

Artificial Intelligence Driving Materials Discovery? Perspective on the Article: Scaling Deep Learning for Materials Discovery

Anthony K. Cheetham* and Ram Seshadri*



Cite This: *Chem. Mater.* 2024, 36, 3490–3495



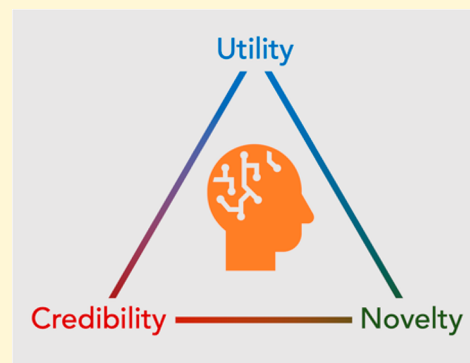
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: The discovery of new crystalline inorganic compounds—novel compositions of matter within known structure types, or even compounds with completely new crystal structures—constitutes an important goal of solid-state and materials chemistry. Some fractions of new compounds can eventually lead to new structural and functional materials that enhance the efficiency of existing technologies or even enable completely new technologies. Materials researchers eagerly welcome new approaches to the discovery of new compounds, especially those that offer the promise of accelerated success. The recent report from a group of scientists at Google who employ a combination of existing data sets, high-throughput density functional theory calculations of structural stability, and the tools of artificial intelligence and machine learning (AI/ML) to propose new compounds is an exciting advance. We examine the claims of this work here, unfortunately finding scant evidence for compounds that fulfill the trifecta of novelty, credibility, and utility. While the methods adopted in this work appear to hold promise, there is clearly a great need to incorporate domain expertise in materials synthesis and crystallography.



INTRODUCTION

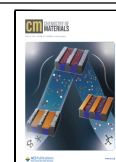
In an article in *Nature* published in November 2023, Merchant et al.¹ describe the application of artificial intelligence and machine learning (AI/ML) techniques such as deep learning of experimental databases and computational data to the discovery of new inorganic materials, including classical inorganic compounds such as oxides and halides, as well as other main group compounds and intermetallics. Their approach claims to “enable the discovery of 2.2 million structures below the current convex hull, many of which escaped previous human chemical intuition...representing an order-of-magnitude expansion in stable materials known to humanity”. Almost 400,000 of the structures are deemed to be stable and have been listed in a Stable Structure database,² while further structural details of more than 2000 of these new compounds have been placed in the GNoME Explorer archive³ within the Materials Project.⁴ The bold claims in the *Nature* paper warrant scrutiny by the materials chemistry community, which has hitherto undertaken the discovery of new materials through the (perhaps) more pedestrian approach of synthesizing new compounds planned around how they relate to what is already known, with experience and empirical knowledge (which is sometimes arguably mislabeled as chemical intuition) serving as guides. In this perspective, we scrutinize the claims by Merchant et al. by looking in detail at a small, randomized subset of the new materials in both the GNoME Explorer archive and the Stable Structure database to evaluate their novelty. We do this from the perspective of experimental materials chemists

with decades of experience in discovery research who welcome new approaches to materials discovery.

We have recently contributed a perspective on the relationship between chemical synthesis and materials discovery wherein we have analyzed the different laboratory approaches that have led to major breakthroughs in the materials area, such as the discovery of high temperature superconductivity and lithium ion batteries.⁵ One of the lessons from that study was that most breakthroughs are not achieved by design or serendipity, but by exploring opportunities in the extensive repository of compounds that are already known. In this respect, we applaud the news that this repository has apparently become ten times larger since this should clearly increase the probability of making important materials breakthroughs in the future. The following sections summarize our views on this potentially important development.

Before progressing further, we point out that the predictions in the contribution of Merchant et al.¹ are solely of crystalline inorganic compounds and should be described as such, rather than using the more generic label “material”. There are many

Received: March 4, 2024
 Revised: March 22, 2024
 Accepted: March 25, 2024
 Published: April 8, 2024



communities who would justifiably be upset by the description “representing an order-of-magnitude expansion in stable materials”: polymers, glasses, metal–organic frameworks, heterostructures, and composites are only a few excluded materials classes that come to mind, and that are each infinite in their scope. Additionally, it is the usual practice in the field that chemical compounds become materials when they demonstrate some utility. We propose that impactful predictions of new materials should lie somewhere within the triangle of being credible, implying that the proposed structure and composition of matter should be experimentally realizable, novel in the sense of not being more than a trivial extension of known compounds, and display some evidence of utility so that they can truly be recognized as materials.

■ GENERAL COMMENTS ABOUT THE GNOME EXPLORER DATABASE

In this first section, we have looked through part of the GNoME Explorer database for any obvious signs of entries that may lack novelty. This has not been done comprehensively since this would be too time-consuming, but we examined the first 250 entries to get a sense of any underlying issues. Each of the 2047 entries in the GNoME Explorer compilation has been presented with chemical composition, space group, atomic coordinates, formation energy, and, where appropriate, tentative assignments of oxidation states. In the spirit of constructive criticism, we make the following observations concerning the compounds in this subset of the GNoME database.

- (i) We believe that experimentalists, who presumably are an important target audience, would find it helpful if the results were presented in a more organized manner, rather than as a seemingly random walk through the periodic table. For example, it would be useful to list all the oxides together, as well as the fluorides, chlorides, bromides, etc. This could no doubt be done by the user with the help of the search function in the database, but it would be helpful if it was done in the parent listing, given the enormous number of entries.
- (ii) The compositions are often not presented in a manner that an experimental materials chemist would find appropriate or helpful, nor are the usual rules of chemical nomenclature followed, notably, anions after cations, alkali and alkaline earth metals before transition metals, etc. For example, $\text{Ac}(\text{ErB}_8)_3$ (mp-3169969) could be more usefully listed as $\text{AcEr}_3\text{B}_{24}$ since there is no special relationship between the Er and the B. It is then easy to see that this is a hexaboride of Ac and Er: $\text{AcEr}_3(\text{B}_6)_4$. Such hexaborides are well-known and are found for a wide range of elements, such as the alkaline earths and rare earths. LaB_6 is a typical example and adopts the cubic space group, $Pm\bar{3}m$, with $a \approx 4.1 \text{ \AA}$.⁶ $\text{AcEr}_3(\text{B}_6)_4$ is predicted to be tetragonal with $a \approx 4.1 \text{ \AA}$ and $c \approx 16.5 \text{ \AA}$, revealing that a proposed ordering of Ac^{3+} and Er^{3+} (which may or may not be experimentally viable) leads to a $4\times$ superstructure along c and lowers the symmetry of the parent LaB_6 structure from cubic to tetragonal. There are other examples of apparently novel and complex borides that are based upon the LaB_6 structure, e.g., $\text{K}_2\text{SmLuB}_{24}$ (mp-3170680), but the unconventional presentation of their compositions makes it difficult to search for them automatically. Perhaps this could be addressed in the next iteration of the database.
- (iii) Oxidation states are often inconsistent with known materials, or structural features are obscure or entirely unprecedented in known materials. For example, TbSmF_{30} (mp-3170019) defies all the usual rules of valence and contains an improbable ordering of Tb^{3+} and Sm^{3+} (see below). Further inspection of the structure reveals that it contains large numbers of F_2 molecules, some isolated and others acting as linkers between LnF_9 polyhedra (where Ln is the lanthanide cation). Neither of these features is found in the literature on known fluoride compounds, and their existence at ambient pressure seems highly improbable through a structural chemist's lens. There are other examples in the database of compositions with a large excess of anions that potentially arise from the difficulty of dealing with the chemical potential of a gas-phase component (in this case F_2) in electronic structure calculations of formation energies and stability.⁷
- (iv) Several of the proposed new structures have obvious analogues, which somewhat undermines the impact of their prediction. $\text{K}_3\text{Nd}(\text{AsO}_4)_2$ (mp-3196061) is a good example. It is predicted to be monoclinic, in space group $P2_1$, with $a = 9.8 \text{ \AA}$, $b = 5.8 \text{ \AA}$, $c = 7.6 \text{ \AA}$, and $\beta = 91.87^\circ$. Since arsenates are often isostructural with phosphates, it is possible to recognize that the analogous phosphate, $\text{K}_3\text{Nd}(\text{PO}_4)_2$, is known with a monoclinic $P2_1/m$ structure ($a = 9.5 \text{ \AA}$, $b = 5.6 \text{ \AA}$, $c = 7.4 \text{ \AA}$, $\beta = 90.95^\circ$).⁸ These structures are virtually identical, aside from the lower symmetry in the GNoME entry, a recurring theme that we shall return to below.
- (v) We would now like to focus on a recurrent issue that is found throughout the GNoME database, which is that many of the entries are based upon the ordering of metal ions that are unlikely to be ordered in the real world. For example, it is well-known that ordering of lanthanide atoms and ions is highly improbable at finite temperatures. However, density functional theory (DFT)-based electronic structure calculations at 0 K will find a way to order them because they do not take entropy into account. Entropic effects, however, typically favor disordered arrangements at the synthesis temperatures used for most inorganic materials. For example, in TbSmSeO_2 (mp-3169970), the Tb^{3+} and Sm^{3+} cations are predicted to be on separate crystallographic sites, even though their charges are the same and their 6-coordinate ionic radii are very similar (0.92 and 0.96 \AA , respectively). Because of the ordering, the compound is proposed to be non-centrosymmetric in space group $P3m1$ with $a \approx 3.91 \text{ \AA}$ and $c \approx 6.92 \text{ \AA}$. Unfortunately, this improbable rare-earth ordering obscures the fact that the disordered structure of $(\text{Tb}/\text{Sm})\text{SeO}_2$ would be isomorphous with the known centrosymmetric $P\bar{3}m1$ structure of La_2SeO_2 , which has $a \approx 4.09 \text{ \AA}$ and $a \approx 7.16 \text{ \AA}$.⁹ As in the case of $\text{Ac}(\text{ErB}_8)_3$, discussed above, the artificial ordering of the cations lowers the symmetry from a centrosymmetric to a non-centrosymmetric structure, albeit within the same crystal system. We regard these as predictions that are highly unlikely to be confirmed because of the virtual certainty that the cations will be disordered.
- (vi) The above comments about cation ordering in inorganic compounds also apply to a very large number of the intermetallic phases, leading to many other examples of unlikely predictions. For example, DyHo_2Rh_9 (mp-

Table 1. Most Abundant Space Groups in the Stable Structure Database

Stable Structure Database (384,870 entries)					
Space group	Occurrence	Percentage	Space group	Occurrence	Percentage
<i>Pm</i>	49037	12.7	$R\bar{3}m$	8834	2.30
<i>P1</i>	39382	10.2	<i>P6</i>	8563	2.22
<i>Amm2</i>	26467	6.88	$P\bar{6}2m$	7652	1.99
<i>Cm</i>	19913	5.17	<i>Imm2</i>	7394	1.92
<i>C2/m</i>	12954	3.37	<i>P3m1</i>	6903	1.79
<i>R3m</i>	11241	2.92	$P\bar{3}m1$	6379	1.66
<i>C2</i>	11201	2.91	$P\bar{6}m2$	6038	1.57
$P\bar{1}$	10463	2.72	<i>Pnma</i>	6005	1.56

3170148) is predicted to adopt space group $P6_3/mmc$ with $a = 5.26 \text{ \AA}$ and $c = 17.58 \text{ \AA}$. This compound clearly has the PuNi_3 structure, which is found in a large number of phases such as CeNi_3 , $P6_3/mmc$ with $a = 4.98 \text{ \AA}$ and $c = 16.94 \text{ \AA}$.¹⁰ Ordered variants of this structure are also known, e.g., YRh_2Si ($P6_3/mmc$ with $a = 5.49 \text{ \AA}$ and $c = 15.03 \text{ \AA}$),¹¹ though it seems highly improbable that Dy and Ho would order in the proposed DyHo_2Rh_9 GNoME structure. $\text{TbHo}_2\text{Tm}_9\text{Co}_4$ (mp-3170547) is an even more striking example among the intermetallics. This proposed alloy between three smaller rare earths and cobalt is predicted to be monoclinic and non-centrosymmetric, space group *Pm*, with $a = 6.14 \text{ \AA}$, $b = 9.23 \text{ \AA}$, $c = 6.89 \text{ \AA}$, and $\beta = 90.1^\circ$. If we rewrite the formula as Ln_2Co_4 , then we can see that it represents an unlikely ordering of the small lanthanides in a compound of composition Ln_3Co . These phases are known for most of the rare earths and yttrium, and they adopt the Fe_3C structure. For example, Ho_3Co adopts a centrosymmetric orthorhombic structure, space group *Pnma*, with $a = 6.92 \text{ \AA}$, $b = 9.29 \text{ \AA}$, and $c = 6.21 \text{ \AA}$;¹² note that the *a* and *c* axes are inverted and that the lowering of symmetry in the AI prediction is again due to the artificial ordering of the small rare-earth metals.

There appear to be countless similar examples in the GNoME database. For example, $\text{Tb}_{16}\text{Ho}(\text{ErIr}_4)_3$ (mp-3170203) could be more usefully written as $\text{Tb}_{16}\text{HoEr}_3\text{Ir}_{12}$. It is predicted to be tetragonal, $P\bar{4}$ (non-centrosymmetric), with $a = 10.92 \text{ \AA}$ and $c = 6.38 \text{ \AA}$. If we assume that very similar rare earths would be disordered, as discussed above, it can then be written as $\text{Ln}_{20}\text{Ir}_{12}$ or Ln_5Ir_3 . This is a known structure type and has been reported for virtually all rare earths with the tetragonal Pu_5Rh_3 structure. For the specific case of Tb_5Ir_3 ,¹³ the space group is centrosymmetric $P4/ncc$ with $a = 10.905 \text{ \AA}$ and $c = 6.299 \text{ \AA}$. Note again that the prediction places the different lanthanides on distinct sites and therefore lowers the symmetry. We will discuss the reasons for the frequent lowering of symmetry that is found in many of the Stable Structure database entries in a later section.

- (vii) Another common claim among the database entries is the prediction of new compounds based upon radioactive elements that do not occur in usable quantities in nature. In the case of the proposed actinium (Ac^{3+}) compounds, of which there are 27 (and approximately 6754 in the larger Stable Structure database), it should be noted that all isotopes of actinium are intensely radioactive with half-lives ranging from days to a few years.¹⁴ The natural abundance of Ac in, for example, uranium ores is so low that a ton of uranium yields less than a milligram of ^{227}Ac (half-life 21.77 years). Milligram quantities of ^{227}Ac can

be made artificially by neutron irradiation of ^{226}Ra , and this isotope has been studied for cancer treatments. However, ionic actinium compounds, which always contain Ac^{3+} , have no industrial applications. Furthermore, the few that are known are isomorphous with their stable lanthanum analogues (e.g., AcPO_4 and LaPO_4). The “discoveries” of new actinium compounds, such as AcPO_3 (mp-3170055), are therefore impractical and not very novel. In fact, there are 18,138 compounds of such radioactive elements in the large Stable Structure database, including those of Pm, Ac, and Pa. We question whether these can be regarded as potential new materials. There are a further 23,529 entries for compounds containing the highly radioactive elements Tc, Np, and Pu.

SPACE GROUP ANALYSIS

Considering the frequent observations of unlikely low symmetry structures among the predictions, we have examined the space group statistics of the 384,870 entries in the Stable Structure database. We have found that the space groups of the predicted compounds (Table 1) have a distribution that is strikingly different from those in the Inorganic Crystal Structure Database (ICSD),¹⁵ which has been analyzed by Urusov and Nadezhina.¹⁶ For example, the top two space groups in the ICSD are centrosymmetric *Pnma* and $P2_1/c$, accounting for approximately 16% of all structures, with $\sim 8\%$ in each case. By contrast, *Pnma* is ranked 16th in Stable Structure database with 1.56% of the structures, while $P2_1/c$ is found in only 0.7% of the structures, which is an order of magnitude lower than in the ICSD. In fact, the top four space groups in the Stable Structure database are all non-centrosymmetric and account for $\sim 34\%$ of all the structures. By contrast, in the ICSD there is only one non-centrosymmetric space group in the top 24 and it accounts for only 1% of all structures.

The striking disparity between the space group distributions for the predicted phases and known ICSD structures is a matter of serious concern. We also note that this issue has been highlighted¹⁷ in a commentary on a different *Nature* article on robotic/AI-based materials discovery.¹⁸ The main reason for the disparity is due to the frequent prediction of structures with atoms ordered on distinct crystallographic sites that—in most cases—are likely to be disordered. In many cases, e.g., $\text{Dy}_6\text{Y}_2\text{Ho}_{11}\text{Lu}(\text{Cd}_3\text{Ru})_2$ (mp-3195118), they are suggestive of high-entropy alloys,¹⁹ where the whole point is to avoid atomic ordering or phase separation. Predictions of ordering usually lead to lower symmetry structures and in some cases to superstructures, as illustrated above. These general trends that are not found in the laboratory could arise because the DFT

Table 2. Comparison of 10 Randomly Selected Compounds from the Stable Structure Database with Appropriate ICSD Entries⁴⁴

	Entry #	Composition	Space Group	Volume	<i>a, b, c</i> (Å)	α, β, γ (°)
Database						
ICSD analog						
1	186808	B ₃ Dy ₄ RuTc ₃	<i>Pm</i>	202.3	5.944, 5.293, 6.430	90.0, 90.1, 90.0
	100774	B ₂ DyRu	<i>Pnma</i>	198.2	5.886, 5.300, 6.352	90, 90, 90
2	308005	Pb ₂ Pd ₃ SnY ₃	<i>P1</i>	205.7	7.794, 7.837, 3.881	90.0, 90.0, 119.8
	230854	PbPtGd	<i>P-62m</i>	200.7	7.637, 7.637, 3.965	90, 90, 120
3	354609	Hf ₄ Ir ₈ N ₄ NbZr ₁₁	<i>Cm</i>	478.3	8.799, 8.799, 8.800	59.7, 59.6, 59.6
	640826	Zr ₄ Ir ₂ N	<i>Fd-3m</i>	1884	12.35, 12.35, 12.35	90, 90, 90
4	162234	B ₂₄ Er ₄ Os ₃ Tb ₄ Tc	<i>P1</i>	385.4	3.656, 9.137, 11.53	89.9, 90.0, 90.0
	603495	B ₈ Er ₂ Os	<i>Pban</i>	377.2	9.097, 11.46, 3.617	90, 90, 90
5	352440	AlTa ₁₆ Te ₁₆ V ₃	<i>P1</i>	1060	4.907, 11.11, 19.44	90.01, 90.0, 90.0
	79796	FeTa ₄ Te ₄	<i>Pbam</i>	925.2	10.51, 18.27, 4.185	90, 90, 90
6	304898	Ga ₄ La ₂ Pt ₃ Rh	<i>P-4n2</i>	194.5	4.321, 4.321, 10.52	90, 90, 90
	621150	Ce ₂ Pt _{1.5} Ga _{6.5}	<i>I4/mmm</i>	196.1	4.315, 4.315, 10.53	90, 90, 90
7	369281	Cu ₈ P ₄ SrTm	<i>Cmmm</i>	204.2	3.815, 7.317, 7.317	88.96, 90, 90
	62553	CaCu ₄ P ₂	<i>P4₂/mnm</i>	203.6	7.290, 7.290, 3.831	90, 90, 90
8	252376	DyPSi ₅ Tb ₉	<i>P1</i>	393.7	8.460, 8.466, 6.344	90.0, 90.0, 120.0
	99641	Gd ₅ Si ₃	<i>P6₃/mcm</i>	403.0	8.513, 8.513, 6.421	90, 90, 120
9	13745	Ac ₄ P ₈ S ₂₈ Tl ₈	<i>C2/c</i>	1339	6.822, 12.29, 17.12	79.5, 78.4, 73.9
	25121	La ₂ P ₄ S ₁₄ Rb ₄	<i>P2₁/c</i>	1276	9.950, 6.846, 19.82	90, 98.2, 90
10	326805	Ga ₂ Nb ₂₀ Os ₈	<i>P4₂/mnm</i>	508.7	9.933, 9.933, 5.156	90, 90, 90
	230694	Ge _{3.9} Nb _{20.4} Rh _{5.9}	<i>P4₂/mnm</i>	497.6	9.848, 9.848, 5.133	90, 90, 90

⁴⁴The high incidence of pseudosymmetry in the Stable Structure database entries is notable.

modeling is performed at effectively 0 K, and on small unit cells, ignoring the effects of configurational entropy, resulting in the predicted structures being fully ordered. In addition, many compounds undergo phase transitions to lower symmetry structures on cooling to low temperatures, although we have no evidence of this from the GNoME entries that we have examined.

■ RANDOMLY SELECTED EXAMPLES FROM THE STABLE STRUCTURE DATABASE

Because our scrutiny of the GNoME database was selective to identify some obvious shortcomings of the methodology, we have also carried out a random examination of some of the entries in the Stable Structure database, selecting 10 compounds from among the 384,870 database entries. The results from this analysis are summarized in Table 2.

The key conclusions from this analysis are as follows:

- We were able to identify the structure of every one of the 10 Stable Structure entries in the ICSD database, albeit usually with a space group of higher symmetry than that in the AI database. This is expected from the discussion above, with the lowered symmetry arising from artificial atomic orderings, and indeed, several of the Stable Structure entries involve unit cell parameters that are incompatible with the stated space groups, i.e., display pseudosymmetry.
- In only one case was the space group the same. This enabled us to find an obvious analogue of Ga₂Nb₂₀Os₈ in the ICSD by searching for structures in space group *P4₂/mnm* with lattice parameters close to *a* = 9.9 Å and *c* = 5.1 Å, pointing us to Ge_{3.9}Nb_{20.4}Rh_{5.9}, which is one of the well-known Frank–Kasper sigma (σ) phases.^{20,21} This is a common and versatile intermetallic structure type that is well-known to accommodate disorder on the five independent metal sites in the unit cell. It is indeed

impressive that the predictive approaches can identify new compositions in this structure space.

- In some cases, the analogy between the Stable Structure database and the ICSD entries is obvious. For example, Cu₈P₄SrTm was identified as having the same structure as CaCu₄P₂,²² noting that the lattice parameters were switched between the two entries and that the predicted compound was pseudo-tetragonal. It is questionable whether Sr and Tm would order on distinct crystallographic sites. If Tm were to substitute as the expected Tm³⁺, this would require electron-doping a Cu¹⁺ compound, which is unlikely.
- In other cases, the identification of the structure type takes a little more knowledge of the periodic table and crystal symmetry. For example, in Hf₄Ir₈N₄NbZr₁₁ our starting point was to recognize that Hf and Zr are two of the most similar elements in the periodic table and are almost certain to be disordered on the same sites (pointing us toward Zr₁₁Ir₈N₄Nb). Furthermore, in a metal-rich compound, Nb and Zr/Hf are also likely to alloy on the same crystallographic site. A search in the ICSD for compounds containing Zr, Ir, and N then quickly led us to Zr₄Ir₂N,²³ whose cubic unit cell is 4× larger than the cell of Hf₄Ir₈N₄NbZr₁₁ in the database. The final step is to recognize that the pseudo-rhombohedral cell from the Stable Structure database with α = 59.7° is also nearly pseudocubic, thus accounting for the 4× discrepancy in the cell volume. The two structures are compared in Figure 1.
- In the case of Ac₄P₈S₂₈Tl₈, the similarity to the structure of La₂P₄S₁₄Rb₄²⁴ is not immediately apparent from the unit cells. However, we would expect Ac compounds would be isomorphous with their La analogues. Tl¹⁺ and Rb¹⁺ are also close in size and identical in their charge states. The similarity of the structures then becomes evident.

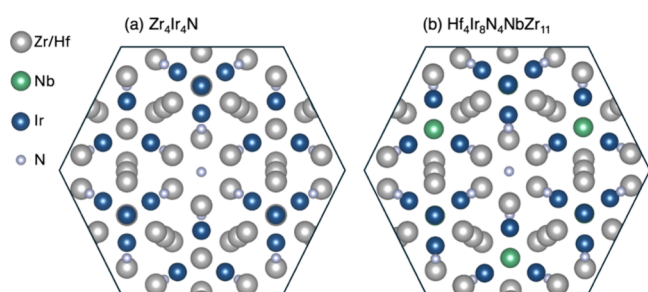


Figure 1. Views of the crystal structures of (a) known Zr_4Ir_4N (ICSD 640826) compared in a similar projection and identical scaling with (b) the proposed structure of $Hf_4Ir_8N_4NbZr_{11}$. The novelty of the structure and composition in (b) would arise only if Zr, Hf, and Nb were ordered on distinct crystallographic sites, which is unlikely.

CONCLUSIONS

Our analysis of the predictions by Merchant et al.¹ has raised several important issues that need to be addressed for AI to have a significant impact on the discovery of new materials. These include the elimination of a large number of radioactive materials that are unlikely to have any utility in the materials world. This point particularly concerns the inclusion of compounds of Pm, Ac, and Pa (more than 18,000 in all), which are only available in minute quantities and in the rarest of circumstances. The challenges of dealing with disorder and behavior at finite temperatures are far more daunting. The computational tools to deal with these issues do exist, but they are computationally intensive and are not scalable in how Merchant et al.¹ have approached their work. Much could be achieved, however, by embedding a knowledge of solid-state chemistry into their methodology. For example, the recognition that the 14 rare-earth elements and yttrium have very similar chemistries could be incorporated as well as the recognition that zirconium and hafnium, among other important metal pairs, have virtually identical chemistries.

There is also much room for improvement in the crystallographic aspects of the work. For example, ICSD has excellent search options that can utilize combinations of elemental composition, lattice parameters, and space groups. We employed these to identify the structures that are listed in Table 2. There are several crystallographic tools available to regularize structural information and enable the comparison of seemingly distinct structure types.²⁵ Some of them would be useful in resolving issues of pseudosymmetry that is present in many of the entries.²⁶ Efforts to relate structure types using their structures and geometries are also ongoing, and clearly need large-scale implementation.²⁷

While the above analysis may seem to be critical, we do believe that many of our points could be adopted in the next version of this work. More scrutiny of the “new” materials needs to be performed prior to putting them into a database and claiming “...an order-of-magnitude expansion in stable materials known to humanity”. In fact, we have yet to find any strikingly novel compounds in the GNoME and Stable Structure listings, although we anticipate that there must be some among the 384,870 compositions. We also note that, while many of the new compositions are trivial adaptations of known materials, the computational approach delivers credible overall compositions, which gives us confidence that the underlying approach is sound. For example, in addition to the hexaborides, there are many examples in the GNoME list that are clearly diborides, which is another important stoichiometry in the boride area. What is now

needed is greater effort to connect the predictions to what is already known in the literature to filter out the many candidates that are not truly novel. It is impractical to do this manually with a list of almost 400,000 new compositions.

It also must be recognized that a large fraction of the 384,870 compositions adopt structures that are already known and can be found in the ICSD database. This is not surprising to an experimental materials chemist, because it is quite rare to find an entirely new structure type in the inorganic world. This can be thought of as being a manifestation of Pauling’s Fifth Rule, the Rule of Parsimony.²⁸ According to this rule, the number of geometrical units that Nature uses to assemble crystals is quite limited and relies heavily on a small number of recurrent polyhedral motifs such as tetrahedra or octahedra. While we are sure that there must be some new structure types predicted, our analysis suggests that there will not be many of them.

This brings us to our final point concerning the claim of “an order-of-magnitude expansion in stable materials known to humanity”. We would respectfully suggest that the work by Merchant et al.¹ does not report any new materials but reports a list of proposed compounds. In our view, a compound can be called a material when it exhibits some functionality and, therefore, has potential utility. Since no functionality has been demonstrated for the 384,870 compositions in the Stable Structure database, they cannot yet be regarded as materials. The few examples of functionality mentioned in the article are associated with Li^+ -ion conductors. While the proposed materials are encouraging, their compositions leave much to be desired since they incorporate chemically soft anions. These anions are usually associated with narrow electrochemical stability windows, which renders materials that incorporate them somewhat pointless as Li^+ solid electrolytes.²⁹ This points to an interesting contraindication in materials design: soft anions such as Te^{2-} and I^- readily permit cation transport but do not possess the requisite electrochemical (redox) stability for use as solid electrolytes. Conversely, hard anions such as O^{2-} and F^- are redox-stable, but they bind cations strongly, and the resulting materials usually display limited ionic conductivity.

In closing, we hope the comments presented here will usefully serve the large community of materials scientists and engineers in their continued quest to develop the next generation of useful materials. While we are confident that the tools of Artificial Intelligence and Machine Learning have a bright future in the field of materials discovery, more work needs to be done before that promise is fulfilled.

AUTHOR INFORMATION

Corresponding Authors

Anthony K. Cheetham – Materials Department and Materials Research Laboratory, University of California, Santa Barbara, California 93106, United States; Department of Materials Science and Engineering, National University of Singapore, Singapore 117575, Singapore; orcid.org/0000-0003-1518-4845; Email: akc30@cam.ac.uk

Ram Seshadri – Materials Department and Materials Research Laboratory, University of California, Santa Barbara, California 93106, United States; orcid.org/0000-0001-5858-4027; Email: seshadri@mrl.ucsb.edu

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.chemmater.4c00643>

Notes

The authors declare no competing financial interest.

Biographies

Anthony K. Cheetham is a Research Professor at UC Santa Barbara and a Distinguished Visiting Professor at NUS Singapore (2017–present). He was previously at Oxford (1974–1991) before moving to UC Santa Barbara to become Professor of Materials. He then held the Goldsmiths' Chair of Materials Science at the University of Cambridge (2007–2017) and was the Treasurer and Vice-President of the Royal Society (2012–2017). His research covers inorganic materials and metal–organic frameworks.

Ram Seshadri has been on the faculty at UC Santa Barbara since 2002 and is currently Distinguished Professor and the Fred and Linda R. Wudl Chair in the Materials Department and the Department of Chemistry and Biochemistry. He is an Executive Editor of *Chemistry of Materials* and is the Editor of *Annual Reviews of Materials Research*. His research addresses functional inorganic (and occasionally hybrid) materials for magnetic, energy storage, and quantum information applications.

ACKNOWLEDGMENTS

We gratefully acknowledge the invaluable help of Farnaz Kaboudvand, Jonathan Li, Anya Mulligan, and Derek Ober during different stages of this work and thank Anton Van Der Ven for useful insights. A.K.C. thanks the Ras al Khaimah Centre for Advanced Materials for financial support. R.S. gratefully acknowledges the US Department of Energy, Office of Science, Basic Energy Sciences, for support through DE-SC0024422.

REFERENCES

- (1) Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. Scaling Deep Learning for Materials Discovery. *Nature* **2023**, *624*, 80–90.
- (2) https://github.com/google-deepmind/materials_discovery (accessed January 27, 2024).
- (3) <https://next-gen.materialsproject.org/materials/gnome> (accessed November 29, 2023).
- (4) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1*, 011002.
- (5) Cheetham, A. K.; Seshadri, R.; Wudl, F. Chemical Synthesis and Materials Discovery. *Nat. Synthesis* **2022**, *1*, 514–520.
- (6) Kiessling, R.; Hagdahl, L.; Sillen, L. G.; Rottenberg, M. The Borides of Some Transition Elements. *Acta Chem. Scand.* **1950**, *4*, 209–227.
- (7) Wang, L.; Maxisch, T.; Ceder, G. Oxidation Energies of Transition Metal Oxides Within the GGA+U Framework. *Phys. Rev. B* **2006**, *73*, 195107.
- (8) Hong, H. Y.-P.; Chinn, S. R. Crystal Structure and Fluorescence Lifetime of Potassium Neodymium Orthophosphate, $K_3Nd(PO_4)_2$, a New Laser Material. *Mater. Res. Bull.* **1976**, *11*, 421–428.
- (9) Eick, H. A. The Crystal Structure and Lattice Parameters of Some Rare Earth Mono-Seleno Oxides. *Acta Crystallogr.* **1960**, *13*, 161.
- (10) Cromer, D. T.; Olsen, C. E. The Crystal Structure of $PuNi_3$ and $CeNi_3$. *Acta Crystallogr.* **1959**, *12*, 689–694.
- (11) Paccard, L.; Paccard, D. Structure of YRh_2Si : An Ordered $CeNi_3$ Type. *J. Less Common Met.* **1985**, *109*, 229–232.
- (12) Buschow, K. H. J.; Van Der Goot, A. S. The Crystal Structure of Rare-Earth Cobalt Compounds of the Type R_3Co . *J. Less Common Metals* **1969**, *18*, 309–311.
- (13) Le Roy, J.; Moreau, J. M.; Paccard, D.; Parthé, E. Rare Earth-Iridium Compounds with Pu_3Rh_3 and Y_3Rh_2 Structure Types: Members of a New Structural Series with Formula $R_{3n+6}T_{3n+5}$. *J. Less Common Met.* **1980**, *76*, 131–135.
- (14) Kirby, H. W.; Morss, L. R. Actinium. In *The Chemistry of the Actinide and Transactinide Elements*; Morss, L. R., Edelstein, N. M., Fuger, J., Eds.; Springer: Dordrecht, The Netherlands, 2008.
- (15) Hellenbrandt, M. The Inorganic Crystal Structure Database (ICSD) — Present and Future. *Crystallography Reviews* **2004**, *10*, 17–22.
- (16) Urusov, V. S.; Nadezhina, T. N. Frequency Distribution and Selection of Space Groups in Inorganic Crystal Chemistry. *J. Struct. Chem.* **2009**, *50*, 22–37.
- (17) Leeman, J.; Liu, Y.; Stiles, J.; Lee, S.; Bhatt, P.; Schoop, L.; Palgrave, R. Challenges in High-Throughput Inorganic Material Prediction and Autonomous Synthesis. *PRX Energy* **2024**, *3*, 011002.
- (18) Szymanski, N. J.; Rendy, B.; Fei, Y.; Kumar, R. E.; He, T.; Milsted, D.; McDermott, M. J.; Gallant, M.; Cubuk, E. D.; Merchant, A.; Kim, H.; Jain, A.; Bartel, C. J.; Persson, K.; Zeng, Y.; Ceder, G. A Laboratory for the Accelerated Synthesis of Novel Materials. *Nature* **2023**, *624*, 86–97.
- (19) George, E. P.; Raabe, D.; Ritchie, R. O. High-Entropy Alloys. *Nat. Rev. Mater.* **2019**, *4*, 515–534.
- (20) Carnicom, E. M.; Kong, T.; Klimczuk, T.; Cava, R. J. The Sigma-Phase Superconductors $Nb_{20.4}Rh_{5.7}Ge_{3.9}$ and $Nb_{20.4}Rh_{5.7}Si_{3.9}$. *Solid State Commun.* **2018**, *284*, 96–101.
- (21) Joubert, J.-M. Crystal Chemistry and Calphad Modeling of the σ Phase. *Prog. Mater. Sci.* **2008**, *53*, 528–583.
- (22) Mewis, A. Eine mit Tetraederketten Aufgefüllte Rutilstuktur: Die Verbindung $CaCu_4P_2 = Ca_2P_2(Cu_4)$. *Z. Anorg. Allg. Chem.* **1987**, *545*, 43–56.
- (23) Holleck, H.; Thuemmler, F. Ternaere Komplex-Carbide, -Nitride und -Oxide mit Teilweise Aufgefüllter Ti_2Ni -Struktur. *Monat. Chemie* **1967**, *98*, 133–134.
- (24) Kutahyali Aslani, C.; Breton, L. S.; Klepov, V. V.; zur Loye, H.-C. A Series of $RbLn_2(P_2S_6)(PS_4)_2$ ($Ln = La, Ce, Pr, Nd, Sm, Gd$) Rare Earth Thiophosphates with Two Distinct Thiophosphate Units $[P^{VS}_4]^{3-}$ and $[P^{IV}_2S_6]^{4-}$. *Dalton Trans.* **2021**, *50*, 1683–1689.
- (25) Gelato, L. M.; Parthé, E. STRUCTURE TIDY - a Computer Program to Standardize Crystal Structure Data. *J. Appl. Crystallogr.* **1987**, *20*, 139–143.
- (26) Spek, A. L. checkCIF Validation ALERTS: What They Mean and How to Respond. *Acta Crystallogr.* **2020**, *76*, 1–11.
- (27) Thomas, J. C.; Natarajan, A. R.; Van der Ven, A. Comparing Crystal Structures with Symmetry and Geometry. *npj Comp. Mater.* **2021**, *7*, 164.
- (28) Pauling, L. The Principles Determining the Structure of Complex Ionic Crystals. *J. Am. Chem. Soc.* **1929**, *51*, 1010–1026.
- (29) Richards, W. D.; Miara, L. J.; Wang, Y.; Kim, J. C.; Ceder, G. Interface Stability in Solid-State Batteries. *Chem. Mater.* **2016**, *28*, 266–273.