

# Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD)

JAMES E. SAAL,<sup>1</sup> SCOTT KIRKLIN,<sup>1</sup> MURATAHAN AYKOL,<sup>1</sup>  
BRYCE MEREDIG,<sup>1</sup> and C. WOLVERTON<sup>1,2</sup>

1.—Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60201, USA. 2.—e-mail: c-wolverton@northwestern.edu

High-throughput density functional theory (HT DFT) is fast becoming a powerful tool for accelerating materials design and discovery by the amassing tens and even hundreds of thousands of DFT calculations in large databases. Complex materials problems can be approached much more efficiently and broadly through the sheer quantity of structures and chemistries available in such databases. Our HT DFT database, the Open Quantum Materials Database (OQMD), contains over 200,000 DFT calculated crystal structures and will be freely available for public use at <http://oqmd.org>. In this review, we describe the OQMD and its use in five materials problems, spanning a wide range of applications and materials types: (I) Li-air battery combination catalyst/electrodes, (II) Li-ion battery anodes, (III) Li-ion battery cathode coatings reactive with HF, (IV) Mg-alloy long-period stacking ordered (LPSO) strengthening precipitates, and (V) training a machine learning model to predict new stable ternary compounds.

## INTRODUCTION

A foundational concept in materials science and engineering is the processing/structure/properties paradigm: processing determines structure, which in turn defines observed properties. The close ties between structure and properties are historically evident in the development of models to describe materials behavior (e.g., vacancy and dislocation theory). Currently, *ab initio* predictive models of properties based on structure are commonplace, which has enabled a growing trend in materials design: experimental synthesis in tandem with predictive modeling to facilitate the optimization of materials properties. The promise of such an approach is to dramatically reduce development time for novel materials with innovative design tools and methods. Indeed, accelerating materials design is the primary goal and motivation for the U.S. Materials Genome Initiative.<sup>1</sup>

A critical innovation toward accelerating materials design has been accurate and efficient first-principles prediction of materials properties with density functional theory (DFT). Employing only

quantum mechanical concepts and little experimental input, DFT allows one to predict properties of crystalline solids such as lattice parameters, magnetic moments, formation energies, band structures, etc. Although the fundamental concepts underlying DFT was developed in the 1960s,<sup>2,3</sup> it took 20 years or so for the practical application of the theory with efficient DFT codes and algorithms.<sup>4-7</sup> Since then, DFT has been one of the great successes in modeling materials behavior.<sup>8</sup> With ever-increasing computational power at lower costs and improvements in computational algorithms, the cpu time to perform DFT calculations has been steadily declining to the point that performing large-scale calculations on the order of tens or hundreds of thousands of structures is possible in a reasonable amount of time. So-called high-throughput (HT) DFT calculations enable the generation of large databases of DFT-predicted materials properties,<sup>9</sup> which can accelerate materials design through direct searches of materials with desired properties or the development of higher-level models (e.g., data mining).

Several efforts are underway to generate large-scale HT DFT databases, including the Open

Quantum Materials Database (OQMD),<sup>10</sup> the Materials Project,<sup>11</sup> the Computational Materials Repository,<sup>12</sup> and AFLOWLIB.<sup>13</sup> We have developed the OQMD, which is an extensive HT DFT database consisting of DFT predicted crystallographic parameters and formation energies for over 200,000 experimentally observed International Crystal Structure Database (ICSD)<sup>14,15</sup> and theoretical prototype structures, discussed in more detail in the section titled “[The OQMD and DFT Accuracy](#)”. An important feature of the OQMD is the open nature of our database, meaning we will provide access to the complete database without limitation for the community to use, strongly in line with the Materials Genome Initiative.<sup>1</sup> In the coming months, we will make the OQMD accessible over the web at <http://oqmd.org>, with a download option available for the database files themselves and the code to use them.

Since the development and application of the earliest HT DFT database,<sup>16</sup> HT DFT has proven to be a successful tool for many and varied materials problems.<sup>17–25</sup> In this article, we summarize our own HT DFT efforts, beginning with a description of the OQMD, including a broad comparison of DFT formation energies to experimental values.<sup>10</sup> We then provide examples from our work of applying HT DFT to several interesting materials problems. First, we employ the OQMD to search for materials with optimum properties for Li-air battery electrodes,<sup>26</sup> Li-ion battery anodes,<sup>27</sup> and Li-ion battery cathode coatings reactive with HF.<sup>28</sup> Then, we use the formation energies of the OQMD to test whether novel Mg-alloy strengthening long-period stacking ordered (LPSO) precipitates are thermodynamically stable structures.<sup>29</sup> Last, we describe the use of OQMD formation energies to train a machine learning model with which potential novel ternary compounds are identified.<sup>30</sup> We then conclude with thoughts on the future development of HT DFT databases.

## THE OQMD AND DFT ACCURACY

The Open Quantum Materials Database is a collection of consistently calculated DFT total energies and relaxed crystal structures. Using the Vienna Ab-initio Simulation Package (VASP),<sup>31,32</sup> DFT calculations have been performed for every unique entry in the ICSD without partial site occupancy and less than 35 atoms in the primitive cell, 32,489 structures as of August 2013.<sup>10</sup> The OQMD serves two primary functions: as a large set of data for known structures from which optimum materials can be searched (such as in the following sections: “[High-Capacity Conversion Anode Screening](#)”, “[Li<sub>2</sub>O Battery Screening](#)”, and “[HF Scavenging Li-Ion Battery Cathode Coatings](#)”) and as an accurate description of the chemical potentials and convex hulls of simple and complex systems from which tests of stability can be readily performed (such as

in the following sections: “[Searching for New Strengthening Precipitates in Lightweight Mg Alloys](#)” and “[Data Mining for Novel Ternary Compounds](#)”). The OQMD is primarily limited by what has been experimentally observed and catalogued in the ICSD (i.e., there may exist novel unexplored systems and compounds which are technologically important, see the “[Data Mining for Novel Ternary Compounds](#)” section). Toward resolving this issue, we also include in the database DFT calculations of many unary, binary, and ternary prototype structures. These include, for example, every possible combination of A<sub>3</sub>B L1<sub>2</sub> and X<sub>2</sub>YZ heusler chemistries for over 80 elements. The inclusion of these prototypes in the OQMD provides an approximation for unexplored convex hulls and possible undiscovered compounds as they sample unexplored compositions and systems. The total number of structures in the OQMD, including both ICSD structures and prototypes, as of August 2013 is over 200,000 and is growing every day.

For many materials applications, thermodynamic stability is an important quantity. The long-term stability of  $\gamma'$  Co<sub>3</sub>(Al,W) L1<sub>2</sub> precipitate in Co-based superalloys,<sup>33</sup> hydrogen storage decomposition pathways in metal borohydrides,<sup>34,35</sup> and spinodal decomposition in IV–VI rock salt thermoelectric semiconductors<sup>36</sup> are several examples where the stability of phases is critical for understanding materials behavior. For a compound to be stable, it must not only be lower in energy than all other compounds at that stoichiometry but also be lower in energy than linear combinations of all other compounds in a given system. Thus, an accurate description of stability requires the calculation of the phase in question (e.g., Co<sub>3</sub>(Al,W)) and all other competing phases in the given system (e.g., Co, CoAl, and Co<sub>3</sub>W<sup>33</sup>). Because both the composition and the free energy are linear as a function of quantity of different phases in a system, the set of phases that has the minimum total free energy at a given composition can be determined by linear programming. We have employed this approach, grand canonical linear programming (GCLP),<sup>37</sup> to study hydrogen storage reactions,<sup>37,38</sup> Li-battery anode conversion materials,<sup>27</sup> and general multiphase ground state stability.<sup>10</sup> We have recently revised GCLP to make it more efficient when examining stability in highly multicomponent systems.<sup>39</sup>

As an example of the broad application of GCLP with the DFT-calculated formation energies in the OQMD, we used GCLP to determine how many ICSD compounds in the OQMD are thermodynamically stable at 0 K and zero pressure. Under these conditions, 23% of the 32,489 calculated ICSD structures are stable. Figure 1 shows the quantity of total ICSD structures and OQMD-predicted stable structures by year of their discovery. Before about 1960, most discovered compounds were also thermodynamically stable. After 1960, the number

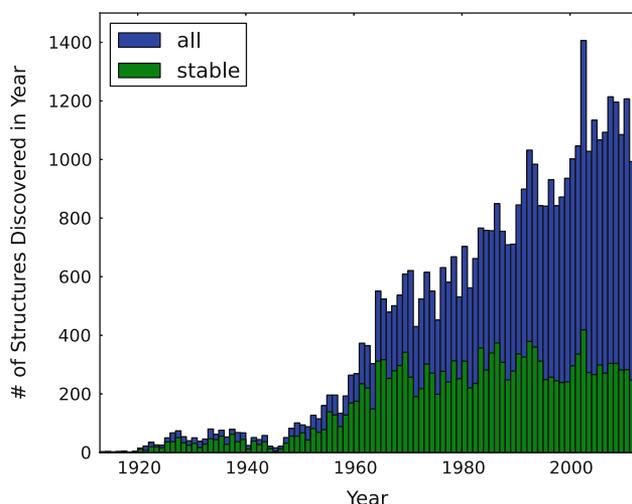


Fig. 1. The number of OQMD-predicted stable compounds and total ICSD compounds by year of their discovery. The year for a structure corresponds to the earliest publication year for ICSD entries at that given structure's composition and symmetry.

of metastable structures per year grew rapidly, outpacing the fairly constant rate of stable structure discoveries. The surge in thermodynamically metastable yet experimentally observed structures is perhaps due to the advent of complex synthesis techniques for strained and/or high-pressure structures, such as thin-film deposition and the diamond anvil cell.

As a large database of DFT calculations, the OQMD can be used to perform a broad comparison of DFT predicted properties to experimental measurements to assess the accuracy of the DFT approach. We have done so for thermodynamic stability by comparing 1290 DFT and experimentally measured formation energies compiled in the Scientific Group Thermodata Europe SSUB thermodynamic database.<sup>40</sup> The SSUB does not contain structural information about a given compound, only a composition. Therefore, we compare an SSUB formation energy to the most stable structure at that composition predicted by DFT, as shown in Fig. 2. The average error between DFT and experimental formation energies is 24 meV/atom and the mean absolute error is 113 meV/atom.

## RESULTS—SUCCESSFUL APPLICATIONS OF OQMD DATABASE

In the following sections, we will discuss five examples from our work for the application of the OQMD to materials design problems covering a range of materials types.

### High-Capacity Conversion Anode Screening

Because of their commercial significance and amenability to simple bulk analysis with DFT, batteries were one of the first areas to be tackled using HT DFT.<sup>41,42</sup> While previous high-throughput

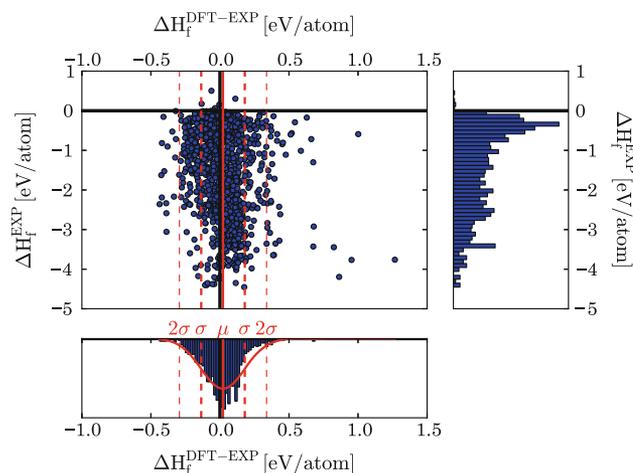


Fig. 2. Comparison of the DFT-predicted formation energies with difference between the experimental<sup>40</sup> and DFT values. The black line indicates perfect agreement between the two, the solid red line indicates the average agreement, and the dashed red lines one and two standard deviations. Histograms are provided for both axes. The thick red line corresponds to a normal distribution fitted to the histogram of the formation energy differences.

studies have focused on the cathode of Li-ion batteries, we have used the OQMD to search for novel high-capacity anodes<sup>27</sup> in three promising classes of materials: transition metal silicides, stannides, and phosphides. Although a patchwork of silicon,<sup>43–46</sup> tin,<sup>47–50</sup> and phosphorus<sup>51–55</sup> compounds has been explored as anode materials in the past, a complete and systematic study of these materials had never been undertaken. We studied this entire class of materials, looking for promising battery reactions based on a series of thermodynamic screens.

Important descriptors for a battery reaction are voltage, gravimetric and volumetric capacity, and volume expansion. The voltage is important in an anode for two reasons: (I) In a low-voltage anode, where the voltage is only slightly higher than that of metallic lithium, lithium metal can form as dendrites, introducing major safety risks, and (II) in a high-voltage anode, the energy density of the total cell is diminished. Finally, it has been observed that internal stresses from the enormous volume expansion associated with extremely high lithium capacity in silicon lead to rapid cell degradation.<sup>56</sup> We speculated that finding reactions which have lower volume product phases may represent an avenue to lower volume expansion—and improved cyclability—in anode materials.

All three anode traits described can be calculated from DFT-predicted ground-state thermodynamics. We explored all possible Li-ion anode reactions within the OQMD to find the optimum anode material based on these quantities, as summarized in Fig. 3. As a result of this screening process, several conversion reactions were found that exhibited high capacities, moderate voltages, and minimal volume expansion. As a promising indicator of the

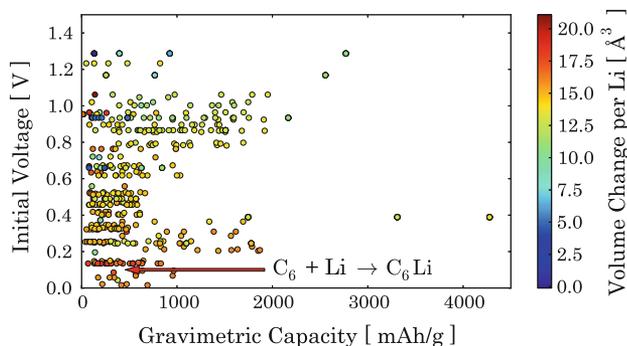


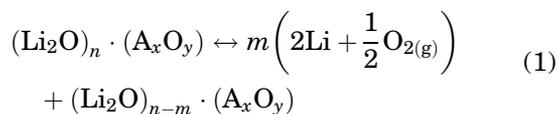
Fig. 3. Initial voltage and gravimetric capacity for transition metal silicides, stannides, and phosphides lithiation reactions in the OQMD. The color of the points is determined by the volume change per lithium atom for the reaction.

selective ability of our screens, one of the anode materials that passed through the screen was CoSn—which is already a known anode material,<sup>57</sup> currently employed as the active component of Sony’s Nexelion battery (Sony Corporation, Tokyo, Japan). Sixteen potential candidate reactions were predicted, three of which stood out as being the most promising: TiP, LiSiNi<sub>2</sub>, and CoSi<sub>2</sub>.

### Li<sub>2</sub>O Battery Screening

Recently, a new approach to lithium-air batteries was described by Johnson et al.<sup>58</sup> and Trahey et al.<sup>59</sup> in which Li<sub>5</sub>FeO<sub>4</sub> was cycled in a cell open to oxygen. In this system, after electrochemically delithiating this material, a surprisingly high voltage was observed. This high voltage covers a large amount of capacity and is attributed to the reinsertion of Li<sub>2</sub>O units. When such novel reaction types or battery architectures are developed, it is tremendously efficient to survey the new field very rapidly using a HT DFT database because the new search space is so open to exploration. To that end, the existing database of calculated structures in the OQMD was employed to screen for other reactions involving Li<sub>2</sub>O units that may follow a similar path.<sup>26</sup>

To begin, the OQMD was used to define every possible reaction of structures which satisfies the equation



where for the completeness of the search, “A” can be any compound and not just an element. For instance, Li<sub>4</sub>KAlO<sub>4</sub> in this reaction can react to form KAlO<sub>2</sub>. According to the search, 255 “A” compounds in the ICSD can satisfy Eq. 1, as summarized in Fig. 4. The stability of the compounds is then taken into account, excluding reactions that do not occur by two-phase equilibria (i.e., indirect).<sup>37,60</sup> Last, several of additional constraints were applied to mimic the

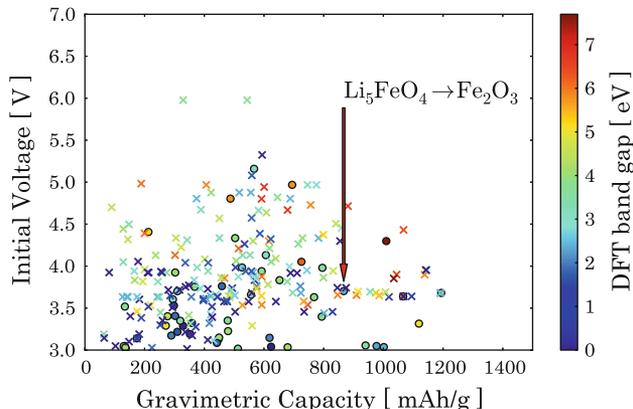


Fig. 4. Initial voltage and gravimetric capacity for HT DFT predicted Li-ion anode reactions (Eq. 1). The color of each point indicates the widest DFT bandgap of any step in the reaction, with red indicating wide bandgap reactions and blue indicating narrow bandgap reactions. The shape of the points indicates either direct (circles) or indirect (crosses) reaction paths.

constraints on real battery materials. Reactions were terminated (I) when the reaction encountered a material with a wide DFT bandgap; (II) when the next decomposition step occurs at a high voltage, such that it might endanger the electrolyte; or (III) if the reaction passes through a region of phase equilibria containing more than two phases.

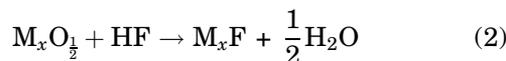
This reaction enumeration process was performed under a set of very stringent requirements and again at a wider tolerance. Once the reactions were determined, the results were screened for the highest possible capacity. The best materials that were identified by the very stringent requirements are being disclosed in another publication,<sup>26</sup> but among the rest of the results were several known expected reactions (which builds our confidence in the screening methodology), as well as several new, promising, and previously unknown reactions. For instance, the reaction that was the inspiration for this search (Li<sub>5</sub>FeO<sub>4</sub>) was recovered, which supports the selection of reaction descriptors and screening criteria. Several more reactions show potential when the less strict screening conditions are used. Among these is LiOH, which is the reaction product of Li-water cells,<sup>35,61,62</sup> which have been observed. Also predicted is Li<sub>2</sub>CO<sub>3</sub>; it is the dead product phase of CO<sub>2</sub> and lithia, which is the reason CO<sub>2</sub> must be scrubbed from Li-air cells, and has also been studied as a possible battery itself.<sup>63</sup> As a result of this study, the list of reactions that are likely worth experimental investigation has been reduced from 255 to the 10 most likely to yield high capacity and cyclability.<sup>26</sup>

### HF Scavenging Li-Ion Battery Cathode Coatings

Degradation of electrodes, induced by their contact with the electrolyte, has to be suppressed to improve the capacity retention and rate capability of Li-ion batteries.<sup>64,65</sup> Coating the cathode material

is an effective remedy to retard the capacity fading upon cycling,<sup>66,67</sup> but experimentally testing all coating candidates for a given cathode is a resource-demanding task that requires fabrication, cycling, and disassembly of the batteries. We have recently performed an HT DFT screening of metal oxide cathode coatings for Li-ion batteries to predict promising coatings a priori.<sup>28</sup> The hydrogen-fluoride (HF) scavenging capability of the coating materials was chosen as the primary design attribute because the highly corrosive HF present in conventional LiPF<sub>6</sub>-based electrolytes is known to attack the cathode and trigger the dissolution of redox-active metal ions into the electrolyte.<sup>64,68</sup> Contemporary battery technology requires functional coatings that can preferentially react with HF in presence of the cathode,<sup>69</sup> beyond what is expected from a simple protective barrier. We employ the OQMD to scan for materials that can serve as HF scavengers.<sup>28</sup>

For an oxide coating ( $M_xO_{\frac{1}{2}}$ ), a generic HF-scavenging reaction producing its conjugate fluoride ( $M_xF$ ) can be written as



The free energy of this reaction, approximated as the DFT enthalpy at  $T = 0$  K ( $\Delta H_{s-HF}$ ), is a natural measure of the HF-scavenging capability of the coating  $M_xO_{\frac{1}{2}}$ . The first screening criterion was selected as  $\Delta H_{s-HF}[\text{coating}] < \Delta H_{s-HF}[\text{cathode}]$  to ensure the preferential reaction of HF with the coating over the cathode material. However, to estimate  $\Delta H_{s-HF}[\text{cathode}]$ , a reaction between the cathode material and HF (i.e., the HF-attack reaction) analogous to Eq. 2 must be devised, such as  $\frac{1}{4}LiCoO_2 + HF \rightarrow \text{Product}(s) + \frac{1}{2}H_2O$  for  $LiCoO_2$ , where one needs to specify the product(s). Our approach to this problem was to find the combination of phases that yields the lowest energy at the chemical composition corresponding to the product(s) in the HF-attack reactions using GCLP along with OQMD. Because the calculated  $\Delta H_{s-HF}[\text{cathode}]$  values is on the order of  $-0.3$  eV/HF for typical cathode materials ( $LiCoO_2$ ,  $LiNiO_2$ ,  $LiMn_2O_4$ ,  $LiFePO_4$ , etc.), the first screen amounts to finding coating materials with  $\Delta H_{s-HF}$  values more negative than  $-0.3$  eV/HF. Volumetric ( $\Omega_V$ ) and gravimetric ( $\Omega_G$ ) HF-scavenging capacities (defined as moles of HF that a coating scavenges per unit volume and gram, respectively) were introduced as additional design parameters. As the volume of the coating increases, it becomes more likely to impede Li and electron transport, and as the mass of the coating increases, specific properties (such as the energy density) degrade. Consequently, to design optimal coatings, both  $\Omega_V$  and  $\Omega_G$  need to be maximized.

Unlike the insertion cathodes with fast kinetics, stable oxides and fluorides are expected to be lithiated mostly via relatively sluggish conversion

reactions<sup>70,71</sup> requiring significant overpotentials to reverse upon charging. The overpotential may lead to entrapment of Li in the coating and impair the rate capability as well as the capacity of the battery. This cyclable-Li loss into a coating may occur if the voltage decreases to a level comparable to the lithiation voltage of the coating upon discharging the battery. Accordingly, we selected a typical Li-ion battery discharge cutoff of  $\sim 3.0$  V as the upper lithiation voltage limit for screening the coatings. In fact, the conjugate fluoride  $M_xF$  of a given oxide  $M_xO_{\frac{1}{2}}$  in Eq. 2 almost always has a conversion voltage higher than  $M_xO_{\frac{1}{2}}$  (except for a few alkali/alkaline earth M).<sup>28</sup> Fluorides are, therefore, more prone to reacting with lithium, and accordingly, the lithiation voltage of  $M_xF$  was chosen as the fourth design attribute.

The high-throughput screening was carried out within the four-dimensional design space of  $\Delta H_{s-HF}$ ,  $\Omega_V$ ,  $\Omega_G$ , and lithiation voltage in a set of 81 *s*-, *p*-, and *d*-block binary metal oxide coating candidates, for which a reaction in the form of Eq. 2 can be devised using the compounds available in the ICSD (see Fig. 5). Materials that are experimentally known to be effective coatings, such as  $Al_2O_3$ ,  $ZrO_2$ , and  $MgO$ ,<sup>67-74</sup> passed through all screens, and  $Al_2O_3$  was found to provide an optimal compromise among all four attributes. Besides this validation of the selected design parameters, we observed that the extent of the experimental capacity retention provided by different coatings on the same cathode is correlated with their HF-scavenging tendencies and capacities, as long as the fluoride produced by the HF-scavenging reaction is a stable solid near room temperature. With an HT thermodynamic analysis of 81 binary oxides, we predicted several new, promising cathode coating candidates with attributes similar to the well-tested coating materials such as  $Al_2O_3$  and  $MgO$  (see Fig. 5).

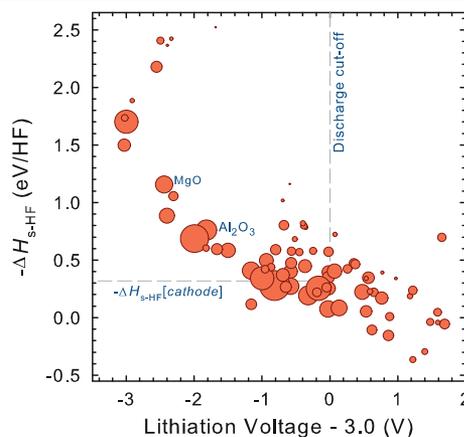


Fig. 5. Calculated HF-scavenging tendency ( $-\Delta H_{s-HF}$ ) of binary metal oxides versus lithiation voltage of the corresponding metal fluoride product layers. Dashed lines enclose the compounds that pass the  $-\Delta H_{s-HF}$  and lithiation voltage screens. Point sizes are proportional to the gravimetric HF-scavenging capacities.

## Searching for New Strengthening Precipitates in Lightweight Mg Alloys

A critical approach to improve the efficiency of transportation systems is to reduce the weight of the vehicles, where a 10% reduction in the weight of a conventional combustion automobile can improve fuel efficiency by 6–8%.<sup>1</sup> As one of the lightest structural metals, Mg and Mg-based alloys offer an attractive alternative to Al and steel vehicle parts. However, issues with poor strength and ductility have limited the use of Mg in automobiles. Therefore, there has been a large effort in recent years to improve the mechanical properties of Mg alloys. Rare earth (RE) solute additions improve Mg alloy strength and ductility through the appearance of novel precipitates and strengthening mechanisms.<sup>75</sup> One such precipitate, long-period stacking ordered (LPSO) structures, is responsible for dramatic increases in yield strength and ductility, 610 MPa at 15% elongation.<sup>76</sup> However, requiring as much as 1 at.% of expensive RE elements, Mg alloys containing LPSO structures are too costly for many industrial applications. Therefore, we employed HT DFT and the OQMD database to search for alternative, more affordable LPSO-forming elements.<sup>29</sup>

LPSO structures are observed to be Mg-rich ternary precipitates in Mg- $X^L$ - $X^S$  ternary systems,<sup>77–87</sup> where  $X^L$  is an element larger than Mg and  $X^S$  is an element smaller than Mg. LPSO structures have been observed in several ternary systems, notably Mg-Y-Zn, as summarized in Fig. 6. LPSO structures, as the name implies, are precipitates that exhibit long-period order of atomic layer stacking along the Mg hexagonal close-packed (hcp)  $c$ -axis. The structures contain order of both the stacking of atomic layers, which alternate between hcp- and

face-centered cubic (fcc)-type stacking, as well as chemical order within the fcc-stacked layers, with fully ordered arrangements of binary and ternary sets of elements.<sup>88,89</sup> 18R and 14H LPSO structures have been observed,<sup>81,90</sup> where the number corresponds to the number of atomic layers in the period of the  $c$ -axis stacking, and the letter refers to whether the structure has rhombohedral or hexagonal symmetry. Only recently has the crystal structure of the LPSO precipitates been fully determined,<sup>88,89</sup> with  $Mg_{71}X_8^L X_6^S$  as the formula for the 14H LPSO “interstitial” structure model.<sup>91</sup> For a DFT study of intermetallic stability, where energy differences are on the order of 0.01 eV/atom, an accurate crystal structure is necessary.

To predict novel non-RE LPSO chemistries, we first tested the ability of DFT to predict the known LPSO-forming Mg- $X^L$ - $X^S$  ternary systems. The stabilities of 85 Mg-RE- $X^S$  LPSO structures were predicted with DFT by comparing the LPSO structure total energy to the OQMD convex hulls for every system as generated by GCLP at the LPSO composition. The results, summarized in Fig. 6, agree perfectly with experimental observations, where all 11 known LPSO forming ternary systems are predicted by DFT to form a stable LPSO structure. Furthermore, 41 novel RE-containing LPSO systems are predicted as well, each representing a ternary system awaiting experimental investigation and confirmation of the DFT prediction. Having proven DFT’s ability to predict LPSO stability, we extended the search to include 11 non-RE  $X^L$  elements, primarily focusing on elements larger than Mg. As shown in Fig. 6, four non-RE elements are shown to form stable, or nearly stable, LPSO structures: Pa, Ca, Th, and Sr. Pa and Th are radioactive, severely limiting their applicability. Ca and Sr are promising, particularly the Mg-Ca-Zn system, which is predicted to form stable LPSO structures. Mg-Ca-Zn alloys have been explored experimentally and LPSO structures have not been observed.<sup>92–95</sup> However, these alloys have different compositions from those that have formed LPSO structures and were not extruded in the manner that has known to readily form LPSO structures. Therefore, using HT DFT and the OQMD, we predict not only 41 novel RE-containing LPSO systems but also Ca and Sr additions as potential affordable alternatives to RE elements to form LPSO structures in Mg alloys.

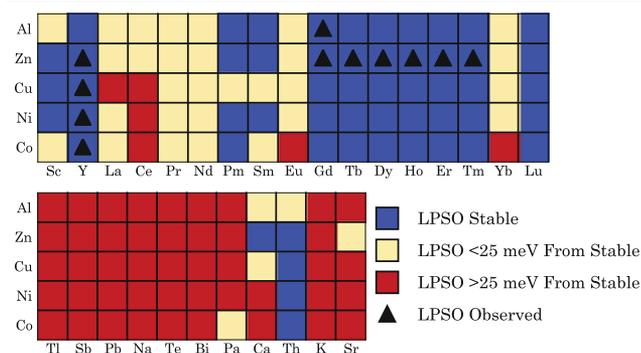


Fig. 6. DFT predicted stability of 14H-i and 18R-i LPSO structures for Mg- $X^L$ - $X^S$  ternary systems.  $X^S$  and  $X^L$  elements are given along the vertical and horizontal axes, respectively. The colors are defined by the stability of the LPSO structure relative to the convex hull: blue if the LPSO structure is on the convex hull, yellow if it is within 25 meV/atom above the convex hull, and red if it is more than 25 meV/atom above the convex hull.  $X^L$  = RE systems are given at top and  $X^L \neq$  RE systems at bottom. Experimentally observed LPSO-forming systems are also indicated with triangles.<sup>77–87</sup> Blue (and possibly yellow) squares without triangles represent predictions of alloy systems where as yet unobserved LPSO structures should be stable.

## Data Mining for Novel Ternary Compounds

An issue with HT DFT materials design is that the set of materials in the database is limited to those experimentally observed because DFT requires, as input, the crystal structure. There are two consequences of this limitation. First, the true convex hull for unexplored systems may be incorrect. To an extent, this is addressed within the OQMD in a limited way by the calculation of many prototype structures,

as discussed in the section titled “The OQMD and DFT Accuracy”. Second, the discovery of novel materials is limited to the set of structures calculated within the database. For example, a search for novel large bandgap materials would be limited to structures observed in the ICSD and common prototypes as opposed to truly novel complex crystal structures. Predicting the ground-state crystal structure for arbitrary compositions remains a heavily studied yet elusive goal.<sup>38,96–98</sup> The primary difficulty is that, at a given composition, many structures (on the order of millions) must be tested to ensure an accurate prediction of the ground state. For sufficient accuracy, DFT must be used to test the quality of potential structures but is still far too costly for this task.

Meredig et al.<sup>30</sup> approached this problem by avoiding the question of structure altogether by training a pair of heuristic and machine learning (ML) models with OQMD data and using a combination of the models to quantitatively predict the formation energies of arbitrary ternary compositions. With such an approach, all possible ternary compositions can be explored at a fraction of the computational time and cost of a single DFT calculation and the most likely compositions with novel ternary compounds can be explored with DFT structure prediction methods or experiments. The heuristic model employs a well-established approach to predict A-B-C ternary formation energies by taking the composition-weighted average of those from the A-B, A-C, and B-C binary systems.<sup>99,100</sup> When applied to the known OQMD data, this heuristic greatly underestimates the DFT formation energies but in a very systematic way. Fitting the heuristic to the DFT calculated OQMD ternary formation energies results in a simple correction:  $\Delta H_F^{\text{heur-corr}} = 1.50\Delta H_F^{\text{heur}} - 0.02$  eV. The ML model<sup>101</sup> incorporates numerous decision trees to learn the behavior of element interactions to quantitatively predict formation energies of arbitrary compositions. Although ML has been previously used to predict stable crystal structures at given compositions,<sup>97,102,103</sup> the application of ML in the current work is unique in that a quantitative, structure-independent property prediction is produced.

The accuracy of the combined heuristic and machine learning model approach has been tested by training the models with 4000 OQMD ternary compound formation energies and then using the models to predict the formation energies of ~8700 others that have been already calculated within OQMD. The agreement with DFT is very close, with mean absolute errors about half of what is typical of DFT compared to experiment (~0.11 eV/atom, see the section titled “The OQMD and DFT Accuracy”). Having demonstrated the ability of these two models to predict the formation energy of ternary compounds without the need for an input structure, Meredig et al.<sup>30</sup> then applied these models to predict the formation energies of over 1.6 million ternary compositions. These results are summarized in

Fig. 7, where the likelihood of a given pair of elements producing a stable ternary compound is plotted, as determined by ranking the stabilities of all the predicted compositions. Such a ranking is made possible because the combined heuristic-ML approach can predict the stabilities of the 1.6 million compositions in minutes, whereas a DFT ground state search of each composition would require tens of thousands of cpu-years. Last, the nine most likely ternary compositions to form new compounds were investigated further with DFT crystal structure prediction to determine possible stable ground states. Of the nine compositions, eight of them form new stable compounds:  $\text{SiYb}_3\text{F}_5$ ,  $\text{Pa}_2\text{O}$  ( $\text{SiO}_6$ ),  $\text{U}_2\text{O}(\text{PO}_4)_2$ ,  $\text{S}_2(\text{VF}_6)$ ,  $\text{Pm}_2\text{S}_3$ ,  $\text{P}_3(\text{BrCs}_4)$ ,  $\text{Te}_3\text{Y}_4\text{N}_2$ , and  $\text{Ba}(\text{TeS}_3)$ . There are, of course, many more potential compositions to be explored (approximately 4500), and this approach, currently demonstrated with formation energy, can be applied to many other materials properties for efficient and broad materials discovery. Thus, machine learning and heuristic models trained on OQMD data have been used to efficiently predict several thousand novel stable ternary compositions.

## FUTURE OUTLOOK

High-throughput density functional theory is fast becoming a powerful tool for approaching complex materials design problems. We have summarized several examples where we successfully employed the OQMD in materials discovery, data mining, and

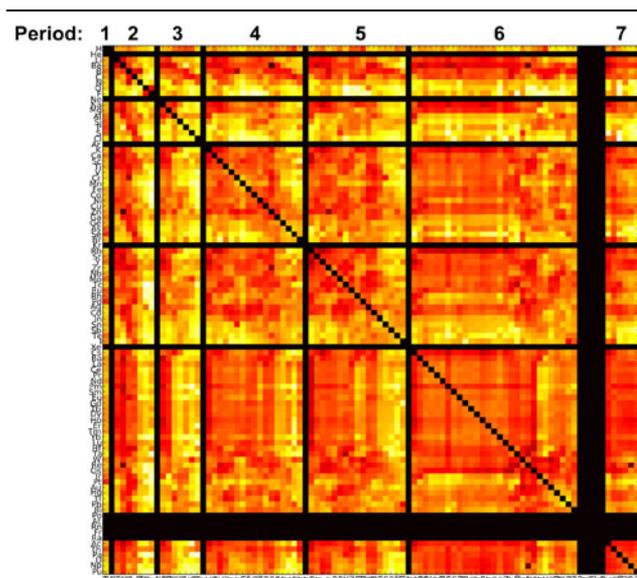


Fig. 7. Heat map of 1.6 M candidate ternary compositions' stability rankings according to the combined heuristic-ML model. Brighter colors imply higher rankings (greater stability). Each point on the heat map corresponds to the average likelihood of stability of all ternary compounds containing the two elements on the plot axes; e.g., the Fe-Cl point gives the average likelihood of stability of all Fe-Cl-X compounds. The black bars on the plot correspond to either noble gases or several exotic heavy elements that were not considered in the survey.

materials optimization. Progress for the creation and use of HT DFT databases will continue along two paths: increasing complexity and increasing understanding. As to the former, the ever-improving speed and efficiency of computing processors will allow for more accurate and more costly calculations to be performed. Both with predictive models more accurate than DFT (such as the GW approximation,<sup>104</sup> the random phase approximation,<sup>105</sup> and quantum Monte Carlo<sup>106</sup>) and more elaborate DFT calculations (such as finite-temperature frozen phonon calculations and larger, more complex prototype structures), the prediction of HT DFT will grow more accurate and be applicable to more materials properties over time. We are advancing OQMD in this regard with the calculation of more complex and higher-order prototype compounds, including perovskite and heusler structures for every possible chemistry, elastic tensor calculations of the ICSD structures, and the thermodynamic and electronic effects of dilute mixing in thermoelectric materials, all of which will be openly available in updates to the OQMD.

The second path of the future of HT DFT, increasing understanding, is likely the more challenging of the two. With the increasing capability to create new data, making sense of it all becomes more difficult. Data mining techniques, such as those demonstrated in the “[Data Mining for Novel Ternary Compounds](#)” section, will become critical to effectively solve materials problems with HT DFT databases such as OQMD. Our own future efforts in this regard include complex chemical potential fits to more accurately describe the formation energy of transition metal oxides and halides (compared to the simple fits employed in “[The OQMD and DFT Accuracy](#)” section). We are also developing additional machine learning models for new types of materials properties, such as bandgap, magnetic moment, and vacancy formation energy. The breadth of HT DFT calculations is also a challenge because tools such as GCLP are only efficient for investigating phase stability in specific, small regions of composition space. We are building tools to efficiently explore and analyze ground state thermodynamics for broad searches of composition space (e.g., 10-component phase diagrams). With improved DFT calculations and data analysis tools, HT DFT will become an even more critical tool in materials science.

## ACKNOWLEDGEMENTS

We would like to acknowledge funding support from the following sources: the Center for Electrical Energy Storage: Tailored Interfaces, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences (to S. K.); The Dow Chemical Company (to M. A.); the Ford-Boeing-Northwestern Alliance for the LPSO study and U.S. Department of Energy, Office of Basic Energy Sciences through grant DE-FG02-98ER45721 for work on the OQMD

(to J. E. S.); and the Department of Defense through the National Defense Science & Engineering Graduate Fellowship Program with further support by DOE under Grant No. DE-FG02-07ER46433 (to B. M. and C. W.). Calculations were performed on the Northwestern University high-performance computing system Quest, as well as on resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

## REFERENCES

1. National Science and Technology Council, Materials Genome Initiative for Global Competitiveness, *Tech. Rep.* (June 2011).
2. P. Hohenberg and W. Kohn, *Phys. Rev.* 136, B864 (1964).
3. W. Kohn and L.J. Sham, *Phys. Rev.* 140, A1133 (1965).
4. J. Perdew and A. Zunger, *Phys. Rev. B* 23, 5048 (1981).
5. J. Ihm, A. Zunger, and M.L. Cohen, *J. Phys. C Solid State* 12, 4409 (1979).
6. D. Ceperley and B. Alder, *Phys. Rev. Lett.* 45, 566 (1980).
7. J. Ihm, M. Yin, and M.L. Cohen, *Solid State Commun.* 37, 491 (1981).
8. J. Hafner, C. Wolverton, and G. Ceder, *MRS Bull.* 31, 659 (2006).
9. S. Curtarolo, G.L.W. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, *Nat. Mater.* 12, 191 (2013).
10. J. Saal, S. Kirklin, B. Meredig, A. Thompson, J. Doak, and C. Wolverton, unpublished research (2013).
11. A. Jain, G. Hautier, C.J. Moore, S. Ping Ong, C.C. Fischer, T. Mueller, K.A. Persson, and G. Ceder, *Comput. Mater. Sci.* 50, 2295 (2011).
12. D.D. Landis, J.S. Hummelshoj, S. Nestorov, J. Greeley, M. Dulak, T. Bli-gaard, J.K. Nørskov, and K.W. Jacobsen, *Comput. Sci. Eng.* 14, 51 (2012).
13. S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R.H. Taylor, L.J. Nelson, G.L. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, and O. Levy, *Comput. Mater. Sci.* 58, 227 (2012).
14. G. Bergerhoff, R. Hundt, R. Sievers, and I.D. Brown, *J. Chem. Inf. Model.* 23, 66 (1983).
15. A. Belsky, M. Hellenbrandt, V.L. Karen, and P. Luksch, *Acta Crystallogr. B* 58, 364 (2002).
16. G. Jóhannesson, T. Bligaard, A. Ruban, H. Skriver, K. Jacobsen, and J. Nørskov, *Phys. Rev. Lett.* 88, 1 (2002).
17. T. Bligaard, G.H. Jóhannesson, A.V. Ruban, H.L. Skriver, K.W. Jacobsen, and J.K. Nørskov, *Appl. Phys. Lett.* 83, 4527 (2003).
18. M. Andersson, T. Bligaard, A. Kustov, K. Larsen, J. Greeley, T. Jóhannesson, C. Christensen, and J. Nørskov, *J. Catal.* 239, 501 (2006).
19. J. Greeley, T.F. Jaramillo, J. Bonde, I.B. Chorkendorff, and J.K. Nørskov, *Nat. Mater.* 5, 909 (2006).
20. A. Jain, S.-A. Seyed-Reihani, C.C. Fischer, D.J. Couling, G. Ceder, and W.H. Green, *Chem. Eng. Sci.* 65, 3025 (2010).
21. G. Hautier, A. Jain, S.P. Ong, B. Kang, C. Moore, R. Doe, and G. Ceder, *Chem. Mater.* 23, 3495 (2011).
22. S. Wang, Z. Wang, W. Setyawan, N. Mingo, and S. Curtarolo, *Phys. Rev. X* 1, 021012 (2011).
23. K. Yang, W. Setyawan, S. Wang, M. Buongiorno Nardelli, and S. Curtarolo, *Nat. Mater.* 11, 614 (2012).
24. L. Yu and A. Zunger, *Phys. Rev. Lett.* 108, 068701 (2012).
25. X. Zhang, L. Yu, A. Zakutayev, and A. Zunger, *Adv. Funct. Mater.* 22, 1425 (2012).
26. S. Kirklin and C. Wolverton, unpublished research (2013).
27. S. Kirklin, B. Meredig, and C. Wolverton, *Adv. Energy Mater.* 3, 252 (2013).
28. M. Aykol, S. Kirklin, and C. Wolverton, unpublished research (2013).

29. J. Saal and C. Wolverton, unpublished research (2013).
30. B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J.W. Doak, A. Thompson, K. Zhang, A. Choudhary, and C. Wolverton, unpublished research (2013).
31. G. Kresse and J. Furthmuller, *Phys. Rev. B* 54, 11169 (1996).
32. G. Kresse and J. Furthmuller, *Comput. Mater. Sci.* 6, 15 (1996).
33. J. Saal and C. Wolverton, *Acta Mater.* 61, 2330 (2013).
34. Y. Zhang, E. Majzoub, V. Ozolinš, and C. Wolverton, *J. Phys. Chem. C* 116, 10522 (2012).
35. Y. Zhang, E. Majzoub, V. Ozolinš, and C. Wolverton, *Phys. Rev. B* 82, 174107 (2010).
36. J.W. Doak and C. Wolverton, *Phys. Rev. B* 86, 144202 (2012).
37. A. Akbarzadeh, V. Ozolinš, and C. Wolverton, *Adv. Mater.* 19, 3233 (2007).
38. C. Wolverton and V. Ozolinš, *Phys. Rev. B* 75, 1 (2007).
39. S. Kirklin and C. Wolverton, unpublished research.
40. SGTE, *Thermodynamic Properties of Inorganic Materials*, vol. 19 (Berlin, Germany: Springer-Verlag, 1999).
41. Y. Meng and M.A.D. Dompablo, *Energy Environ. Sci.* 2, 589 (2009).
42. G. Ceder, *MRS Bull.* 35, 693 (2010).
43. M.D. Fleischauer, J.M. Topple, and J.R. Dahn, *Electrochem. Solid State* 8, A137 (2005).
44. H. Lee, Y. Kim, M. Hong, and S. Lee, *J. Power Sources* 141, 159 (2005).
45. C.-M. Hwang and J.-W. Park, *Surf. Coat. Technol.* 205, S439 (2010).
46. Z. Wang, W. Tian, X. Liu, Y. Li, and X. Li, *Mater. Chem. Phys.* 100, 92 (2006).
47. A.D.W. Todd, R.E. Mar, and J.R. Dahn, *J. Electrochem. Soc.* 154, A597 (2007).
48. M. Xue and Z. Fu, *Solid State Ionics* 177, 1501 (2006).
49. P. Ferguson and M. Martine, *J. Power Sources* 194, 794 (2009).
50. X. Wang and W. Han, *ACS Appl. Mater. Int.* 2, 1548 (2010).
51. J. Xiang, X. Wang, X. Xia, J. Zhong, and J. Tu, *J. Alloy. Compd.* 509, 157 (2011).
52. D. Souza, V. Pralong, A. Jacobson, and L. Nazar, *Science* 296, 2012 (2002).
53. V. Pralong, *Electrochem. Commun.* 4, 516 (2002).
54. J.L. Tirado, J.C. Jumas, L. Monconduit, and J. Olivier-Fourcade, *J. Power Sources* 109, 308 (2002).
55. S. Boyanov, F. Gillot, and L. Monconduit, *Ionics* 14, 125 (2008).
56. L.Y. Beaulieu, K.W. Eberman, R.L. Turner, L.J. Krause, and J.R. Dahn, *Electrochem. Solid State* 4, A137 (2001).
57. J. Wolfenstine, J.L. Allen, J. Read, and D. Foster, *Tech. Rep.* June (2006).
58. C.S. Johnson, S.-H. Kang, J.T. Vaughey, S.V. Pol, M. Balasubramanian, and M.M. Thackeray, *Chem. Mater.* 22, 1263 (2010).
59. L. Trahey, C. Johnson, and J. Vaughey, *Electrochem. Solid State* 14, A64 (2011).
60. C. Wolverton, D.J. Siegel, A.R. Akbarzadeh, V. Ozolinš, *J. Phys.-Condens. Mater.* 20, 064228 (2008).
61. E. Littauer, W. Momyer, and K. Tsai, *J. Power Sources* 2, 163 (1977).
62. M. Urquidi-Macdonald, J. Flores, D. Macdonald, O. Pensado-Rodriguez, and D. Vanvoorhis, *Electrochim. Acta* 43, 3069 (1998).
63. K. Takechi, T. Shiga, and T. Asaoka, *Chem. Commun.* 47, 3463 (2011).
64. Z. Chen, Y. Qin, K. Amine, and Y.-K. Sun, *J. Mater. Chem.* 20, 7606 (2010).
65. K. Edstrom, T. Gustafsson, and J. Thomas, *Electrochim. Acta* 50, 397 (2004).
66. Y.J. Kim, J. Cho, T.-J. Kim, and B. Park, *J. Electrochem. Soc.* 150, A1723 (2003).
67. J. Cho, Y.J. Kim, and B. Park, *Chem. Mater.* 4, 3788 (2000).
68. D. Aurbach, B. Markovskiy, A. Rodkin, and E. Levi, *Electrochim. Acta* 47, 4291 (2002).
69. N.V. Landschoot, E. Kelder, P. Kooyman, C. Kwakernaak, and J. Schoonman, *J. Power Sources* 138, 262 (2004).
70. J. Cabana, L. Monconduit, D. Larcher, and M.R. Palacin, *Adv. Mater.* 22, E170 (2010).
71. F. Wang, *J. Am. Chem. Soc.* 133, 18828 (2011).
72. J. Cho, T.-G. Kim, C. Kim, J.-G. Lee, Y.-W. Kim, and B. Park, *J. Power Sources* 146, 58 (2005).
73. H. Zhao, L. Gao, W. Qiu, and X. Zhang, *J. Power Sources* 132, 195 (2004).
74. C. Li, H. Zhang, L. Fu, H. Liu, Y. Wu, E. Rahm, R. Holze, and H. Wu, *Electrochim. Acta* 51, 3872 (2006).
75. J.-F. Nie, *Metall. Mater. Trans. A* 43, 3891 (2012).
76. Y. Kawamura, K. Hayashi, A. Inoue, and T. Masumoto, *Metall. Mater. Trans.* 42, 1172 (2001).
77. K. Amiya, T. Ohsuna, and A. Inoue, *Mater. Trans.* 44, 2151 (2003).
78. M. Yamasaki, T. Anan, S. Yoshimoto, and Y. Kawamura, *Scripta Mater.* 53, 799 (2005).
79. Y. Kawamura, T. Kasahara, S. Izumi, and M. Yamasaki, *Scripta Mater.* 55, 453 (2006).
80. K. Yamada, Y. Okubo, M. Shiono, and H. Watanabe, *Mater. Trans.* 47, 1066 (2006).
81. Y. Kawamura and M. Yamasaki, *Metall. Mater. Trans.* 48, 2986 (2007).
82. T. Itoi, K. Takahashi, H. Moriyama, and M. Hirohashi, *Scripta Mater.* 59, 1155 (2008).
83. J. Nie, K. Ohishi, X. Gao, and K. Hono, *Acta Mater.* 56, 6061 (2008).
84. H. Yokobayashi, K. Kishida, H. Inui, M. Yamasaki, and Y. Kawamura, *Acta Mater.* 59, 7287 (2011).
85. S.-B. Mi and Q.-Q. Jin, *Scripta Mater.* 68, 635 (2013).
86. Q.-Q. Jin, C.-F. Fang, and S.-B. Mi, *J. Alloy Compd.* 7 (2013).
87. Z. Leng, J. Zhang, T. Yin, L. Zhang, S. Liu, M. Zhang, and R. Wu, *Mater. Sci. Eng.: A* 580, 196 (2013).
88. Y. Zhu, A. Morton, and J. Nie, *Acta Mater.* 58, 2936 (2010).
89. D. Egusa and E. Abe, *Acta Mater.* 60, 166 (2012).
90. T. Itoi, T. Seimiya, Y. Kawamura, and M. Hirohashi, *Scripta Mater.* 51, 107 (2004).
91. D. Egusa and E. Abe (Paper presented at LPSO Conference, Sapporo, Japan, 2 October 2012).
92. H. Somekawa and T. Mukai, *Mater. Sci. Eng.: A* 459, 366 (2007).
93. Y.-N. Zhang, D. Kevorkov, J. Li, E. Essadiqi, and M. Medraj, *Intermetallics* 18, 2404 (2010).
94. K. Oh-ishi, R. Watanabe, C. Mendis, and K. Hono, *Mater. Sci. Eng.: A* 526, 177 (2009).
95. T. Zhou, M. Yang, Z. Zhou, J. Hu, and Z. Chen, *J. Alloy Compd.* 560, 161 (2013).
96. G. Trimarchi and A. Zunger, *Phys. Rev. B* 75, 104113 (2007).
97. C.C. Fischer, K.J. Tibbetts, D. Morgan, and G. Ceder, *Nat. Mater.* 5, 641 (2006).
98. C.W. Glass, A.R. Oganov, and N. Hansen, *Comput. Phys. Commun.* 175, 713 (2006).
99. Y.M. Muggianu, M. Gambino, and J.P. Bros, *J. Chim. Phys.* 72, 83 (1975).
100. M. Hillert, *Phase Equilibria, Phase Diagrams and Phase Transformations* (Cambridge, U.K.: Cambridge University Press, 1998).
101. J.J. Rodríguez, L.I. Kuncheva, and C.J. Alonso, *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 1619 (2006).
102. G. Hautier, C.C. Fischer, A. Jain, T. Mueller, and G. Ceder, *Chem. Mater.* 22, 3762 (2010).
103. S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, and G. Ceder, *Phys. Rev. Lett.* 91, 135503 (2003).
104. M. Shishkin and G. Kresse, *Phys. Rev. B* 75, 235102 (2007).
105. J. Harl and G. Kresse, *Phys. Rev. Lett.* 103, 056401 (2009).
106. W. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, *Rev. Mod. Phys.* 73, 33 (2001).