

# Commentary on the recent publication “Scaling deep learning for materials discovery” by Merchant *et al.* *Nature* 624, 80–90 (2023)

Anthony K. Cheetham<sup>a,b</sup> and Ram Seshadri<sup>a</sup>

<sup>a</sup>Materials Department and Materials Research Laboratory  
University of California, Santa Barbara, California 93106, United States

<sup>b</sup>Department of Materials Science and Engineering  
National University of Singapore, Singapore 117575, Singapore

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 fraction 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 fulfil 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 better incorporate domain expertise in materials synthesis and crystallography. We suggest that not doing so results in predictions that are premature, with exaggerated claims and unfulfilled promises.

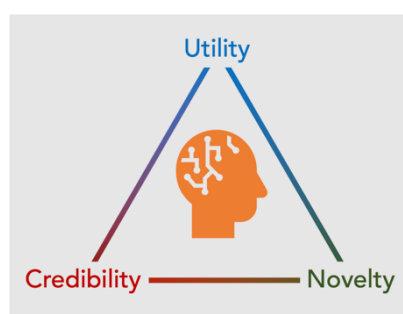
## Introduction

In an article in *Nature* published in November 2023, Merchant *et al.*<sup>1</sup> 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 both classical inorganic compounds such as oxides and halides, as well as 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 Structures Database,<sup>2</sup> while further structural details of more than 2000 of these new compounds have been placed in the GNoME Explorer archive<sup>3</sup> within the Materials Project.<sup>4</sup> 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 (that is sometimes arguably mislabeled as chemical intuition) serving as guides. In this commentary, 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 Structures 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 to materials discovery, such as the discovery of high temperature superconductors and lithium ion batteries.<sup>5</sup> 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.

Exploring and Exploiting?



**Figure 1:** Predictions of new materials should ideally combine three qualities depicted in this triangle: The prediction should be credible, with evidence that the compound would exist; the compound should be truly novel, in the sense that it should not be an obvious extension of the rules of crystal chemistry; finally, to be called a material, some evidence of utility should be demonstrated.

Before progressing further, we point out that the predictions in the contribution of Merchant *et al.*<sup>1</sup> are solely of crystalline inorganic compounds and should be described as such, rather than using the more generic label “material”. There are many 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 displayed in **Figure 1**, 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 displaying 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 have 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 would like to 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, and 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, and 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.  $\text{LaB}_6$  is a typical example and adopts the cubic space group,  $Pm\bar{3}m$ , with  $a \approx 4.1 \text{ \AA}$ .<sup>6</sup>  $\text{AcEr}_3(\text{B}_6)_4$  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  $\text{Ac}^{3+}$  and  $\text{Er}^{3+}$  (which may or may not be experimentally viable) leading to a  $4\times$  superstructure along  $c$  and lowering the symmetry of the parent  $\text{LaB}_6$  structure from cubic to tetragonal. There are other examples of apparently novel and complex borides that are based upon the  $\text{LaB}_6$  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,  $\text{TbSmF}_{30}$  (mp-3170019) defies all the usual rules of valence and contains an improbable ordering of  $\text{Tb}^{3+}$  and  $\text{Sm}^{3+}$  (see below). Further inspection of the structure reveals that it contains large numbers of  $\text{F}_2$  molecules, some isolated and others acting as linkers between  $\text{LnF}_9$  polyhedra (where  $\text{Ln}$  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  $\text{F}_2$ ) in electronic structure calculations of formation energies and stability.<sup>7</sup>

(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  $a = 9.8 \text{ \AA}$ ,  $b = 5.8 \text{ \AA}$ ,  $c = 7.6 \text{ \AA}$ ,  $\beta = 91.87^\circ$ . 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 ( $a = 9.5 \text{ \AA}$ ,

$b = 5.6 \text{ \AA}$ ,  $c = 7.4 \text{ \AA}$ ,  $\beta = 90.95^\circ$ ).<sup>8</sup> These structures are virtually identical, aside from the lower symmetry in the GNoME entry, a recurring theme that we shall return 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  $\text{TbSmSeO}_2$  (mp-3169970), the  $\text{Tb}^{3+}$  and  $\text{Sm}^{3+}$  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 ( $0.92 \text{ \AA}$  and  $0.96 \text{ \AA}$ , 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 improbably 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*  $P-3m1$  structure of  $\text{La}_2\text{SeO}_2$ , which has  $a \approx 4.09 \text{ \AA}$  and  $a \approx 7.16 \text{ \AA}$ .<sup>9</sup> 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,  $\text{DyHo}_2\text{Rh}_9$  (mp-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  $\text{PuNi}_3$  structure, which is found in a large number of phases such as  $\text{CeNi}_3$ ,  $P6_3/mmc$  with  $a = 4.98 \text{ \AA}$  and  $c = 16.94 \text{ \AA}$ .<sup>10</sup> Ordered variants of this structure are also known, e.g.  $\text{YRh}_2\text{Si}$  ( $P6_3/mmc$  with  $a = 5.49 \text{ \AA}$  and  $c = 15.03 \text{ \AA}$ ),<sup>11</sup> though it seems highly improbable that Dy and Ho would order in the proposed  $\text{DyHo}_2\text{Rh}_9$  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  $\text{Ln}_{12}\text{Co}_4$ , then we can see that it represents an unlikely ordering of the small lanthanides in a compound of composition  $\text{Ln}_3\text{Co}$ . These phases are known for most of the rare-earths and yttrium, and they adopt the  $\text{Fe}_3\text{C}$  structure. For example,  $\text{Ho}_3\text{Co}$  adopts a centrosymmetric orthorhombic structure, space group  $Pnma$ , with  $a = 6.92 \text{ \AA}$ ,  $b = 9.29 \text{ \AA}$ ,  $c = 6.21 \text{ \AA}$ ;<sup>12</sup> 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,  $P-4$  (non-centrosymmetric), with  $a = 10.92 \text{ \AA}$  and  $c = 6.38 \text{ \AA}$ . If we assume that the very similar rare earths would be disordered, as discussed above, it can then be written  $\text{Ln}_{20}\text{Ir}_{12}$  or  $\text{Ln}_5\text{Ir}_3$ . This is a known structure type and has been reported for virtually all the rare earths with the tetragonal  $\text{Pu}_5\text{Rh}_3$  structure. For the specific case of  $\text{Tb}_5\text{Ir}_3$ ,<sup>13</sup> the space group is centrosymmetric  $P4/ncc$  with  $a = 10.905 \text{ \AA}$  and  $c = 6.299 \text{ \AA}$ . Note again that the prediction puts

the different lanthanides on distinct sites and therefore lowers the symmetry. We shall 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 ( $\text{Ac}^{3+}$ ) compounds, of which there are 27 (and approximately 6,754 in the larger Stable Structures Database), it should be noted that all isotopes of actinium are intensely radioactive with half-lives ranging from days to a few years.<sup>14</sup> The natural abundance of Ac in, for example, uranium ores is so low that a ton of uranium yields less than a milligram of  $^{227}\text{Ac}$  (half-life 21.77 years). Milligram quantities of  $^{227}\text{Ac}$  can be made artificially by neutron irradiation of  $^{226}\text{Ra}$  and this isotope has been studied for cancer treatments. However, ionic actinium compounds, which always contain  $\text{Ac}^{3+}$ , have no industrial applications. Furthermore, the few that are known are isomorphous with their stable lanthanum analogs (e.g.  $\text{AcPO}_4$  and  $\text{LaPO}_4$ ). The “discoveries” of new actinium compounds, such as  $\text{AcPO}_3$  (mp-3170055), are therefore impractical and not very novel. In fact, there are 18,138 compounds of such radioactive elements in the large Stable Structures 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

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       | <i>R-3m</i>  | 8834       | 2.30       |
| <i>P1</i>                                   | 39382      | 10.2       | <i>P-6</i>   | 8563       | 2.22       |
| <i>Amm2</i>                                 | 26467      | 6.88       | <i>P-62m</i> | 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       | <i>P-3m1</i> | 6379       | 1.66       |
| <i>C2</i>                                   | 11201      | 2.91       | <i>P-6m2</i> | 6038       | 1.57       |
| <i>P-1</i>                                  | 10463      | 2.72       | <i>Pnma</i>  | 6005       | 1.56       |

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),<sup>15</sup> which has been analyzed by Urusov and Nadezhina.<sup>16</sup> For example, the top two space groups in the ICSD are centrosymmetric *Pnma* and *P2<sub>1</sub>/c*, accounting for approximately 16 % of all structures, with  $\approx 8$  % in each case. By contrast, *Pnma* is ranked 16<sup>th</sup> in Stable Structures Database with 1.56 % of the structures, while *P2<sub>1</sub>/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 Structures database are all non-centrosymmetric and account for  $\approx 34$  % 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 the known ICSD structures is a matter of serious concern. We also note that this issue has been highlighted<sup>17</sup> in a commentary on a different *Nature* article on robotic/AI-based materials discovery.<sup>18</sup> 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.* Dy<sub>6</sub>Y<sub>2</sub>Ho<sub>11</sub>Lu(Cd<sub>3</sub>Ru)<sub>2</sub> (mp-3195118), they are suggestive of high-entropy alloys,<sup>19</sup> 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 arise because the DFT calculations have been performed at 0 K, where entropic effects are ignored, resulting in the predicted structures being fully ordered. In addition, many compounds undergo phase transitions to lower symmetry structures on cooling to low temperatures, though we have no evidence of this from the GNoME entries that we have examined.

## Randomly selected examples from the Stable Structures Database

Table 2: Comparison of 10 randomly selected compounds from the Stable Structure database with appropriate ICSD entries. The high incidence of pseudosymmetry in the Stable Structure database entries is notable.

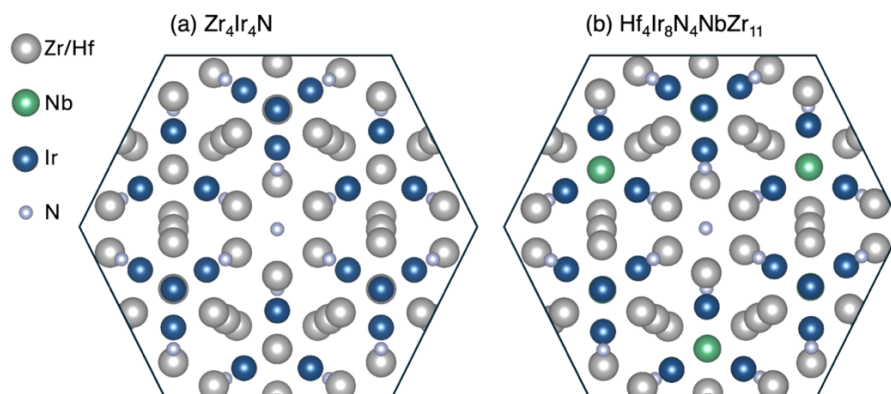
|    | Entry #     | Composition   | Space Group               | Volume | <i>a</i> , <i>b</i> , <i>c</i> (Å) | $\alpha$ , $\beta$ , $\gamma$ (°) |
|----|-------------|---|---------------------------|--------|------------------------------------|-----------------------------------|
|    | Database    |   |                           |        |                                    |                                   |
|    | ICSD analog |   |                           |        |                                    |                                   |
| 1  | 186808      | BsDy <sub>4</sub> RuTc <sub>3</sub>                               | <i>Pm</i>                 | 202.3  | 5.944, 5.293, 6.430                | 90.0, 90.1, 90.0                  |
|    | 100774      | B <sub>2</sub> DyRu   | <i>Pnma</i>               | 198.2  | 5.886, 5.300, 6.352                | 90, 90, 90                        |
| 2  | 308005      | Pb <sub>2</sub> Pd <sub>3</sub> SnY <sub>3</sub>                  | <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 <sub>4</sub> Ir <sub>8</sub> N <sub>4</sub> NbZr <sub>11</sub> | <i>Cm</i>                 | 478.3  | 8.799, 8.799, 8.800                | 59.7, 59.6, 59.6                  |
|    | 640826      | Zr <sub>4</sub> Ir <sub>2</sub> N                                 | <i>Fd-3m</i>              | 1884   | 12.35, 12.35, 12.35                | 90, 90, 90                        |
| 4  | 162234      | B <sub>2</sub> Er <sub>4</sub> Os <sub>3</sub> Tb <sub>4</sub> Tc | <i>P1</i>                 | 385.4  | 3.656, 9.137, 11.53                | 89.9, 90.0, 90.0                  |
|    | 603495      | B <sub>6</sub> Er <sub>2</sub> Os                                 | <i>Pban</i>               | 377.2  | 9.097, 11.46, 3.617                | 90, 90, 90                        |
| 5  | 352440      | AlTa <sub>16</sub> Te <sub>16</sub> V <sub>3</sub>                | <i>P1</i>                 | 1060   | 4.907, 11.11, 19.44                | 90.01, 90.0, 90.0                 |
|    | 79796       | FeTa <sub>4</sub> Te <sub>4</sub>                                 | <i>Pbam</i>               | 925.2  | 10.51, 18.27, 4.185                | 90, 90, 90                        |
| 6  | 304898      | Ga <sub>4</sub> La <sub>2</sub> Pt <sub>3</sub> Rh                | <i>P-4n2</i>              | 194.5  | 4.321, 4.321, 10.52                | 90, 90, 90                        |
|    | 621150      | Ce <sub>2</sub> Pt <sub>1.5</sub> Ga <sub>6.5</sub>               | <i>I4/mmm</i>             | 196.1  | 4.315, 4.315, 10.53                | 90, 90, 90                        |
| 7  | 369281      | Cu <sub>8</sub> P <sub>4</sub> SrTm                               | <i>Cmmm</i>               | 204.2  | 3.815, 7.317, 7.317                | 88.96, 90, 90                     |
|    | 62553       | CaCu <sub>4</sub> P <sub>2</sub>                                  | <i>P4<sub>2</sub>/mnm</i> | 203.6  | 7.290, 7.290, 3.831                | 90, 90, 90                        |
| 8  | 252376      | DyPSi <sub>5</sub> Tb <sub>9</sub>                                | <i>P1</i>                 | 393.7  | 8.460, 8.466, 6.344                | 90.0, 90.0, 120.0                 |
|    | 99641       | Gd <sub>5</sub> Si <sub>3</sub>                                   | <i>P6<sub>3</sub>/mcm</i> | 403.0  | 8.513, 8.513, 6.421                | 90, 90, 120                       |
| 9  | 13745       | Ac <sub>4</sub> P <sub>8</sub> S <sub>28</sub> Tl <sub>8</sub>    | <i>C2/c</i>               | 1339   | 6.822, 12.29, 17.12                | 79.5, 78.4, 73.9                  |
|    | 25121       | La <sub>2</sub> P <sub>4</sub> Si <sub>14</sub> Rb <sub>4</sub>   | <i>P2<sub>1</sub>/c</i>   | 1276   | 9.950, 6.846, 19.82                | 90, 98.2, 90                      |
| 10 | 326805      | Ga <sub>2</sub> Nb <sub>20</sub> Os <sub>8</sub>                  | <i>P4<sub>2</sub>/mnm</i> | 508.7  | 9.933, 9.933, 5.156                | 90, 90, 90                        |
|    | 230694      | Ge <sub>3.9</sub> Nb <sub>20.4</sub> Rh <sub>5.9</sub>            | <i>P4<sub>2</sub>/mnm</i> | 497.6  | 9.848, 9.848, 5.133                | 90, 90, 90                        |

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 Structures 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:

- (i) We were able to find 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.
- (ii) In only one case was the space group the same. This enabled us to find an obvious analogue of  $\text{Ga}_2\text{Nb}_{20}\text{Os}_8$  in the ICSD by searching for structures in space group  $P4_2/mnm$  with lattice parameters close to  $a = 9.9 \text{ \AA}$  and  $c = 5.1 \text{ \AA}$ , pointing us to  $\text{Ge}_{3.9}\text{Nb}_{20.4}\text{Rh}_{5.9}$ , which is one of the well-known Frank-Kasper sigma ( $\sigma$ ) phases.<sup>20,21</sup> 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.
- (iii) In some cases, the analogy between the Stable Structure database and the ICSD entries is obvious. For example,  $\text{Cu}_8\text{P}_4\text{SrTm}$  was identified as having the same structure as  $\text{CaCu}_4\text{P}_2$ ,<sup>22</sup> 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  $\text{Tm}^{3+}$ , this would require electron-doping a  $\text{Cu}^{1+}$  compound, which is unlikely.



**Figure 2:** Views of the crystal structures of (a) known  $\text{Zr}_4\text{Ir}_4\text{N}$  (ICSD 640826) compared in a similar projection, and identical scaling with (b) the proposed structure of  $\text{Hf}_4\text{Ir}_8\text{N}_4\text{NbZr}_{11}$ . 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.

- (iv) In other cases, the identification of the structure type takes a little more knowledge of the periodic table and crystal symmetry. For example, in  $\text{Hf}_4\text{Ir}_8\text{N}_4\text{NbZr}_{11}$  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 towards  $\text{Zr}_{11}\text{Ir}_8\text{N}_4\text{Nb}$ ). 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  $\text{Zr}_4\text{Ir}_2\text{N}$ ,<sup>23</sup> whose cubic unit cell is 4× larger than the cell of  $\text{Hf}_4\text{Ir}_8\text{N}_4\text{NbZr}_{11}$  in the database. The final step is to recognize that the pseudo rhombohedral cell from the Stable Structures Database with  $\alpha = 59.7^\circ$  is also nearly pseudo cubic, thus accounting for the 4× discrepancy in the cell volume. The two structures are compared in **Figure 2**.
- (v) In the case of  $\text{Ac}_4\text{P}_8\text{S}_{28}\text{Tl}_8$ , the similarity to the structure of  $\text{La}_2\text{P}_4\text{Si}_{14}\text{Rb}_4$ <sup>24</sup> is not immediately apparent from the unit cells. However, we would expect Ac compounds would be isomorphous with their La analogues.  $\text{Tl}^{1+}$  and  $\text{Rb}^{1+}$  are similar close in size and identical in their charge states. The similarity of the structures becomes evident.

## Conclusions

Our analysis of the predictions by Merchant *et al.*<sup>1</sup> 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 the 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.*<sup>1</sup> 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, the 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.<sup>25</sup> Some of them would be useful in resolving issues of pseudosymmetry that is present in many of the entries.<sup>26</sup> Efforts to relate structure types using their structures and geometries are also ongoing, and clearly need large-scale implementation.<sup>27</sup>

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 is delivering 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 a greater effort to connect the predictions with 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.<sup>28</sup> 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 very 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.*<sup>1</sup> 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<sup>+</sup>-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<sup>+</sup> solid electrolytes.<sup>29</sup> This points to an interesting contraindication in materials design: soft anions such as Te<sup>2-</sup> and I<sup>-</sup> readily permit cation transport but do not possess the requisite electrochemical (redox) stability for use as solid electrolytes. Conversely, hard anions such as O<sup>2-</sup> and F<sup>-</sup> 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.

**Acknowledgements:** 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, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon and E. D. Cubuk, Scaling Deep Learning for Materials Discovery, *Nature* **624**, 80–90 (2023). DOI: 10.1038/s41586-023-06735-9.
2. [https://github.com/google-deepmind/materials\\_discovery](https://github.com/google-deepmind/materials_discovery)
3. <https://next-gen.materialsproject.org/materials/gnome>
4. A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation, *APL Mater.* **1**, 011002 (2013). DOI: 10.1063/1.4812323
5. A. K. Cheetham, R. Seshadri and F. Wudl, Chemical Synthesis and Materials Discovery, *Nat. Synthesis* **1**, 514–520 (2022). DOI: 10.1038/s44160-022-00096-3
6. R. Kiessling, The Borides of Some Transition Elements, *Acta Chem. Scand* **4**, 209–227 (1950). DOI: 10.3891/acta.chem.scand.04-0209
7. L. Wang, T. Maxisch, and G. Ceder, Oxidation Energies of Transition Metal Oxides Within the GGA+*U* Framework, *Phys. Rev. B* **73**, 195107 (2006). DOI: 10.1103/PhysRevB.73.195107
8. H. Y.-P. Hong and S. R. Chinn, Crystal Structure and Fluorescence Lifetime of Potassium Neodymium Orthophosphate,  $\text{K}_3\text{Nd}(\text{PO}_4)_2$ , a New Laser Material, *Mater. Res. Bull.* **11**, 421–428 (1976). DOI: 10.1016/0025-5408(76)90091-X
9. H. A. Eick, The Crystal Structure and Lattice Parameters of Some Rare Earth Mono-Seleno Oxides, *Acta Crystallog.* **13**, 161 (1960). DOI: 10.1107/S0365110X60000339
10. D. T. Cromer and C. E. Olsen, The Crystal Structure of  $\text{PuNi}_3$  and  $\text{CeNi}_3$ , *Acta Crystallogr.* **12**, 689–694 (1959). DOI: 10.1107/S0365110X59002006
11. L. Paccard and D. Paccard, Structure of  $\text{YRh}_2\text{Si}$ : An Ordered  $\text{CeNi}_3$  Type, *J. Less Common Met.* **109**, 229–232 (1985). DOI: 10.1016/0022-5088(85)90054-2
12. K. H. J. Buschow and A. S. Van Der Goot, The Crystal Structure of Rare-Earth Cobalt Compounds of the Type  $\text{R}_3\text{Co}$ , *J. Less Common Metals* **18**, 309–311 (1969). DOI: 10.1016/0022-5088(69)90169-6
13. J. Le Roy, J. M. Moreau, D. Paccard, and E. Parthé, Rare Earth-Iridium Compounds with  $\text{Pu}_5\text{Rh}_3$  and  $\text{Y}_3\text{Rh}_2$  Structure Types: Members of a New Structural Series with Formula  $\text{R}_{5n+6}\text{T}_{3n+5}$ , *J. Less Common. Met.* **76**, 131-135 (1980). DOI: 10.1016/0022-5088(80)90016-8
14. H. W. Kirby and L. R. Morss, Actinium, in L. R. Morss, N. M. Edelstein, and J. Fuger, (eds.), *The Chemistry of the Actinide and Transactinide Elements*, Springer, Dordrecht (2008). DOI: 10.1007/1-4020-3598-5\_2
15. M. Hellenbrandt, The Inorganic Crystal Structure Database (ICSD) — Present and Future, *Crystallography Reviews*, **10**, 17–22 (2004). DOI: 10.1080/08893110410001664882
16. V. S. Urusov and T. N. Nadezhina, *J. Struct. Chem.* **50**, S22-S37 (2009). DOI: 10.1007/s10947-009-0186-9
17. J. Leeman, Y. Liu, J. Stiles, S. Lee, P. Bhatt, L. Schoop, and R. Palgrave, Challenges in High-Throughput Inorganic Material Prediction and Autonomous Synthesis, DOI: 10.26434/chemrxiv-2024-5p9j4
18. N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng, and G. Ceder, *Nature* **624**, 86–97 (2023). DOI: 10.1038/s41586-023-06734-w

19. E. P. George, D. Raabe, and R. O. Ritchie, High-Entropy Alloys, *Nat. Rev. Mater.* **4**, 515–534 (2019). DOI: 10.1038/s41578-019-0121-4
20. E. M. Carnicom, T. Kong, T. Klimczuk, and R. J. Cava, The Sigma-Phase Superconductors Nb<sub>20.4</sub>Rh<sub>5.7</sub>Ge<sub>3.9</sub> and Nb<sub>20.4</sub>Rh<sub>5.7</sub>Si<sub>3.9</sub>, *Solid State Commun.* **284**, 96–101 (2018). DOI: 10.1016/j.ssc.2018.09.014
21. J.-M. Joubert, Crystal Chemistry and Calphad Modeling of the  $\sigma$  Phase, *Prog. Mater. Sci.* **53**, 528–583 (2008). DOI: 10.1016/j.pmatsci.2007.04.001
22. Mewis, Eine mit Tetraederketten Aufgefüllte Rutilstuktur: Die Verbindung CaCu<sub>4</sub>P<sub>2</sub> = Ca<sub>2</sub>P<sub>2</sub>(Cu<sub>4</sub>), *Z. Anorg. Allg. Chemie*, **545**, 43–56 (1987). DOI: 10.1002/zaac.19875450205
23. H. Holleck and F. Thuemmler, Ternäre Komplex-Carbide, -Nitride und -Oxide mit Teilweise Aufgefüllter Ti<sub>2</sub>Ni-Struktur, *Monat. Chemie* **98**, 133–134 (1967). DOI: 10.1007/BF00901107
24. C. T. Aslani, L. S. Breton, V. V. Klepov, and H. C. zur Loye, A Series of RbLn<sub>2</sub>(P<sub>2</sub>S<sub>6</sub>)(PS<sub>4</sub>)<sub>2</sub> (Ln = La, Ce, Pr, Nd, Sm, Gd) Rare Earth Thiophosphates with Two Distinct Thiophosphate Units [P<sup>V</sup>S<sub>4</sub>]<sup>3-</sup> and [P<sup>IV</sup><sub>2</sub>S<sub>6</sub>]<sup>4-</sup> *Dalton Trans.* **50**, 1683–1689 (2021). DOI: 10.1039/D0DT03718D
25. L. M. Gelato and E. Parthé, STRUCTURE TIDY – a Computer Program to Standardize Crystal Structure Data, *J. Appl. Cryst.* **20**, 139–143 (1987). DOI: 10.1107/S0021889887086965
26. A. L. Spek, checkCIF Validation ALERTS: What They Mean and How to Respond, *Acta Cryst.* **E76**, 1–11 (2020). DOI: 10.1107/S2056989019016244 1
27. J. C. Thomas, A. R. Natarajan, and A. Van der Ven, Comparing Crystal Structures with Symmetry and Geometry, *npj Comp. Mater.* **7**, 164 (2021) 164. DOI: 10.1038/s41524-021-00627-0
28. L. Pauling, The Principles Determining the Structure of Complex Ionic Crystals, *J. Am. Chem. Soc.* **51**, 1010–1026 (1929). DOI: 10.1021/ja01379a006
29. W. D. Richards, L. J. Miara, Y. Wang, J C Kim and G. Ceder, Interface Stability in Solid-State Batteries, *Chem. Mater.* **28**, 266–273 (2016). DOI: 10.1021/acs.chemmater.5b04082