition with the SISSO approach to identify the best combination of elemental descriptors, a structure map was developed that reveals the hidden pattern of CsCl-, TlI-, and FeB-type structures among *RX* intermetallics. This model serves as a powerful tool to classify the structures of known compounds and to predict missing ones, some of which were experimentally confirmed here. It provides convincing rationalization for many previous observations about these compounds that have long eluded a unified explanation, including the occurrence of polymorphs and structural instability at boundary conditions. It also suggests strategies for stabilizing desired structures by introduction of a third component. Compared to DFT calculations and more opaque machine learning methods, it performs as well or better, while having the merits of being chemically interpretable. This case study demonstrates "proof of principle" for relatively simple binary phases and serves as a first step to extending this idea to more complex multinary phases.

4. EXPERIMENTAL SECTION

4.1. Synthesis and Powder XRD. Starting materials were pieces of rare-earth metals (Dy, Er, Lu; 99.99%, HEFA), palladium (99.95%, Alfa-Aesar), copper (99.9%, Anachemia), and silver (99.9%, Sigma-Aldrich). Various binary rare-earth intermetallics *RX* were prepared by combining the elemental components in equimolar ratios with a total mass of 0.2 g. The mixtures were arc melted on a water-cooled copper hearth under high-purity argon gettered with titanium in an Edmund Bühler MAM-1 compact arc melter. To ensure homogeneity, the samples were melted three times.

Powder XRD patterns of the ground polycrystalline samples were collected on a Bruker D8 Advance powder diffractometer equipped with a Cu *Kα* radiation source operated at 40 kV and 40 mA. The samples were placed on a zero-background plate holder. Phases were checked against entries in Pearson's Crystal Data.[16] Full-profile Le Bail fitting was applied to the patterns, and cell parameters were refined using the FullProf software suite.[79,80] Crystal structures were visualized using the program VESTA 3.[81]

4.2. Computational Methods. SVC, RF, and DT were used as implemented in the Scikit-learn (version 0.24.2) Python package,[76] and SISSO (version 3.0.2) was used for the SISSO calculations.[55] For SVC and RF, features were standardized by subtracting the mean and then dividing by their standard deviation. For DT, the features were used without any modification for easier interpretability. The hyperparameters for the machine learning models were optimized via grid search with fivefold cross-validation. The optimized features are given in Table S2. The feature importance was calculated using the permutation importance method, as implemented in Scikit-learn. To find the *y*-axis parameter of Figure 3b, numerical values corresponding to the order of the *X* elements (Si = 1, Au = 2, Cu = 3, Ag = 4, Rh = 5, Pd = 6, Ni = 7, and Pt = 8) were assigned, and a SISSO regression was then performed using Pauling electronegativity, atomic number, group, period, covalent radii (Å), and metallic radii (Å) of the elements as features.

All input and output files for the machine learning and SISSO calculations are available in electronic form at figshare.com (doi = 10.6084/m9.figshare.19895788.v5).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemmater.2c02425.

Relative DFT formation energies for *RX* intermetallics, hyperparameters used for machine learning models, literature review of papers on equiatomic *AB* compounds, crystal structures of rare-earth aluminides *R*Al and polymorphs of TbNi, other models for classification of *RX* intermetallics, and permutation importance of features (PDF)

Average metrics and feature importance (PDF)

Classification of *RX* compounds by machine learning algorithms (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Volodymyr Gvozdetskyi − *Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada*; Email: gvozdets@ualberta.ca

Arthur Mar − *Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada*; orcid.org/0000-0003-0474-5918; Email: amar@ualberta.ca

Authors

Balaranjan Selvaratnam − *Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada*

https://doi.org/10.1021/acs.chemmater.2c02425
Chem. Mater. 2023, 35, 879−890

Anton O. Oliynyk − *Department of Chemistry, University of Alberta, Edmonton, Alberta T6G 2G2, Canada; Department of Chemistry and Biochemistry, Manhattan College, Riverdale, New York 10471, United States;* orcid.org/0000-0003-0732-7340

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemmater.2c02425

Author Contributions

The manuscript was written through contributions of all authors. All authors have given their approval for the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC, through Discovery Grant RGPIN-2018-04294), the Canada First Research Excellence Fund (CFREF, through the Future Energy Systems Research Institute at the University of Alberta, Project no. T12-P01), and resources available through Digital Research Alliance of Canada (www.computecanada.ca). We thank Dr. Lawrence Adutwum for the helpful discussions.

■ REFERENCES

(1) Oliynyk, A. O.; Mar, A. Discovery of intermetallic compounds from traditional to machine-learning approaches. *Acc. Chem. Res.* **2018**, *51*, 59−68.
(2) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547−555.
(3) Alberi, K.; Nardelli, M. B.; Zakutayev, A.; Mitas, L.; Curtarolo, S.; Jain, A.; Fornari, M.; Marzari, N.; Takeuchi, I.; Green, M. L.; Kanatzidis, M.; Toney, M. F.; Butenko, S.; Meredig, B.; Lany, S.; Kattner, U.; Davydov, A.; Toberer, E. S.; Stevanovic, V.; Walsh, A.; Park, N.-G.; Aspuru-Guzik, A.; Tabor, D. P.; Nelson, J.; Murphy, J.; Setlur, A.; Gregoire, J.; Li, H.; Xiao, R.; Ludwig, A.; Martin, L. W.; Rappe, A. M.; Wei, S.-H.; Perkins, J. The 2019 materials by design roadmap. *J. Phys. D: Appl. Phys.* **2019**, *52*, 013001.
(4) Chen, C.; Zuo, Y.; Ye, W.; Li, X.; Deng, Z.; Ong, S. P. A critical review of machine learning of energy materials. *Adv. Energy Mater.* **2020**, *10*, 1903242.
(5) Lotfi, S.; Brgoch, J. Discovering intermetallics through synthesis, computation, and data-driven analysis. *Chem. Eur. J.* **2020**, *26*, 8689−8697.
(6) Ribeiro, M. T.; Singh, S.; Guestrin, C.Why should I trust you?. Explaining the Predictions of Any Classifier. arXiv preprint **2016**, arXiv:1602.04938v3 [cs.LG].
(7) Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206−215.
(8) Hoffmann, R.; Malrieu, J.-P. Simulation vs. understanding: A tension, in quantum chemistry and beyond. Part B. The march of simulation, for better or worse. *Angew. Chem., Int. Ed.* **2020**, *59*, 13156−13178.
(9) Oviedo, F.; Ferres, J. L.; Buonassisi, T.; Butler, K. T.Interpretable and Explainable Machine Learning for Materials Science and Chemistry. arXiv preprint **2021**, arXiv:2111.01037v2 [cond-mat.mtrl-sci].
(10) Rudin, C.; Chen, C.; Chen, Z.; Huang, H.; Semenova, L.; Zhong, C. Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statist. Surv.* **2022**, *16*, 1−85.
(11) Pauling, L.*The Nature of the Chemical Bond*, 3rd ed.; Cornell University Press: Ithaca, NY, 1960; pp 505−562.
(12) George, J.; Waroquiers, D.; Di Stefano, D.; Petretto, G.; Rignanese, G.-M.; Hautier, G. The limited predictive power of the Pauling rules. *Angew. Chem., Int. Ed.* **2020**, *59*, 7569−7575.
(13) Goldschmidt, V. M. Die Gesetze der Krystallochemie. *Naturwissenschaften* **1926**, *14*, 477−485.
(14) Bartel, C. J.; Sutton, C.; Goldsmith, B. R.; Ouyang, R.; Musgrave, C. B.; Ghiringhelli, L. M.; Scheffler, M. New tolerance factor to predict the stability of perovskite oxides and halides. *Sci. Adv.* **2019**, *5*, No. eaav0693.
(15) Graf, T.; Felser, C.; Parkin, S. S. P. Simple rules for the understanding of Heusler compounds. *Prog. Solid State Chem.* **2011**, *39*, 1−50.
(16) Villars, P.; Cenzual, K.*Pearson's Crystal Data − Crystal Structure Database for Inorganic Compounds (On DVD)*; ASM International, 2021Materials Park, OH, USA. Release 2021/22.
(17) Bergerhoff, G.; Brown, I. D.Inorganic crystal structure database. In *Crystallographic Databases*; Allen, F. H., Bergerhoff, G., Sievers, R., Eds.; International Union of Crystallography: Chester, U.K., 1987.
(18) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr. Sect. B* **2016**, *72*, 171−179.
(19) Gates-Rector, S. D.; Blanton, T. N. The Powder Diffraction File: a quality materials characterization database. *Powder Diffr.* **2019**, *34*, 352−360.
(20) Villars, P.; Berndt, M.; Brandenburg, K.; Cenzual, K.; Daams, J.; Hulliger, F.; Okamoto, H.; Osaki, K.; Prince, A.; Putz, H.; Iwata, S. The Pauling File. *Mater. Sci. Forum* **2004**, *443-444*, 357−360.
(21) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials design and discovery with high-throughput density functional theory: The Open Quantum Materials Database (OMQD). *JOM* **2013**, *65*, 1501−1509.
(22) Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R. H.; Nelson, L.; Sanvito, G. L. W.; Buongiorno-Nardelli, S.; Mingo, M.; Levy, N.; Levy, O. AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **2012**, *58*, 227−235.
(23) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.
(24) Ortiz, C.; Eriksson, O.; Klintenberg, M. Data mining and accelerated electronic structure theory as a tool in the search for new functional materials. *Comput. Mater. Sci.* **2009**, *44*, 1042−1049.
(25) Landis, D. D.; Hummelshøj, J. S.; Nestorov, S.; Greeley, J.; Dułak, M.; Bligaard, T.; Nørskov, J. K.; Jacobsen, K. W. The Computational Materials Repository. *Comput. Sci. Eng.* **2012**, *14*, 51−57.
(26) Põhls, J.-H.; Heyberger, M.; Mar, A. Comparison of computational and experimental inorganic crystal structures. *J. Solid State Chem.* **2020**, *290*, 121557.
(27) Therrien, F.; Jones, E. B.; Stevanović, V. Metastable materials discovery in the age of large-scale computation. *Appl. Phys. Rev.* **2021**, *8*, 031310.
(28) Pettifor, D. G. A chemical scale for crystal structure maps. *Solid State Commun.* **1984**, *51*, 31−34.
(29) Pettifor, D. G.; Podloucky, R. Microscopic theory of the structural stability of pd-bonded AB compounds. *Phys. Rev. Lett.* **1984**, *53*, 1080−1083.
(30) Pettifor, D. G. The structures of binary compounds: I. Phenomenological structure maps. *J. Phys. C: Solid State Phys.* **1986**, *19*, 285−313.
(31) Pettifor, D. G.; Podloucky, R. The structures of binary compounds: II. Theory of the pd-bonded AB compounds. *J. Phys. C: Solid State Phys.* **1986**, *19*, 315−331.
(32) Zunger, A. Systematization of the stable crystal structure of all AB-type binary compounds: A pseudopotential orbital-radii approach. *Phys. Rev. B* **1980**, *22*, 5839−5872.
(33) Sproud, G. D. Evaluation of electronegativity scales. *ACS Omega* **2020**, *5*, 11585−11594.

(34) Dschemuchadse, J.; Steurer, W. Some statistics on intermetallic compounds. *Inorg. Chem.* **2015**, *54*, 1120−1128.
(35) Dschemuchadse, J.; Steurer, W. More statistics on intermetallic compounds. *Acta Crystallogr., Sect. A* **2015**, *71*, 335−345.
(36) Hever, A.; Oses, C.; Curtarolo, S.; Levy, O.; Natan, A. The structure and composition statistics of 6A binary and ternary crystalline materials. *Inorg. Chem.* **2018**, *57*, 653−667.
(37) Saal, J. E.; Oliynyk, A. O.; Meredig, B. Machine learning in materials discovery: Confirmed predictions and their underlying approaches. *Annu. Rev. Mater. Res.* **2020**, *50*, 49−69.
(38) Kauwe, S. K.; Graser, J.; Murdock, R.; Sparks, T. D. Can machine learning find extraordinary materials? *Comput. Mater. Sci.* **2020**, *174*, 109498.
(39) Villars, P. A three-dimensional structural stability diagram for 998 binary AB intermetallic compounds. *J. Less-Common Met.* **1983**, *92*, 215−238.
(40) Saad, Y.; Gao, D.; Ngo, T.; Bobbitt, S.; Chelikowsky, J. R.; Andreoni, W. Data mining for materials: Computational experiments with AB compounds. *Phys. Rev. B* **2012**, *85*, 104104.
(41) Pilania, G.; Gubernatis, J. E.; Lookman, T. Structure classification and melting temperature prediction in octet AB solids via machine learning. *Phys. Rev. B* **2015**, *91*, 214302.
(42) Oliynyk, A. O.; Adutwum, L.; Mar, J. J.; Mar, A. Classifying crystal structures of binary compounds AB through cluster resolution feature selection and support vector machine analysis. *Chem. Mater.* **2016**, *28*, 6672−6681.
(43) Gutfleisch, O.; Willard, M. A.; Brück, E.; Chen, C. H.; Sankar, S. G.; Liu, J. P. Magnetic materials and devices for the 21st century: Stronger, lighter, and more energy efficient. *Adv. Mater.* **2011**, *23*, 821−842.
(44) Singh, P.; Del Rose, T.; Vazquez, G.; Arroyave, R.; Mudryk, Ya. Machine-learning enabled thermodynamic model for the design of new rare-earth compounds. *Acta Mater.* **2022**, *229*, 117759.
(45) Meyer, G.; Hwu, S.-J.; Wijeyesekera, S.; Corbett, J. D. Synthetic study of some rare-earth-metal monohalide hydrides MXH_x and their alkali-metal intercalates. *Inorg. Chem.* **1986**, *25*, 4811−4818.
(46) Petit, L.; Szotek, Z.; Lüders, M.; Svane, A. Rare-earth pnictides and chalcogenides from first-principles. *J. Phys.: Condens. Matter* **2016**, *28*, 223001.
(47) Buschow, K. H. J.; van Vucht, J. H. N. Systematic arrangement of the binary rare-earth aluminium systems. *Philips Res. Rep.* **1967**, *22*, 233−245.
(48) "We should totally buy a bar." "Our bar would be awesome." "And dude, dude, dude, DUDE ... the name of our bar ... Puzzles. People will be, like, 'Why is it called Puzzles?' That's the puzzle!" in How I Met Your Mother, Season 4, *Episode13*, 2009.
(49) Gladyshevskii, E. I.; Kripyakevich, P. I. Monosilicides of rare earth metals and their crystal structures. *Zh. Strukt. Khim.* **1964**, *5*, 853−859.
(50) Schob, O.; Parthé, E. AB compounds with Sc, Y, and rare earth metals. I. Scandium and yttrium compounds with CrB and CsCl structure. *Acta Crystallogr* **1965**, *19*, 214−224.
(51) Hohnke, D.; Parthé, E. AB compounds with Sc, Y, and rare earth metals. II. FeB and CrB type structures of monosilicides and germanides. *Acta Crystallogr* **1966**, *20*, 572−582.
(52) Raman, A. Ternary FeB- and CrB-type phases with yttrium, lanthanum, and some rare earths. *Inorg. Chem.* **1968**, *7*, 973−976.
(53) Palenzona, A.; Cirafici, S. Thermodynamic and crystallographic properties of REPd intermetallic compounds. *Thermochim. Acta* **1975**, *12*, 267−275.
(54) Lemaire, R.; Paccard, D. Polymorphisme du composé TbNi. *J. Less-Common Met.* **1970**, *21*, 403−413.
(55) Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M. SISSO. A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Mater.* **2018**, *2*, 083802.
(56) Klepp, K.; Parthé, E. Phase relationship of ternary rare earth-transition metal alloys with CrB and FeB structures or stacking variants. *J. Less-Common Met.* **1982**, *85*, 181−194.
(57) Castets, A.; Gignoux, D.; Gomez-Sal, J. C. Magnetic properties and structures of equiatomic rare earth-platinum compounds RPt (R = Gd, Tb, Dy, Ho, Er, Tm). *J. Solid State Chem.* **1980**, *31*, 197−207.
(58) Pospíšil, J.; Poltierová Vejpravová, J.; Diviš, M.; Sechovský, V. Electronic structure and magnetism of $PrNi_xPt_{1-x}$ compounds. *J. Alloys Compd.* **2008**, *450*, 118−127.
(59) Paudyal, D.; Mudryk, Ya.; Lee, Y. B.; Pecharsky, V. K.; Gschneidner, K. A.; Harmon, B. N. Understanding the extraordinary magnetoelastic behavior in GdNi. *Phys. Rev. B* **2008**, *78*, 184436.
(60) Dwight, A. E.; Conner Jnr, R. A., Jr.; Downey, J. W. Equiatomic compounds of the transition and lanthanide elements with Rh, Ir, Ni and Pt. *Acta Crystallogr* **1965**, *18*, 835−839.
(61) Baenziger, N. C.; Moriarty Jnr, J. L., Jr. Gadolinium and dysprosium intermetallic phases. I. The crystal structures of DyGa and GdPt and their related compounds. *Acta Crystallogr* **1961**, *14*, 946−947.
(62) Zhuang, Y.; Ding, R.; Liu, J.; Ou, X.; Deng, H. The isothermal section of the Gd−Mn−Ni ternary system at 803 K. *J. Alloys Compd.* **2002**, *346*, 181−183.
(63) Inagawa, K.; Watanabe, K.; Kaneko, T.; Ido, H.; Watanabe, H. Crystal Structure of RPd (R=Tb, Ho, Yb). *J. Phys. Soc. Jpn.* **1974**, *36*, 1709.
(64) Longworth, G.; Harris, I. R. Mössbauer and X-ray studies of the magnetic properties and crystal structure of the phase PdEu. *J. Less-Common Met.* **1973**, *33*, 83−90.
(65) Schobinger-Papamantellos, P.; Buschow, K. H. J. Magnetic structure changes observed by neutron diffraction in the system $TbGe_{1-x}Si_x$ ($0.4 \leq x \leq 1.0$). *J. Magn. Magn. Mater.* **1988**, *71*, 134−146.
(66) Marciniak, B.; Pavlyuk, V.; Rozycka-Sokolowska, E.; Karwowski, L.; Bak, Z. The isothermal section of the phase diagram of Sm−Cu−Mg ternary system at 670 K. *J. Alloys Compd.* **2015**, *652*, 254−259.
(67) Dürr, I.; Bauer, B.; Rőhr, C. Lanthan-Triel/Tetrel-ide $La(Al,Ga)_x(Si,Ge)_{1-x}$. Experimentelle und theoretische Studien zur Stabilität intermetallischer 1:1-Phasen. *Z. Naturforsch. B* **2011**, *66*, 1107−1121.
(68) Giovannini, M.; Saccone, A.; Delfino, S.; Rogl, P. A comparative investigation of isothermal sections of rare earth-Pd−In systems. *Intermetallics* **2003**, *11*, 1237−1243.
(69) Pukas, S.; Łasocha, W.; Gladyshevskii, R. Phase equilibria in the Er−Al−Si system at 873 K. *CALPHAD: Comput. Coupling Phase Diagrams Thermochem* **2009**, *33*, 23−26.
(70) Nevitt, M. V.Alloy chemistry of transition elements. In *Electronic Structure and Alloy Chemistry of the Transition Elements*; Beck, P. A., Ed.; Interscience: New York, 1963, pp 101−178.
(71) Iandelli, A.; Palenzona, A. Behavior of ytterbium with metals of the eighth group of periodic table. *Rev. Chim. Miner.* **1976**, *13*, 55−61.
(72) Mattausch, H.; Oeckler, O.; Simon, A. Eine neue Modifikation von Lanthanmonosilicid − lT-LaSi. *Z. Anorg. Allg. Chem.* **1999**, *625*, 1151−1154.
(73) Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater.* **2015**, *1*, 15010.
(74) Tolba, S. A.; Gameel, K. M.; Ali, B. A.; Almossalami, H. A.; Allam, N. K.*The DFT+U: Approaches, Accuracy, and Applications*. In *Density Functional Calculations - Recent Progresses of Theory and Application*; Yang, G., Ed.; IntechOpen: London, 2018.
(75) Himmetoglu, B.; Floris, A.; de Gironcoli, S.; Cococcioni, M. Hubbard-corrected DFT energy functionals: The LDA+U description of correlated systems. *Int. J. Quantum Chem.* **2014**, *114*, 14−49.
(76) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(77) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.
(78) Box, G. E. P.; Draper, N. R.*Empirical Model-Building and Response Surfaces*; Wiley: New York, 1987, p 424.

https://doi.org/10.1021/acs.chemmater.2c02425
Chem. Mater. 2023, 35, 879−890

(79) Le Bail, A. Whole powder pattern decomposition methods and applications: A retrospection. *Powder Diffr.* **2005**, *20*, 316−326.
(80) Rodríguez-Carvajal, J. Recent developments of the program FULLPROF. *Commission on Powder Diffraction (IUCr) Newsletter* **2001**, *26*, 12−19.
(81) Momma, K.; Izumi, F. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Crystallogr.* **2011**, *44*, 1272−1276.

https://doi.org/10.1021/acs.chemmater.2c02425
Chem. Mater. 2023, 35, 879−890