

ARTICLE OPEN

The Open Quantum Materials Database (OQMD):
assessing the accuracy of DFT formation energiesScott Kirklin¹, James E Saal¹, Bryce Meredig¹, Alex Thompson¹, Jeff W Doak¹, Muratahan Aykol¹, Stephan Rühl² and Chris Wolverton¹

The Open Quantum Materials Database (OQMD) is a high-throughput database currently consisting of nearly 300,000 density functional theory (DFT) total energy calculations of compounds from the Inorganic Crystal Structure Database (ICSD) and decorations of commonly occurring crystal structures. To maximise the impact of these data, the entire database is being made available, without restrictions, at www.oqmd.org/download. In this paper, we outline the structure and contents of the database, and then use it to evaluate the accuracy of the calculations therein by comparing DFT predictions with experimental measurements for the stability of all elemental ground-state structures and 1,670 experimental formation energies of compounds. This represents the largest comparison between DFT and experimental formation energies to date. The apparent mean absolute error between experimental measurements and our calculations is 0.096 eV/atom. In order to estimate how much error to attribute to the DFT calculations, we also examine deviation between different experimental measurements themselves where multiple sources are available, and find a surprisingly large mean absolute error of 0.082 eV/atom. Hence, we suggest that a significant fraction of the error between DFT and experimental formation energies may be attributed to experimental uncertainties. Finally, we evaluate the stability of compounds in the OQMD (including compounds obtained from the ICSD as well as hypothetical structures), which allows us to predict the existence of ~3,200 new compounds that have not been experimentally characterised and uncover trends in material discovery, based on historical data available within the ICSD.

npj Computational Materials (2015) 1, 15010; doi:10.1038/npjcompumats.2015.10; published online 11 December 2015

INTRODUCTION

The development of new materials is critical to continued technological advancement, a fact that has spurred the creation of the Materials Genome Initiative.¹ One component required to build material innovation infrastructure and accelerate material development is the creation of large sets of shared and comprehensive data. In the 1960s, the development of density functional theory (DFT)^{2,3} created a theoretical framework for accurately predicting the electronic-scale properties of a crystalline solid from first principles. However, it was many years before the first practical DFT algorithms were constructed and calculations performed,^{4–7} and even then it was an impressive and noteworthy accomplishment to describe the electronic structure of a single compound. Since then, computational resources have advanced to the point where it is now feasible to predict the properties of many thousands of compounds in an efficient, high-throughput manner.^{8–16} We extend the promise of high-throughput DFT to its logical extreme, calculating in a consistent and accurate manner the properties of a significant fraction of known crystalline solids. Using experimentally measured crystal structures obtained from our partnership with the Inorganic Crystal Structure Database (ICSD),^{17,18} we have created a new database of DFT-relaxed structures and total energies, which is called the Open Quantum Materials Database (OQMD). The OQMD has already been used to perform several high-throughput DFT analyses for a variety of material applications.^{14,19–23}

There are other efforts in the high-throughput calculation of compounds from large crystal structure databases, including the

Materials Project,^{11,24} the Computational Materials Repository¹⁶ and AFLOWLIB.²⁵ We intend the current data set to be freely available, in its entirety, to the scientific community without any conditions or limitations. It is currently available for download at www.oqmd.org/download. We envision three important benefits from making the entire data set available. First, the availability of such a large data set of DFT data enables new and creative uses of these results by others in the field who lack the resources to create their own database. This outcome is strongly in line with the goals of the Materials Genome Initiative. Second, this data set may serve as a nucleus from which external—and unaffiliated—projects can grow. By providing calculated electronic structures for a large fraction of known materials, along with a utility for performing new calculations, new projects can begin more quickly. Third, multiple calculations of the same data set (e.g. the ICSD structures) enable confirmation of the accuracy of the calculations across all databases. Minor differences in the approach, i.e., the use of a slightly different set of potentials or a different choice for GGA+U parameters, makes it possible to see whether a particular choice gives systematically better results.

This paper is composed of two main sections. First, we provide the details of the construction of the OQMD—a description of the calculation settings and chemical-potential fitting approach—and review the current state of the database, which contains the DFT-predicted 0-K relaxed ground-state structures and total energies for every calculable ICSD compound with 34 atoms or less, a total of over 32,559 compounds. In addition, the database contains 259,511 hypothetical compounds described in the

¹Department of Materials Science and Engineering, Northwestern University, Evanston, IL, USA and ²FIZ Karlsruhe—Leibniz Institute for Information Infrastructure, Eggenstein-Leopoldshafen, Germany.

Correspondence: C Wolverton (c-wolverton@northwestern.edu)

Received 17 September 2015; revised 31 October 2015; accepted 2 November 2015

Table 1. GGA+U $U-J$ values and their corresponding fitted chemical-potential corrections employed in OQMD calculations for oxides of the listed elements

Element	$U-J$ (eV)	Correction (eV/atom)
V	3.1	2.675
Cr	3.5	2.818
Mn	3.8	1.987
Fe	4.0	2.200
Co	3.3	1.987
Ni	6.4	2.530
Cu	4.0	1.381
Th	4.0	0.999
U	4.0	2.614
Np	4.0	2.705
Pu	4.0	2.177

Abbreviation: OQMD, Open Quantum Materials Database.

section ‘Structures in the OQMD’, based on decorations of commonly occurring crystal structures. The database also contains a growing number of additional structures (currently ~5,000) calculated for ongoing material discovery projects such as for structural alloys and energy materials.¹⁴ These numbers give OQMD a total size of 297,099 DFT calculations to date. Second, we employ the OQMD to investigate a fundamental question of DFT: how accurately can we use DFT to reproduce experimentally known elemental ground states and compound-formation energies? We use these comparisons as a basis for establishing confidence in the database, before using the database to calculate the stability of every compound in the database. With this large library of DFT calculations of hypothetical compounds included in the OQMD, we are able to make predictions of compositions where new, previously unknown compounds are likely to exist; in this work we identify 3,231 such compositions in which one of our hypothetical structures is predicted to be stable. Finally, we use the breadth of our database to examine historical trends in material discovery.

RESULTS AND DISCUSSION

The OQMD Methodology

A critical component of any high-throughput DFT database is the infrastructure to create and access the contained knowledge, e.g., pymatgen,²⁶ ASE²⁷ and AFLOW.¹⁵ We have developed an infrastructure for such high-throughput DFT calculations and database management, dubbed qmpy. qmpy is written in python, and it uses the django web framework as an interface to a MySQL database. In the same ‘Open’ spirit as the database itself, qmpy is freely available for download as well. It is our goal to develop tools that any research group can use to catalogue, access and analyse large sets of calculations. To this end, qmpy is designed with a decentralised model—any user can download and use it to build a database (e.g., PostgreSQL, MySQL, sqlite or Oracle), and have simple, programmatic access to their calculations. The package has a built-in web interface, and, because we utilise a django backend, it is very easy to customise the web interface depending on the specific needs of any user. Details of qmpy and its analysis algorithms can be found at www.oqmd.org/static/docs.

The calculation of many thousands of compounds within DFT in a reasonable timeframe demands that optimal efficiency within a constrained standard of convergence be found. Furthermore, the comparison of calculation results across many different types of materials (e.g., metals, semiconductors and oxides) requires that all the calculations be performed at a consistent level of theory, which is acceptable for all classes of materials, e.g., consistent plane-wave cutoff, smearing schemes and k -point densities.

To that end, extensive testing on a sample of ICSD structures has resulted in the calculation flow described in Materials and Methods, which ensures converged results in an efficient manner for a variety of material classes. Furthermore, the settings are consistent across all the calculations, ensuring that results between different compounds are directly comparable (e.g., predictions of energetic stability). Using DFT and DFT+U calculations (with the parameters listed in Table 1) and the scheme described in the Materials and Methods section, at the time of the writing of this paper we have calculated 32,559 compounds from the ICSD and 259,511 hypothetical compounds based on decorations of prototype structures.

Structures in the OQMD

The structures in the OQMD come from two sources. The first source, and the origin of the majority of our lowest-energy structures, is the ICSD. For the ICSD structures, we start with a list of 148,279 entries. Of those, 64,412 structures contain atomic positions with partial occupancy, which are substantially more difficult to treat in a high-throughput manner and thus are not included in the current study. A further 32,202 are found to be duplicates of other entries and have been discarded. Structural uniqueness is determined with respect to the lattice and the internal coordinates. Two lattices are compared by finding the reduced primitive cells of each structure and comparing all lattice parameters. Internal coordinates of different structures are compared by testing all rotations allowed by each lattice, and searching for a rotation + translation that maps atoms of the same species onto one another within a given tolerance. Here, any two structures in which all atoms can be mapped to within 0.2 Å of an identical atom are considered identical. Of the remaining structures, 13,934 have incomplete entries in the database, missing either atomic coordinates or spacegroup information. The removal of these structures leaves a pool of 44,506 unique, eligible structures to be calculated. Having started with the structures with the fewest number of atoms, the database currently consists of calculations for all ICSD entries consisting of less than 34 atoms and passing the above filters, a total of 32,559 structures.

The differences between what we identify as calculable, unique structures and the total set of structures in the ICSD must be understood in the context of the challenges of creating a repository of measured crystal structures. There is no general way to judge the quality of a crystal structure. Moreover, it is not simple to determine which of the several measurements/refinements is the best for a given structure, because there are many ways to manipulate the typical criteria used (r -values, goodness-of-fit and low estimated standard deviation). These challenges are compounded by the fact that every measured crystal structure is an average in time (i.e., the time over which data are collected) and space (size of the crystal). DFT energetics can help resolve some of these difficulties by providing another criterion for determining the physicalness of a refined structure, as well as differentiating between structures that give similar fits to experimental data.²⁸

In addition to ICSD structures, we have also calculated decorations of many simple prototype structures over a wide range of compositions. We define a prototype structure to be a crystal structure commonly observed in nature for a variety of chemical compositions, e.g., A1 FCC and L1₂. Table 2 gives a complete listing of the prototype structures that we have calculated. For every elemental prototype, we have calculated every element for which Vienna *Ab-initio* Simulation Package (VASP) includes projected augmented wave-Perdew, Burke and Ernzerhof (PAW-PBE) potentials (89 elements). For all of the binary prototypes, we have calculated each structure with every combination of elements excluding the noble gases (84 elements).

Table 2. Prototype structures calculated in the OQMD

Designation	Number of atoms	SG	Formula	Description	Example
<i>Unary</i>					
A1	1	Fm $\bar{3}$ m	A	FCC	Cu
A2	1	Im $\bar{3}$ m	A	BCC	W
A3	2	P6 $_3$ /mmc	A	HCP	Mg
A3'	4	P6 $_3$ /mmc	A	Distorted HCP	α -La
A4	2	Fd $\bar{3}$ m	A	Diamond	C
A5	2	I4 $_1$ /amd	A		β -Sn
A6	1	I4/mmm	A		In
A7	2	R $\bar{3}$ m	A		α -As
A8	3	P3 $_1$ 21	A		γ -Se
A9	4	P6 $_3$ /mmc	A	Graphite	C
A10	1	R $\bar{3}$ m	A	Simple rhombohedral	α -Hg
A11	4	Cmca	A	BCO	α -Ga
A12	29	I4 $\bar{3}$ m	A		α -Mn
A13	20	P4 $_1$ 32	A		β -Mn
A15	8	Pm $\bar{3}$ m	A	W $_3$ O	β -W
A17	4	Cmca	A	Black phosphorus	P
A20	2	Cmcm	A		α -U
A $_a$	1	I4/mmm	A	BCT	α -Pa
A $_h$	1	Pm $\bar{3}$ m	A	Simple cubic	α -Po
C19	3	R $\bar{3}$ m	A		α -Sm
<i>Binary</i>					
B2	2	Pm $\bar{3}$ m	AB	BCC superstructure	CsCl
D0 $_3$	4	Fm $\bar{3}$ m	A $_3$ B	BCC superstructure	AlFe $_3$
L1 $_0$	2	P4/mmm	AB	FCC superstructure along [100]	AuCu
L1 $_1$	2	R $\bar{3}$ m	AB	FCC superstructure along [111]	CuPt
L1 $_2$	4	Pm $\bar{3}$ m	A $_3$ B	FCC superstructure with maximised unlike bonds	Cu $_3$ Au
B $_h$	2	P6m2	AB	HCP superstructure along [0001]	WC
B19	4	Pmma	AB	HCP superstructure	AuCd
D0 $_{19}$	8	P6 $_3$ /mmc	A $_3$ B	HCP superstructure with maximised unlike bonds	Ni $_3$ Sn
E2 $_1$	5	Pm $\bar{3}$ m	ABO $_3$	Perovskite	CaTiO $_3$
defect-E2 $_1$	9	P4/mmm	A $_2$ B $_2$ O $_5$	Defect-perovskite	
D5 $_1$	10	R $\bar{3}$ c	A $_2$ B $_3$	Corundum	Al $_2$ O $_3$
C4	6	P4 $_2$ /mnm	AB $_2$	Rutile	TiO $_2$
B1	2	Fm $\bar{3}$ m	AB	Rocksalt	NaCl
B3	2	F4 $\bar{3}$ m	AB	Zinblende	ZnS
B4	4	P6 $_3$ mc	AB	Wurtzite	ZnS
D0 $_{22}$	4	I4/mmm	A $_3$ B	FCC superstructure along [012]	Al $_3$ Ti
<i>Ternary</i>					
L2 $_1$	4	Fm $\bar{3}$ m	A $_2$ BC	Heusler BCC superstructure	Cu $_2$ MnAl

Abbreviation: OQMD, Open Quantum Materials Database.

For each structure the number of atoms in the primitive cell, general formula, a simple description and an example compound are given.

This results in 3,486 compounds for structures with symmetrically equivalent sites and 6,972 compounds for structures with symmetrically distinct sites (e.g., AB $_2$ or A $_3$ B structures). Perovskite and defect-perovskite prototypes are only calculated for oxides, and thus are treated as a binary, e.g., ABO $_3$. We also calculated one ternary prototype, the Heusler or L2 $_1$ structure. We have calculated 186,596 compounds in the Heusler structure. A relatively small number of additional hypothetical compounds (~5,000 to date) computed for projects that utilised OQMD are also included in the database.

We included these prototype structures for two purposes. First, by including all compositions, we have a more complete picture of the energetic landscape of phase space. It is important to understand the energy landscape over a comprehensive range of compositions and structures, as, in order to reliably assess the stability of any individual compound, its energy must be compared with the energy of all possible competing phases and combinations of phases. These hypothetical compounds based on common prototype structures are useful because they ensure that our stability calculations are reasonable, even at compositions where limited experimental data are available. Furthermore, at

compositions where prototype compounds are predicted to be stable, it is likely that new compounds are waiting to be discovered. Although in general the prototype compounds that are predicted to be stable are not likely to be the true stable ground-state structures, they indicate a region of composition space where *some* new stable compounds must exist, but which has not been found (or at least not in the OQMD set of ICSD structures). Thus, these ground states represent predictions of new ordered compounds that should be validated experimentally. Second, internal interest in particular crystal structures has motivated the calculation of some prototypes for specific applications—a demand that is easily accomplished using the qmpy framework.

Elemental Ground-State Prediction

We begin by examining the phase stability of all 89 elements included in OQMD in a variety of structure types. We determine the lowest DFT energy structure for every element with a VASP PAW-PBE potential (89 elements) and compare them with the experimentally observed low-temperature structures.^{29,30} In addition to the elemental structures in the ICSD, we also

Table 3. Elemental ground-state structures and chemical potentials predicted by DFT at 0 K

Potential	DFT ground state μ_i (eV/atom)					Exp. LT			Exp. RT			
	ID	SG	Fit-none	Fit-partial	Fit-all	ID	SG	ΔE (eV/atom)	ID	SG	ΔE (eV/atom)	
H		I4/mmm	-3.327	-3.394	-3.434						Gas	
He	A5	I41/amd	-0.004	-0.004	-0.004	A3	P6 ₃ /mmc	0.006			Gas	
LLsv	C19	R $\bar{3}$ m	-1.907	-1.907	-1.731				A2		Im $\bar{3}$ m	0.003
Be	A3	P6 ₃ /mmc	-3.755	-3.755	-3.653							
B		R $\bar{3}$ m	-6.678	-6.678	-6.656							
C	A9	P6 ₃ /mmc	-9.217	-9.217	-9.044							
N		Pa $\bar{3}$	-8.235	-8.122	-8.195						Gas	
O		C2/m	-4.844	-4.485	-4.523						Gas	
F	A11	Cmca	-1.666	-1.429	-1.443		C2/c	0.006			Gas	
Ne	A1	Fm $\bar{3}$ m	-0.029	-0.029	-0.029						Gas	
Na_pv	A3	P6 ₃ /mmc	-1.303	-1.212	-1.196				A1		Fm $\bar{3}$ m	0.003
Mg	A3	P6 ₃ /mmc	-1.542	-1.542	-1.417							
Al	A1	Fm $\bar{3}$ m	-3.746	-3.746	-3.660							
Si	A4	Fd $\bar{3}$ m	-5.425	-5.425	-5.386							
P		P $\bar{1}$	-5.405	-5.161	-5.175	A17	Cmca	0.031				
S		P21	-4.114	-3.839	-3.868							
Cl		Cmca	-1.820	-1.465	-1.479						Gas	
Ar	A3	P6 ₃ /mmc	-0.006	-0.006	-0.006	A1	Fm $\bar{3}$ m	0.009			Gas	
K_sv	A7	R $\bar{3}$ m	-1.097	-1.097	-0.987	A2	Im $\bar{3}$ m	0.000				
Ca_pv	A1	Fm $\bar{3}$ m	-1.978	-1.978	-1.780							
Sc_sv	A3	P6 ₃ /mmc	-6.328	-6.328	-6.344							
Ti		P6/mmm	-7.776	-7.712	-7.702				A3		P6 ₃ /mmc	0.014
V	A2	Im $\bar{3}$ m	-8.941	-8.941	-8.898							
Cr	A2	Im $\bar{3}$ m	-9.508	-9.508	-9.463							
Mn	A12	I $\bar{4}$ 3m	-9.027	-9.027	-8.898							
Fe	A2	Im $\bar{4}$ 3m	-8.308	-8.308	-8.499							
Co	A3	P6 ₃ /mmc	-7.090	-7.090	-7.078							
Ni	A1	Fm $\bar{3}$ m	-5.567	-5.567	-5.587							
Cu	A1	Fm $\bar{3}$ m	-3.716	-3.716	-3.710							
Zn	A3	P6 ₃ /mmc	-1.266	-1.266	-1.157							
Ga_d	A11	Cmca	-3.032	-3.032	-2.902							
Ge_d	A4	Fd $\bar{3}$ m	-4.624	-4.624	-4.522							
As	A7	R $\bar{3}$ m	-4.652	-4.652	-4.593							
Se	A8	P3 ₁ 21	-3.481	-3.481	-3.374							
Br		Cmca	-1.606	-1.317	-1.333						Liquid	
Kr	A3	P6 ₃ /mmc	-0.004	-0.004	-0.004	A1	Fm $\bar{3}$ m	0.003			Gas	
Rb_sv	C19	R $\bar{3}$ m	-0.963	-0.963	-0.881	A2	Im $\bar{3}$ m	0.001				
Sr_sv	A1	Fm $\bar{3}$ m	-1.683	-1.683	-1.549							
Y_sv	A3	P6 ₃ /mmc	-6.464	-6.464	-6.449							
Zr_sv	A3	P6 ₃ /mmc	-8.547	-8.547	-8.438							
Nb_pv	A2	Im $\bar{3}$ m	-10.094	-10.094	-10.017							
Mo_pv	A2	Im $\bar{3}$ m	-10.848	-10.848	-10.921							
Tc_pv	A3	P6 ₃ /mmc	-10.361	-10.361	-10.457							
Ru	A3	P6 ₃ /mmc	-9.202	-9.202	-9.210							
Rh	A1	Fm $\bar{3}$ m	-7.269	-7.269	-7.319							
Pd	A1	Fm $\bar{3}$ m	-5.177	-5.177	-5.197							
Ag	A3'	P6 ₃ /mmc	-2.822	-2.822	-2.907	A1	Fm $\bar{3}$ m	0.000				
Cd	A3	P6 ₃ /mmc	-0.900	-0.900	-0.861							
In_d	A6	I4/mmm	-2.720	-2.720	-2.609							
Sn_d	A4	Fd $\bar{3}$ m	-4.007	-3.895	-3.938				A5		I41/amd	0.042
Sb	A7	R $\bar{3}$ m	-4.118	-4.118	-4.155							
Te	A8	P3 ₁ 21	-3.142	-3.142	-3.027							
I		Cmca	-1.509	-1.344	-1.365							
Xe	A6	I4/mmm	0.003	0.003	-0.640	A1	Fm $\bar{3}$ m	0.004			Gas	
Cs_sv	A3	P6 ₃ /mmc	-0.855	-0.855	-0.744	A2	Im $\bar{3}$ m	0.002				
Ba_sv	A2	Im $\bar{3}$ m	-1.924	-1.924	-1.479							
La	A3'	P6 ₃ /mmc	-4.935	-4.935	-4.959							
Ce_3	A3'	P6 ₃ /mmc	-4.777	-4.777	-4.564	A1	Fm $\bar{3}$ m	0.006	A3'		P6 ₃ /mmc	0.000
Pr_3	A3'	P6 ₃ /mmc	-4.775	-4.775	-4.627							
Nd_3	A3'	P6 ₃ /mmc	-4.763	-4.763	-4.697							
Pm_3	A3'	P6 ₃ /mmc	-4.745	-4.745	-4.716							
Sm_3	A3'	P6 ₃ /mmc	-4.715	-4.715	-4.606				C19		R $\bar{3}$ m	0.004
Eu_2	A2	Im $\bar{3}$ m	-1.888	-1.888	-1.708							
Gd_3	A3'	P6 ₃ /mmc	-4.655	-4.655	-4.718	A3	P6 ₃ /mmc	0.017				
Tb_3	C19	R $\bar{3}$ m	-4.629	-4.629	-4.731	A20	Cmcm	0.017	A3		P6 ₃ /mmc	0.012
Dy_3	C19	R $\bar{3}$ m	-4.602	-4.602	-4.660	A20	Cmcm	0.013	A3		P6 ₃ /mmc	0.008
Ho_3	C19	R $\bar{3}$ m	-4.577	-4.577	-4.573	A3	P6 ₃ /mmc	0.003				
Er_3	A3	P6 ₃ /mmc	-4.563	-4.563	-4.582							

Table 3. (Continued)

Potential	DFT ground state μ_i (eV/atom)					Exp. LT			Exp. RT		
	ID	SG	Fit-none	Fit-partial	Fit-all	ID	SG	ΔE (eV/atom)	ID	SG	ΔE (eV/atom)
Tm_3	A3	P6 ₃ /mmc	-4.475	-4.475	-4.451						
Yb_2	A1	Fm $\bar{3}$ m	-1.513	-1.513	-1.125	A3	P6 ₃ /mmc	0.007	A1	Fm $\bar{3}$ m	0.000
Lu_3	A3	P6 ₃ /mmc	-4.524	-4.524	-4.549						
Hf_pv	A3	P6 ₃ /mmc	-9.955	-9.955	-9.902						
Ta_pv	A2	Im $\bar{3}$ m	-11.853	-11.853	-11.941						
W_pv	A2	Im $\bar{3}$ m	-12.960	-12.960	-13.130						
Re	A3	P6 ₃ /mmc	-12.423	-12.423	-12.378						
Os_pv	A3	P6 ₃ /mmc	-11.226	-11.226	-11.374						
Ir	A1	Fm $\bar{3}$ m	-8.855	-8.855	-8.953						
Pt	A1	Fm $\bar{3}$ m	-6.056	-6.056	-6.162						
Au	A1	Fm $\bar{3}$ m	-3.267	-3.267	-3.283						
Hg	A12	I $\bar{4}$ 3m	-0.298	-0.376	-0.374	A _a	I4/mmm	0.074		Liquid	
Tl_d	A3	P6 ₃ /mmc	-2.359	-2.359	-2.480						
Pb_d	A1	Fm $\bar{3}$ m	-3.704	-3.704	-3.951						
Bi_d	A7	R $\bar{3}$ m	-4.039	-4.039	-4.199						
Ac	A3'	P6 ₃ /mmc	-4.106	-4.106	-4.106	A1	Fm $\bar{3}$ m	0.012			
Th	A1	Fm $\bar{3}$ m	-7.413	-7.413	-7.237						
Pa	A1	Fm $\bar{3}$ m	-9.496	-9.496	-9.497	A _a	I4/mmm	0.017			
U	A20	Cmcm	-11.292	-11.292	-11.032						
Np		Pnma	-12.940	-12.940	-12.797						
Pu		P2 ₁ /m	-14.298	-14.298	-13.950						

Abbreviations: DFT, density functional theory; Exp. LT, experimentally observed lowest temperature; Exp. RT, experimentally observed room temperature PAW-PBE, projected augmented wave-Perdew, Burke and Ernzerhof; VASP, Vienna *Ab-initio* Simulation Package.

The 'Potential' column corresponds to VASP PAW-PBE potential names for the elements used in the current work. The ID and SG columns list the Strukturbericht designation and the space group of the crystal structures, respectively. The 'fit-none' chemical potentials are the DFT ground-state total energies of each element, whereas the 'fit-partial' and 'fit-all' correspond to the chemical-potential correction schemes described in the section 'Formation Energy Calculation'. The 'Exp. LT' and/or 'Exp. RT' ground states^{29,30} are given if they differ from the DFT-predicted 0-K ground state (RT only provided if it differs from LT), along with the difference between the DFT-predicted total energy of the Exp. LT or RT structure and DFT ground-state energy (energy difference only provided if the DFT ground state differs from Exp. LT or Exp. RT).

calculate each element in 20 elemental ground-state structures, listed in Table 2. For each element, therefore, we have the energetics of a large number and geometric variety of (≥ 20) crystal structure types, and we refer to the structure with the lowest energy as the DFT-predicted $T=0$ K ground state. Previous efforts at a comprehensive comparison of experimentally observed elemental ground-state structures and DFT-predicted elemental ground-state structures have been limited to HCP, FCC and BCC structures.³¹ Hence, our present study provides a more complete, systematic investigation of the ability of DFT (at the settings of the OQMD) to predict the correct 0-K elemental ground-state crystal structures.

Table 3 shows the DFT-predicted ground-state structure and the experimentally observed ground-state structure^{29,30} for all 89 elements, for which a potential is available in VASP. During DFT calculations of the 89 elements in the wide spectrum of crystal structure types tested, it is possible for an element to relax to a higher-symmetry parent structure starting from a lower-symmetry prototype-based crystal structure. Therefore, for all elements in Table 3, we analysed the OQMD-relaxed structures of elements that are almost degenerate with the ground-state structure (taken as within ~ 2 meV/atom) using the structure comparison approach outlined in the section 'Structures in the OQMD'. We found that only a small set of structures (Al, Ni, Rh, Ir and Th in A6, K, Sr, Al and Th in A7, Sr in A8 and Pa in A10 structures) relaxed to the higher-symmetry A1 (FCC) structure, and only Mg and Er in the A20 structure relaxed to the higher-symmetry A3 (HCP) structure.

Of the 89 elements in Table 3, the OQMD prediction agrees with the experimentally observed low-temperature structure for 82 elements to within 12 meV/atom, and 77 elements to within 5 meV/atom. For the 12 elements with a discrepancy between the OQMD-predicted ground state and the experimentally observed

ground state greater than 5 meV/atom, we look for possible sources of error. The elements for which we fail to correctly predict the ground-state structure are He, F, P, Ar, Ce, Gd, Tb, Dy, Yb, Hg, Ac and Pa. There are three distinct groups among these elements: first are noble gases and molecular solids, i.e., He, Ar, F and P, second are solids of elements with *f*-electrons, i.e., Ce, Gd, Tb, Dy, Yb, Ac and Pa and third is Hg. For most of these elements, we found no change in the predicted ground states with calculations at ~ 30 – 100% higher plane-wave basis-set cutoffs. The exceptions are He, F and Ar for which the higher-cutoff calculations can reproduce the experimental ground states. However, for such noble gas and molecular solids, it is expected that van der Waals interactions contribute a significant portion of the total energy and therefore this change in stabilities cannot be substantiated. For phosphorous, which has the second largest discrepancy between the OQMD-predicted ground state and the experimental ground state, the experimental ground state consists of layers of covalently bonded atoms that interact primarily through van der Waals forces, and increasing the cutoff has no effect on the predicted relative stability of P allotropes. For all elements in the first group above, the ground-state structure cannot be determined reliably with (semi)-local exchange-correlation functionals used in DFT because of the lack of van der Waals interactions, and can likely be corrected by their inclusion.³²

The second group—elements that contain *f*-electrons—presents its own set of known challenges for DFT.^{33–38} There are two sources of error for these elements. For Ce, Gd, Tb, Dy and Yb we use the 'frozen' potentials, meaning that the *f*-electrons are frozen in the core and not treated explicitly as valence. These potentials may introduce errors because freezing the *f*-electrons into the core neglects interactions involving these electrons. For example, for Ce, we found that the experimental low-

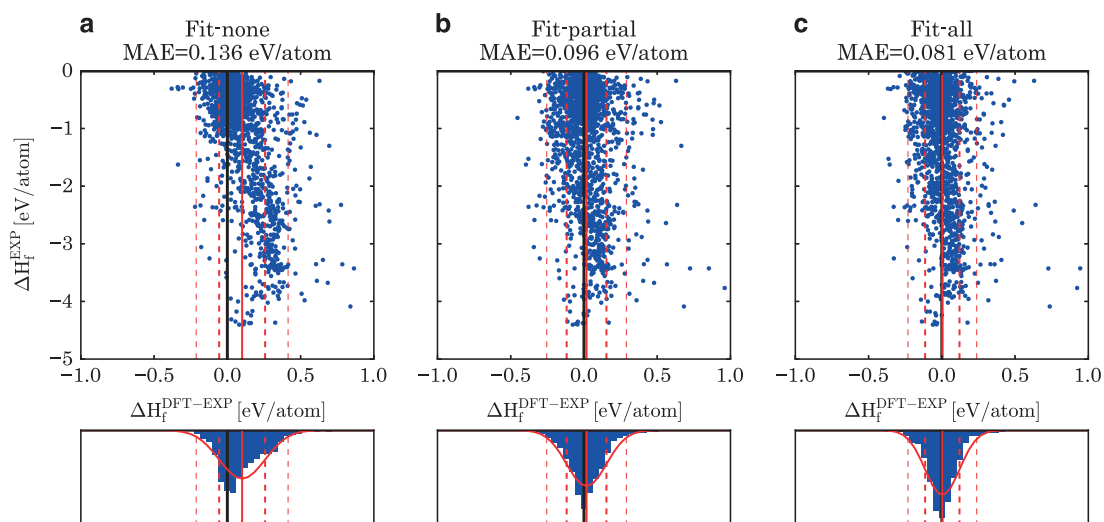


Figure 1. Comparison between the OQMD and 1,670 experimentally measured formation energies for three different sets of elemental chemical potentials. **(a)** Fit-none reference states: the DFT energy of the OQMD ground-state structure is taken as the chemical potential for each element. **(b)** Fit-partial reference states: chemical potentials for 13 elements where the DFT ground-state structures are known to poorly represent the STP elemental chemical potentials are fit to experimental formation energies (for all other elements, the reference state is the DFT energy). **(c)** Fit-all reference states: the chemical potentials for all elements are fit to experimental formation energies. The solid red line in each plot corresponds to the average error between DFT and experiment. The dashed red lines indicate the first and second s.d.'s. The curves in the lower plots correspond to normal distributions computed from the mean and s.d. of each data set.

temperature structure (α -Ce) can be reproduced as the ground state using the Ce potential with f -electrons in the valence. However, in the OQMD, frozen potentials are used because the high degree of correlation of f -electrons and related self-interaction error when they are included as valence is hard to accurately treat with DFT. Pa and Ac are examples where frozen potentials are not available, and therefore the 'free' f -electrons lead to errors. The issues with f -electrons in DFT, as well as the trade-offs that come with using frozen potentials, are thoroughly discussed elsewhere.^{33–37} Lastly, it is already known that the relative stabilities of allotropes of Hg cannot be reproduced with local density approximation or generalized gradient approximation (GGA),³⁹ and relativistic effects such as spin-orbit coupling (excluded from our DFT calculations in the OQMD) are essential for accurate treatment of Hg.⁴⁰

Formation Energies of Compounds

One of the most useful quantities that we have calculated for each compound is its formation energy—the energy required to form (positive formation energy), or given off by forming (negative formation energy), a compound from its constituent elements. Compound-formation energies are required to predict compound stability, generate phase diagrams, calculate reaction enthalpies and voltages and determine many other material properties. Because this quantity is so ubiquitous, it is important to determine the trustworthiness of our predictions. Although previous large-scale investigations of DFT's accuracy in predicting formation energies have been performed, these investigations have either been limited in the scope of material type considered^{41–44} or in the quantity of structures assessed.⁴⁵ The database of formation energies for 297,099 structures in the OQMD allows for the most comprehensive assessment of the ability of DFT to predict formation energies for solids to date.

As DFT calculations are performed at 0K and experimental formation energies are typically measured at room temperature, we must consider how much we can expect the formation energy to change between 0 and 300 K. The largest source of differences between 0 and 300 K formation energies is the existence of phase transformations in this temperature range. These phase transformations can take the form of solid–liquid, solid–gas or

solid–solid transformations and lead to significant changes in energetics. In addition to the elements that are gaseous or liquid at room temperature, at least five elements are known to exhibit a solid–solid transformation below 300 K: Ce, Na, Li, Ti and Sn.²⁹ As shown in Table 3, the energy differences between the 0-K- and room-temperature structures of Ce, Na and Li are less than 7 meV/atom. For Ti and Sn, however, the energy differences are 14 and 42 meV/atom, respectively, which will introduce systematic errors in the OQMD-predicted formation energies when comparing with experimental formation energies. For this reason, the chemical potentials of Ti and Sn have been fit to experimental data, as discussed in the section 'Formation Energy Calculation'.

In the following sections we make extensive use of a large collection of experimental formation energies. These experimental formation energies come from two sources: the SGTE Solid SUBstance (SSUB) database,⁴⁶ from which we obtain 1,702 compound-formation energies, and the thermodynamic database at the Thermal Processing Technology Center at the Illinois Institute of Technology (IIT),⁴⁷ from which we obtain 994 compound-formation energies. The SSUB contains many oxides (680), nitrides (75), hydrides (102) and halides (369), and relatively few intermetallics (272). In contrast, the IIT database is exclusively intermetallic compounds. By combining these databases we have a total of 2,712 experimental formation energies to compare with. For Th, Pa and Np oxides, which do not appear in either of these databases, we also include formation energies reported in a review of actinide thermodynamics,⁴⁸ which includes measured formation energies for ThO₂, Np₂O₅ and NpO₂ and estimated formation energies for PaO₂ and NpO₃ from trends among similar actinide oxides.

Because the different experimental data sources specialise in different types of materials, they have relatively limited overlap, but they do have some compositions in common. As a result, after controlling for double counting of compositions within and between databases, we still have 2,233 distinct compositions with experimental formation energies. The number of comparisons is further reduced because some of the compositions for which we have experimental formation energies have no known corresponding crystal structure. One might wish to ascertain whether, for a given stoichiometry, a given experimental formation energy

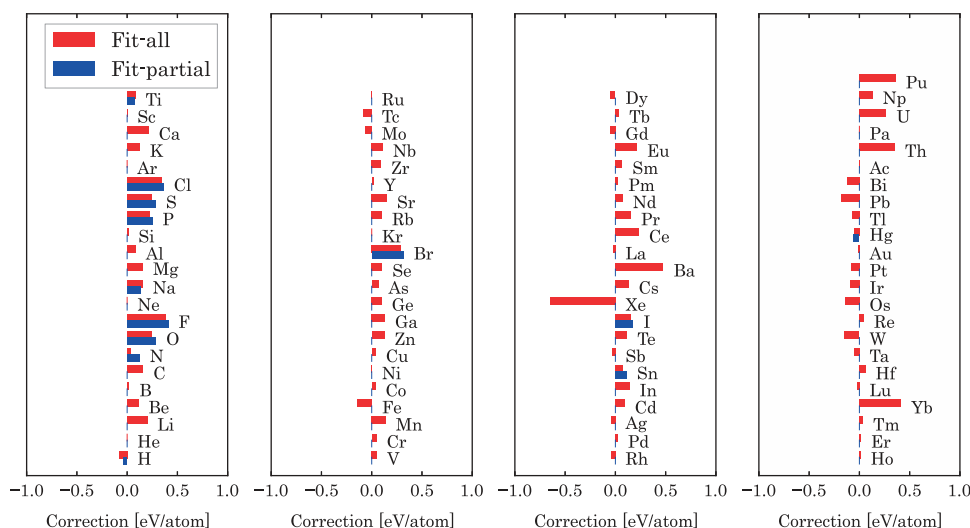


Figure 2. Corrections to chemical potentials ($\mu_{\text{fit}} - \mu_{\text{DFT}}$) as determined by fitting OQMD formation energies to experimental formation energies. The corrections in blue were obtained by fitting only the chemical potentials of elements whose STP phase differs significantly from the 0-K phase, while the corrections in red were obtained by fitting the chemical potentials of all elements simultaneously.

corresponds precisely to the phase with crystal structure reported in the ICSD. However, this comparison is not possible because the thermodynamic databases often do not have any explicit crystal structure information. We therefore assume that the existence of a corresponding crystal structure in the ICSD is sufficient to include a composition in the formation energy comparison, and, if more than one crystal structure is available, we always compare with the lowest-energy OQMD structure. As a result, we have a total of 1,670 formation energies to compare between the OQMD and experiment—the largest comparison of this kind that has been performed to date.

Formation Energy Calculation. In general, the formation energy for a compound is given by,

$$\Delta H_f = E_{\text{tot}} - \sum_i \mu_i x_i \quad (1)$$

where E_{tot} is the DFT total energy of the compound, μ_i is the chemical potential of element i and x_i is the quantity of element i in the compound. The standard convention is to take the chemical potential of each species to be the DFT total energy of the elemental ground state. With this choice, the computed formation energy is valid only for 0 K. However, as the available experimental formation energies are typically measured at room temperature or above, it is useful to understand how well equation 1 approximates the standard temperature and pressure (STP) formation energy. In order to answer this question, we calculate the formation energy of all OQMD-calculated compounds and compare the results with experimentally measured formation energies.

Using the elemental DFT total energies as chemical potentials, we find that the OQMD-calculated formation energies have an average error of 0.105 eV/atom with respect to experimentally measured formation energies and a mean absolute error (MAE) of 0.136 eV/atom. The 1,670 formation energies used in this comparison include a wide range of compositions, including oxides, semiconductors and intermetallics. This error is plotted against experimental formation energy in Figure 1a. There is a clear systematic error: compounds with more-negative formation energies underestimate formation energies compared with those having less-negative formation energies. This trend is easily understood; highly stable compounds, such as oxides and halides, frequently contain elements whose 0-K ground states differ from the STP stable phases, e.g., gases. In order to improve these

formation energies, we fit chemical potentials to experimental formation energies.

The use of experimental data to fit elemental reference energies has been shown to reduce systematic error in DFT formation energies.^{11,44,45,49,50} In this work we perform simultaneous least squares fitting in the manner of the Fitted Elemental Reference Energies method,⁴⁵ which was developed for finding chemical potentials for GGA+U oxides; however, in our work we extend this method to a wider set of elements involving a much wider set of compounds. In our approach we apply two distinct corrections; therefore, we perform two independent fits. The first fit is to calculate chemical potentials for elements with ground states that are not applicable to STP, and the second fit is to find corrections to chemical potentials for elements to which we apply GGA+U (Table 1). In the first step, we correct all the elemental chemical potentials simultaneously, fitting to the experimental formation energies (binary, ternary, quaternary and so on) of every applicable compound that does not contain a GGA+U element, and in the second step we determine values for the chemical potentials of GGA+U elements (described below).

To evaluate the efficacy of chemical-potential fitting, we define three fit sets that we will compare throughout this section and the next. The first fit set is empty, which we will label ‘fit-none,’ and corresponds to completely uncorrected DFT. The second fit set we will label ‘fit-partial,’ which refers to the elements for which we can rationally argue that the DFT energy of the $T=0$ K ground state is not an accurate reference state for STP formation energies. We have identified five groups of elements to include in this fit set. These groups are room-temperature diatomic gases (H, N, O, F and Cl), room-temperature liquids (Br and Hg), molecular solids (P, S and I), several elements with structural phase transformations between 0 and 298 K (Na, Ti and Sn) and elements employing GGA+U for oxide-formation energies (Table 1). The third fit set, which we will label ‘fit-all,’ has the chemical potential for every element fit to experiment; none of the chemical potentials come from DFT.

For the case of GGA+U elements, the chemical potentials of the elements in structures with the U correction (e.g., Fe in Fe_3O_4) and without the U correction (e.g., Fe in BCC Fe) are different.¹¹ To address this, we fit corrections for the GGA+U elements (Table 1) using the method of ref. 11 *after* all other corrections described in this section; no compounds for which GGA+U was applied were used in the fitting of chemical potentials for non-GGA+U species. For GGA+U compounds, all corrected chemical potentials (e.g., O_2)

Table 4. Comparison of errors between experimental and OQMD-predicted formation energies for a variety of material classes, under different sets of elemental chemical potentials

Category	$\langle \Delta H_f^{\text{EXP}} \rangle$	Number of comps.	Fit-none		Fit-partial		Fit-all	
			Err	MAE	Err	MAE	Err	MAE
<i>Magnetism</i>								
Magnetic	-1.231	327	0.135	0.151	0.052	0.102	0.023	0.097
Non-magnetic	-1.234	1343	0.098	0.132	0.011	0.093	-0.005	0.076
<i>Bandgap</i>								
Metallic ($E_g = 0$)	-0.706	921	0.056	0.098	0.031	0.086	0.005	0.077
Semi-conductor ($0 < E_g < 2$)	-1.104	249	0.137	0.155	0.017	0.115	0.007	0.093
Wide bandgap ($E_g > 2$)	-2.175	491	0.180	0.195	0.003	0.101	-0.006	0.083
<i>Number of components</i>								
Binary	-0.935	1376	0.089	0.124	0.025	0.097	0.002	0.082
Ternary	-2.187	259	0.179	0.191	0.002	0.094	0.002	0.082
Quaternary	-2.439	33	0.207	0.214	0.019	0.085	0.020	0.071
<i>Bonding type</i>								
I-VII (ionic)	-2.225	18	0.224	0.224	0.090	0.090	0.052	0.057
III-V (covalent)	-0.577	15	0.144	0.146	0.095	0.123	0.056	0.090
Intermetallic binary (metallic)	-0.434	278	0.006	0.071	-0.002	0.069	-0.013	0.069
<i>Binary compounds that contain a/an.^a</i>								
Alkali metal	-1.047	87	0.120	0.126	0.052	0.081	0.017	0.056
Alkali earth metal	-1.256	92	0.161	0.191	0.098	0.128	0.009	0.092
Transition metal	-0.727	855	0.051	0.094	0.002	0.079	-0.006	0.074
f-block metal	-1.106	414	0.098	0.128	0.060	0.098	0.010	0.074
Semi-metal ^b	-0.573	317	0.046	0.093	0.027	0.093	-0.009	0.082
Post-transition metal ^c	-0.651	233	0.078	0.101	0.028	0.093	-0.005	0.075
Halide	-1.717	258	0.199	0.214	0.015	0.135	0.014	0.122
Compounds w/o corrections ^d	-0.514	780	0.031	0.081	0.031	0.081	-0.009	0.066
Overall	-1.232	1670	0.105	0.136	0.020	0.096	0.002	0.081

Abbreviations: comps, compounds; Err, average error; MAE, mean absolute error; OQMD, Open Quantum Materials Database; w/o, with out.

The definitions of the different fits are given in the section 'Formation Energy Calculation'. Average experimental formation energies ($\langle \Delta H_f^{\text{EXP}} \rangle$), Err and MAE are given in eV/atom.

^aAt least one of the components in the binary compound belongs to a given subcategory of elements, and the second component can be any element.

^bWe take semimetal elements to be B, Si, Ge, As, Sb and Te.

^cWe take post-TM elements to be Al, Ga, In, Sn, Tl, Pb and Bi.

^dIncludes all compounds that do not contain any elements that require GGA+U or chemical-potential corrections.

were applied first, and then the GGA+U correction was determined.

The difference between the fitted chemical potential (for both fit-partial and fit-all sets) and DFT ground-state energy for each element is shown in Figure 2, and the values of the chemical potentials for each element in each fit set are provided in Table 3. The fit-partial corrections are consistently larger than the fit-all corrections for the same elements. However, compared with the magnitude of the corrections, the difference between fit-partial and fit-all corrections is quite small in all cases as observed in Figure 2. Adding more fitting parameters in fit-all does not substantially alter the fitted chemical potentials from fit-partial, suggesting that the corrections in fit-partial capture the majority of the error associated with those particularly problematic elemental chemical potentials.

In the fit-all set, some elements appear to have surprisingly large corrections. These are elements that have relatively few compounds to fit to—the actinides and some lanthanides have comparatively large corrections. The small pool of formation energies may lead to overcorrection, and less-predictive calculated formation energies. *For this reason, the formation energies that are available through www.oqmd.org/download are all computed using the fit-partial correction scheme.* In the section 'Comparison of OQMD to Experiment', we next provide a detailed analysis of the agreement between the OQMD and experimental

formation energies, and how the agreement with experiment depends on the fit set being used.

Comparison of OQMD to Experiment. To understand the limits of applicability of our large database of calculated thermochemical data, it must be comprehensively validated against a known reliable source. In the case of the OQMD, we attempt to validate the predicted energies of formation by comparing with as many experimental formation energies as possible. In this section, we compare OQMD values for formation energies to experiment in several ways. First, we will look at the statistics of the agreement between OQMD and experiment and investigate the outliers in the data set. For this analysis we will also consider the effects of different fit sets, described in the section 'Formation Energy Calculation', and determine which set of chemical potentials we find to be most trustworthy. Then, we explore variation in the error for various material classes. Finally, we consider how to correctly distribute error between DFT and experiment, and how to improve the predictive power of DFT-formation energies.

In Figure 1 we compare the OQMD formation energies to experimental formation energies, and in Table 4 we present detailed statistics for the same data. In Figure 1a–c, we show the difference between OQMD and experiment for the fit-none, fit-partial and fit-all chemical-potential sets, respectively. In the case of fit-none, the average difference is 0.105 eV/atom, with a MAE of

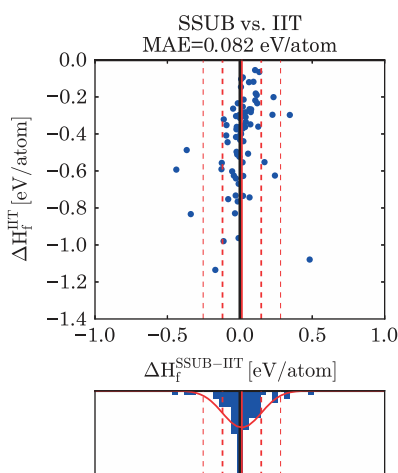


Figure 3. Illustration of the lack of agreement between the IIT⁴⁷ and SSUB⁴⁶ experimental thermochemical databases. Plots average error ($\Delta H_f^{\text{SSUB-IIT}}$) against IIT formation energy (ΔH_f^{IIT}), with the distribution of errors summarised in the histogram at the bottom. The significant range of ΔH_f^{IIT} values demonstrates the surprising degree of disagreement between these experimental formation energy databases. This experiment to experiment comparison for the 75 intermetallic compounds common to both databases gives a MAE of 0.082 eV/atom, whereas the MAE of OQMD formation energies for intermetallic compounds is 0.071 eV/atom (using ‘fit-none’ chemical potentials).

0.136 eV/atom. Using chemical potentials from the fit-partial set, we find that the average error is reduced to 0.020 eV/atom and the MAE is 0.096 eV/atom. Finally, we find that, by fitting the chemical potentials of all elements, the average error is 0.002 eV/atom and MAE is 0.081 eV/atom, a slight improvement compared with the difference between fit-none and fit-partial chemical potentials.

With such a diverse database of experimental formation energies to compare with, we are able to look for trends in the errors across various material classes as a function of chemical-potential fitting. Table 4 compares the errors between experiment and the OQMD formation energies for a variety of material classes for the fit-none, fit-partial and fit-all chemical-potential sets. Note that for all classes of compounds in the fit-none set, the average error (OQMD–EXP) is positive, which agrees with the expectation that, in the generalised gradient approximation, on average DFT underbinds or underestimates the stability of compounds.

Below, we discuss in more detail the comparison between the OQMD and experimental formation energies for the specific classes of materials in Table 4:

Magnetism: The MAE of magnetic-structure formation energies is larger than that of non-magnetic structures regardless of chemical-potential choice—0.097 vs. 0.076 eV/atom, using the fit-all chemical potentials. Including more complicated magnetic ordering beyond the ferromagnetism assumed here could lead to lower-energy structures, reducing the MAE of magnetic-compound-formation energies.

Bandgap: Without any fitting, the difference in the formation energy of a compound between the OQMD and experiment is largest for wide bandgap insulators, smaller for semiconductors and smallest for metals. As expected, formation energy error magnitudes follow the trend in absolute formation energy magnitudes for these classes of compounds. Wide bandgap compounds are expected to be found in material classes such as nitrides, fluorides or oxides, for which the uncorrected chemical potentials are unreasonable, which results in systematically skewed formation energies. Once corrections to the chemical

potentials are applied, however, the error becomes largely independent of band gap.

Number of components: We find that the average and the MAE between OQMD-calculated and experimental formation energies (using the fit-none chemical-potential set) increase as the number of elements in the compound increases from 0.124 to 0.191 to 0.214 eV/atom for the MAE of binary, ternary and quaternary compounds, respectively. This trend disappears when either the fit-partial or fit-all chemical potentials are used. There are only 33 quaternary compound-formation energies to compare between OQMD and experiment, and all of them are oxides, causing the effect of chemical-potential fitting to be large for this category.

Bonding type: We compare DFT accuracy for three groups of binary compounds: alkali-metal–halide (I–VII), III–V and intermetallic (both elements are metals, including alkali metals, alkaline earth, lanthanides, actinides, transition metals and poor metals). These groups correspond roughly to ionic, covalent and metallic bonding characters, respectively. Without any fitting, intermetallic binary compound-formation energies have the smallest MAE with respect to experimental formation energies, followed by covalent binary compound-formation energies, and finally ionic compound-formation energies have the largest MAE of the three bonding types. In addition, the accuracy of intermetallic formation energies is almost completely unaffected by chemical-potential fitting. In contrast, ionic compounds have formation energies that are systematically more positive than experiment (Err. equals MAE for fit-none and fit-partial chemical potential sets in Table 4), and ionic compound-formation energies were significantly improved by both chemical-potential corrections, with their MAE shrinking from 0.224 to 0.090 to 0.057 eV/atom for fit-none, fit-partial and fit-all, respectively.

Binary compounds that contain a/an:

Alkali metal: Binary compounds containing alkali metals have formation energies with MAEs slightly below the overall average for all chemical-potential sets, and have reductions in MAE from fit-none to fit-partial and from fit-partial to fit-all of 0.045 and 0.025 eV/atom, respectively.

Alkaline earth metal: Alkaline earth binaries have the largest reduction in MAE between fit-partial and fit-all—an improvement of 0.036 eV/atom. Most categories have less than a 0.02-eV/atom improvement over the same range, suggesting that the alkaline earth elements have systematic errors in their DFT reference state energies.

Transition metal: As the largest binary compound data set with 855 compounds, transition metal containing binary compounds have agreement with experiment that is slightly below the overall average, and show little improvement between the fit-partial and fit-all chemical potentials.

f-block element: Binary compounds containing f-block elements show consistent improvement in MAE across all fitting levels and have MAEs comparable to the overall MAE.

Semi-metal: Semi-metal containing binaries have one of the lowest MAEs before fitting and one of the smallest changes in MAE between the fit-none and fit-all chemical-potential sets, suggesting that the largest error component in these compounds cannot be addressed by simply adjusting chemical potentials.

Post-transition metal: These compounds show consistent improvement in accuracy across all levels of fitting, and overall average accuracy.

Halide: Across all fitting levels, binary halides have the largest MAE of all binary compound categories.

Next, we consider several possible sources of error between the OQMD and experiment. First, there are several ways in which the OQMD formation energies can be expected to be improved. One significant approximation made during our calculations is the assumption that all magnetic structures are ferromagnetic. Another potential improvement we could make would be to assign *U*-values for more elements and systems. In this study we

applied DFT+U correction only to transition metal oxides and actinide oxides; however, by applying this correction to more compositions—both applying GGA+U corrections to additional cations and applying those corrections in the presence of additional anions—further improvements in formation energies may be achieved. For example, local-environment (anion and oxidation state)-dependent GGA+U calculations have recently been shown to provide improved thermochemical accuracy in transition metal oxides and fluorides.⁵¹ For systems where dispersion interactions are important, such as molecular or layered crystals, more accurate predictions may require van der Waals inclusive methods beyond GGA.^{32,52} Lastly, we should note that DFT-predicted bulk formation energies serve as a $T=0$ K starting point for further thermodynamic analysis. Enthalpic and entropic contributions at finite temperatures such as lattice dynamics, configurational defects and order–disorder transitions can be captured more accurately with relevant statistical, mechanical and DFT methods.

Assessing the Accuracy of Experimental Formation Energies. Improvements to the DFT calculation scheme may lead to some reductions in the discrepancy between the OQMD and experimental formation energies. However, some of the errors can also be attributed to the experimental formation energies themselves. We wish to ascertain the size of this error or uncertainty. With multiple experimental data sources to draw from, we can compare experimental measurements of a given compound with one another. Figure 3 shows the discrepancies between formation energies from different experimental sources for the same compounds. The resulting MAE of the experimental values is surprisingly large, 0.082 eV/atom. This experimental error is calculated based on a comparison between the 75 compounds common to both of the experimental databases used in this study. Note that this comparison is limited by the fact that the IIT database is strictly intermetallic compounds, and therefore all of the energies compared are for intermetallics. Of course we acknowledge that not all experimental data are equally reliable or accurate and that advances in techniques can yield more accurate data. However, these comparisons are for curated databases, and therefore might reasonably be expected to represent a high degree of experimental accuracy.

The OQMD formation energies (using the ‘fit-none’ chemical potentials, i.e., uncorrected DFT formation energies) have a MAE relative to experiment of 0.071 eV/atom for similar compounds, i.e., intermetallics, which is slightly less than the experimental error with a second experiment. For this same set of 75 compounds, we compare the OQMD formation energy with each of the experimental values. The minimum MAE between experiment and DFT (i.e., comparing DFT with the experimental formation energy closer in energy to the DFT formation energy for each compound) is 0.057 eV/atom, while the maximum MAE between experiment and DFT (i.e., comparing DFT with the experimental formation energy farther away in energy from the DFT formation energy for each compound) is 0.116 eV/atom. From this result it is clear that, where experiments disagree, DFT is often significantly closer to one experiment than the other.

Given this level of disparity in experimental formation energies, it is highly unlikely that *all* of the errors should be attributed to DFT. In fact, without additional information, it is impossible to fairly determine which values are in error. *We conclude from these comparisons that there remains a need for additional experimental thermochemical data.* This is particularly urgent as many computational schemes^{45,49,50} rely on these experimental values to obtain elemental correction factors.

To explore the source of disparity between DFT and experiment further, we looked at several of the compounds that have the poorest agreement with experiment, and searched the literature for alternative values for their formation energies. For the cases

Table 5. Comparison of SSUB database⁴⁶ and alternative sources for experimental formation energies

Composition	SSUB	ΔH_f (eV/atom) alternative	OQMD
LiNbO ₃	−5.660	−2.774 ^a	−2.739
Eu ₂ O ₃	−3.437	−3.423 ^b	−2.574
SiB ₆	−0.173	−0.181 ^c	0.455
EuF ₃	−4.104	−4.195 ^d	−3.483
MnSe	−0.889	−0.922 ^e	−0.363
InN	−0.500	−0.148 ^f	−0.021
AlB ₂	−0.521	−0.055 ^g	−0.044
PtZr	−1.561	−0.99 ^d	−1.086
Ba ₂ Pb	−1.012	−0.694 ^h	−0.540
PuCl ₃	−2.614	−2.486 ⁱ	−2.176
PrO ₂	−3.280	−3.316 ^j	−2.846
CaAl ₂	−0.759	−0.346 ^k	−0.335
BaSiO ₃	−3.354	−3.330 ^l	−2.945
CeN	−1.715	−1.763 ^m	−1.326

Abbreviations: OQMD, Open Quantum Materials Database; SSUB, Solid Substance.

Comparison made for outlying compounds that show a large disagreement between SSUB data and OQMD-predicted formation energies (using fit-partial chemical potentials).

Differences between SSUB and alternative-source formation energies ranging from 0.008 eV/atom (SiB₆) to 2.885 eV/atom (LiNbO₃).

^aReference 53.

^bReference 54.

^cReference 55.

^dReference 56.

^eReference 57.

^fReference 58.

^gReference 59.

^hReference 60.

ⁱReference 61.

^jReference 62.

^kReference 63.

^lReference 64.

^mReference 65.

where we were able to find another value, we show the composition, SSUB⁴⁶ formation energy and the literature formation energy in (Table 5) these compounds with very large discrepancies between experimental formation energies—the second value we found in literature is often closer to the OQMD-predicted formation energy. As a result, we conclude that some of the *very significant disagreements* between DFT and experiment are more likely to be due to experimental or transcription errors than to problems in DFT. On the basis of these findings, we believe that for many other compounds with large formation-energy errors, and for which no alternative formation energies could be found, the source of error might also be the experimental measurement, rather than only the DFT calculation.

Comparison of OQMD to Other DFT Databases and the Miedema Model. The Miedema model has historically been widely used to provide estimates of formation enthalpies of solid alloys and intermetallic compounds.⁶⁶ The Miedema model is a semi-empirical approach wherein atoms are conceptually treated as space-filling polyhedra. Chemical bonding is treated by considering the overlap of the surface areas of neighbouring atomic polyhedra, weighted by the difference in charge density of each atom at the boundary and the electronegativity difference between the atoms.⁶⁶ The model contains several element-dependent parameters that Miedema fit to trends in the formation energies of a range of binary intermetallic compounds, as well as elemental properties (bulk modulus, molar volume and work function), which were adjusted to give the best fit to experimental

Table 6. Comparison of predicted formation energies with experimental values for two DFT databases (the OQMD with ‘fit-partial’ corrections and Materials Project) and an empirical model for intermetallic compounds (Miedema model)

		Materials Project ^a	Miedema
Number of compounds		1,386	820
ΔH_f^{X-EXP}	Avg Err	0.006	0.033
	MAE	0.133	0.199
$\Delta H_f^{OQMD-EXP}$	Avg Err	0.032	0.029
	MAE	0.108	0.090

Abbreviations: API, application programming interface; Avg Err, average error; DFT, density functional theory; MAE, mean absolute error; OQMD, Open Quantum Materials Database.

Average error (ΔH_f^{X-EXP}) and MAE ($|\Delta H_f^{X-EXP}|$) are given in eV/atom.

^aData from the Materials Project used in this comparison were retrieved via the Materials Project API^{24,26} on 16/12/2013.

formation energies. A comparison of the accuracy of the Miedema model with the accuracy of the OQMD is important as the Miedema model is still actively employed.^{67,68} Table 6 contains a comparison of Miedema model predictions for formation energy to experiment. The MAE between the Miedema model and experiment is 0.199 eV/atom, greater than twice that of the OQMD for the same set of compounds, 0.090 eV/atom. This result indicates that, in addition to the inherent drawbacks of the Miedema model (i.e., applicable only to binary intermetallics), the OQMD is a much more accurate predictor than the Miedema model for formation energies.

The Materials Project^{11,24} was one of the first high-throughput databases to be developed. As the Materials Project database uses slightly different calculation parameters from OQMD, and a different chemical potential correction scheme, there is an interesting opportunity to directly compare the results of different DFT databases with one another. Table 6 shows the statistics of the agreement between Materials Project and experimental formation energies. The average error of formation energies for the Materials Project is 0.006 eV/atom, smaller than the average error of the OQMD formation energies, 0.032 eV/atom. However, the MAE for the Materials Project is 0.133 eV/atom, which is larger than that of the OQMD over the same compounds, which is 0.108 eV/atom.

We attribute the difference in MAE with experiment between OQMD and Materials Project to the difference in chemical-potential fitting procedures for the two data sets.^{11,24} The chemical-potential fitting used in the OQMD is performed on the same set of compounds on which the computed accuracy is based, which gives the OQMD a ‘natural advantage.’ Further evidence to support this argument can be found by calculating the mean absolute difference between the OQMD and Materials Project formation energies for all compounds for which we do not fit any chemical potentials. For this set of 563 compounds, the mean absolute difference between the OQMD and Materials Project is 0.028 eV/atom, much smaller than the difference between the OQMD and experiment (0.093 eV/atom for these 563 compounds) or between Materials Project and experiment (0.086 eV/atom for these 563 compounds). As a result, we conclude that in general the two databases contain very similar results, and that different choices for DFT parameters have a much smaller impact on compound-formation energies than do the different approaches to chemical-potential fitting. Finally, as new calculations are continuously added to both OQMD and Materials Project, the analysis above corresponds only to certain *snapshots* of each database; however, we expect this conclusion to be valid as long as the chemical-potential fitting approaches are not significantly revised.

Historical Trends in Material/Compound Discovery

A large thermodynamic database of energetics and phase stability of the type presented here can be used to address many interesting general trends. For instance, we leverage the fact that we have evaluated a significant fraction of known ground-state compounds to answer several questions about trends and patterns in material discovery and stability. Without a large database such as the OQMD it would otherwise be impossible to answer many of these questions.

How many stable compounds are in the database? How many of these are experimentally known versus theoretically predicted? In order to answer these questions, first we determine phase stability. Phase stability is determined by constructing the energy convex hull of a given region of composition space.⁶⁹ Once this has been determined using existing computational geometry algorithms,⁷⁰ (Kirklin, S. & Wolverton, C. (2015, unpublished)) every phase that lies on the convex hull is stable at $T=0$ K (i.e., it is lower in energy than any other phase or combination of phases in the database). Of the 297,099 calculated compounds in the OQMD, we find that 19,757 are thermodynamically stable at $T=0$ K. Of these, 16,526 were from the ICSD with the remaining 3,231 being prototype structures.

All the 3,231 compounds that we predict to be stable, but are not in the ICSD, represent new compounds to be discovered. The prototype compounds were constructed from commonly occurring, simple, crystal structures, and do not represent an exhaustive crystal structure determination for each predicted compound. For this reason, we do not assert that in all cases the predicted compounds are stable in the crystal structure we list. Rather, in these cases, our predicted convex hull is an upper bound to the true ground-state hull. Thus, for all the 3,231 cases, we predict that some new compound(s) are awaiting experimental discovery in these (binary and ternary) systems. A detailed crystal structure search in such systems can be made using evolutionary⁷¹ or minima hopping⁷² methods. Furthermore, because the prototype compounds have identified so many holes in our knowledge of ground-state phase stability, we expect that by including the prototype structures in our list of ground-state structures we are providing a better estimate of the energy landscape where we do not have experimentally measured structures.

What is the rate of stable material discovery? New materials and compounds are being discovered all the time, some of which are stable and some of which are not. Utilising the publication data associated with ICSD records, we can study the historical rate of material discovery. In Figure 4a we plot the total rate of material discovery as the number of new compounds reported per year since 1910. Each compound only appears in the year in which it was *first* reported. In Figure 4b, we plot the number of *stable* ICSD compounds discovered per year. The ‘material discovery’ data in Figure 4a come directly from the ICSD (no DFT is required). However, to classify a compound as ‘stable’ is not possible from the ICSD alone but requires some measures of energetics, as well as those of competing (combinations of) phases. The latter is possible only with a large material database containing formation energies, such as the OQMD.

We find that the rate of compound discovery is increasing with time—in most years, more compounds are discovered than the year before. In contrast, the rate of *stable* material discovery has been fairly constant since the 1960s. By decomposing the number of discovered materials into binary, ternary, quaternary and pentanary compounds, we observe that the number of stable binary compounds discovered each year has been dwindling since the 1970s, when the number of ternary compounds discovered began to significantly rise. As of the 1990s, the number of stable ternaries has also stagnated, while the number of stable quaternary compounds began to increase.

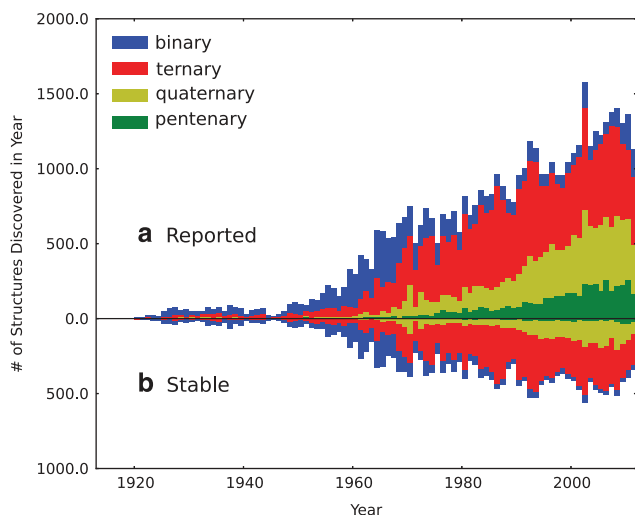


Figure 4. (a) The total number of compound discovery within the ICSD by year. (b) The number of *stable* ($T=0$ K) compound discovery in the ICSD, where the stability is assessed by the OQMD energies. The year for a structure corresponds to the earliest publication year for ICSD entries at that given structure's composition.

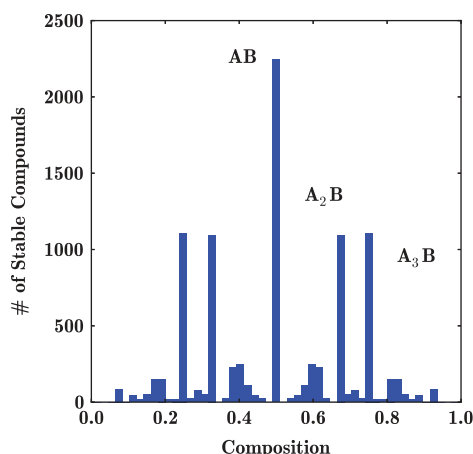


Figure 5. Distribution of stable binary compounds as a function of composition. Note the presence of large peaks at low integer ratios, i.e., 1:1, 2:1 and 3:1.

What compositions are most likely to be stable? On the basis of the same stability data, we can also look at which compositions are most frequently stable. In Figures 5 and 6 we provide histograms showing the frequency at which compositions are stable for binary and ternary compositions, respectively. We find that the most commonly stable binary composition is at 1:1 (AB), followed by 3:1 (A_3B) and 2:1 (A_2B). For ternary compounds, the most common composition is 2:1:1 (A_2BC), which is the composition of the L_2_1 (Heusler) prototype. We believe that the preponderance of newly predicted L_2_1 compounds is primarily because L_2_1 is the only ternary prototype that we have calculated in a wide range of compositions. Many ternary systems that have favourable ordering, but for which the true ground state is unknown, could yield predictions that the L_2_1 compound is stable (the upper bound to the convex hull), demonstrating the need for further exploration of a wide range of composition spaces. Following this interpretation, we checked every stable ternary L_2_1 prototype in the OQMD and searched for any ternary ICSD compound in that system—that is, if Ca_2GaLi is a stable Heusler, is there any ternary Ca-Ga-Li compound reported in the ICSD? Of the

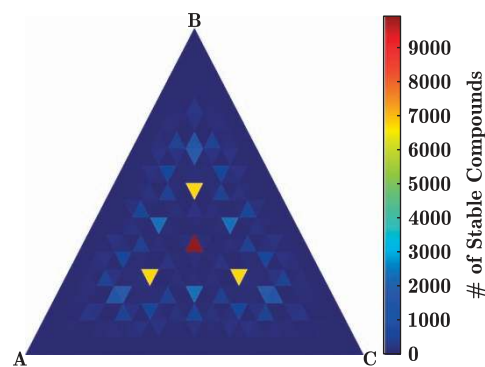


Figure 6. Distribution of stable ternary compounds as a function of composition. Plotted on log scale to account for the extremely high density of phases at A_2BC compositions because of the calculation of over 180,000 decorations of the L_2_1 structure.

2,290 stable Heusler prototypes in the OQMD, only 781, or 22%, have any ICSD structure in that region of phase space. In the remaining $\sim 1,500$ ternary systems, we predict the existence of stable compounds waiting to be discovered.

In order to facilitate the discovery of new, stable compounds in the thousands of regions of composition space where we predict stable compounds to exist, we provide a full list of compositions where we predict a prototype to be stable online. We break this list into (i) prototypes that are more stable than an experimentally measured structure and (ii) prototypes that have no experimental structure at that composition. In the first case, finite-temperature effects may cause the formation energy of the experimental structure to lower relative to the formation energy of the prototype. However, in the second case, *some* compounds should be found at the listed composition, although possibly not the prototype. A current list of predicted compounds can be found online at www.oqmd.org/materials/discovery where one can obtain the entire list or filter by composition.

CONCLUSIONS

The OQMD is a high-throughput database of DFT calculations of 32,559 ICSD structures and 259,511 prototypical structures, growing steadily as new structures are added continuously. OQMD is available for download without restrictions at www.oqmd.org/download. Included in the download is a complete framework for performing additional calculations that are commensurate with the database. We use the breadth of the OQMD to compare DFT-calculated structures and formation energies with experiment at an unprecedented scale. We find the following:

Elemental Ground-State Structures. Using the capabilities of qmpy and the OQMD, we find that for 77 out of the 89 elements DFT as implemented in VASP at the settings of the OQMD is able to correctly predict the observed ground-state structure as being lowest in energy out of 20 possible structures, chosen from among the known ground states of all elements. In all cases where DFT finds a lower-energy structure, the observed ground state is nearly degenerate with the lowest-energy structure (by far, the largest errors being phosphorous and mercury, with 0.036 and 0.074 eV/atom error, respectively).

Formation Energies. In order to most accurately determine compound-formation energies, we evaluate the effects of three different choices of chemical potentials: using DFT ground-state reference energies, fitting chemical potentials to experiment for elements where the DFT ground state differs significantly from the room temperature stable phase, and fitting chemical potentials for

all elements (labelled fit-none, fit-partial and fit-all). We find that fit-partial exhibits significant improvements over fit-none, with a reduction in MAE against experiment from 0.136 to 0.096 eV/atom. In comparison, fit-all has only marginal gains over fit-partial, with the MAE reducing to 0.081 eV/atom. However, by increasing the number of fit parameters, we also increase the risk of over-fitting, and, as a result, take fit-partial to be the optimal choice of chemical potentials for predicting formation energies.

To put the discrepancy between the OQMD and experiment into appropriate context, we also compared formation energies between the two database experimental formation energies used in this work (the SSUB database,⁴⁶ and the thermodynamic database at the Thermal Processing Technology Center at the IIT).⁴⁷ We find that the MAE between these two databases is 0.082 eV/atom, a value that is similar to the error between the OQMD and experiment. As a result, it is impossible to assign all of the errors between DFT and experiment solely to DFT, and leads us to conclude that, in order to establish a more accurate evaluation of the accuracy of formation energies based on DFT total energies, more accurate measurements of formation energies should be undertaken in future.

We also compare the OQMD with two other databases of calculated formation energies: the Miedema model and the Materials Project. We compare the Miedema model with experiment for all compositions for which we have (i) an experimental formation energy, (ii) an OQMD formation energy, and (iii) the Miedema model is applicable. Over the resulting pool of 820 compounds, the OQMD has a MAE of 0.090 eV/atom, less than half the error found in the Miedema model, 0.199 eV/atom. We made an identical comparison with the Materials Project, which had a comparison pool of 1,386 compounds. For this set of compounds, we found that the Materials Project had an MAE of 0.133 eV/atom, which is slightly larger than the OQMD error for the same set of compounds, 0.108 eV/atom. By comparing cases in which the formation energy is calculated without any chemical-potential fitting with cases in which the chemical potentials have been fit to experiment, we determine that the majority of the difference between the error in the OQMD and Materials Project is attributed to differences in the chemical-potential-fitting approach.

Historical Trends in Material/Compound Discovery. The OQMD allows us to explore trends, both historical and stoichiometric, in compound discovery and stability. From a historical perspective, the number of reported structures has been increasing linearly with time, while the number of *stable* structures reported annually has remained roughly constant since the 1960s. In addition, the scope of compound discovery has progressed from binary to ternary to quaternary compounds over time. This trend has now been disrupted by recent advances in structure prediction, structure determination and high-throughput structure calculation. In particular, in this study we predict the existence of ~3,200 new compounds. Using our predictions of the compositions at which new phases should be found, experimentalists can now more efficiently discover and characterise new materials.

In this work we demonstrate a few examples of how a large database of DFT calculations can be used to extract information beyond what can be gleaned from many distributed collections containing similar data. We believe that there is much more that can be understood by looking at large-scale material property databases, and in order to facilitate such discovery, we make the entire database available for download, without restriction.

MATERIALS AND METHODS

All DFT calculations are performed with the VASP^{73,74} (v5.3.2). The electron exchange and correlation are described with the GGA of PBE,⁷⁵ using the

potentials supplied by VASP with the PAW method.⁷⁶ PAW-PBE potentials for 89 elements are supplied with VASP, and those employed in the OQMD are listed in Table 3. We follow the VASP guidelines concerning the optimum choice of potentials.⁷⁷ Potentials where electrons have been moved from the core and treated as valence are appended with *_sv*, *_pv* and *_d* in Table 3 for *s*-, *p*- and *d*-electrons, respectively. For the 4*f*-elements, we employ potentials where the valence *f*-electrons are treated as core electrons (appended with *_2* and *_3*). For all calculations, Γ -centred *k*-point meshes are constructed to have the same relative ratios of mesh points to reciprocal lattice vector length, and with the number of *k*-points such that the *k*-points per reciprocal atom (calculated from $N_{k\text{-points}} \times N_{\text{atoms}}$) is as close to a target value as possible. The electronic self-consistency (for a given set of ion positions) is converged to within 10^{-4} eV/atom.

Any calculation containing *d*-block or actinide elements are spin-polarised with a ferromagnetic alignment of spins to capture possible magnetism, with initial magnetic moments of 5 and 7 μ_B for the *d*-block and actinide elements, respectively. It should be noted that this approach will not capture more complex magnetic ordering, such as antiferromagnetism. For many *d*-block oxides, a typical difference in total energy between ferromagnetic and antiferromagnetic states is on the order of 10–20 meV/atom.⁴⁵ However, given that in this work the range of compounds being calculated is extremely broad, it is likely that in some cases this error will be larger.

For several *d*- and *f*-block elements listed in Table 1, the GGA+U approach is implemented to improve the exchange and correlation description of the localised charge density when these elements are in compounds with oxygen. We employed the Dudarev approach to GGA+U,⁷⁸ where the only input parameter is *U*–*J*. For several transition metals, previously determined *U*–*J* values were used.⁴⁹ For actinide elements in oxides, we apply a *U*–*J* parameter of 4 eV. This value was chosen because no reliable values had been reported in the literature when these calculations were begun, and we found that the formation energies and band gaps of compounds containing these elements are relatively insensitive to the exact value of *U*–*J*; therefore, we elected to use a moderate *U*–*J* value of 4 eV for these elements. Recent lines of work have identified *U*–*J* values for a few of these elements,^{51,79} with *U*–*J* values close to those used herein. All *U*–*J* values are given in Table 1.

All calculations were completed in a two-step scheme. First, the structures were fully relaxed, followed by a static calculation. In relaxing an ICSD structure, we begin with the given ICSD structure parameters and perform several relaxation runs sequentially, until the volume change within the last relaxation run is less than 10%. The relaxation calculations are performed at a plane-wave basis-set energy cutoff at the energy recommended in the VASP potentials of the elements in the structure, and 6,000 *k*-points per reciprocal atom. The quasi-Newton scheme is used to optimise the structure to within 10^{-3} eV/atom. In these relaxation steps, Gaussian smearing is applied with a width of 0.2 eV. The final static calculation of the structure is performed at an energy cutoff of 520 eV using tetrahedral *k*-point integration.⁷⁶ The 520-eV cutoff is chosen because it is 25% higher than the highest recommended energy cutoff over all of the potentials used (including Li_{sv}, for which we use the version with lower recommended cutoff). This constant cutoff for all calculations ensures that all the energies calculated in OQMD are compatible, and can be used to evaluate the formation energies of compounds and *T*=0 ground-state phase diagrams.

ACKNOWLEDGEMENTS

We 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 (SK); the Revolutionary Materials for Solid State Energy Conversion, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC00010543 (JWD); the Ford-Boeing-Northwestern Alliance (JES); the Department of Defense through the National Defense Science and Engineering Graduate Fellowship Program with further support by DOE under Grant No. DE-FG02-07ER46433 (BM and AT); The Dow Chemical Company (MA) and the National Science Foundation under grant DRL-1348800 (CW). We acknowledge the openness of the Materials Project for making it possible to obtain a large set of energies with which to compare with the OQMD and experiment. We acknowledge a large allocation on Northwestern University's Quest high-performance computing system. Calculations were also performed 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.

CONTRIBUTIONS

SK developed DFT automation and thermodynamic analysis tools, and had primary writing responsibilities. JWD carried out comparison with Miedema model. JES carried out comparison with Materials Project, evaluated SSUB formation energies at STP. BM and AT selected DFT+U parameterisations and VASP potentials. MA analysed the elemental ground states. BM and SR organised the partnership with the ICSD. MA, JWD, JES and CW extensively edited the manuscript. CW conceptualised, guided all aspects and led the project.

COMPETING INTERESTS

The authors declare no conflict of interest.

REFERENCES

1. National Science and Technology Council. Materials Genome Initiative for Global Competitiveness Tech. Rep. <http://www.whitehouse.gov/blog/2011/06/24/materials-genome-initiative-renaissance-american-manufacturing> (2011).
2. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, 864–871 (1964).
3. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, 1133–1138 (1965).
4. Perdew, J. P. & Zunger, A. Self-Interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048–5079 (1981).
5. Ihm, J., Zunger, A. & Cohen, M. L. Momentum-space formalism for the total energy of solids. *J. Phys. C Solid State Phys.* **12**, 4409–4422 (1979).
6. Ceperley, D. M. & Alder, B. J. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.* **45**, 566–569 (1980).
7. Ihm, J., Yin, M. T. & Cohen, M. L. Quantum mechanical force calculations in solids: the phonon spectrum of Si. *Solid State Commun.* **37**, 491–494 (1981).
8. Levy, O., Hart, G. L. W. & Curtarolo, S. Uncovering compounds by synergy of cluster expansion and high-throughput methods. *J. Am. Chem. Soc.* **132**, 4830–4833 (2010).
9. Curtarolo, S. et al. The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191–201 (2013).
10. Hautier, G., Fischer, C., Ehrlacher, V., Jain, A. & Ceder, G. Data mined ionic substitutions for the discovery of new compounds. *Inorg. Chem.* **50**, 656–663 (2010).
11. Jain, A. et al. A high-throughput infrastructure for density functional theory calculations. *Comput. Mater. Sci.* **50**, 2295–2310 (2011).
12. Armiento, R., Kozinsky, B., Fornari, M. & Ceder, G. Screening for high-performance piezoelectrics using high-throughput density functional theory. *Phys. Rev. B* **84**, 014103 (2011).
13. Setyawan, W., Gaume, R. M., Lam, S., Feigelson, R. S. & Curtarolo, S. High-throughput combinatorial database of electronic band structures for inorganic scintillator materials. *ACS Comb. Sci.* **13**, 382–390 (2011).
14. Saal, J., Kirklın, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: the Open Quantum Materials Database (OQMD). *JOM* **65**, 1501–1509 (2013).
15. Curtarolo, S. et al. AFLOW: an automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226 (2012).
16. Landis, D. D. et al. The computational materials repository. *Comput. Sci. Eng.* **14**, 51–57 (2012).
17. Bergerhoff, G. & Brown, I. D. *Crystallographic Databases. chap. Inorganic* 147–156 (International Union of Crystallography: Chester, 1987).
18. Belsky, A., Hellenbrandt, M., Karen, V. L. & Luksch, P. New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Crystallogr. B* **58**, 364–369 (2002).
19. Saal, J. E. & Wolverton, C. Thermodynamic stability of Mg-Y-Zn long-period stacking ordered structures. *Scr. Mater.* **67**, 798–801 (2012).
20. Kirklın, S., Meredig, B. & Wolverton, C. High-throughput computational screening of new Li-ion battery anode materials. *Adv. Energy Mater.* **3**, 252–262 (2013).
21. Meredig, B. et al. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **89**, 094104 (2014).
22. Aykol, M., Kirklın, S. & Wolverton, C. Thermodynamic aspects of cathode coatings for lithium-ion batteries. *Adv. Energy Mater.* **4**, 1400690 (2014).
23. Kirklın, S., Chan, M., Trahey, L., Thackeray, M. M. & Wolverton, C. M. High-throughput screening of high-capacity electrodes for hybrid Li-ion/Li-O₂ cells. *Phys. Chem. Chem. Phys.* **16**, 22073–22082 (2014).
24. Jain, A. et al. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
25. Curtarolo, S. et al. AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012).
26. Ong, S. P. et al. Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
27. Bahn, S. R. & Jacobsen, K. W. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.* **4**, 56–66 (2002).
28. Meredig, B. & Wolverton, C. A hybrid computational-experimental approach for automated crystal structure solution. *Nat. Mater.* **12**, 123–127 (2013).
29. Young, D. A. *Phase Diagrams of the Elements* (University of California Press: Berkeley and Los Angeles, 1991).
30. King, H. W. in *CRC Handbook of Chemistry and Physics* 95th edn (ed. Haynes W. M.) Ch. 12, 15–18 (CRC Press, Taylor & Francis, 2012).
31. Wang, Y. et al. Ab initio lattice stability in comparison with CALPHAD lattice stability. *Calphad* **28**, 79–90 (2004).
32. Klimeš, J. & Michaelides, A. Perspective: advances and challenges in treating van der Waals dispersion forces in density functional theory. *J. Chem. Phys.* **137**, 120901 (2012).
33. Tao, X. et al. Phase stability of magnesium-rare earth binary systems from first-principles calculations. *J. Alloys Compd.* **509**, 6899–6907 (2011).
34. Gao, M. C., Rollett, A. D. & Widom, M. First-principles calculation of lattice stability of C15M2R and their hypothetical C15 variants (M = Al, Co, Ni; R = Ca, Ce, Nd, Y). *Calphad* **30**, 341–348 (2006).
35. Mao, Z., Seidman, D. N. & Wolverton, C. First-principles phase stability, magnetic properties and solubility in aluminum-rare-earth (AlRE) alloys and compounds. *Acta Mater.* **59**, 3659–3666 (2011).
36. Temmerman, W. et al. *Handbook on the Physics and Chemistry of Rare Earths* vol. 39 (Elsevier, 2009); <http://www.sciencedirect.com/science/article/pii/S0168127308000019>.
37. Zhou, F. & Ozoliņš, V. Obtaining correct orbital ground states in f-electron systems using a nonspherical self-interaction-corrected LDA+U method. *Phys. Rev. B* **80**, 125127 (2009).
38. Duthie, J. C. & Pettifor, D. G. Correlation between d-band occupancy and crystal structure in the rare earths. *Phys. Rev. Lett.* **38**, 564–567 (1977).
39. Biering, S. & Schwerdtfeger, P. High-pressure transitions in bulk mercury: a density functional study. *Theor. Chem. Acc.* **130**, 455–462 (2011).
40. Mishra, V., Gyanchandani, J., Chaturvedi, S. & Sikka, S. Effect of spin-orbit coupling on the ground state structure of mercury. *Solid State Commun.* **186**, 38–41 (2014).
41. Wolverton, C. Crystal structure and stability of complex precipitate phases in Al-Cu-Mg-(Si) and Al-Zn-Mg alloys. *Acta Mater.* **49**, 3129–3142 (2001).
42. Wolverton, C., Ozoliņš, V. & Asta, M. Hydrogen in aluminum: First-principles calculations of structure and thermodynamics. *Phys. Rev. B* **69**, 144109 (2004).
43. Lany, S. Semiconductor thermochemistry in density functional calculations. *Phys. Rev. B* **78**, 245207 (2008).
44. Hautier, G., Ong, S. P., Jain, A., Moore, C. J. & Ceder, G. Accuracy of density functional theory in predicting formation energies of ternary oxides from binary oxides and its implication on phase stability. *Phys. Rev. B* **85**, 155208 (2012).
45. Stevanović, V., Lany, S., Zhang, X. & Zunger, A. Correcting density functional theory for accurate predictions of compound enthalpies of formation: fitted elemental-phase reference energies. *Phys. Rev. B* **85**, 115104 (2012).
46. SGTE. *Thermodynamic Properties of Inorganic Materials* Vol. 19. (Springer-Verlag: Berlin, Heidelberg, 1999).
47. Nash, P. Thermodynamic database <https://tptc.iit.edu/index.php/thermo-database> (2013).
48. Konings, R. J. M., Morss, L. R., Fuger, J. in *The Chemistry of the Actinide and Transactinide Elements* 3rd edn (eds Morss L. R. et al.) Ch. 19, 2113–2224 (Springer: Dordrecht, 2006); <http://www.springerlink.com/index/117213068108mk22.pdf>.
49. Wang, L., Maxisch, T. & Ceder, G. Oxidation energies of transition metal oxides within the GGA+U framework. *Phys. Rev. B* **73**, 195107 (2006).
50. Grindy, S., Meredig, B., Kirklın, S., Saal, J. E. & Wolverton, C. Approaching chemical accuracy with density functional calculations: diatomic energy corrections. *Phys. Rev. B* **87**, 075150 (2013).
51. Aykol, M. & Wolverton, C. Local environment dependent GGA+U method for accurate thermochemistry of transition metal compounds. *Phys. Rev. B* **90**, 115105 (2014).
52. Aykol, M., Kim, S. & Wolverton, C. van der Waals interactions in layered lithium cobalt oxides. *J. Phys. Chem. C* **119**, 19053–19058 (2015).
53. Pozdnyakova, I., Navrotsky, A., Shilkina, L. & Reznitschenko, L. Thermodynamic and structural properties of sodium lithium niobate solid solutions. *J. Am. Ceram. Soc.* **85**, 379–384 (2004).
54. CRC. *Handbook of Chemistry and Physics*, 93rd edn 2012–2013 <http://www.hbcpnetbase.com/> (2012).
55. Gordienko, S. P. Enthalpies of formation for boron silicides. *Powder Metall. Met. Ceram.* **34**, 660–662 (1995).
56. Kubaschewski, O., Alcock, C. B. & Spencer, P. J. *Materials Thermochemistry* 6th edn (Pergamon Press, 1993).

57. Olin, Å., Nolång, B., Öhman, L.-O., Osadchii, E. & Rosén, E. *Chemical Thermodynamics* (Elsevier Science, 2005).
58. Ranade, M. R., Tessier, F., Navrotsky, A. & Marchand, R. Calorimetric determination of the enthalpy of formation of InN and comparison with AlN and GaN. *J. Mater. Res.* **16**, 2824–2831 (2001).
59. van Setten, M. & Fichtner, M. On the enthalpy of formation of aluminum diboride, AlB₂. *J. Alloys Compd.* **477**, L11–L12 (2009).
60. Sommer, F., Borzone, G., Parodi, N. & Ferro, R. Enthalpy of formation of CaPb and BaPb alloys. *Intermetallics* **14**, 287–296 (2006).
61. Lemire, R. J. *Chemical Thermodynamics of Neptunium and Plutonium* Vol. C (Elsevier Science, 2001).
62. Gramsch, S. A. & Morss, L. R. Standard molar enthalpies of formation of PrO₂ and SrPrO₃: the unusual thermodynamic stability of APrO₃ (A = Sr, Ba). *J. Chem. Thermodyn.* **27**, 551–560 (1995).
63. Notin, M., Mejbar, J., Bouhajib, A., Charles, J. & Hertz, J. The thermodynamic properties of calcium intermetallic compounds. *J. Alloys Compd.* **220**, 62–75 (1995).
64. Huntelaar, M. E., Cordfunke, E. H. P. & Ouweltjes, W. The standard molar enthalpies of formation of BaSiO₃ (s) and Ba₂SiO₄ (s). *J. Chem. Thermodyn.* **24**, 1099–1102 (1992).
65. Tessier, F. *et al.* Thermodynamics of formation of binary and ternary nitrides in the system Ce/Mn/N. *Z. Anorg. Allg. Chem.* **627**, 194–200 (2001).
66. de Boer, F. R., Boom, R., Mattens, W. C. M., Miedema, A. R. & Niessen, A. K. *Cohesion in Metals: Transition Metal Alloys* (North-Holland, Amsterdam 1988).
67. Das, N. *et al.* Miedema model based methodology to predict amorphous-forming-composition range in binary and ternary systems. *J. Alloys Compd.* **550**, 483–495 (2013).
68. Ray, P. K., Akinc, M. & Kramer, M. J. Applications of an extended Miedema's model for ternary alloys. *J. Alloys Compd.* **489**, 357–361 (2010).
69. Akbarzadeh, A. R., Ozoliņš, V. & Wolverton, C. First-principles determination of multicomponent hydride phase diagrams: application to the Li-Mg-N-H system. *Adv. Mater.* **19**, 3233–3239 (2007).
70. Barber, C., Dobkin, D. & Huhdanpaa, H. The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.* **22**, 469–483 (1996).
71. Oganov, A. R. & Glass, C. W. Crystal structure prediction using *ab initio* evolutionary algorithms: principles and applications. *J. Chem. Phys.* **124**, 244704 (2006).
72. Amsler, M. & Goedecker, S. Crystal structure prediction using the minima hopping method. *J. Chem. Phys.* **133**, 224104 (2010).
73. Kresse, G. & Furthmüller, J. Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).
74. Kresse, G. & Furthmüller, J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
75. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
76. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
77. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
78. Dudarev, S. L., Botton, G. A., Savrasov, S. Y., Humphreys, C. J. & Sutton, A. P. Electron-energy-loss spectra and the structural stability of nickel oxide. *Phys. Rev. B* **57**, 1505–1509 (1998).
79. Dorado, B. & García, P. First-principles DFT+U modeling of actinide-based alloys: application to paramagnetic phases of UO₂ and (U,Pu) mixed oxides. *Phys. Rev. B* **87**, 195139 (2013).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>