

High-Throughput Machine-Learning-Driven Synthesis of Full-Heusler Compounds

Anton O. Oliynyk,[†] Erin Antono,[‡] Taylor D. Sparks,[§] Leila Ghadbeigi,[§] Michael W. Gaultois,[¶]
Bryce Meredig,^{‡,*} and Arthur Mar^{†,*}

[†] *Department of Chemistry, University of Alberta, Edmonton, AB T6G 2G2 Canada*

[‡] *Citrine Informatics, Redwood City, CA 94063 USA*

[§] *Department of Materials Science and Engineering, University of Utah, Salt Lake City, UT
84112 USA*

[¶] *Department of Chemistry, University of Cambridge, Cambridge CB2 1EW United Kingdom*

Abstract

A machine-learning model has been trained to discover Heusler compounds, which are intermetallics exhibiting diverse physical properties attractive for applications in thermoelectric and spintronic materials. Improving these properties requires knowledge of crystal structures, which occur in three subtle variations (Heusler, inverse Heusler, and CsCl-type structures) that are difficult, and at times impossible, to distinguish by diffraction techniques. Compared to alternative approaches, this Heusler discovery engine performs exceptionally well, making fast and reliable predictions of the occurrence of Heusler vs non-Heusler compounds for an arbitrary combination of elements with no structural input on over 400,000 candidates. The model has a true positive rate of 0.94 (and false positive rate of 0.01). It is also valuable for data sanitizing, by flagging questionable entries in crystallographic databases. It was applied to screen candidates with the formula AB_2C and predict the existence of 12 novel gallides MRu_2Ga and RuM_2Ga ($M = Ti-Co$) as Heusler compounds, which were confirmed experimentally. One member, $TiRu_2Ga$, exhibited diagnostic superstructure peaks that confirm the adoption of an ordered Heusler as opposed to a disordered CsCl-type structure.

1. Introduction

Heusler compounds form a large class of intermetallics that exhibit versatile properties. The first compound, Cu_2MnAl , was discovered in 1898 by Friedrich (“Fritz”) Heusler, a German mining engineer, and attracted attention – before its structure or composition was understood – because it is ferromagnetic even though it is formed from metals that are nonferromagnetic.^{1,2} There seem to be simple rules for relating the electron count to the physical properties,³ permitting the prediction of half-metallic ferromagnets,⁴ ferrimagnets,⁵ semiconductors,⁶ and superconductors.^{7,8} Given their tunable semiconducting properties (made possible by adjusting the chemical composition to attain a desired electron count so that the band gap varies from 0 to a few eV),³ these compounds are currently being heavily investigated for sustainable technologies such as solar energy and thermoelectric conversion.^{9,10} Exciting new applications for Heusler compounds include spintronics,^{11,12} superconductivity,^{7,8} magnetocalorics,¹³ and topological insulators.¹⁴ Thus, these compounds are advancing the frontiers of science and providing solutions to materials engineering challenges in the future.

To gauge interest in this area, a plot of Heusler compounds reported structurally (culled from Pearson’s Database¹⁵) shows a peak in the 1980’s, when magnetic properties were examined, followed by a recent renaissance, when exotic properties were discovered (Figure 1). In Heusler’s time, these compounds were thought to be solid solutions adopting the structure of one of the metal components.² More than 20 years passed before the first crystal structure was elucidated,¹⁶ and many years still before an appreciation of the subtleties was attained. There are two families of Heusler compounds: half-Heusler compounds ABC , and (full-)Heusler compounds AB_2C . The components are metals, where typically A is a large electropositive metal, B is a transition metal, and C is an electronegative metal (usually a p-block metalloid).

We focus our attention on the latter, referred to simply as Heusler compounds. The Heusler structure (also called Cu_2MnAl -type) is a superstructure of the CsCl-type (Figure 2). In AB_2C , the B atoms form a primitive cubic sublattice; the A and C atoms are arranged in a disordered fashion within body centres in the CsCl-type but in an ordered fashion in the Heusler structure. The CsCl-type structure contains two sites within a primitive cubic lattice (space group $Pm\bar{3}m$) whereas the Heusler structure contains three sites within a face-centered cubic lattice (space group $Fm\bar{3}m$). There is a complication: inverse Heusler compounds A_2BC are formed with the content of A doubled and that of B halved relative to the normal Heusler compounds AB_2C . The inverse Heusler structure (called Hg_2CuTi - or Li_2AgSb -type) consists of four sites within a face-centered cubic lattice and has lower symmetry (space group $F\bar{4}3m$). Many other variants of these basic structures are possible, involving distortions, split positions, and more complex ordering, which entice chemists and physicists to frolic in this rich playground of compounds.

Like golf, these structures are deceptively simple and endlessly complicated.¹⁷ Indeed, their very simplicity makes it difficult to detect the subtle differences. For example, LiAg_2Al was reported as an inverse Heusler compound,¹⁸ but the simulated powder X-ray diffraction (XRD) patterns for LiAg_2Al in CsCl-type, Heusler, and inverse Heusler structures are nearly identical (Figure 2). The presence of weak superstructure peaks (111, 311, 331), barely observable except perhaps with synchrotron radiation XRD, distinguishes CsCl-type from Heusler or inverse Heusler structures. In turn, Heusler and inverse Heusler structures have identical sets of peaks differing almost imperceptibly in intensities, or for LiAg_2Al , these intensities are numerically identical. It is difficult to measure powder XRD patterns accurately, given that uncertainties in the intensities (e.g., arising from preferred orientation, severe absorption, or detector noise) can exceed the difference needed to distinguish these structures.

Sometimes, assumptions are made about the structure but they may not be correct (as is the case for LiAg_2Al , discussed later). A more accurate method is single-crystal XRD, but out of 1415 reports of Heusler compounds, only 31 (or 2%) have been characterized this way. Even single-crystal XRD is not infallible, because very weak intensities may still be hard to detect. Determining the correct structures of these compounds is vitally important to improving their performance in materials applications, because subtle structural differences can greatly influence their physical properties.

Given this difficulty in structure determination, prior work has exploited first-principles quantum mechanical calculations, in which the structural preference is dictated by small energetic differences;¹² this could be viewed as a “brute force” approach to addressing the problem. For some categories of Heusler compounds, simple and chemically intuitive rules have been developed that relate electron count to their structures and properties,³ but these rules are not necessarily applicable to the whole set of compounds. Here, we propose to apply data-mining and machine-learning techniques with these aims: (1) assign the correct structure of Heusler vs. inverse Heusler compounds for some arbitrary combination of elements, (2) predict the existence of new Heusler compounds, and (3) evaluate the reliability of structure assignments for the entire set of compounds AB_2C reported in crystallographic databases (“data sanitizing”). Importantly, we also test these predictions through experimental methods (synthesis and structure determination). In general, predicting what structure will form for a given combination of elements is one of the “grand challenges” of chemistry.¹⁹ This goal has broader implications for advancing materials science, because compounds with specific properties can be discovered more quickly without having to explore vast reaches of chemical space.

2. Experimental Section

2.1. Structure Prediction. The Heusler prediction engine was built using a materials informatics approach from three integral components: training data, chemical descriptors, and a machine-learning algorithm. The key idea behind materials informatics is that, given enough examples (training data) and informative numerical representation of the materials (descriptors), a machine-learning algorithm can determine patterns to predict how unseen examples will behave. Careful development and selection of all three parts are necessary to create a useful predictive engine.

Training data can be characterized by their quality and quantity. A sufficient quantity is required to be able to statistically detect a pattern: more complex patterns demand more examples. High quality data are also crucial, because errors create extra noise that can obscure underlying patterns. The training data used in this engine are a compilation of experimentally confirmed structures for compounds with a formula of AB_2C . Crystallographic data for all such compounds were extracted from Pearson's Crystal Data¹⁵ and the ASM Alloy Phase Diagram Database,²⁰ with the following criteria imposed for the input set of compounds used in the structure predictor: (i) the phases do not contain hydrogen, noble gases, and elements with $Z > 83$ (radioactive and actinide elements), and (ii) the phases exhibit exact 1:2:1 stoichiometry, contain 3 components, and are thermodynamically stable. A total of 1948 compounds crystallizing in 208 unique structure types were found to satisfy these criteria and were confirmed to exist experimentally under ambient temperatures and pressures. The input file contained information about the composition and structure type of these compounds. Out of these, the most populous are Heusler compounds (341 entries, or 18%) and the second most

populous are NaFeO₂-type compounds (255 entries, or 13%). There are 94 compounds each having only one reported entry, meaning that they crystallize in their own unique structure types.

Once a training set is compiled, the next challenge is to represent these examples in a machine-understandable way. Typically, descriptors are properties of a material that can be used to compare one compound to another, such as crystal structure or average atomic mass. Choosing a good set of descriptors is an essential part of materials informatics. The relationship between descriptor choice and model quality,²¹ compounded with the lack of a standard representation, has led to a growing body of literature emphasizing the importance of descriptors.^{22,23} Moreover, materials datasets tend to be much smaller than traditional machine-learning datasets, which makes complex patterns harder to detect with a suboptimal collection of descriptors. Descriptors allow for integration of chemical knowledge to help the model by describing materials in dimensions where patterns are likely to be found, so that the patterns can be detected with fewer data. Careful choice of a descriptor set takes advantage of prior knowledge to identify *where* the pattern is, allowing the algorithm to then determine *what* the pattern is. For example, because atomic size is known to be an important factor in determining chemical structure, descriptors such as the difference in atomic radii are included to capture this information for the model in the structure predictor. Table 1 lists the 22 descriptors and their weights used in the machine-learning model.

Finally comes the choice of machine-learning algorithm. Although many algorithms are possible, we have chosen the random forest algorithm²⁴ because it has been successfully applied to materials informatics and requires no external validation step.^{25,26} This algorithm is an example of an ensemble method, which trains multiple predictors and combines their results to make a single final prediction. In the case of random forest, each of these sub-predictors is a

decision tree trained on a fraction of the training data, where the possible descriptors for each branch point is a random subset of the descriptor list. An example of one such decision tree is illustrated in Figure 3. The decision tree structure is able to capture complex interactions between descriptors. By averaging over the predictions of an ensemble of these decision trees, the random forest algorithm incorporates the different trends found by each tree, resulting in a complex and robust model. We use the standard k -fold cross-validation technique to characterize the performance of the models. This technique evaluates the model on examples that were unseen during training in order to accurately gauge model performance. To perform k -fold cross-validation, the data are randomly split into k equal segments, called folds, such that each has approximately the same proportion of Heusler and non-Heusler compounds. For each of these folds, a model is trained on all data except the selected fold. Predictions are made for the data in the fold that were not included in the model training. The predictions are then compared to the known values for those examples.

2.2. Synthesis and Structural Characterization. From the recommendations offered above, the most promising Heusler compositions were identified for experimental validation. However, candidates involving substitution of similar elements were excluded to apply a fair test of the model (e.g., if FeB_2C and NiB_2C are known Heusler compounds, the unknown but chemically obvious compound CoB_2C containing the intervening transition metal was not considered). The ternary gallides MRu_2Ga and RuM_2Ga (where M is a first-row transition metal) were targeted for synthesis because their probability of forming Heusler compounds is predicted to be high, their existence cannot be extrapolated simply through periodic trends from previously known compounds, and they are dissimilar to previous $M-M'$ -Ga phases which mostly contain first-row transition metals for both M and M' components. These gallides were

also chosen because they are predicted by a thermoelectrics recommendation engine²⁷ (using a similar algorithm as the one presented here) to exhibit low thermal conductivity, a property that has been associated with half-Heusler but not Heusler compounds. For example, this thermoelectrics engine predicts that TiRu_2Ga , if it formed, would have a high probability of exhibiting a thermal conductivity κ of less than $10 \text{ W m}^{-1} \text{ K}^{-1}$. Conversely, to test for false negative predictions, the series $\text{La}M_2\text{Ga}$ (where M is again a first-row transition metal) was also targeted for synthesis because their probability of forming Heusler compounds is predicted to be low.

Mixtures of Ru powder (99.95%, Alfa-Aesar) or La pieces (99.9%, Hefa), powders of various first-row transition metals M (Ti to Ni, >99.5%, Alfa-Aesar or Cerac), and Ga pieces (99.95%, Alfa-Aesar) were combined in ratios according to the formulas indicated above with a total mass of 0.2 g, pressed into pellets, and melted on a copper hearth under argon atmosphere in an Edmund Bühler MAM-1 arc melter. The ingots were placed in fused-silica tubes, which were evacuated and sealed. Annealing was done in one step at $800 \text{ }^\circ\text{C}$, followed by quenching in cold water. The samples were ground to powders and examined by powder XRD on an Inel diffractometer equipped with a $\text{Cu } K\alpha_1$ radiation source and a curved position-sensitive detector.

Single crystals of TiRu_2Ga were selected and confirmed by energy-dispersive X-ray (EDX) analysis, performed on a JEOL JSM-6010LA scanning electron microscope, to have a composition ($\text{Ti}_{24(2)}\text{Ru}_{49(3)}\text{Ga}_{27(3)}$) in good agreement with the formula of a Heusler compound. Single-crystal diffraction data were collected at room temperature on a Bruker PLATFORM diffractometer equipped with a SMART APEX II CCD area detector and a graphite-monochromated $\text{Mo } K\alpha$ radiation source, using ω scans at 8 different ϕ angles with a frame width of 0.3° and an exposure time of 15 s per frame. The structure was solved and refined with

use of the SHELXTL (version 6.12) program package.²⁸ Face-indexed absorption corrections were applied. The cubic space group $Fm\bar{3}m$ was chosen on the basis of Laue symmetry, intensity statistics, and systematic absences.

3. Results and Discussion

Predicting the existence and structure of unknown compounds is an ongoing challenge in chemistry. At one extreme, semiclassical approaches assume that the structure depends on chemical concepts like atomic size and electron count, which are used to create structure maps;²⁹⁻³¹ although intuitively appealing, there is a risk that the choice of parameters is biased or too simple. At another extreme, first-principles quantum mechanical calculations are performed to determine the total energies of alternative structures;^{32,33} however, the gain in accuracy is offset by a loss in easy understanding through general chemical concepts and by a need for powerful computational resources. Intermediate approaches such as principal component analysis and machine-learning methods are now being applied to this general problem.^{34,35}

3.1. Assignment of Heusler, Inverse Heusler, and Non-Heusler Compounds. We have constructed a prediction engine that suggests the structure of compounds AB_2C by exploiting a random forest algorithm. The distinguishing feature of our approach is that *the prediction is made from the descriptors based on the composition alone*. The model required less than 1 minute to train the dataset, and about 45 minutes to make the full set of predictions on over 400,000 candidates, *viz.* <0.01 s per compound. The candidates with the highest and lowest probability of forming Heusler compounds are listed (Tables S1 and S2 in Supplementary Information). We performed 20-fold cross-validations for this model, and the distribution of probability values (color-coded by the actual class) is shown in Figure 4. The procedure was

repeated with 10-fold cross-validation, and the results were unchanged (Figure S1(a) in Supplementary Information). The model is exceptionally successful in making correct predictions, giving high probabilities for candidates observed to be Heusler compounds experimentally, and low probabilities for those that are not; it is even able to correctly predict inverse Heusler compounds to be non-Heusler compounds.

Analyzing the descriptors (variables) that were most effective in separating Heusler structures from others in the prediction engine yields interesting insight (Table 1). The most important such descriptors, as quantified by their weights as described in the model, are the position of element B (in the formula AB_2C) in the periodic table, the total number of electrons (and especially number of p-electrons), and the difference in atomic sizes for the A and B components. Because the electron configuration of element B is given by its group number or position in the periodic table, the first two descriptors essentially correspond to electron count, which is known to strongly influence the energetic preference for Heusler compounds³ and their various physical properties (magnetism,^{4,5} superconductivity,^{7,8} topological insulator behaviour¹⁴), depending on the combination of elements. A size factor, given by the third descriptor, dictates formation of Heusler vs inverse Heusler structures, in which the occupations of A and B atoms are partly interchanged. (As noted earlier, Heusler and inverse Heusler structures are not easy to distinguish experimentally from their powder XRD patterns.) However, it would be misleading to conclude that these are the only important factors. The whole point of the machine-learning algorithm is to capture complex, nonlinear relationships that cannot be reduced to a small number of factors influencing the formation of Heusler compounds (inherent in semiclassical approaches), while not having to expend costly computational effort (inherent in first-principles quantum calculations). Specifically, Heusler compounds often

contain d- and f-block elements that can be tricky to model with density functional theory. ~~Even though the other descriptors have lower weights individually, their combined weights exceed that of the first three descriptors.~~ An appropriately weighted combination of all these descriptors leads to a much better discrimination of Heusler vs non-Heusler compounds than would be possible from the simpler models prevalent in the literature, such as one based solely on the valence electron count for which the probability values of forming Heusler compounds tend to cluster near 0.50 (Figure S1(a) in Supplementary Information). The accuracy of the machine-learning model with the full set of descriptors is also greatly improved over that of the model based on electron count, as seen in the receiver operating characteristic (ROC) curves (Figure S1(b) in Supplementary Information). The full-feature model has a true positive rate of 0.94 (and false positive rate of 0.01), in contrast to the one-descriptor model, which has a true positive rate of 0.49 (and false positive rate of 0.07).

The Heusler prediction engine offers several key advantages compared to the previous approaches: (1) it is fast, giving predictions with fractions of a second; (2) it requires no structural information, the very thing that needs to be predicted; (3) it uses only descriptors based on elemental properties (or combinations thereof), which are well tabulated in the literature; and (4) it evaluates quantitative probabilities for the formation of a hypothetical compound. With this tool, experimentalists can exploit machine-learning guidance to complement their chemical intuition in designing compounds. In this way, they can accelerate the search for new materials, they can reduce the risks when the syntheses are difficult, costly, or dangerous to perform, and they can get new ideas to “think outside the box.”

3.2. Prediction of New Heusler Compounds. For experimental validation, several compounds were selected belonging to two series of unknown gallides MRu_2Ga and RuM_2Ga

($M = \text{Ti-Ni}$), which have probabilities of >50% of being Heusler compounds according to the machine-learning model, including both positive and some non-obvious positive predictions (Figure 5). In contrast, a model based on valence electron count predicts that these gallides have probabilities of <60% of being Heusler compounds, except for CoRu_2Ga , which has a good chance (probability of 92%) of success; three of these – RuTi_2Ga , RuV_2Ga , and RuCr_2Ga – have zero probabilities of forming Heusler compounds (Table S3 in Supplementary Information). This stark difference in predicted outcomes provides a good test of the machine-learning model *vis-à-vis* conventional heuristic models.

As vindication for the value of the machine-learning model, arc-melting and annealing reactions (at 800 °C) led to the successful preparation of all members of $M\text{Ru}_2\text{Ga}$ and $\text{Ru}M_2\text{Ga}$ (except for $M = \text{Ni}$) with Heusler structures, as confirmed by powder XRD (Figure 6 and Table S4 in Supplementary Information) on all samples and single-crystal XRD on TiRu_2Ga as a representative member (Table S5 in Supplementary Information). The phase analysis is complicated by severe X-ray absorption caused by the presence of the large proportion of heavy elements. Heusler compounds AB_2C (where A and C atoms are ordered) can be misidentified as CsCl-type (where A and C atoms are disordered), or vice versa. Sometimes, it is possible to apply the rule that if two binary alloys AB and BC exist with CsCl-type structures, then they can form an ordered Heusler compound AB_2C . The occurrence of weak superstructure peaks, observed in both powder and single-crystal XRD (Figures S2–S3 in Supplementary Information), provides definitive evidence that these gallides adopt an ordered Heusler rather than a disordered CsCl-type structure. SEM/EDX analysis reveals that the compositions agree well with the formula of a Heusler compound and that the crystallites are chemically homogeneous (Figure S4 in Supporting Information). The two Ni-containing members (NiRu_2Ga , RuNi_2Ga) were not

confirmed experimentally, but neither could the powder XRD patterns of these samples be assigned to any existing phase. In fact, Ni-containing intermetallic systems are often anomalous in that they exhibit a much richer variety of phases with diverse compositions,^{36,37} not captured within the purview of this prediction engine. It is possible that under the synthetic conditions used, which were not optimized, equilibria were at play involving formation of neighbouring phases with compositions close to the Heusler structure.

Experimental validation for negative predictions is also important, to ensure that the prediction engine does not give false negative results. To test for these, compounds belonging to a third series $\text{La}M_2\text{Ga}$ ($M = \text{Ti-Ni}$) were selected, which have probabilities of <20% of being Heusler compounds based on the machine-learning model (Figure 5). (For comparison, the valence electron count model also indicates that these compounds have low probabilities, <40%, of being Heusler compounds, except for LaCo_2Ga , with a probability of 57% (Table S3 in Supplementary Information)). The reactions attempted did not lead to formation of any ternary compounds, but rather to binary phases (LaGa , La_5Ga_3 , LaGa_2 , and elemental M). Indeed, as *post hoc* rationalization, inspection of the few phase diagrams experimentally investigated here (La-V-Ga ,³⁸ La-Mn-Ga ,³⁹ La-Fe-Ga ⁴⁰) reveals no ternary phases in these systems.

3.3. Data Sanitizing. Structural confirmation of Heusler compounds is exceedingly tricky, as illustrated earlier by the nearly identical powder XRD patterns of the CsCl-type, Heusler, and inverse Heusler structures (Figure 2), and rarely performed by single-crystal diffraction. Thus, there are often uncertain or even incorrect assignments when structural investigation is deficient or absent. This could pose problems if conclusions about materials properties are made based on erroneous assumptions. The Heusler prediction engine can be

applied to identify suggest correct structural assignments of existing compounds reported in the literature and in databases.

Heusler AB_2C and inverse Heusler compounds A_2BC have the same overall composition (1:2:1). In databases,¹⁵ LiAg_2Al was listed as an inverse Heusler compound on the basis of powder XRD,¹⁸ but this assignment is suspect because the prediction engine suggests that it should be a Heusler compound at an 85% probability (the anomalously high entry in Figure 4b). The experimental ambiguity is understandable because, as discussed earlier, the simulated powder XRD patterns for LiAg_2Al are fortuitously identical for Heusler and inverse Heusler structures (Figure 2). There is also a generalization that Heusler structures AB_2C tend to contain a transition metal as the B component, whereas inverse Heusler structures A_2BC tend to contain a large electropositive metal as the A component. Given that Ag is a transition metal and Li is not a large or strongly electropositive metal atom, the assignment of LiAg_2Al as a Heusler compound is more chemically sensible. We have attempted DFT calculations to ascertain which of these two alternative structures is more stable, but the energy differences are too small to be conclusive; in any case, experimental evidence is more conclusive.

In addition to identifying potential false positives as in the case of LiAg_2Al above, inspection of the probability distributions also reveals several potential false negatives (the anomalously low entries in Figure 4a); i.e., compounds that are predicted to be non-Heusler, but have been reported in the literature as Heusler compounds. Most of these compounds are Li-rich intermetallics (Li_2IrSn , Li_2AgSn , Li_2PdGe , Li_2CuSn , Li_2PdPb , Li_2CoSb), with probabilities ranging from 3 to 26% of being Heusler compounds. It is no accident that all these problematic examples contain Li, which is difficult to detect using X-ray diffraction methods because of its

poor scattering ability. Although beyond the scope of this paper, further structural investigation of these compounds, perhaps by neutron diffraction, would be worthwhile.

4. Conclusions

A machine-learning prediction engine exploiting a random forest algorithm was applied to evaluate the probabilities at which compounds with the formula AB_2C will adopt Heusler structures, from descriptors based on the composition alone. This approach was exceptionally successful in distinguishing between Heusler and non-Heusler compounds, including the prediction of heretofore unknown compounds and flagging erroneously assigned entries in the literature and in crystallographic databases. Compared to approaches which were limited in scope (semiclassical) or computationally demanding (quantum calculations), the Heusler predictor is fast (<0.01 s per compound), requires no structural input, uses descriptors based on elemental properties, and evaluates quantitative probabilities. Novel predicted candidates MRu_2Ga and RuM_2Ga ($M = Ti-Co$) were synthesized and confirmed to be Heusler compounds. Because these compounds may also be good candidates for thermoelectric materials,²³ it will be worthwhile to optimize the synthesis of these compounds so their physical properties can be measured. Preliminary measurements suggest that $TiRu_2Ga$ exhibits low thermal conductivity but this requires further confirmation. The results have significant broader impact in accelerating the search for not only Heusler compounds (which have diverse applications for sustainable energy, among many), but also materials candidates for other applications, by offering ideas “outside the box.”

Table 1. Descriptors Used in Machine-Learning Model for Heusler Compounds

descriptor	weight
1. Group number of element <i>B</i>	0.1718
2. Total number of p valence electrons	0.1568
3. Radius difference <i>A/B</i>	0.0959
4. Electronegativity value of element <i>B</i>	0.0543
5. Group number of element <i>A</i>	0.0536
6. Group number of element <i>C</i>	0.0466
7. Radius difference <i>A/C</i>	0.0465
8. Radius value of element <i>C</i>	0.0361
9. Molar mass value of element <i>C</i>	0.0353
10. Radius difference <i>B/C</i>	0.0343
11. Total number of d valence electrons	0.0317
12. Molar mass value of element <i>B</i>	0.0284
13. Electronegativity value of element <i>A</i>	0.0284
14. Electronegativity value of element <i>C</i>	0.0281
15. Radius value of element <i>B</i>	0.0266
16. Total number of s valence electrons	0.0251
17. Total number of valence electrons	0.0243
18. Molar mass value of element <i>C</i>	0.0195
19. Period number of element <i>A</i>	0.0179
20. Radius value of element <i>C</i>	0.0175
21. Period number of element <i>B</i>	0.0119
22. Period number of element <i>C</i>	0.0092

Figure Captions

- Figure 1.** Number of structural reports of Heusler compounds found in Pearson's Database (Release 2015/2016) per year.¹⁵
- Figure 2.** Simulated powder XRD patterns for LiAg_2Al in (a) CsCl-type, (b) Heusler (Cu_2MnAl -type), and (c) inverse Heusler (Hg_2CuTi - or Li_2AgSb -type) structures. The difference plot between the XRD patterns for CsCl-type and Heusler structures is shown in red at the bottom. Note that the patterns for Heusler and inverse Heusler structures have the same sets of peaks differing only slightly in intensities (here, for LiAg_2Al , the intensities are identical).
- Figure 3.** Schematic representation of a single decision tree, with a fraction of descriptors applied to a fraction of data, for prediction of Heusler structure in a candidate AB_2C .
- Figure 4.** Distribution of probabilities for forming Heusler compounds evaluated in the validation process of the Heusler prediction engine for (a) Heusler vs. non-Heusler structures and (b) inverse Heusler structures, as assigned in crystallographic databases and literature. The machine-learning model used here clearly discriminates between Heusler and other structures.
- Figure 5.** Machine-learning-predicted probability of forming Heusler compounds for three series of gallides, and experimental confirmation through arc-melting and annealing at 800 °C (check marks indicate successful preparation of Heusler compound, and crosses indicate absence of Heusler compound).
- Figure 6.** Powder XRD patterns for (a) $M\text{Ru}_2\text{Ga}$ and (b) $\text{Ru}M_2\text{Ga}$ ($M = \text{Ti-Co}$) series with Heusler structures. Red asterisks indicate small amounts of secondary phases (typically Ru and binary gallides).

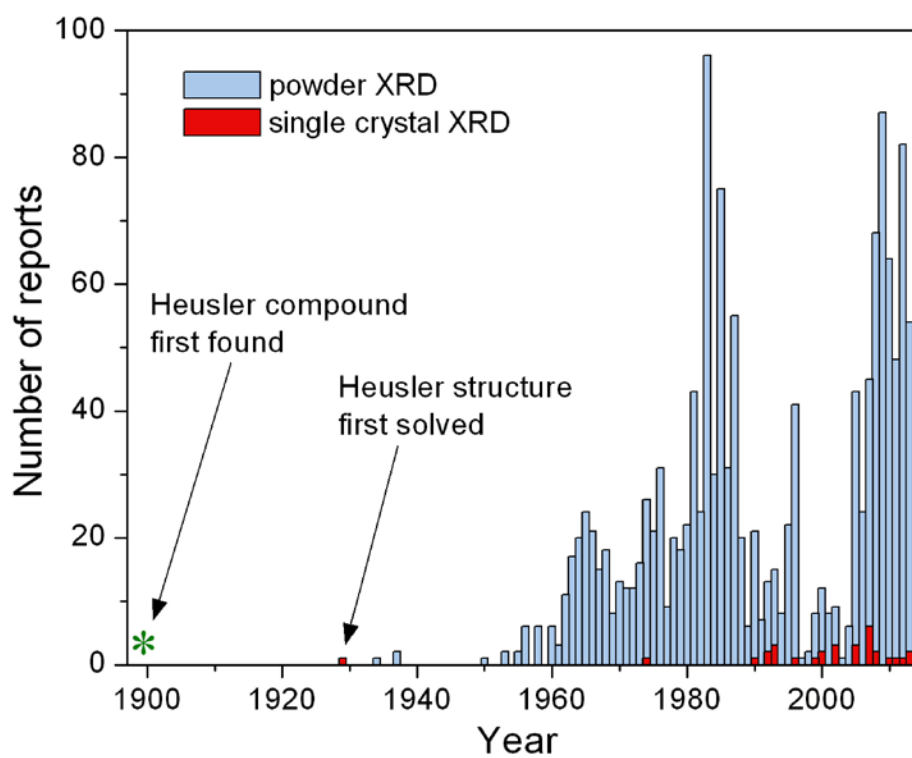


Figure 1. Number of structural reports of Heusler compounds found in Pearson's Database (Release 2015/2016) per year.¹⁵

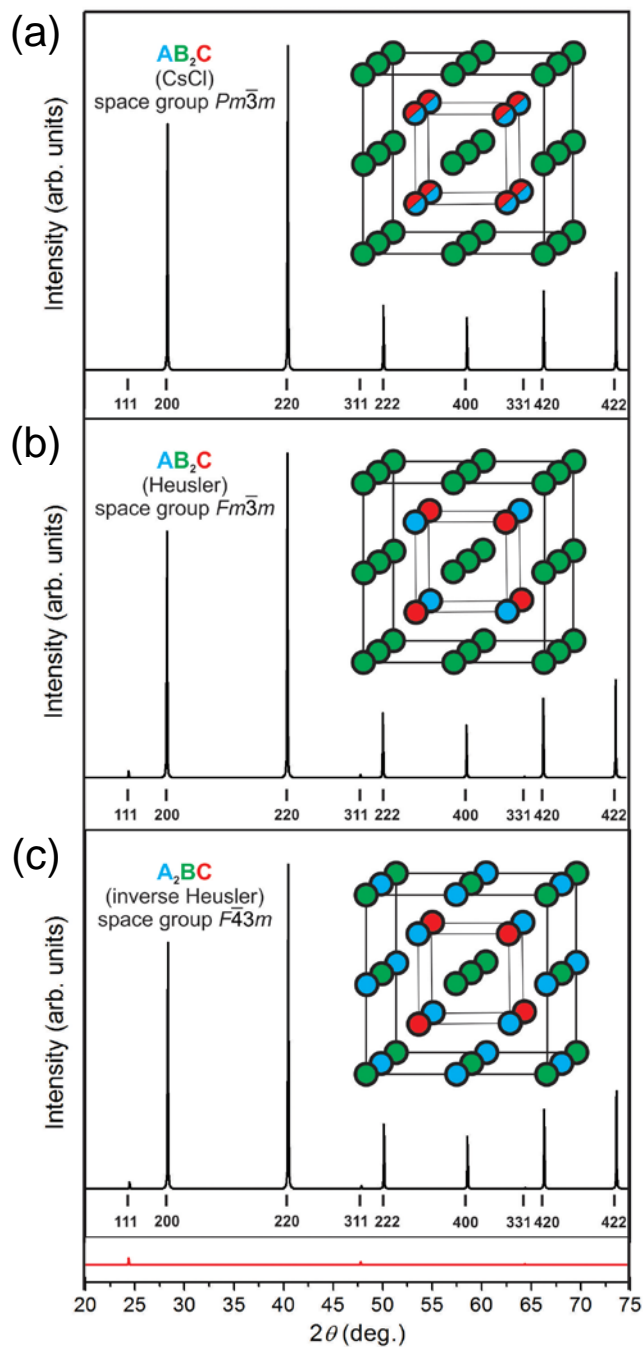


Figure 2. Simulated powder XRD patterns for $LiAg_2Al$ in (a) CsCl-type, (b) Heusler (Cu_2MnAl -type), and (c) inverse Heusler (Hg_2CuTi - or Li_2AgSb -type) structures. The difference plot between the XRD patterns for CsCl-type and Heusler structures is shown in red at the bottom. Note that the patterns for Heusler and inverse Heusler structures have the same sets of peaks differing only slightly in intensities (here, for $LiAg_2Al$, the intensities are identical).

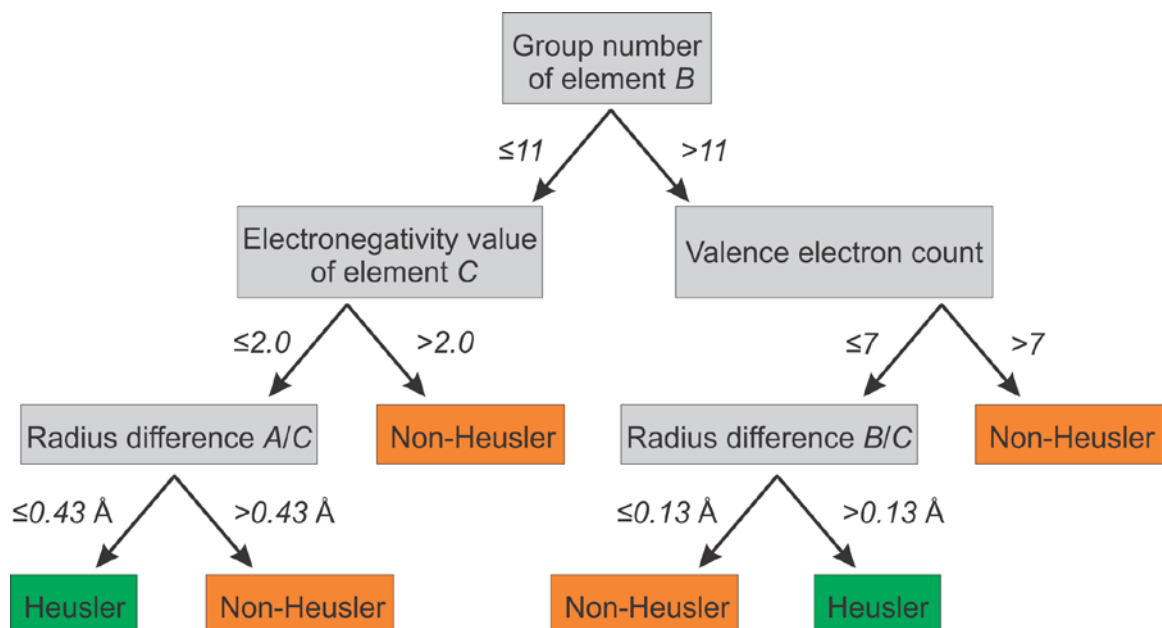


Figure 3. Schematic representation of a single decision tree, with a fraction of descriptors applied to a fraction of data, for prediction of Heusler structure in a candidate AB_2C .

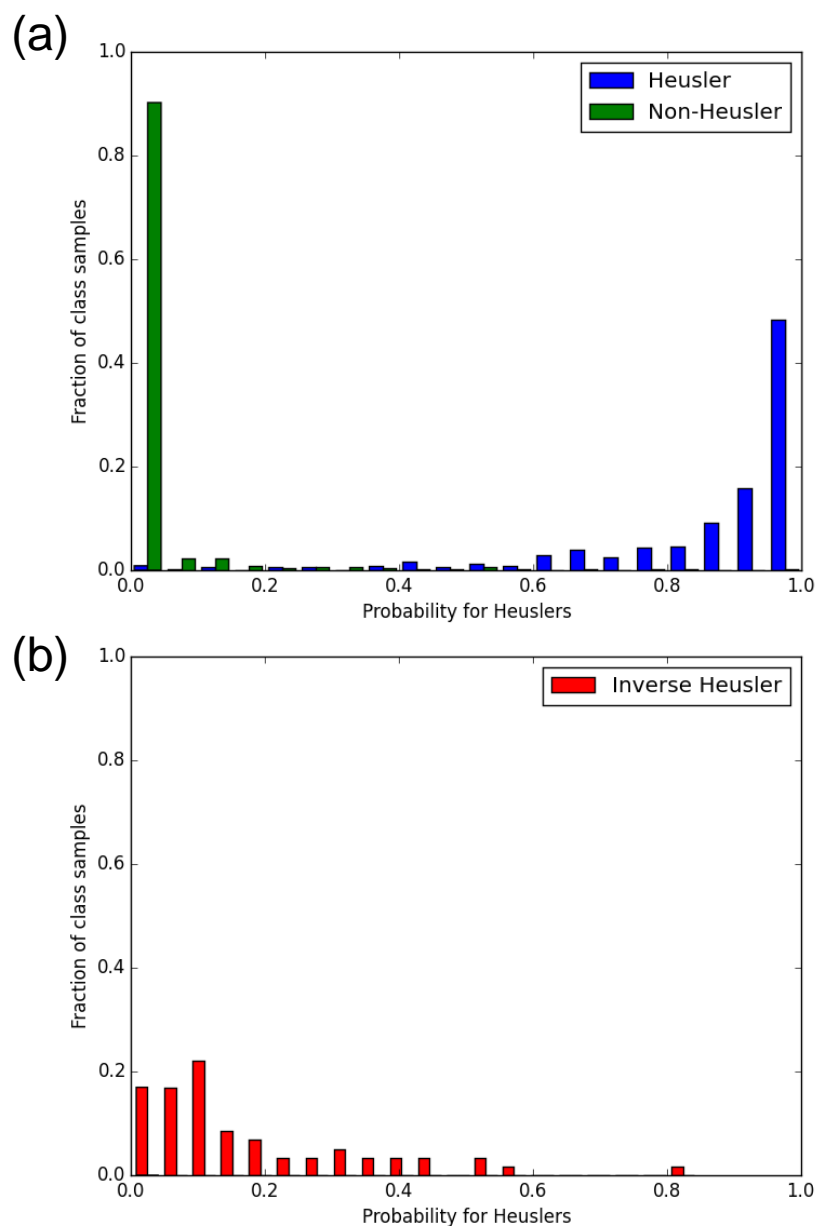


Figure 4. Distribution of probabilities for forming Heusler compounds evaluated in the validation process of the Heusler prediction engine for (a) Heusler vs. non-Heusler structures and (b) inverse Heusler structures, as assigned in crystallographic databases and literature. The machine-learning model used here clearly discriminates between Heusler and other structures.

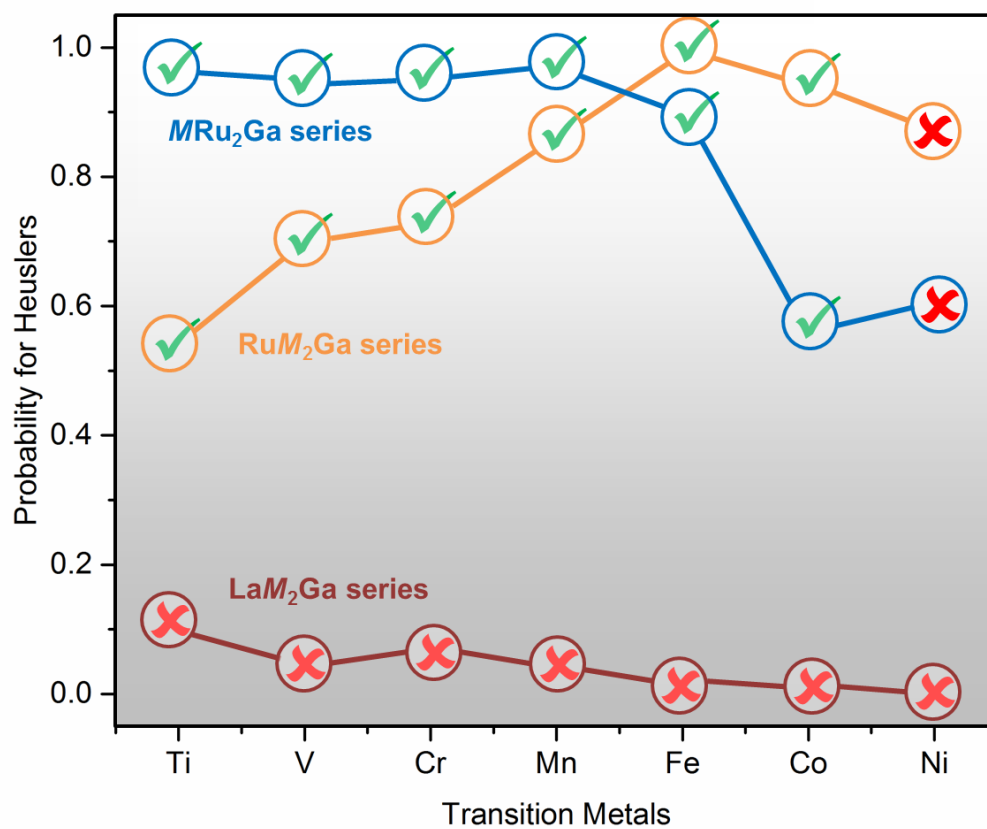


Figure 5. Machine-learning-predicted probability of forming Heusler compounds for three series of gallides, and experimental confirmation through arc-melting and annealing at 800 °C (check marks indicate successful preparation of Heusler compound, and crosses indicate absence of Heusler compound).

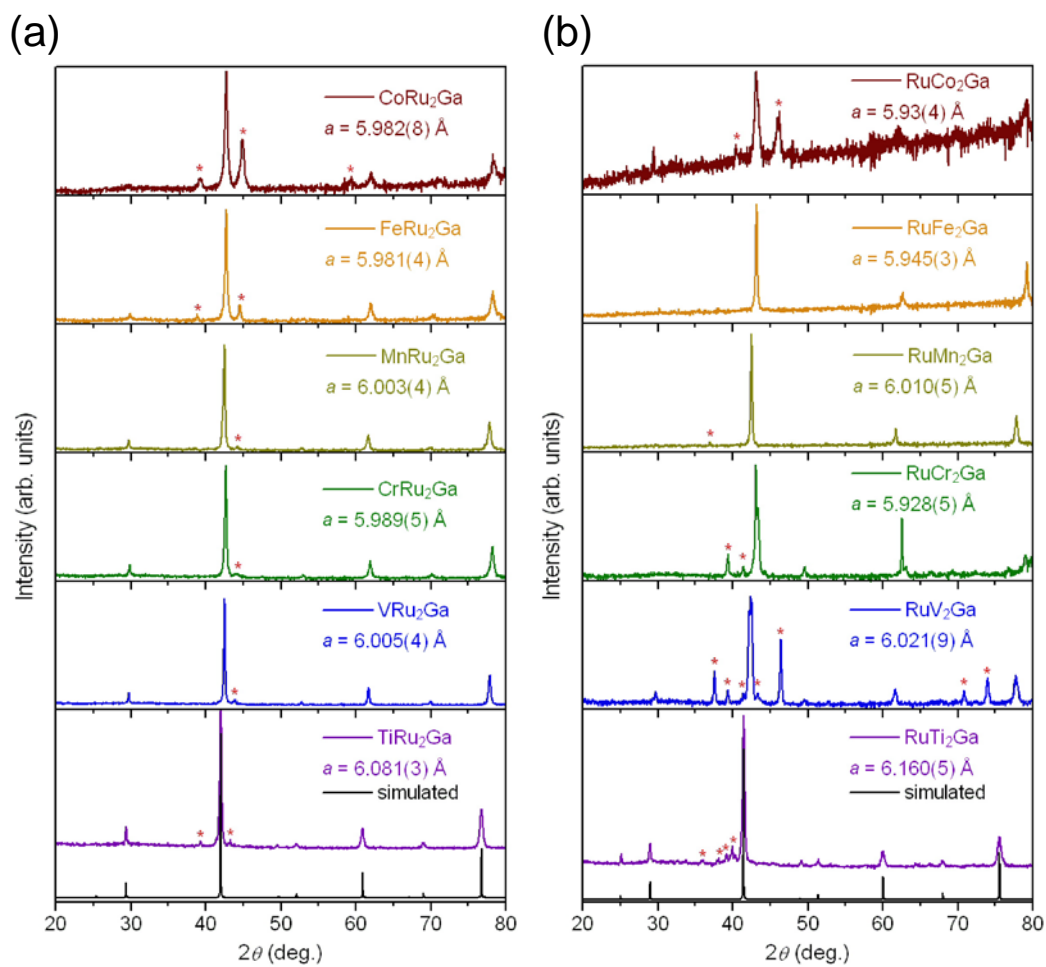


Figure 6. Powder XRD patterns for (a) MRu_2Ga and (b) RuM_2Ga ($M = Ti-Co$) series with Heusler structures. Red asterisks indicate small amounts of secondary phases (typically Ru and binary gallides).

Associated Content

Supporting Information. Machine-learning-predicted probabilities for the most likely Heusler and non-Heusler compounds; predicted probabilities for MRu_2Ga , RuM_2Ga , and LaM_2Ga ; probability distribution plots using 10-fold cross-validation and receiver operating characteristic curves; powder XRD analyses for ternary gallides; single-crystal XRD analysis for $TiRu_2Ga$; and SEM/EDX analysis for $TiRu_2Ga$.

Author Information

Corresponding Authors

* E-mail: bryce@citrine.io, arthur.mar@ualberta.ca

Notes

The authors declare no competing financial interest.

Acknowledgment

This work was supported by the Natural Sciences and Engineering Research Council of Canada through the Discovery Grants and the Collaborative Research and Training Experience Programs (A.O.O., A.M.). M.W.G. is grateful for support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 659764. T.D.S. acknowledges resources from the DARPA SIMPLEX program N66001-15-C-4036.

References

- (1) Heusler, F.; Stark, W.; Haupt, E. Über die Synthese ferromagnetischer Manganlegierungen. *Verh. Deut. Phys. Ges.* **1903**, *5*, 220–223.
- (2) Bradley, A. J.; Rodgers, J. W. The Crystal Structure of the Heusler Alloys. *Proc. R. Soc. London, Ser. A* **1934**, *144*, 340–359.
- (3) Graf, T.; Felser, C.; Parkin, S. S. P. Simple rules for the understanding of Heusler compounds. *Prog. Solid State Chem.* **2011**, *39*, 1–50.
- (4) Kübler, J.; Williams, A. R.; Sommers, C. B. Formation and coupling of magnetic moments in Heusler alloys. *Phys. Rev. B* **1983**, *28*, 1745–1755.
- (5) Wurmehl, S.; Kandpal, H. C.; Fecher, G. H.; Felser, C. Valence electron rules for prediction of half-metallic compensated-ferrimagnetic behaviour of Heusler compounds with complete spin polarization. *J. Phys.: Condens. Matter* **2006**, *18*, 6171–6181.
- (6) Ouardi, S.; Fecher, G. H.; Felser, C.; Kübler, J. Realization of Spin Gapless Semiconductors: The Heusler Compound Mn_2CoAl . *Phys. Rev. Lett.* **2013**, *110*, 100401-1–100401-5.
- (7) Winterlik, J.; Fecher, G. H.; Felser, C. Electronic and structural properties of palladium-based Heusler superconductors. *Solid State Commun.* **2008**, *145*, 475–478.
- (8) Aoki, Y.; Sato, H. R.; Sugawara, H.; Sato, H. Anomalous magnetic properties of Heusler superconductor YbPd_2Sn . *Physica C: Supercond.* **2000**, *333*, 187–194.
- (9) Yang, J.; Li, H.; Wu, T.; Zhang, W.; Chen, L.; Yang, J. Evaluation of Half-Heusler Compounds as Thermoelectric Materials Based on the Calculated Electrical Transport Properties. *Adv. Funct. Mater.* **2008**, *18*, 2880–2888.

- (10) Larson, P.; Mahanti, S. D.; Sportouch, S.; Kanatzidis, M. G. Electronic structure of rare-earth nickel pnictides: Narrow-gap thermoelectric materials. *Phys. Rev. B* **1999**, *59*, 15660–15668.
- (11) Galanakis, I.; Dederichs, P. H.; Papanikolaou, N. Origin and properties of the gap in the half-ferromagnetic Heusler alloys. *Phys. Rev. B* **2002**, *66*, 134428-1–134428-10.
- (12) Kandpal, H. C.; Fecher, G. H.; Felser, C. Calculated electronic and magnetic properties of the half-metallic, transition metal based Heusler compounds. *J. Phys. D: Appl. Phys.* **2007**, *40*, 1507–1523.
- (13) Ren, J.; Li, H.; Feng, S.; Zhai, Q.; Fu, J.; Luo, Z.; Zheng, H. Giant magnetocaloric effect in a Heusler $\text{Mn}_{50}\text{Ni}_{40}\text{In}_{10}$ unidirectional crystal. *Intermetallics* **2015**, *65*, 10–14.
- (14) Chadov, S.; Qi, X.; Kübler, J.; Fecher, G. H.; Felser, C.; Zhang, S. C. Tunable multifunctional topological insulators in ternary Heusler compounds. *Nat. Mater.* **2010**, *9*, 541–545.
- (15) Villars, P.; Cenzual, K. *Pearson's Crystal Data – Crystal Structure Database for Inorganic Compounds (on DVD), Release 2015/16*, ASM International, Materials Park, Ohio, USA.
- (16) Harang, L. On the Crystal Structure of the Heusler Alloys. *Z. Kristallogr.* **1927**, *65*, 261–285.
- (17) Palmer, A. “*Golf is deceptively simple and endlessly complicated; it satisfies the soul and frustrates the intellect. It is at the same time rewarding and maddening – and it is without a doubt the greatest game mankind has ever invented.*”
- (18) Lacroix-Orio, L.; Tillard, M.; Belin, C. Exploration of the lithium-aluminum-silver system. Crystal and electronic structure analysis of new phases $\text{Li}_{6.98}\text{Al}_{4.15}\text{Ag}_{0.87}$ and LiAlAg_2 . *Solid State Sci.* **2004**, *6*, 1429–1437.

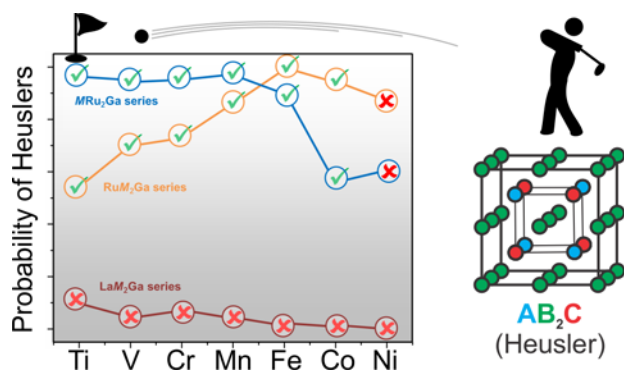
- (19) U.S. Department of Energy, Grand Challenge #3: How do remarkable properties of matter emerge from complex correlations of the atomic or electronic constituents and how can we control these properties? (<http://science.energy.gov/bes/efrc/research/grand-challenges/>).
- (20) Villars, P.; Okamoto, H.; Cenzual K. (Eds.), *ASM Alloy Phase Diagrams Database*, ASM International, Materials Park, OH, 2016 (<http://www.asminternational.org>).
- (21) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Muller, K.-R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6*, 2326–2331.
- (22) Curtarolo, S.; Hart, G. L. W.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. The high-throughput highway to computational materials design. *Nat. Mater.* **2013**, *12*, 191–201.
- (23) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114*, 105503-1–105503-5.
- (24) Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5–32.
- (25) Carrete, J.; Li, W.; Mingo, N.; Wang, S.; Curtarolo, S. Finding Unprecedentedly Low-Thermal-Conductivity Half-Heusler Semiconductors via High-Throughput Materials Modeling. *Phys. Rev. X* **2014**, *4*, 011019-1–011019-9.
- (26) Meredig, B.; Agrawal, A.; Kirklin, S.; Saal, J. E.; Doak, J. W.; Thompson, A.; Zhang, K.; Choudhary, A.; Wolverton, C. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **2014**, *89*, 094104-1–094104-7.

- (27) Gaultois, M. W.; Oliynyk, A. O.; Mar, A.; Sparks, T. D.; Mulholland, G. J.; Meredig, B. Perspective: Web-based machine learning models for real-time screening of thermoelectric materials. *APL Mater.* **2016**, *4*, 053213-1–053213-11.
- (28) Sheldrick, G. M. *SHELXTL*, version 6.12; Bruker AXS Inc.: Madison, WI, 2001.
- (29) Phillips, J. C. Structure and Properties: Mooser-Pearson plots. *Helv. Phys. Acta* **1985**, *58*, 209–215.
- (30) Pettifor, D. G. A chemical scale for crystal-structure maps. *Solid State Commun.* **1984**, *51*, 31–34.
- (31) Villars, P. A three-dimensional structural stability diagram for 998 binary AB intermetallic compounds. *J. Less-Common Met.* **1983**, *92*, 215–238.
- (32) Oganov, A. R.; Lyakhov, A. O.; Valle, M. How Evolutionary Crystal Structure Prediction Works – and Why. *Acc. Chem. Res.* **2011**, *44*, 227–237.
- (33) Hautier, G.; Fischer, C. C.; Jain, A.; Mueller, T.; Ceder, G. Finding Nature’s Missing Ternary Oxide Compounds Using Machine Learning and Density Functional Theory. *Chem. Mater.* **2010**, *22*, 3762–3767.
- (34) Rajan, K. Materials informatics. *Mater. Today* **2005**, *8*, 38–45.
- (35) Broderick, S.; Rajan, K. Informatics derived materials databases for multifunctional properties. *Sci. Technol. Adv. Mater.* **2015**, *16*, 013501-1–013501-8.
- (36) Oryshchyn, S. V.; Le Sénéchal, C.; Députier, S.; Bauer, J.; Guérin, R.; Akselrud, L. G. New Ternary Phases in the Mo–Ni–P system: Synthesis and Crystal Structures. *J. Solid State Chem.* **2001**, *160*, 156–166.

- (37) Zelinska, M.; Zhak, O.; Oryshchyn, S.; Polianska, T.; Pivan, J.-Y. Solid State Phase Equilibria in the Er–Ni–P and Er–Ni–As Systems at 800 °C. *Z. Naturforsch. B* **2007**, *62*, 1143–1152.
- (38) Svechnikov, V. N.; Pan, V. M.; Spektor, A. Ts. Alloying the superconducting compound V₃Ga with group III–VIII elements. *Russ. Metall.* **1971**, *1*, 133–137.
- (39) Markiv, V. Ya.; Belyavina, N. N.; Shevchenko, I. P. Isothermal sections of the La–Mn–Ga and Ce–Mn–Ga systems. *Russ. Metall.* **1994**, *3*, 148–151.
- (40) Raghavan, V. The Fe–Ga–La (Iron–Gallium–Lanthanum) System, Phase Diagrams Ternary Iron Alloys. *Indian Inst. Met.* **1992**, *6B*, 823–826.

For Table of Contents Only

On the basis of a machine-learning prediction engine that gives probabilities for the occurrence of Heusler compounds, the existence and nonexistence of three series of gallides were experimentally confirmed.



SUPPORTING INFORMATION

Deceptively Simple and Endlessly Complicated: Machine Learning Prediction and Experimental Confirmation of Novel Heusler Compounds

Anton O. Oliynyk,[†] Erin Antono,[‡] Taylor D. Sparks,[§] Leila Ghadbeighi,[§] Michael W. Gaultois,[¶]
Bryce Meredig,^{‡,*} and Arthur Mar^{†,*}

[†] *Department of Chemistry, University of Alberta, Edmonton, AB T6G 2G2 Canada*

[‡] *Citrine Informatics, Redwood City, CA 94063 USA*

[§] *Department of Materials Science and Engineering, University of Utah, Salt Lake City, UT
84112 USA*

[¶] *Department of Chemistry, University of Cambridge, Cambridge CB2 1EW United Kingdom*

Table S1. Phases Present in Samples MRu_2Ga and RuM_2Ga ($M = Ti-Co$) Identified by Powder XRD

loading composition	phases	loading composition	phases
TiRu ₂ Ga	TiRu ₂ Ga (95%), Ru (5%) ^a	RuTi ₂ Ga	RuTi ₂ Ga (73%), Ti ₅ Ga ₃ (27%)
VRu ₂ Ga	VRu ₂ Ga (97%), Ru (3%) ^a	RuV ₂ Ga	RuV ₂ Ga (48%), V ₃ Ga (43%), RuGa ₂ (9%)
CrRu ₂ Ga	CrRu ₂ Ga (96%), Ru (4%) ^a	RuCr ₂ Ga	RuCr ₂ Ga (79%), RuGa ₂ (21%)
MnRu ₂ Ga	MnRu ₂ Ga (96%), Ru (4%) ^a	RuMn ₂ Ga	RuMn ₂ Ga (100%) ^b
FeRu ₂ Ga	FeRu ₂ Ga (87%), Ru (13%) ^a	RuFe ₂ Ga	RuFe ₂ Ga (100%)
CoRu ₂ Ga	CoRu ₂ Ga (67%), Ru (33%) ^a	RuCo ₂ Ga	RuCo ₂ Ga (62%), Ru (38%) ^a

^a Elemental ruthenium forms a solid solution with other transition metals up to 50 at. % (e.g., Ru_{0.5}Co_{0.5} with the same structure as Ru).

^b One unidentified peak suggests trace amount of unknown phase.

Table S2. Crystallographic Data for TiRu₂Ga

<i>Data collection and refinement</i>	
formula	TiRu ₂ Ga
fw (amu)	319.76
space group	<i>Fm</i> $\bar{3}$ <i>m</i> (No. 225)
<i>a</i> (Å)	6.0736(14)
<i>V</i> (Å ³)	224.05(9)
<i>Z</i>	4
ρ_{calcd} (g cm ⁻³)	9.480
<i>T</i> (K)	296(2)
crystal dimensions (mm)	0.05 × 0.05 × 0.04
radiation	graphite monochromated Mo <i>K</i> α, λ = 0.71073 Å
$\mu(\text{Mo } K\alpha)$ (mm ⁻¹)	28.101
transmission factors	0.348–0.498
2θ limits	13.44–66.11°
data collected	−8 ≤ <i>h</i> ≤ 8, −8 ≤ <i>k</i> ≤ 8, −8 ≤ <i>l</i> ≤ 8
no. of data collected	846
no. of unique data, including $F_o^2 < 0$	22 (<i>R</i> _{int} = 0.0202)
no. of unique data, with $F_o^2 > 2\sigma(F_o^2)$	22
no. of variables	5
<i>R</i> (<i>F</i>) for $F_o^2 > 2\sigma(F_o^2)$ ^a	0.0086
<i>R</i> _w (F_o^2) ^b	0.0202
goodness of fit	1.054
(Δρ) _{max} , (Δρ) _{min} (e Å ⁻³)	0.369, −0.478
<i>Positional and displacement parameters</i> ^c	
Ti at 4 <i>a</i> (0, 0, 0)	
<i>U</i> _{eq} (Å ²) ^c	0.01(2)
Ru at 8 <i>c</i> (1/4, 1/4, 1/4)	
<i>U</i> _{eq} (Å ²) ^c	0.0077(4)

Ga at 4b (1/2, 1/2, 1/2)

$$U_{\text{eq}} (\text{\AA}^2)^c \quad 0.007(12)$$

Interatomic distances (Å)

$$\text{Ti-Ru} (\times 4) \quad 2.6299(6)$$

$$\text{Ru-Ga} (\times 4) \quad 2.6299(6)$$

$$\text{Ru-Ru} (\times 4) \quad 3.0368(7)$$

$$\text{Ti-Ga} (\times 4) \quad 3.0368(7)$$

$$^a R(F) = \frac{\sum |F_o| - |F_c|}{\sum |F_o|}$$

$$^b R_w(F_o^2) = \left[\frac{\sum [w(F_o^2 - F_c^2)]^2}{\sum w F_o^4} \right]^{1/2}; \quad w^{-1} = [\sigma^2(F_o^2) + (Ap)^2 + Bp]$$

where $p = [\max(F_o^2, 0) + 2F_c^2] / 3$.

^c U_{eq} is defined as one-third of the trace of the orthogonalized U_{ij} tensor.

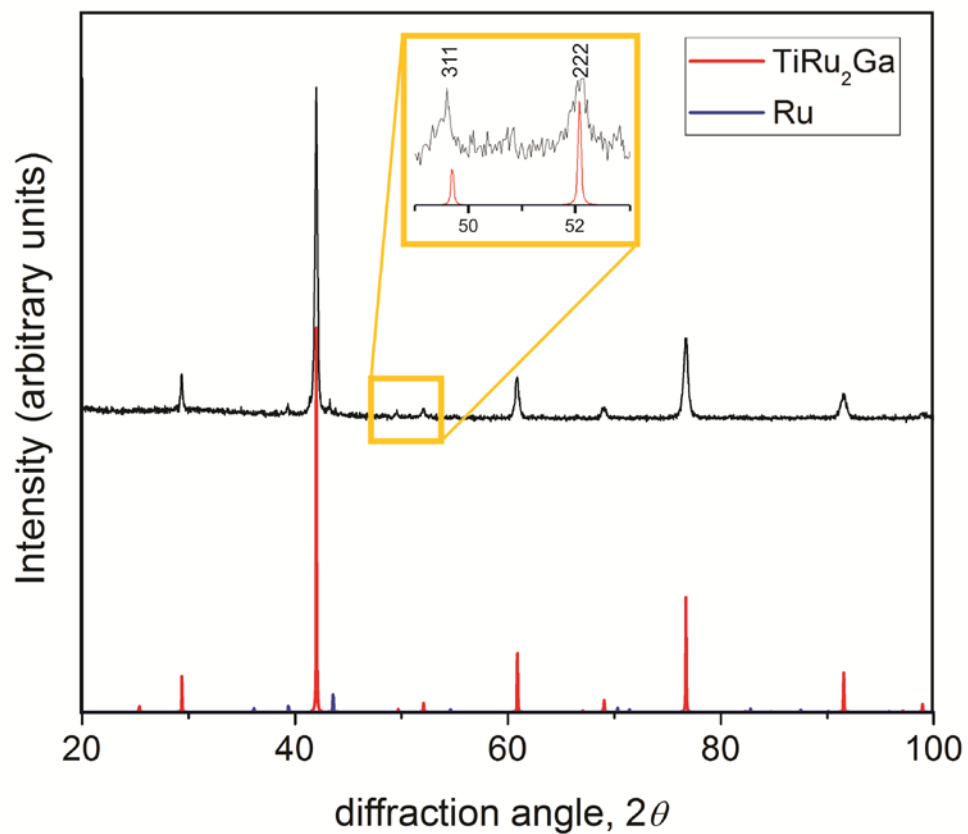


Figure S1. Powder XRD pattern for TiRu₂Ga sample, showing weak superstructure reflections diagnostic of a Heusler structure instead of a CsCl-type structure.

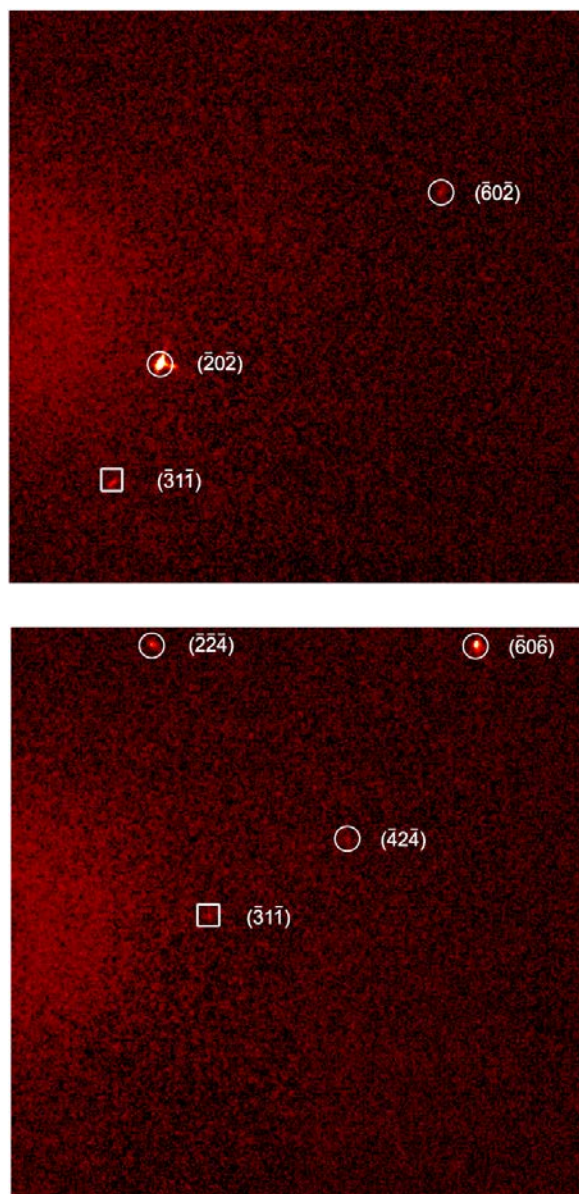


Figure S2. CCD frames from single-crystal diffraction data collection of TiRu_2Ga . The observation of weak superstructure reflections $\bar{3}1\bar{1}$ (enclosed in squares) provides evidence for a Heusler structure instead of a CsCl-type structure. The relative intensities in the simulated powder XRD pattern cannot be directly compared with those on these images which also depend on the diffractometer angles.

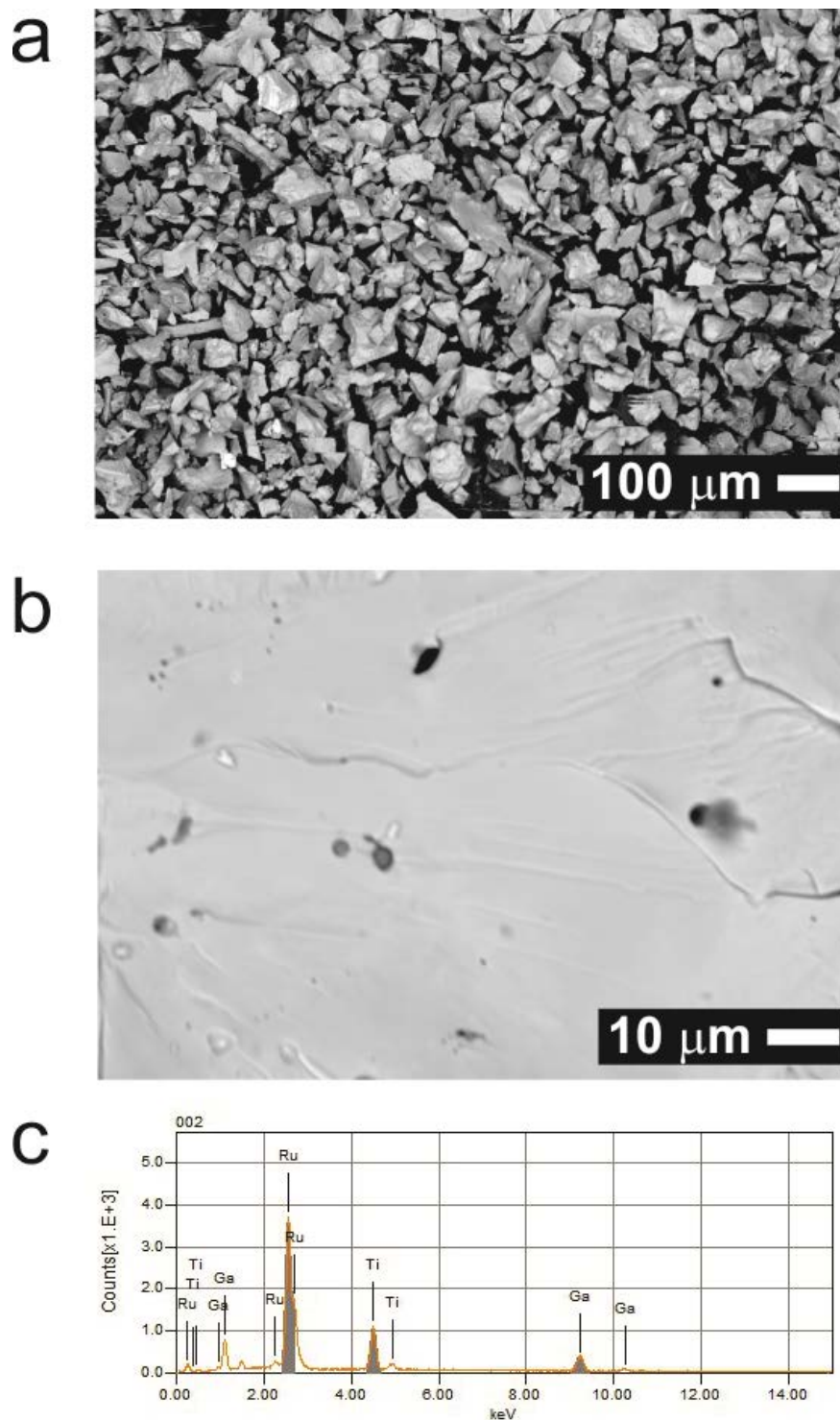


Figure S3. EDX/SEM analysis for TiRu_2Ga sample: (a) Backscattered electron image of TiRu_2Ga powder; (b) surface of TiRu_2Ga crystallite, with no other phases visible on a mass contrast image; (c) EDX spectrum indicating the composition $\text{Ti}_{24(2)}\text{Ru}_{49(3)}\text{Ga}_{27(3)}$, in good agreement with the formula of a Heusler compound.