



Thermodynamic assessment of machine learning models for solid-state synthesis prediction

Cite this: DOI: 10.1039/d6mh00209a

Received 3rd February 2026,
Accepted 15th June 2026

DOI: 10.1039/d6mh00209a

rsc.li/materials-horizonsJane Schlesinger, Simon Hjaltason, Nathan J. Szymanski and
Christopher J. Bartel *

Machine learning models have recently emerged to predict whether hypothetical solid-state materials can be synthesized. These models aim to circumvent direct first-principles modeling of solid-state phase transformations, instead learning from large databases of successfully synthesized materials. Here, we assess the alignment of several recently introduced synthesis prediction models with material and reaction thermodynamics, quantified by the energy with respect to the convex hull and a metric accounting for thermodynamic selectivity of enumerated synthesis reactions. A dataset of successful synthesis recipes was used to determine the likely bounds on both quantities beyond which materials can be deemed unlikely to be synthesized. With these bounds as context, thermodynamic quantities were computed using the CHGNet foundation potential for thousands of new hypothetical materials generated using the Chemeleon generative model. Four recently published machine learning models for synthesizability prediction were applied to this same dataset, and the resultant predictions were considered against computed thermodynamics. We find these models generally overpredict the likelihood of synthesis, but some model scores do trend with thermodynamic heuristics, assigning lower scores to materials that are less stable or do not have an available synthesis recipe that is calculated to be thermodynamically selective. In total, this work identifies existing gaps in machine learning models for materials synthesis and introduces a new approach to assess their quality in the absence of extensive negative examples (failed syntheses).

Introduction

Inorganic crystalline solids underpin numerous technologies, motivating the discovery of new materials with optimized properties to support further innovation. Density functional theory (DFT) calculations are now regularly applied to compute materials properties in high throughput, allowing candidates to be screened prior to synthesis.^{1–4} Meanwhile, generative artificial intelligence

New concepts

Synthesis remains the primary bottleneck in the materials discovery pipeline. Machine learning models that can rapidly predict whether hypothetical materials can be synthesized would be extremely valuable, but their efficacy remains difficult to quantify due to challenges in proving materials cannot be made. Without these negative examples, the success of models in identifying unsynthesizable materials cannot be determined. In this work, we propose a framework for model evaluation in the absence of known negative examples by identifying thermodynamic bounds on successfully synthesized materials. Using this framework, we show that existing models generally overpredict the synthesizability of hypothetical materials. The deployment of these models is therefore likely to lead to many false positives – materials predicted to be synthesizable that cannot be made by solid-state synthesis. However, some model predictions do align with thermodynamic trends, suggesting machine learning models hold promise for synthesis prediction. Given the relative sparsity of materials synthesis data, especially failed syntheses, our work shows how calculated thermodynamic quantities can be incorporated into model evaluation.

(AI) models are being used to propose thousands of novel materials with desired properties, further expanding the pool of hypothetical materials.⁵ To support this effort, researchers often leverage calculable thermodynamic heuristics as a filter to identify viable and realistic synthesis candidates. Most commonly used is thermodynamic stability, which can be quantified by the energy of a material above the convex hull (E_{hull}) defined by known ground states. E_{hull} is a relative measure obtained by constructing a phase diagram for the chemical space of interest. The energies used in this construction can be drawn from existing databases^{6–9} or computed using density functional theory (DFT) or interatomic potentials for the materials of interest.^{10–14} If a given material at a particular composition cannot lower its energy by phase separation or phase transition, it is stable ($E_{\text{hull}} = 0$) and therefore likely to be synthesizable. Thermodynamically unstable materials can often still be synthesized, but those near the hull (low E_{hull}) are generally considered more accessible.¹⁵ Sun *et al.* found that about 50% of all experimentally reported materials in the Inorganic Crystal Structure Database (ICSD)¹⁶ are stable and 90% of those that are

Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA. E-mail: cbartel@umn.edu

unstable (above the hull) exhibit $E_{\text{hull}} \leq 67$ meV per atom as computed with DFT at 0 K.¹⁷

Thermodynamic stability is effective as a preliminary filter of candidate materials, but turning these feasible candidates into experimentally made materials remains a major challenge. As such, the number of hypothetical materials (e.g., those appearing in open DFT databases^{4,18}) far outpaces those that have been experimentally realized (e.g., those appearing in the ICSD). The core bottleneck is our inability to predict whether and how a material can be synthesized.¹⁹ Thermodynamically stable materials should form at equilibrium if the conditions used for synthesis are not too dissimilar from those accounted for computationally. However, an actual solid-state reaction may not reach equilibrium, or it is possible that synthesis conditions (e.g., temperature) have a significant influence on the resulting thermodynamics. The product that does form is determined by the reaction pathway, which is itself dependent on the concerted motion of many atoms through defect formation, ion transport, nucleation, and crystal growth, among other processes. The complex mechanism by which crystalline materials transform precludes direct atomistic modeling of the synthesis process, and solid-state reaction outcomes are therefore challenging to directly predict. As a result, costly trial-and-error experimentation is often an unavoidable, and sometimes fruitless, step in the discovery of novel materials.

The successful synthesis of a new material depends on the chosen synthesis recipe, defined by the precursors and reaction conditions. There have been several recent efforts to apply first-principles thermodynamic calculations to anticipate how the choice of synthesis recipe influences the likely phase evolution sequence, mostly in the context of solid-state synthesis.^{20–24} Towards the goal of identifying “optimal” solid-state synthesis recipes, McDermott *et al.* formalized an approach to enumerate and quantify the plausibility of many pathways to a given target. This approach enables the identification of precursors that minimize a user-specified cost function (e.g., related to the anticipated reaction energies) to form the desired target.²⁵ Building on this work, the authors posited that optimal precursors have both a strong driving force (i.e., a large and negative Gibbs free energy of reaction, ΔG) to form the target material and minimal driving force for “competing” impurity-forming reactions.²⁶ In this context, primary competition (C_1) is defined as the difference between ΔG to form the target and ΔG to form the most favorable competing product from a given precursor set. Secondary competition (C_2) measures the tendency of the target material (once formed) to react with unconsumed precursors and produce unwanted impurities. The authors proposed an aggregate metric (Γ) to account for both C_1 and C_2 :

$$\Gamma = w_0 \Delta G_{\text{rxn}} + w_1 C_1 + w_2 C_2 \quad (1)$$

ΔG_{rxn} measures the reaction energy associated with target formation from the precursors, and the weights (w_0 , w_1 , w_2) can be chosen to weight the influence of reaction energy, primary competition, and secondary competition as desired. The choice of precursors that minimizes Γ for a given target is then deemed the optimally selective recipe for that target from

a thermodynamic perspective. Unlike E_{hull} , which is defined for a material, Γ is defined for a chosen synthesis recipe, so the two quantities should be complementary in the discovery of new materials.

Machine learning (ML) methods are emerging as an alternative approach to predict the synthesizability of hypothetical materials without the need for first-principles thermodynamic calculations. ML models can capture complex, often elusive, physical relationships making them a promising tool for the multifaceted challenge of predictive synthesis. Existing applications of ML for synthesis prediction have focused on predicting whether a hypothetical material can be synthesized by any approach, irrespective of available synthesis recipes. In this way, models are tasked with labeling materials as either synthesizable or unsynthesizable. The success of ML is generally contingent on sufficient data availability. While the literature provides extensive data on successfully synthesized materials and the synthesis recipes used to make them,^{16,27} it is more difficult to obtain reliable “negative” data.²⁸ The absence of a material from these databases does not imply it is unsynthesizable; rather, it might not have been attempted or reported yet. Without clear negative examples, conventional classification models cannot learn the distinction between synthesizable and unsynthesizable materials.

Jang *et al.* proposed a positive-unlabeled (PU) learning approach to address the limited availability of negative synthesis data. The authors employed PU learning to assign labels to unreported materials with a transductive bagging approach, aggregating the labels assigned by models trained on different subsets of the data to ultimately label all unlabeled materials.²⁹ Experimentally reported materials are treated as positive examples, and the method for labeling assumes that materials differing substantially from positive examples with respect to a learned representation are less likely to be synthesizable and are labeled accordingly. Synthesizability classification is then treated as a classic supervised learning problem trained on the now fully labeled dataset. Several related efforts have since followed suit, using PU learning to train synthesizability classifiers with varied training data and model architecture.^{30–33} The models use structural and/or compositional information to predict “likelihood scores,” indicating the probability that each material can be synthesized (using any synthetic approach). If this score is above a designated threshold (often 0.5), the material is classified as synthesizable.

ML models are appealing for their streamlined and low-cost quantification of synthesizability. However, their utility to synthetic chemists is contingent on their reliability, which is difficult to quantify without access to extensive negative examples (failed synthesis experiments). In the absence of negative data, the thermodynamics of successful syntheses can inform the likelihood of accessing a material. In this work, we evaluate previously published ML model predictions in the context of thermodynamic heuristics. We find that synthesizability predictions from these models do correlate with thermodynamic heuristics to some degree, often assigning lower likelihood scores to materials far above the convex hull. However, the predicted scores remain too high in general, overestimating the

synthetic accessibility of materials that are unlikely to be synthesized based on a thermodynamic assessment.

Results

Synthesizability assessments from thermodynamic and data-driven approaches

Γ quantifies the thermodynamic selectivity of a chosen synthesis recipe (eqn (1)), whereas ML models for synthesis prediction assess the synthesizability of a chosen material. To align these two approaches, we define a material-centric quantity, Γ_{opt} as the lowest (most favorable) Γ across all possible precursors for a given target (details in Methods). We consider four recently published ML models for synthesizability prediction to compare with Γ_{opt} , all of which employ some form of PU learning. Of these four models, two use only compositional information (PU-CGNF³⁰ and SynthNN³¹) while two also use crystal structure information (PU-CGCNN²⁹ and SynCoTrain³²). The distinguishing features of each classifier are summarized in Table 1 and further detailed in the Methods section. Given the lack of data on unsynthesizable materials, ML synthesizability predictors are generally assessed on the basis of true positive rate (TPR) or recall. In the context of synthesis prediction, TPR indicates the fraction of synthesized materials that are predicted by the model to be synthesizable. While the specific details associated with training/testing protocols vary across the respective studies, the models employed here report test set TPRs of ~ 80 – 90% .

Calibration on text-mined synthesis data

A text-mined dataset of successful solid-state synthesis reactions²⁷ was used to identify the range of Γ consistent with known (successful) reactions. Using DFT-calculated thermodynamic data in MP, each published reaction in the text-mined dataset was mapped to a computed E_{hull} associated with the target material and a computed Γ associated with the experimentally observed recipe (Γ_{obs}). In Fig. 1a, Γ_{obs} is plotted as a function of the corresponding E_{hull} for 2495 reported reactions targeting ternary metal oxides. Some correlation between Γ_{obs} and E_{hull} is observed, with less stable materials often having higher Γ_{obs} and therefore synthesis recipes that are predicted to be less selective. However, the mapping between these two quantities is not one-to-one. There is substantial variance in Γ_{obs} even at fixed E_{hull} , meaning thermodynamic selectivity can differ substantially for two materials that are comparably close to

the convex hull, supporting the notion that these metrics provide complementary information.

In Fig. 1b, we illustrate how Γ can vary dramatically at relatively constant E_{hull} by comparing LaNb_3O_9 ($E_{\text{hull}} = 31$ meV per atom, $\Gamma_{\text{obs}} = 102$ meV per atom) and LaGaO_3 ($E_{\text{hull}} = 33$ meV per atom, $\Gamma_{\text{obs}} = 40$ meV per atom). The much larger Γ for LaNb_3O_9 can be understood by examining the interface reaction hulls in Fig. 1b, which represent syntheses listed in the text-mined dataset. Whereas the two products that compete with LaGaO_3 (right-hand hull, blue) are comparably “shallow” on the hull – meaning they have small reaction energies to form – the products that compete with LaNb_3O_9 lie much deeper on the interface hull. As a result, primary competition for LaNb_3O_9 formation is high ($C_1 = 115$ meV per atom) since the most competitive product (LaNbO_4) has a substantially greater driving force to form from Nb_2O_5 and La_2O_3 . Secondary competition is similarly quite high ($C_2 = 186$ meV per atom) since it is highly favorable for the target to react with remaining precursors to form LaNbO_4 and others on the La_2O_3 -rich side of the hull. Meanwhile, the reaction energy to form LaGaO_3 is comparable to competing targets, resulting in much lower primary competition ($C_1 = 57$ meV per atom). The driving forces associated with LaGaO_3 reactions with either precursor are also lower than in the La–Nb–O case, leading to lower secondary competition ($C_2 = 55$ meV per atom). Despite the similar E_{hull} for the two targets, the choice of precursors and distribution of competing targets in each chemical space leads to distinct Γ , demonstrating that Γ offers additional insight beyond E_{hull} when screening for synthesizability. Note also in Fig. 1b that the two targets are different distances from the interface hull at their reported synthesis temperatures. Although they have very similar E_{hull} at 0 K, LaNb_3O_9 stabilizes and LaGaO_3 destabilizes with increasing temperature. This behavior further exemplifies the distinction between stability and selectivity. LaNb_3O_9 is much closer to the hull at its listed reaction temperature than LaGaO_3 , but is still a less favorable target from a selectivity perspective due to the presence of more thermodynamically competitive compounds in the La–Nb–O chemical space.

The reaction-centric Γ_{obs} for each text-mined entry is also compared to the material-centric Γ_{opt} associated with the listed target to validate the efficacy of Γ_{opt} as a proxy for Γ_{obs} . This becomes important for hypothetical materials where Γ_{obs} is unknown (as no synthesis has been reported). A hexbin parity plot comparing Γ_{obs} to Γ_{opt} for the 2495 listed reactions (spanning 938 unique targets) is shown in Fig. 2a. The highest density of points lies near the parity line, indicating that observed (successful) reactions often approach the optimal

Table 1 Distinguishing features of the four ML synthesizability classifiers evaluated in this work. MP refers to the Materials Project⁶ and OQMD refers to the Open Quantum Materials Database.⁷ SVM refers to support vector machine

Model	Material encoding	Source of positive data	Source of unlabeled data	Labeling method
PU-CGNF	Composition	MP and OQMD known materials	MP and OQMD hypothetical materials	Transductive bagging SVM
SynthNN	Composition	ICSD	Artificially generated chemical formulae	Probabilistic weighted labeling
PU-CGCNN	Structure	MP known materials	MP hypothetical materials	Transductive bagging SVM
SynCoTrain	Structure	MP known oxide materials	MP hypothetical oxide materials	Co-training with alternating iterations of bagging SVMs

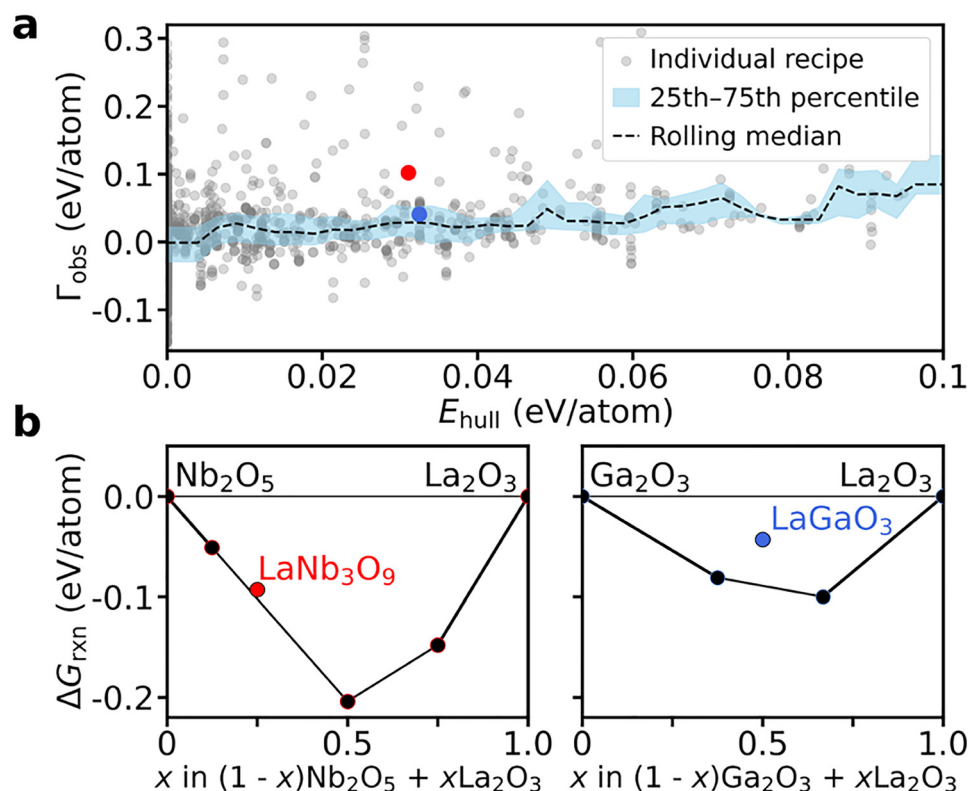


Fig. 1 (a) Γ_{obs} vs. target E_{hull} for each reaction listed in the text-mined dataset (grey points) and a rolling median of Γ_{obs} as a function of target E_{hull} , with the median evaluated over a range of 8 meV per atom every 2.5 meV per atom. The distance from the median to the 25th and 75th percentile values of Γ_{obs} is shaded in light blue. The examples shown in (b) are indicated with red and blue points. Γ_{obs} was calculated at the reported synthesis temperature. E_{hull} was calculated at 0 K. (b) Two interface reaction hulls depicting the relative driving forces to form different products from a given precursor pair. The left-hand hull represents the reaction between Nb_2O_5 and La_2O_3 targeting LaNb_3O_9 , and the right-hand hull represents the reaction between Ga_2O_3 and La_2O_3 targeting LaGaO_3 . These interface hulls are presented at the observed synthesis temperatures of 1273 K and 723 K for LaNb_3O_9 ³⁴ and LaGaO_3 ,³⁵ respectively.

(minimum) selectivity. In fact, 41% of the experimentally reported reactions in the text-mined dataset use the exact precursors that yield Γ_{opt} , and 63% of the experimentally reported reactions have Γ_{obs} within 20 meV per atom of Γ_{opt} . These results indicate that successful synthesis routes are typically associated

with selectivities close to the optimum, supporting the use of Γ_{opt} as a practical thermodynamic heuristic for quantifying the synthesizability of hypothetical materials.

In Fig. 2b, we show cumulative distribution functions for both observed reactions (Γ_{obs}) and their corresponding optimal values

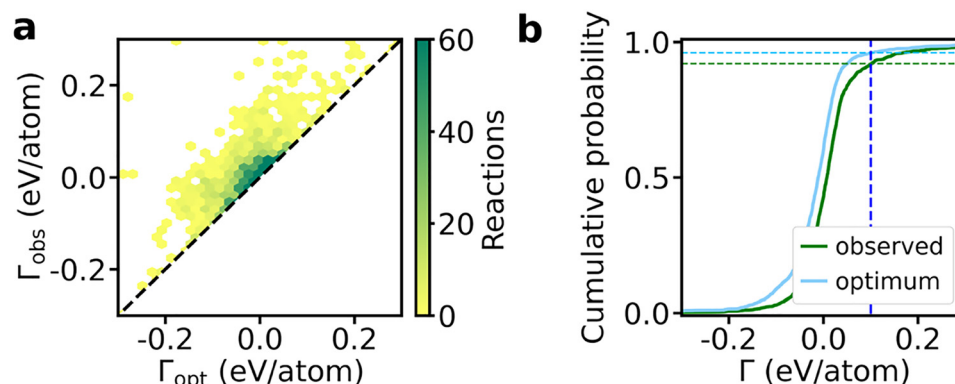


Fig. 2 (a) A pairwise comparison of Γ for 2495 observed reactions²⁷ (Γ_{obs}) with the optimum value (Γ_{opt}) that is possible from known precursors to make the same target. Bins are colored by the number of reactions falling within each region. (b) Cumulative distributions of Γ_{obs} (green) and Γ_{opt} (light blue) for text-mined reactions. The vertical blue dashed line indicates $\Gamma = 100$ meV per atom, corresponding to the 92nd percentile of Γ_{obs} (green dashed horizontal line) and the 96th percentile of Γ_{opt} (light blue dashed horizontal line). Tabulated values corresponding to these plots are available in Table S1.

(Γ_{opt}), showing the fraction of reactions (Γ_{obs}) and targets (Γ_{opt}) with Γ below a given threshold. The closeness of the two distributions further underscores the similarity of Γ_{obs} and Γ_{opt} . Examining the distribution for Γ_{obs} , 92% of observed (successful) reactions have $\Gamma_{\text{obs}} < 100$ meV per atom, as demarcated by the dashed vertical line. Materials with $\Gamma_{\text{opt}} > 100$ meV per atom are therefore considered unlikely to be synthesizable, since even their most selective pathways fall outside the bounds of the vast majority of this set of reactions. Applying similar logic to E_{hull} , the aforementioned study by Sun *et al.* found that $\sim 95\%$ of materials listed in the ICSD have $E_{\text{hull}} < 67$ meV per atom.¹⁷ Allowing for some outliers, materials with $E_{\text{hull}} > 100$ meV per atom are similarly considered unlikely to be synthesizable. These results suggest that hypothetical materials having either $\Gamma_{\text{opt}} > 100$ meV per atom or $E_{\text{hull}} > 100$ meV per atom are unlikely to be synthesized, rendering these useful metrics for assessing ML models in the absence of explicit negative examples. While these proposed bounds are informed by a statistical analysis of known materials and their synthesis recipes, they should not be interpreted as strict cutoffs. It is likely that some materials within these bounds ($\Gamma_{\text{opt}} < 100$ meV per atom and $E_{\text{hull}} < 100$ meV per atom) cannot be synthesized and some materials outside these bounds ($\Gamma_{\text{opt}} > 100$ meV per atom or $E_{\text{hull}} > 100$ meV per atom) can be synthesized. However, given the analysis of known materials, we should expect the likelihood of synthesis to decrease as these thermodynamic heuristics increase and approach zero as these thermodynamic heuristics increase significantly beyond these bounds (*e.g.*, E_{hull} or $\Gamma_{\text{opt}} \gg 0.1$ eV per atom).

Data-driven *versus* thermodynamic predictions on hypothetical materials

Using the bounds on Γ_{opt} established with the text-mined synthesis recipes, we sought to assess ML synthesizability predictors on a new dataset of hypothetical materials. This dataset is comprised of 2673 hypothetical materials generated using the generative AI model, Chemeleon.³⁶ These materials span the same ternary

oxide chemical spaces included in the text-mined dataset but do not appear in the ICSD or MP, meaning they were unlikely to have been used during the training of any models studied here. Further details on the generation procedure are provided in Methods. Each generated structure was relaxed using CHGNet,¹¹ and the resulting energies were used to estimate E_{hull} and Γ_{opt} in the same manner as for the text-mined data (leveraging MP data for other materials in each relevant chemical space).

The resulting E_{hull} distribution for the 2673 generated materials is shown in Fig. 3a. The majority of generated materials ($\sim 60\%$) have $E_{\text{hull}} > 100$ meV per atom, meaning they are unlikely to be synthesizable based on a stability assessment. For the remaining 1057 materials that are stable or plausibly metastable ($E_{\text{hull}} < 100$ meV per atom), we show the distribution of Γ_{opt} in Fig. 3b. Even for these materials that cannot be ruled out from a stability perspective, $\sim 60\%$ have thermodynamic selectivities (Γ_{opt}) greater than 100 meV per atom, falling outside the range where successful synthesis is expected to be likely. This dataset presents a wide spectrum of new hypothetical materials where synthesis appears plausible (low E_{hull} and low Γ_{opt}), and importantly, many examples where ML models should assign low likelihood scores (high E_{hull} or high Γ_{opt}). It should be noted that previous work found CHGNet to slightly overestimate the stability (underestimate E_{hull}) of materials generated by a variety of generative models, suggesting the thermodynamic predictions reported here may themselves overestimate the synthesizability of these hypothetical materials.³⁷

The four aforementioned synthesizability predictors (Table 1) were applied to these newly generated materials. In Table 2, we list the number of hypothetical materials predicted synthesizable by each model and the number of materials within specified thermodynamic bounds. While 40% of materials meet our bound on thermodynamic stability ($E_{\text{hull}} < 100$ meV per atom), only 16% of materials fall within both proposed thermodynamic bounds ($E_{\text{hull}} < 100$ meV per atom and $\Gamma_{\text{opt}} < 100$ meV per atom), indicating the value of thermodynamic selectivity metrics

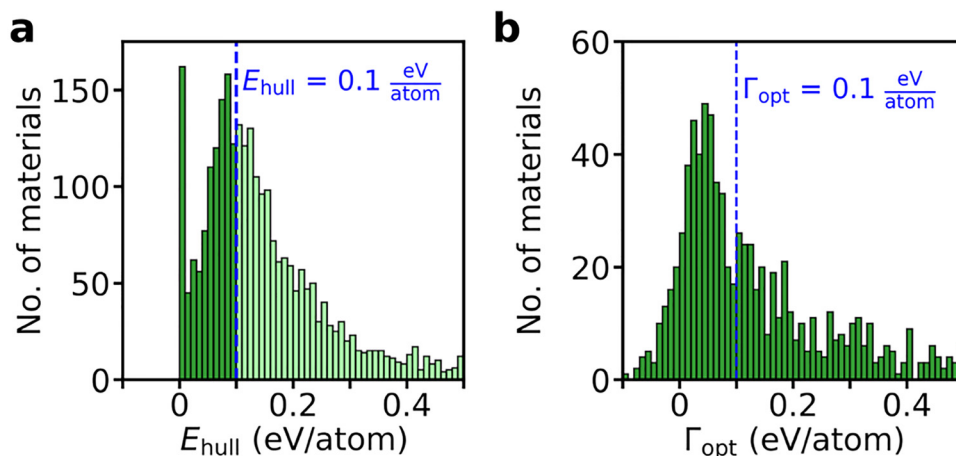


Fig. 3 Distributions of E_{hull} (a) and Γ_{opt} (b) for Chemeleon-generated materials, with proposed bounds on thermodynamic feasibility ($E_{\text{hull}} < 100$ meV per atom and $\Gamma_{\text{opt}} < 100$ meV per atom), demarcated by the blue dashed lines. The Γ_{opt} distribution reflects only materials within the dark green bars in (a) – those with $E_{\text{hull}} < 100$ meV per atom. The energies of generated materials were computed with CHGNet.¹¹ The calculation of E_{hull} and Γ_{opt} from these energies are detailed further in Methods.

Table 2 Number and fraction of Chameleon-generated hypothetical materials predicted to be synthesizable by four ML classifiers, along with subsets of materials that fall within realistic thermodynamic bounds

	# of materials	% of materials
Total hypothetical materials considered	2673	N/A
Predicted synthesizable (PU-CGNF)	2050	77%
Predicted synthesizable (SynthNN)	377	14%
Predicted synthesizable (PU-CGCNN)	1326	50%
Predicted synthesizable (SynCoTrain)	1929	72%
$E_{\text{hull}} < 100$ meV per atom	1057	40%
$E_{\text{hull}} < 100$ meV per atom and $\Gamma_{\text{opt}} < 100$ meV per atom	427	16%

in decreasing the breadth of hypothetical materials that should be targeted for synthesis. All ML models except SynthNN predict at least 50% of generated materials to be synthesizable, indicating a propensity for overprediction. It is notable that the fraction of materials predicted synthesizable by SynthNN (14%) is similar

to the fraction that look promising based on thermodynamic heuristics (16%).

Beyond aggregated statistics, synthesizability scores predicted by each model for each material were compared to their thermodynamic heuristics, E_{hull} and Γ_{opt} . In Fig. 4, we show heatmap distributions of synthesizability scores as a function of E_{hull} for each model. Dark green regions indicate a high density of materials with a given E_{hull} and synthesizability score. Overlain on each heatmap is a moving average of the fraction of materials predicted synthesizable as a function of E_{hull} (indicated on the left-hand y-axis).

Stable ($E_{\text{hull}} = 0$) and weakly unstable materials ($E_{\text{hull}} < 100$ meV per atom) are considered here to be feasible candidates for synthesis, given the distribution of E_{hull} determined by Sun *et al.*¹⁷ for known materials. In contrast, high- E_{hull} materials are unlikely to be synthesized by traditional approaches such as solid-state synthesis. A predictor that successfully aligns with thermodynamic trends should therefore yield a relatively high

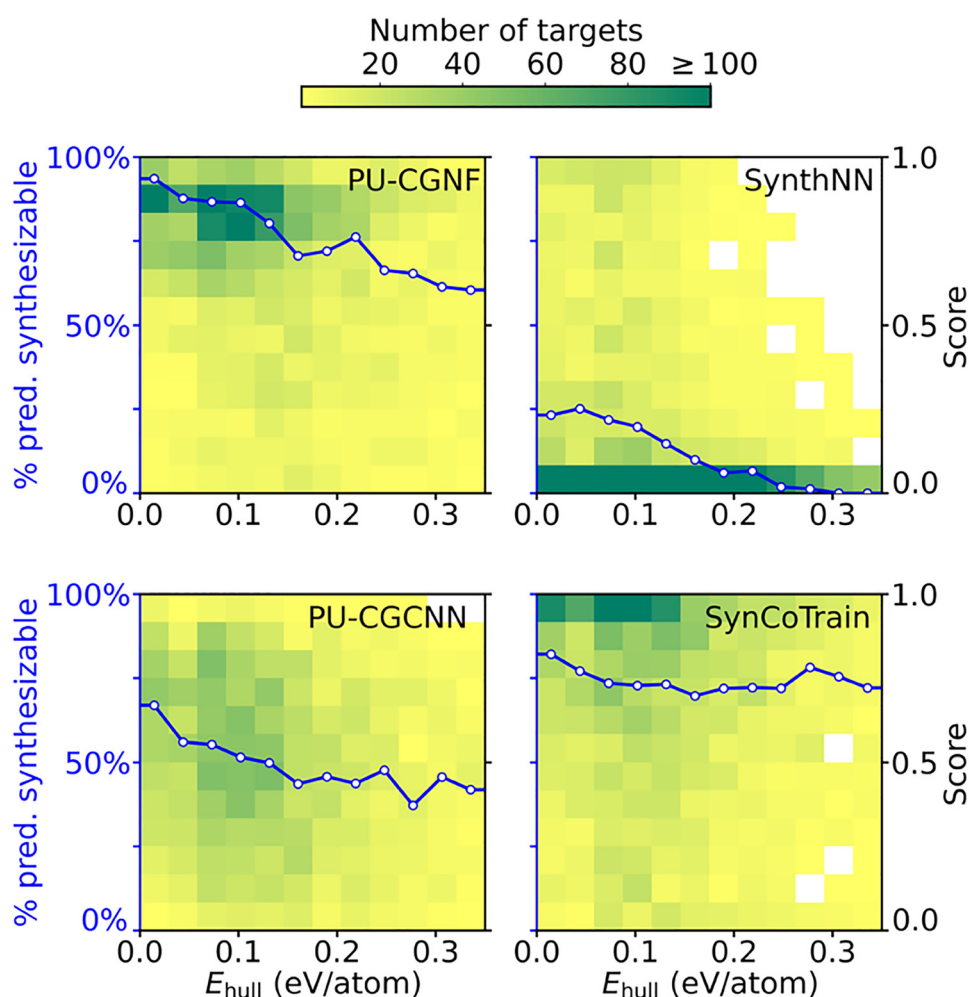


Fig. 4 Model predictions and CHGNet-calculated E_{hull} for 2673 Chameleon-generated materials. The heatmap color indicates the number of these target materials having a certain synthesizability score (right-hand y-axis) and E_{hull} (x-axis). A moving average of the fraction of materials predicted synthesizable (score > 0.5) as a function of E_{hull} is shown by the blue line and corresponding left-hand y-axis. The moving average was computed over 30 meV per atom intervals. White squares indicate no targets falling in a given region. The model name is given in the top right of each subplot. Tabulated values corresponding to these plots are available in Table S2.

density of scores > 0.5 for materials with $E_{\text{hull}} \ll 100$ meV per atom, and progressively lower scores for materials far above the hull. From Fig. 4, it is clear that each of the models predicts stable materials to be synthesizable with the greatest frequency, as expected. However, even for stable materials ($E_{\text{hull}} = 0$), there is a wide range of outcomes depending on the model. For example, of the 136 stable materials, SynthNN predicts 25% to be synthesizable whereas PU-CGNF predicts 93% to be synthesizable. As materials get less stable (larger E_{hull}), models should predict exponentially fewer can be synthesized. While there is a general decreasing trend in the synthesizability scores from PU-CGNF, SynthNN, and PU-CGCNN, SynCoTrain predictions remain relatively constant regardless of E_{hull} . As E_{hull} increases beyond 100 meV per atom, we would expect all models to rapidly decrease in predicted synthesizability score, but only SynthNN predictions go to zero even for the most unstable materials ($E_{\text{hull}} \gg 100$ meV per atom). Among the 679 generated materials with $E_{\text{hull}} > 200$ meV per atom, PU-CGCNN, PU-CGNF, and

SynCoTrain predict 39%, 56%, and 66% to be synthesizable, respectively, but the synthesis of any of these highly unstable materials would be exceptional.

Among generated materials with $E_{\text{hull}} < 100$ meV per atom, synthesis is plausible but may not be straightforward. Here, Γ_{opt} is used as an additional thermodynamic heuristic to probe the synthesizability of materials in the low- E_{hull} regime. In Fig. 5, we show heatmap distributions of synthesizability score as a function of Γ_{opt} for each model. Only the subset of materials with $E_{\text{hull}} < 100$ meV per atom are considered here. Dark green regions indicate a high density of materials with a given Γ_{opt} and synthesizability score. Overlain on the heatmaps are moving averages of the fraction of materials predicted synthesizable (score > 0.5) by each model as a function of Γ_{opt} .

Even for materials with low E_{hull} , the likelihood of synthesis should decrease with increasing Γ_{opt} since a high (positive) Γ_{opt} indicates an absence of selective pathways to the target in question. Recall that just 40% of generated materials with

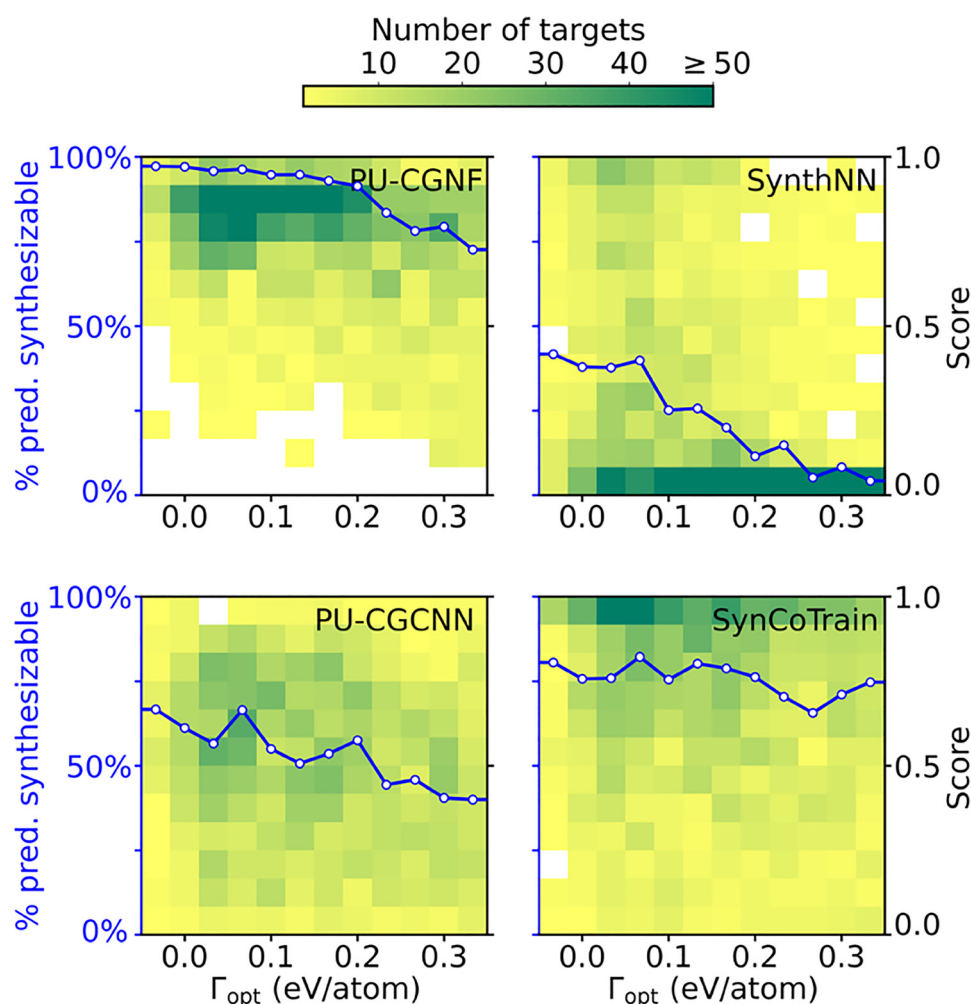


Fig. 5 Model predictions and CHGNet-calculated Γ_{opt} for 1057 Chemeleon-generated materials having $E_{\text{hull}} < 100$ meV per atom. The heatmap color indicates the number of these target materials having a certain synthesizability score (right-hand y-axis) and Γ_{opt} (x-axis). A moving average of the fraction of materials predicted synthesizable (score > 0.5) as a function of Γ_{opt} is shown by the blue line and corresponding left-hand y-axis. The moving average was computed over 33 meV per atom intervals. White squares indicate no targets falling in a given region. The model name is given in the top right of each subplot. Tabulated values corresponding to these plots are available in Table S2.

$E_{\text{hull}} < 100$ meV per atom have associated $\Gamma_{\text{opt}} < 100$ meV per atom. However, PU-CGNF, PU-CGCNN, and SynCoTrain predict 89%, 58%, and 77% of this low- E_{hull} subset to be synthesizable, indicating that synthesizability is overestimated by all models other than SynthNN (23%) with respect to Γ_{opt} as well as E_{hull} . Remarkably, as with E_{hull} , SynthNN shows a marked decrease in synthesizability score as Γ_{opt} increases. While 23% of $E_{\text{hull}} < 100$ meV per atom materials are predicted synthesizable by SynthNN, this fraction increases to 40% if $\Gamma_{\text{opt}} < 100$ meV per atom and decreases to 18% if $\Gamma_{\text{opt}} > 100$ meV per atom. This shift cannot be explained solely by variation in E_{hull} within the smaller range $E_{\text{hull}} < 100$ meV per atom as Γ_{opt} and E_{hull} are only weakly correlated over this subset of generated materials (Fig. S1). In contrast, PU-CGNF, PU-CGCNN, and SynCoTrain exhibit weak downward trends with increasing Γ_{opt} or general insensitivity to Γ_{opt} for materials in this range of E_{hull} .

Comparison among ML models

The synthesizability predictors were also examined relative to one another to understand whether they are learning similar underlying information about the generated materials. The histograms in Fig. 6a show the distribution of synthesizability scores associated with the generated materials for each model. As indicated by the summary statistics in Table 2, SynthNN is heavily skewed toward classifying materials as unsynthesizable, while PU-CGNF and SynCoTrain are inclined to classify materials as synthesizable. PU-CGCNN predictions are distributed more evenly, with an inclination towards low-certainty scores (~ 0.5). While the “true” distribution is unknown since we considered only hypothetical (not-yet-synthesized) materials, SynthNN appears most true to the thermodynamic distributions shown in Fig. 3. However, SynthNN predicts substantially fewer materials to be synthesizable than the number that do

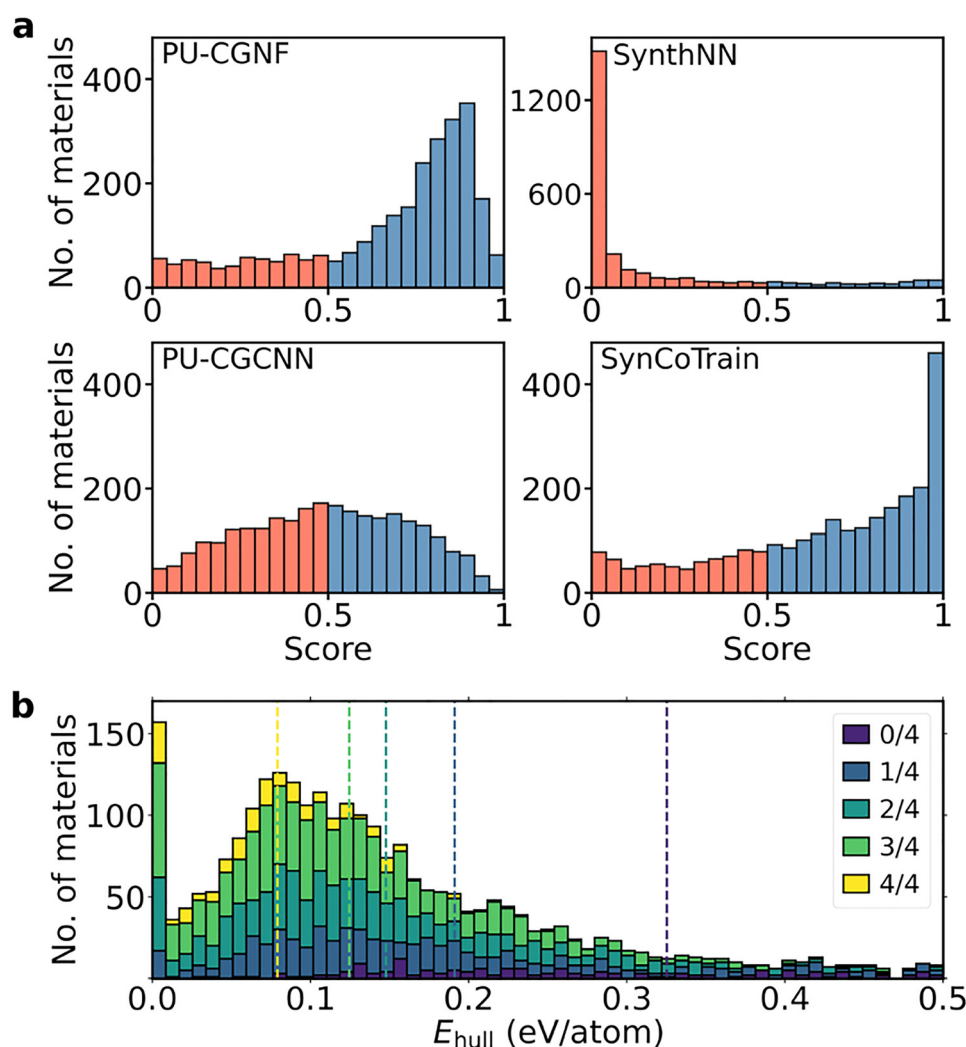


Fig. 6 (a) Distributions of predicted synthesizability scores from each model for the 2673 Chemeleon-generated materials. Scores corresponding to a synthesizable label are colored in blue (> 0.5) while those corresponding to an unsynthesizable label (< 0.5) are colored in red. Note that the y-axis of SynthNN (upper right) is plotted on a different scale than the other models due to the dramatic skew of scores toward low values. (b) A stacked histogram of materials predicted synthesizable by a specified fraction of the models as a function of E_{hull} . For example, “4/4” indicates that all four models (PU-CGNF, SynthNN, PU-CGCNN, and SynCoTrain) predict a material to be synthesizable. The dashed lines in the corresponding colors indicate the mean E_{hull} of materials predicted synthesizable by that fraction of models.

reside within the bounds, suggesting it may be overly pessimistic in some cases.

In Fig. 6b, model agreement is assessed by indicating how many models predict each material to be synthesizable as a function of E_{hull} . Within a given range of E_{hull} , the number of materials classified synthesizable by a specified fraction of the models (from 0/4 to 4/4) is indicated by the height of the bar with the associated color. The distributions corresponding to materials classified as synthesizable by 0/4 models and 4/4 models (purple and yellow bars, respectively) indicate universal agreement among models. These are the least prominent distributions with just 7% of materials being predicted synthesizable and 7% predicted unsynthesizable by all four models. The other distributions (1/4, 2/4, 3/4) all indicate disagreement by at least one model. The most common occurrence, indicated by the bars with the greatest height across E_{hull} , is of materials predicted synthesizable by 3/4 models, with SynthNN being the lone model classifying the material as unsynthesizable in 87% of these instances. Materials predicted synthesizable by all four models have the lowest mean hull energy (79 meV per atom), as indicated by the yellow dashed line. This reflects the propensity of SynthNN to classify only lower E_{hull} materials as synthesizable as well as the decreasing number of materials classified as synthesizable by PU-CGNF and PU-CGCNN with increasing E_{hull} (Fig. 4). Meanwhile, PU-CGNF and SynCoTrain together comprise 90% of cases where a material is predicted synthesizable by only one model. However, the mean E_{hull} is 70 meV per atom higher for materials predicted synthesizable only by SynCoTrain than for materials predicted synthesizable only by PU-CGNF. This reflects the propensity of PU-CGNF to classify fewer materials as synthesizable with increasing E_{hull} , while SynCoTrain scores do not vary substantially with E_{hull} .

While these distributions paint an aggregated picture, comparing the models on a target basis shows that model scores vary substantially for individual materials. A pairwise plot of scores for each pair of models is shown in Fig. S2. The greatest correlation exists among PU-CGNF and SynCoTrain, both having a general propensity for positive classification. The general lack of correlation between these models indicates significant sensitivity with respect to model choices such as training data, labeling strategy, and architecture. If each model's predictions are reliable, it is possible that ensembling may yield complementary information. However, the general disagreement between model predictions and thermodynamic trends does not support this being a viable strategy with current models.

Discussion

ML models have become reliable tools for many materials science problems, including thermodynamic stability prediction.³⁸ However, the task of predictive synthesis is more challenging due to the complexity of phase transformations and relative lack of available data for training. Among other challenges, the efficacy of these models cannot be fully quantified in the absence of known unsynthesizable materials (negative examples). In this

work, we bypass this data shortage by using thermodynamic heuristics (E_{hull} , Γ_{opt}) to identify which materials are more or less likely to be synthesizable. Sun *et al.* reported 95% of synthesized materials in the ICSD exhibit $E_{\text{hull}} \leq 67$ meV per atom.¹⁷ In this work, we find that 96% of studied reactions exhibit $\Gamma_{\text{opt}} \leq 100$ meV per atom. Although particular values of these thermodynamic quantities cannot guarantee (un)synthesizability, they reflect the probability distributions associated with the known materials and syntheses in this analysis. As either E_{hull} or Γ_{opt} increase, these findings suggest that synthesis should become less likely.

Four ML models for synthesizability prediction – PU-CGNF,³⁰ SynthNN,³¹ PU-CGCNN,²⁹ and SynCoTrain³² – were evaluated against these computed thermodynamic heuristics. Three of the four models systematically overestimate synthesizability compared with the bounds on thermodynamic stability and selectivity established on known materials. SynthNN was the only model with drastically decreasing synthesizability scores as E_{hull} increases beyond 100 meV per atom (Fig. 4). For materials with plausible thermodynamic stability ($E_{\text{hull}} < 100$ meV per atom), thermodynamic selectivity as assessed by Γ_{opt} was shown to further distinguish between likely and unlikely candidates for synthesis. Again, SynthNN was the only model to display a strong trend towards unsynthesizable predictions as Γ_{opt} increased (Fig. 5).

The relative success of SynthNN might be explained by the choice of unlabeled materials during training. While all other models use hypothetical materials from materials databases (MP or OQMD) as unlabeled materials, SynthNN generates entirely new hypothetical materials (chemical formulas) spanning a similar distribution to the known materials in the ICSD. While the provenance of hypothetical materials in materials databases varies, these materials may exhibit greater similarity to known materials than the artificially generated formulas in the SynthNN work. Models trained on materials databases may therefore be biased towards classifying more materials as synthesizable. SynthNN may also benefit from a much larger quantity of unlabeled data relative to the other models since the unlabeled data is not sourced from existing databases. Our results indicate that SynthNN exhibits unique behavior among the four models studied, and future work could isolate the effect of training data on this finding.

In addition, PU-CGNF, PU-CGCNN, and SynCoTrain employ hard labeling, which treats borderline-class materials (scores ≈ 0.5) the same as materials with very high (≈ 1) and very low (≈ 0) scores. In contrast, SynthNN privileges materials with scores close to 0 or 1 while down-weighting more ambiguous unlabeled materials in training. This labeling strategy targets precision (the fraction of labeled positives that are truly positive) rather than recall (the fraction of positives correctly identified as positive). As a result, positive labels are assigned more conservatively, resulting in generally lower positive class probabilities. Exposure to more unrealistic materials along with confidence-based labeling may be responsible for better alignment between SynthNN predictions and computed thermodynamic heuristics.

The models considered here are distinguished not only by training data and labeling strategy, but also by material representation. To determine the effect of material representation on model performance, two structural models, PU-CGCNN and SynCoTrain, were compared to two compositional models, SynthNN and PU-CGNF. Prior deep learning models have found substantial improvement with structural representations over compositional representations, which is not surprising given the ubiquity of structure–property relations in materials science.¹⁰ However, the results in this work indicate greater ambiguity for the problem of synthesizability prediction as a compositional model (SynthNN) exhibits the greatest alignment with thermodynamic heuristics. Among the other models, it is difficult to discern a notable benefit (or detriment) of including structure in the representation. A third structural model, TSDNN³³ was examined for further insight on representation effects. TSDNN predicted 98% of the generated materials to be synthesizable, with little variation in score as a function of E_{hull} or Γ_{opt} , leading to the likely erroneous prediction of many thermodynamically unrealistic materials as synthesizable (see Fig. S3). The limited performance gains when including structure in the representation stand in stark contrast to ML models that predict other material properties (e.g., formation energy) and may arise from the relatively smaller datasets available for synthesizability prediction or the disconnect between structural features and synthesis.

The general challenge of predicting synthesizability also stems from the inherently non-intrinsic nature of “synthesis”. Throughout this work, synthesizability is framed as a material property with materials being labeled as either synthesizable or unsynthesizable. In reality, whether or not a material can be synthesized depends critically on the synthesis recipe (precursors, temperature, approach, *etc.*).^{15,19,39} From an ML perspective, this presents a challenge compared with the prediction of calculable, intrinsic material properties (formation energy, band gap, *etc.*). Thermodynamic heuristics that link materials to the plausibility of synthesis are useful in this regard, with E_{hull} being the most widely applied and well understood. Enumerating reaction networks and computing thermodynamic selectivities associated with synthesis recipes adds complementary information to E_{hull} by exploring the landscape of possible synthetic routes to access stable or weakly unstable materials.

While useful, these thermodynamic heuristics are also limited predictors of synthesizability. E_{hull} generally describes the 0 K bulk energetics of pristine materials at equilibrium. In contrast, materials are synthesized at finite temperatures, likely contain defects, and may not reach equilibrium. While Γ_{opt} accounts for some aspects of possible synthesis recipes, this metric is also limited to bulk thermodynamics, neglecting relevant kinetic processes like diffusion and nucleation as well as the thermodynamics of interfaces (where reactions begin). In addition, both Γ_{opt} and E_{hull} are limited to the completeness of computed databases. This could lead to an underestimation of both Γ_{opt} and E_{hull} if important competing phases are not yet tabulated in these databases. Alternatively, Γ_{opt} may be overestimated if certain precursors that are not yet tabulated could lead to thermodynamically selective routes to a given target. In addition, the existing implementation of Γ_{opt} is

limited to synthesis recipes with just two solid precursors and no solid byproducts, but more creative recipes (e.g., metathesis) may improve the selectivity.⁴⁰ Further, the methods employed in this work focus on the internal energy (as computed using DFT or CHGNet) and a model for the vibrational entropy of solids,⁴¹ so other contributions to the Gibbs energy (e.g., configurational entropy) are neglected. For the targeted synthesis of materials where other effects are important and calculable with first-principles methods (e.g., configurational entropy of high-entropy materials), these quantities can be included in the expressions for the Gibbs energy of each relevant compound before computing E_{hull} or Γ_{opt} . E_{hull} and Γ_{opt} as presented here therefore function best as preliminary and general screening tools and do not definitively indicate that a material is or is not synthesizable. The synthetic approach is also an important consideration as materials that cannot be made by solid-state synthesis may be synthesizable through other routes (e.g., solution-based synthesis).^{28,42} As the thermodynamic and kinetic factors dictating reaction pathways differ as the synthetic approach is varied, it is important to consider how the chosen thermodynamic heuristics may or may not apply across synthetic approaches.⁴³

Given the multi-faceted challenge of making new materials, ML synthesizability predictors present an appealing tool for candidate screening as predictions could be made on thousands of materials nearly instantly once these models are trained. Ideally, these models should reflect the kinetic and thermodynamic interplay underpinning synthesis. At a minimum, they should reflect our current understanding of the relationship between synthetic plausibility and thermodynamic quantities. While thermodynamic bounds are not a guarantee of (un)synthesizability, they serve as a strong indicator of likely or unlikely candidates for synthesis. They are leveraged here to identify qualitative and quantitative trends with ML synthesizability predictors as a means of evaluating the performance of these models. We find that the ML models in question produce scores that often overestimate the likelihood that materials can be synthesized compared to established thermodynamic heuristics. Moving forward, we speculate that directly incorporating thermodynamic information into the training and validation process has the potential to improve the ability for models to capture the underlying thermodynamics. The relative success of SynthNN in this regard indicates that training on a wide variety of unlabeled materials and a soft labeling approach may be effective strategies to build on for synthesizability prediction. Ultimately, the efficacy of any synthesis prediction model needs to be validated with experimental synthesis attempts. Currently, synthesizability prediction models likely do not exceed the utility of computable thermodynamic heuristics. While synthesizability predictions are nearly free from a computational standpoint (once models are trained), the time required to perform first-principles estimates of stability and selectivity (~hours per material) remains far lower than the time required to perform experiments (~days per material). Replacing DFT with foundation potentials further reduces the time required to compute thermodynamic heuristics to ~minutes per material.

As generative models rapidly expand the space of proposed materials and large-scale libraries of hypothetical compounds continue to grow, the need for accurate and efficient synthesis prediction has never been greater. Along with calculable thermodynamic quantities, the increasing prevalence of self-driving labs for inorganic synthesis presents an opportunity to systematically validate these models by generating a significant number of both positive and negative examples.⁴⁴ Continued efforts to carefully collect and curate synthesis information from the literature will also be paramount to the development of improved models for synthesis.

Methods

Restructuring text-mined data

We examined a subset of the text-mined dataset of solid-state inorganic reactions constructed by Kononova *et al.* to calibrate Γ on successful reactions.²⁷ The dataset itself required some restructuring and additional computation to this end. We filtered for reactions targeting ternary oxides in MP taking place in an air environment. To simplify the calculation of Γ_{opt} , we considered only reactions having no more than two solid precursors (with all precursors listed in MP) and only O₂, CO₂, and H₂O as allowed byproducts or gaseous precursors. If no environment was given, the environment was assumed to be air. The synthesis temperature was identified as the maximum temperature used in a heating or sintering step. If no temperature was given, the temperature was assumed to be 1073 K.

We used pre-calculated DFT energies (GGA/GGA+*U*)⁴⁵ from MP for the enthalpies of formation and E_{hull} , assuming that the composition listed as a precursor or target in the text-mined dataset was in the lowest energy structure for that composition in MP. We used the model proposed by Bartel *et al.* to determine the Gibbs energy of solid phases at the listed reaction temperatures.⁴¹ NIST data was used for the participating gasses, O₂, CO₂, and H₂O.⁴⁶

Calculating thermodynamic selectivity

We implemented a modified version of the reaction-network package,^{25,26} available at <https://github.com/materialsproject/reaction-network>. Our own reaction generation workflow takes in as arguments target material and temperature and requires three steps. First, a reaction network is generated at 300 K for all materials in the target's chemical space that are in MP and labeled as experimentally observed, specifying the set of allowed precursors as all precursors used in the filtered text-mined reactions. The thermodynamics associated with reactions in the network are then updated based on the synthesis temperature of interest. Finally, target-forming reactions in the network are identified as all reactions that lead to the desired target without solid byproducts. The resulting reaction networks were used to compute ΔG_{rxn} , C_1 , and C_2 as described in ref. 25. These quantities were combined into the aggregate quantity Γ according to eqn (1).

The weights (w_0 , w_1 , w_2) in this expression were chosen arbitrarily in ref. 25. Under the assumption that the optimum (lowest) Γ should correspond with the most likely reaction, we re-optimized the weights w_0 , w_1 , and w_2 by minimizing the L₂ loss between the lowest Γ (Γ_{opt}) and the Γ associated with the observed reaction (Γ_{obs}). The sequential least-squares programming (SLSQP) algorithm, as implemented in *scipy*,⁴⁷ was applied to this end, with multiple initializations. Constraints requiring that the three weights sum to 1 and that each lies within [0, 0.99] were imposed, as prescribed by ref. 25. The set of weights yielding the lowest objective value was selected, resulting in the updated functional form of Γ given by eqn (2).

$$\Gamma = 0.16\Delta G_{\text{rxn}} + 0.56C_1 + 0.28C_2 \quad (2)$$

As this reweighting approach remains somewhat arbitrary, other sets of weights were considered to ensure that the proposed bounds are not highly sensitive to the weighting choice. In Fig. S4, the cumulative distributions of Γ_{obs} and Γ_{opt} are shown for three cases: (a) the weights proposed by McDermott *et al.* ($w_0 = 0.1$, $w_1 = 0.45$, $w_2 = 0.45$), (b) those associated with equal weighting ($w_0 = w_1 = w_2 = 0.33$), and (c) those used in this work ($w_0 = 0.16$, $w_1 = 0.56$, $w_2 = 0.28$). Regardless of weight choice, 90% or more of the text-mined reactions exhibit $\Gamma < 0.1$ eV per atom.

For novel materials with unknown synthesis recipes, we generated reactions at 600, 900, 1200, 1500, and 1800 K and chose the “optimum” reaction as that with the minimum Γ across all temperatures.

We sought to further corroborate two approximations in this workflow on a small subset of the generated materials: (1) the consideration of only text-mined precursors in reaction enumeration and (2) the consideration of only the aforementioned temperatures. We validated (1) by assessing Γ_{opt} with an expanded precursor set for the 44 generated targets in the Ba-Cu-O chemical space. In addition to text-mined precursors, we allowed for experimentally realized materials listed in the Materials Project with $E_{\text{hull}} < 50$ meV per atom. This expanded our allowed precursor set from 14 to 18 materials while still ensuring that the chosen set of precursors were plausibly accessible materials rather than materials where a new synthesis recipe would need to be developed to obtain them. In Fig. S5a, we compare the Γ_{opt} associated with the expanded precursor set vs. that associated with the text-mined precursor set. The addition of new precursors does not change Γ_{opt} for 23 of 44 targets. For the remaining targets, Γ_{opt} changes negligibly for most cases, and importantly never decreases by more than the proposed 100 meV per atom bound. We validated (2) by assessing Γ_{opt} for the same 44 targets in the Ba-Cu-O chemical space using a finer discretization of 100 K. In Fig. S5b, we show a parity plot of Γ_{opt} using each of the two temperature intervals and find that discretizing every 300 K gives nearly identical results.

Novel materials generation

Novel materials were generated using the generative model Chemeleon as detailed at <https://github.com/hspark1212/chemeleon>.³⁶ We considered novel materials in the same chemical

spaces listed in the text-mined dataset so that we could apply the known text-mined precursors for reaction generation. Hundreds of formulas were enumerated for each chemical space, and 50 formulas per chemical space were randomly selected from this larger pool of compositions. Chameleon takes a formula string as input and generates structures accordingly. We relaxed the output structures using pre-trained CHGNet¹¹ version 0.3.0 to a force cutoff of 0.1 eV Å⁻¹. Relaxations were performed using the atomic simulation environment.⁴⁸ The ground-state structure associated with each formula was identified accordingly and used for the calculation of E_{hull} and Γ_{opt} . The chemical spaces with the greatest number of materials having $E_{\text{hull}} < 100$ meV per atom were considered for this examination, resulting in 2673 generated materials spanning 88 ternary oxide chemical systems. Note that while we used CHGNet for novel material energetics, the text-mined material energies were drawn from GGA(+*U*) calculations in MP. Prior work has shown that CHGNet may slightly underestimate E_{hull} when applied in this manner for generated materials.⁴⁹ We employed pymatgen to generate phase diagrams for stability determination.⁵⁰

For validation of thermodynamic quantities derived from CHGNet, 100 generated materials were randomly selected for relaxation with DFT using MP-compatible settings at the GGA(+*U*) level of theory.^{41,47} The 0 K formation energies and temperature-dependent reaction energies were subsequently recomputed. In Fig. S6, we show the performance of CHGNet relative to DFT for these 100 materials, comparing formation energies, E_{f} (a) and reaction energies $\Delta G_{\text{rxn}}(T)$ (b) computed with CHGNet total energies for the target material vs. with DFT total energies for the target material. The corresponding distributions of errors are shown in Fig. S6c and d along with the mean absolute error (MAE) and mean error (ME). The energies from the two sources are highly correlated ($R^2 \geq 0.98$), but CHGNet underpredicts relative to DFT, with mean errors of ~ 25 – 35 meV per atom. Errors of this magnitude do not alter the overall trends reported in this work. Further, the more negative energies predicted by CHGNet suggest that the reported E_{hull} and Γ_{opt} would be slightly larger if recomputed with DFT, exacerbating the discrepancies between most model predictions and these thermodynamic heuristics.

Data-driven synthesizability predictors

For all four ML synthesizability prediction models, we used the pre-trained models provided with each publication. New materials were evaluated *via* a forward pass through the trained models. Consistent with these publications, we considered a classification threshold score of 0.5 for each model, wherein materials assigned scores > 0.5 are classified as synthesizable and those assigned scores < 0.5 are classified as unsynthesizable.

PU-CGNF,³⁰ a compositional model, was trained and validated on previously synthesized (positively labeled) and as-yet-unsynthesized (unlabeled) materials from MP and OQMD. Unlabeled materials were assigned hard labels *via* a transductive bagging support vector machine (SVM). With this method, in each iteration a different subset of the unlabeled materials is

assigned “negative” labels to train an SVM together with the positive-labeled materials, and the SVM is subsequently applied to the rest of the unlabeled materials. SVM predictions from all iterations are aggregated to assign final labels to the entire set of unlabeled materials. A formula string given to PU-CGNF for classification is encoded with Roost⁵¹ and passed through the neural networks associated with each bagging iteration. The scores output from each classifier are averaged, yielding a final score corresponding to the likelihood of falling into the synthesizable class. The implementation for PU-CGNF is further detailed at <https://github.com/kaist-amsg/Synthesizability-stoi-CGNF/>.

SynthNN,³¹ also compositional, was trained on (positive-label) experimentally synthesized materials from the ICSD¹⁶ and unlabeled data generated artificially. During training, unlabeled formulas are first treated as negative examples such that the likelihood of each formula falling into the positive or negative class can be determined. Then, unlabeled materials are duplicated such that each is labeled as both positive and negative. The formula associated with each classification (positive or negative) is weighted during subsequent training by the newly determined likelihood of it falling into that class, such that the higher-likelihood label has a greater influence in training. A formula string given to SynthNN for classification is encoded using the Atom2Vec compositional representation,⁵² wherein composition is represented as a vector of the relative atomic fraction associated with each element type. This material encoding is passed through a deep neural network trained on the positive and newly labeled materials, which yields a score corresponding to the likelihood of falling into the synthesizable class. The implementation for SynthNN is further detailed at <https://github.com/antoniuk1/SynthNN/>.

PU-CGCNN²⁹ employed the same PU-learning approach as PU-CGNF during training, but takes crystal structure as input, rather than composition alone. PU-CGCNN uses a crystal graph convolutional neural network (CGCNN) architecture⁵³ for material encoding and classification. In addition, PU-CGCNN is trained on data from MP but not OQMD, resulting in $\sim 300\,000$ fewer unlabeled materials compared with PU-CGNF. A material, as defined by its composition and structure, given to PU-CGCNN for classification is passed through the CGCNN associated with each bagging iteration. The scores output from each classifier are averaged, yielding a final score corresponding to the likelihood of falling into the synthesizable class. The implementation for PU-CGCNN is further detailed at <https://github.com/kaist-amsg/Synthesizability-PU-CGCNN/>.

SynCoTrain³² is another structure-based model trained on oxide materials from MP. SynCoTrain employs a co-training strategy with two distinct end-to-end material representation and property prediction architectures, ALIGNN⁵⁴ and SchNetPack,⁵⁵ alternating between architectures in each iteration. In each iteration, a bagging SVM is applied to assign labels. Then, newly assigned positive labels are assimilated into the existing positive dataset for subsequent training steps. Ultimately, remaining labels are assigned from the model iteration with the best recall. A material, as defined by its composition and structure, given to SynCoTrain for classification is encoded and classified using a new SchNet model trained on the

positive and newly labeled materials, which yields a score corresponding to the likelihood of falling into the synthesizable class. The implementation for SynCoTrain is further detailed at <https://github.com/BAMeScience/SynCoTrainMP>.

The predictions made by all models on the Chemeleon-generated materials are provided in the supplementary information (SI) along with their computed E_{hull} and Γ_{opt} .

Conflicts of interest

There are no conflicts to declare.

Data availability

The workflow applied in this work, as well as the data used in our analysis, are available at the following GitHub repository: https://github.com/Bartel-Group/synth_assess.

Supplementary information (SI) is available. See DOI: <https://doi.org/10.1039/d6mh00209a>.

Acknowledgements

JS and CJB acknowledge support from National Science Foundation Award No. 2435911 and the University of Minnesota College of Science and Engineering Data Science Initiative. SH acknowledges support from the University of Minnesota Undergraduate Research Opportunities Program (UROP). The authors acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported herein.

References

- 1 A. Jain, Y. Shin and K. A. Persson, Computational Predictions of Energy Materials Using Density Functional Theory, *Nat. Rev. Mater.*, 2016, **1**(1), 15004, DOI: [10.1038/natrevmats.2015.4](https://doi.org/10.1038/natrevmats.2015.4).
- 2 A. R. Oganov, C. J. Pickard, Q. Zhu and R. J. Needs, Structure Prediction Drives Materials Discovery, *Nat. Rev. Mater.*, 2019, **4**(5), 331–348, DOI: [10.1038/s41578-019-0101-8](https://doi.org/10.1038/s41578-019-0101-8).
- 3 A. Zunger, Inverse Design in Search of Materials with Target Functionalities, *Nat. Rev. Chem.*, 2018, **2**(4), 0121, DOI: [10.1038/s41570-018-0121](https://doi.org/10.1038/s41570-018-0121).
- 4 M. K. Horton, P. Huck, R. X. Yang, J. M. Munro, S. Dwaraknath, A. M. Ganose, R. S. Kingsbury, M. Wen, J. X. Shen, T. S. Mathis, A. D. Kaplan, K. Berket, J. Riebesell, J. George, A. S. Rosen, E. W. C. Spotte-Smith, M. J. McDermott, O. A. Cohen, A. Dunn, M. C. Kuner, G.-M. Rignanese, G. Petretto, D. Waroquiers, S. M. Griffin, J. B. Neaton, D. C. Chrzan, M. Asta, G. Hautier, S. Cholia, G. Ceder, S. P. Ong, A. Jain and K. A. Persson, Accelerated Data-Driven Materials Science with the Materials Project, *Nat. Mater.*, 2025, **24**, 1522–1532, DOI: [10.1038/s41563-025-02272-0](https://doi.org/10.1038/s41563-025-02272-0).
- 5 H. Park, Z. Li and A. Walsh, Has Generative Artificial Intelligence Solved Inverse Materials Design, *Matter*, 2024, **7**(7), 2355–2367, DOI: [10.1016/j.matt.2024.05.017](https://doi.org/10.1016/j.matt.2024.05.017).
- 6 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.*, 2013, **1**(1), 011002, DOI: [10.1063/1.4812323](https://doi.org/10.1063/1.4812323).
- 7 S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl and C. Wolverton, The Open Quantum Materials Database (OQMD): Assessing the Accuracy of DFT Formation Energies, *npj Comput. Mater.*, 2015, **1**(1), 15010, DOI: [10.1038/npjcompumats.2015.10](https://doi.org/10.1038/npjcompumats.2015.10).
- 8 M. Scheidgen, L. Himanen, A. N. Ladines, D. Sikter, M. Nakhaee, Á. Fekete, T. Chang, A. Golparvar, J. A. Márquez, S. Brockhauser, S. Brückner, L. M. Ghiringhelli, F. Dietrich, D. Lehmberg, T. Denell, A. Albino, H. Näsström, S. Shabih, F. Dobener, M. Kühbach, R. Mozumder, J. F. Rudzinski, N. Daelman, J. M. Pizarro, M. Kuban, C. Salazar, P. Ondračka, H.-J. Bungartz and C. Draxl, NOMAD: A Distributed Web-Based Platform for Managing materials Science Research Data, *J. Open Source Softw.*, 2023, **8**(90), 5388, DOI: [10.21105/joss.05388](https://doi.org/10.21105/joss.05388).
- 9 C. E. Calderon, J. J. Plata, C. Toher, C. Oses, O. Levy, M. Fornari, A. Natan, M. J. Mehl, G. Hart, M. Buongiorno Nardelli and S. Curtarolo, The AFLOW Standard for High-Throughput Materials Science Calculations, *Comput. Mater. Sci.*, 2015, **108**, 233–238, DOI: [10.1016/j.commatsci.2015.07.019](https://doi.org/10.1016/j.commatsci.2015.07.019).
- 10 C. J. Bartel, A. Trewartha, Q. Wang, A. Dunn, A. Jain and G. Ceder, A Critical Examination of Compound Stability Predictions from Machine-Learned Formation Energies, *npj Comput. Mater.*, 2020, **6**(1), 97, DOI: [10.1038/s41524-020-00362-y](https://doi.org/10.1038/s41524-020-00362-y).
- 11 B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel and G. Ceder, CHGNet as a Pretrained Universal Neural Network Potential for Charge-Informed Atomistic Modeling, *Nat. Mach. Intell.*, 2023, **5**(9), 1031–1041, DOI: [10.1038/s42256-023-00716-3](https://doi.org/10.1038/s42256-023-00716-3).
- 12 B. M. Wood, M. Dzamba, X. Fu, M. Gao, M. Shuaibi, L. Barroso-Luque, K. Abdelmaqsoud, V. Gharakhanyan, J. R. Kitchin, D. S. Levine, K. Michel, A. Sriram, T. Cohen, A. Das, A. Rizvi, S. J. Sahoo, Z. W. Ulissi and C. L. Zitnick, UMA: A Family of Universal Models for Atoms, *arXiv*, 2026, preprint, arXiv:2506.23971, DOI: [10.48550/arXiv.2506.23971](https://doi.org/10.48550/arXiv.2506.23971).
- 13 M. Neumann, J. Gin, B. Rhodes, S. Bennett, Z. Li, H. Choubisa, A. Hussey and J. Godwin, Orb: A Fast, Scalable Neural Network Potential, *arXiv*, 2024, preprint, arXiv:2410.22570, DOI: [10.48550/arXiv.2410.22570](https://doi.org/10.48550/arXiv.2410.22570).
- 14 H. Yang, C. Hu, Y. Zhou, X. Liu, Y. Shi, J. Li, G. Li, Z. Chen, S. Chen, C. Zeni, M. Horton, R. Pinsler, A. Fowler, D. Zügner, T. Xie, J. Smith, L. Sun, Q. Wang, L. Kong, C. Liu, H. Hao and Z. Lu, MatterSim: A Deep Learning Atomistic Model Across Elements, Temperatures and Pressures, *arXiv*, 2024, preprint, arXiv:2405.04967, DOI: [10.48550/arXiv.2405.04967](https://doi.org/10.48550/arXiv.2405.04967).
- 15 C. J. Bartel, Review of Computational Approaches to Predict the Thermodynamic Stability of Inorganic Solids, *J. Mater. Sci.*, 2022, **57**(23), 10475–10498, DOI: [10.1007/s10853-022-06915-4](https://doi.org/10.1007/s10853-022-06915-4).
- 16 D. Zagorac, H. Müller, S. Ruehl, J. Zagorac and S. Rehme, Recent Developments in the Inorganic Crystal Structure Database: Theoretical Crystal Structure Data and Related

- Features, *J. Appl. Crystallogr.*, 2019, **52**(5), 918–925, DOI: [10.1107/s160057671900997x](https://doi.org/10.1107/s160057671900997x).
- 17 W. Sun, S. T. Dacek, S. P. Ong, G. Hautier, A. Jain, W. D. Richards, A. C. Gamst, K. A. Persson and C. Gerbrand, The Thermodynamic Scale of Inorganic Crystalline Metastability, *Sci. Adv.*, 2016, **2**(11), e1600225, DOI: [10.1126/sciadv.1600225](https://doi.org/10.1126/sciadv.1600225).
- 18 T. Cavnac, J. Schmidt, P.-P. De Breuck, A. Loew, T. F. T. Cerqueira, H.-C. Wang, A. Bochkarev, Y. Lysogorskiy, A. H. Romero, R. Drautz, S. Botti and M. A. L. Marques, AI-Driven Expansion and Application of the Alexandria Database, *J. Phys. Mater.*, 2026, **9**(2), 025014, DOI: [10.1088/2515-7639/ae6620](https://doi.org/10.1088/2515-7639/ae6620).
- 19 K. Kovnir, Predictive Synthesis, *Chem. Mater.*, 2021, **33**(13), 4835–4841, DOI: [10.1021/acs.chemmater.1c01484](https://doi.org/10.1021/acs.chemmater.1c01484).
- 20 M. Aykol, J. H. Montoya and J. Hummelshøj, Rational Solid-State Synthesis Routes for Inorganic Materials, *J. Am. Chem. Soc.*, 2021, **143**(24), 9244–9259, DOI: [10.1021/jacs.1c04888](https://doi.org/10.1021/jacs.1c04888).
- 21 J. R. Neilson, M. J. McDermott and K. A. Persson, Modernist Materials Synthesis: Finding Thermodynamic Shortcuts with Hyperdimensional Chemistry, *J. Mater. Res.*, 2023, **38**(11), 2885–2893, DOI: [10.1557/s43578-023-01037-2](https://doi.org/10.1557/s43578-023-01037-2).
- 22 M. Aykol, V. I. Hegde, L. Hung, S. Suram, P. Herring, C. Wolverton and J. S. Hummelshøj, Network Analysis of Synthesizable Materials Discovery, *Nat. Commun.*, 2019, **10**(1), 2018, DOI: [10.1038/s41467-019-10030-5](https://doi.org/10.1038/s41467-019-10030-5).
- 23 N. J. Szymanski, P. Nevatia, C. J. Bartel, Y. Zeng and G. Ceder, Autonomous and Dynamic Precursor Selection for Solid-State Materials Synthesis, *Nat. Commun.*, 2023, **14**(1), 6956, DOI: [10.1038/s41467-023-42329-9](https://doi.org/10.1038/s41467-023-42329-9).
- 24 N. J. Szymanski, Y.-W. Byeon, Y. Sun, Y. Zeng, J. Bai, M. Kunz, D.-M. Kim, B. A. Helms, C. J. Bartel, H. Kim and G. Ceder, Quantifying the Regime of Thermodynamic Control for Solid-State Reactions during Ternary Metal Oxide Synthesis, *Sci. Adv.*, 2024, **10**(27), eadp3309, DOI: [10.1126/sciadv.adp3309](https://doi.org/10.1126/sciadv.adp3309).
- 25 M. J. McDermott, S. S. Dwaraknath and K. A. Persson, A Graph-Based Network for Predicting Chemical Reaction Pathways in Solid-State Materials Synthesis, *Nat. Commun.*, 2021, **12**(1), 3097, DOI: [10.1038/s41467-021-23339-x](https://doi.org/10.1038/s41467-021-23339-x).
- 26 M. J. McDermott, B. C. McBride, C. E. Regier, G. T. Tran, Y. Chen, A. A. Corrao, M. C. Gallant, G. E. Kamm, C. J. Bartel, K. W. Chapman, P. G. Khalifah, G. Ceder, J. R. Neilson and K. A. Persson, Assessing Thermodynamic Selectivity of Solid-State Reactions for the Predictive Synthesis of Inorganic Materials, *ACS Cent. Sci.*, 2023, **9**(10), 1957–1975, DOI: [10.1021/acscentsci.3c01051](https://doi.org/10.1021/acscentsci.3c01051).
- 27 O. Kononova, H. Huo, T. He, Z. Rong, T. Botari, W. Sun, V. Tshitoyan and G. Ceder, Text-Mined Dataset of Inorganic Materials Synthesis Recipes, *Sci. Data*, 2019, **6**(1), 203, DOI: [10.1038/s41597-019-0224-1](https://doi.org/10.1038/s41597-019-0224-1).
- 28 V. Chung, A. Walsh and D. J. Payne, Solid-State Synthesizability Predictions Using Positive-Unlabeled Learning from Human-Curated Literature Data, *Digital Discovery*, 2025, **4**(9), 2439–2453, DOI: [10.1039/D5DD00065C](https://doi.org/10.1039/D5DD00065C).
- 29 J. Jang, G. H. Gu, J. Noh, J. Kim and Y. Jung, Structure-Based Synthesizability Prediction of Crystals Using Partially Supervised Learning, *J. Am. Chem. Soc.*, 2020, **142**(44), 18836–18843, DOI: [10.1021/jacs.0c07384](https://doi.org/10.1021/jacs.0c07384).
- 30 J. Jang, J. Noh, L. Zhou, G. H. Gu, J. M. Gregoire and Y. Jung, Synthesizability of Materials Stoichiometry Using Semi-Supervised Learning, *Matter*, 2024, **7**(6), 2294–2312, DOI: [10.1016/j.matt.2024.05.002](https://doi.org/10.1016/j.matt.2024.05.002).
- 31 E. R. Antoniuk, G. Cheon, G. Wang, D. Bernstein, W. Cai and E. J. Reed, Predicting the Synthesizability of Crystalline Inorganic Materials from the Data of Known Material Compositions, *npj Comput. Mater.*, 2023, **9**(1), 155, DOI: [10.1038/s41524-023-01114-4](https://doi.org/10.1038/s41524-023-01114-4).
- 32 S. Amari Amir, J. George and P. Benner, SynCoTrain: A Dual Classifier PU-Learning Framework for Synthesizability Prediction, *Digital Discovery*, 2025, **4**(6), 1437–1448, DOI: [10.1039/D4DD000394B](https://doi.org/10.1039/D4DD000394B).
- 33 D. Gleaves, N. Fu, E. M. Dilanga Siriwardane, Y. Zhao and J. Hu, Materials Synthesizability and Stability Prediction Using a Semi-Supervised Teacher-Student Dual Neural Network, *Digital Discovery*, 2023, **2**(2), 377–391, DOI: [10.1039/D2DD00098A](https://doi.org/10.1039/D2DD00098A).
- 34 I. Animitsa, A. Iakovleva and K. Belova, Electrical Properties and Water Incorporation in A-Site Deficient Perovskite $\text{La}_{1-x}\text{Ba}_x\text{Nb}_3\text{O}_{9-0.5x}$, *J. Solid State Chem.*, 2016, **238**, 156–161, DOI: [10.1016/j.jssc.2016.03.023](https://doi.org/10.1016/j.jssc.2016.03.023).
- 35 X. Liu, R. Pang, Z. Quan, J. Yang and J. Lin, Tunable Luminescence Properties of Tb^{3+} -Doped LaGaO_3 Nanocrystalline Phosphors, *J. Electrochem. Soc.*, 2007, **154**(7), J185, DOI: [10.1149/1.2735912](https://doi.org/10.1149/1.2735912).
- 36 H. Park, A. Onwuli and A. Walsh, Exploration of Crystal Chemical Space Using Text-Guided Generative Artificial Intelligence, *Nat. Commun.*, 2025, **16**(1), 4379, DOI: [10.1038/s41467-025-59636-y](https://doi.org/10.1038/s41467-025-59636-y).
- 37 N. J. Szymanski and C. J. Bartel, Establishing Baselines for Generative Discovery of Inorganic Crystals, *Mater. Horiz.*, 2025, **12**(19), 8000–8011, DOI: [10.1039/D5MH00010F](https://doi.org/10.1039/D5MH00010F).
- 38 J. Riebesell, R. E. A. Goodall, P. Benner, Y. Chiang, B. Deng, G. Ceder, M. Asta, A. A. Lee, A. Jain and K. A. Persson, A Framework to Evaluate Machine Learning Crystal Stability Predictions, *Nat. Mach. Intell.*, 2025, **7**(6), 836–847, DOI: [10.1038/s42256-025-01055-1](https://doi.org/10.1038/s42256-025-01055-1).
- 39 J. R. Chamorro and T. M. McQueen, Progress toward Solid State Synthesis by Design, *Acc. Chem. Res.*, 2018, **51**(11), 2918–2925, DOI: [10.1021/acs.accounts.8b00382](https://doi.org/10.1021/acs.accounts.8b00382).
- 40 J. R. Neilson, M. J. McDermott and K. A. Persson, Modernist Materials Synthesis: Finding Thermodynamic Shortcuts with Hyperdimensional Chemistry, *J. Mater. Res.*, 2023, **38**(11), 2885–2893, DOI: [10.1557/s43578-023-01037-2](https://doi.org/10.1557/s43578-023-01037-2).
- 41 C. J. Bartel, S. L. Millican, A. M. Deml, J. R. Rumptz, W. Tumas, A. W. Weimer, S. Lany, V. Stevanović, C. B. Musgrave and A. M. Holder, Physical Descriptor for the Gibbs Energy of Inorganic Crystalline Solids and Temperature-Dependent Materials Chemistry, *Nat. Commun.*, 2018, **9**(1), 4168, DOI: [10.1038/s41467-018-06682-4](https://doi.org/10.1038/s41467-018-06682-4).
- 42 Z. Wang, O. Kononova, K. Cruse, T. He, H. Huo, Y. Fei, Y. Zeng, Y. Sun, Z. Cai, W. Sun and G. Ceder, Dataset of Solution-Based Inorganic Materials Synthesis Procedures Extracted from the Scientific Literature, *Sci. Data*, 2022, **9**(1), 231, DOI: [10.1038/s41597-022-01317-2](https://doi.org/10.1038/s41597-022-01317-2).

- 43 Z. Wang, Y. Sun, K. Cruse, Y. Zeng, Y. Fei, Z. Liu, J. Shangguan, Y.-W. Byeon, K. Jun, T. He, W. Sun and G. Ceder, Optimal Thermodynamic Conditions to Minimize Kinetic By-Products in Aqueous Materials Synthesis, *Nat. Synth.*, 2024, **3**(4), 527–536, DOI: [10.1038/s44160-023-00479-0](https://doi.org/10.1038/s44160-023-00479-0).
- 44 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, An Autonomous Laboratory for the Accelerated Synthesis of Novel Materials, *Nature*, 2023, **624**(7990), 86–91, DOI: [10.1038/s41586-023-06734-w](https://doi.org/10.1038/s41586-023-06734-w).
- 45 A. Jain, G. Hautier, S. P. Ong, C. J. Moore, C. C. Fischer, K. A. Persson and G. Ceder, Formation Enthalpies by Mixing GGA and GGA + U Calculations, *Phys. Rev. B*, 2011, **84**(4), 045115, DOI: [10.1103/PhysRevB.84.045115](https://doi.org/10.1103/PhysRevB.84.045115).
- 46 M. Chase, *NIST-JANAF Thermochemical Tables*, American Institute of Physics, 4th edn, 1998, vol. 1.
- 47 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko and Y. Vázquez-Baeza, SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nat. Methods*, 2020, **17**(3), 261–272, DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- 48 A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dulak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, The Atomic Simulation Environment—a Python Library for Working with Atoms, *J. Phys.: Condens. Matter*, 2017, **29**(27), 273002, DOI: [10.1088/1361-648X/aa680e](https://doi.org/10.1088/1361-648X/aa680e).
- 49 B. Deng, Y. Choi, P. Zhong, J. Riebesell, S. Anand, Z. Li, K. Jun, K. A. Persson and G. Ceder, Systematic Softening in Universal Machine Learning Interatomic Potentials, *npj Comput. Mater.*, 2025, **11**(1), 9, DOI: [10.1038/s41524-024-01500-6](https://doi.org/10.1038/s41524-024-01500-6).
- 50 S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and G. Ceder, Python Materials Genomics (Pymatgen): A Robust, Open-Source Python Library for Materials Analysis, *Comput. Mater. Sci.*, 2013, **68**, 314–319, DOI: [10.1016/j.commatsci.2012.10.028](https://doi.org/10.1016/j.commatsci.2012.10.028).
- 51 R. E. A. Goodall and A. A. Lee, Predicting Materials Properties without Crystal Structure: Deep Representation Learning from Stoichiometry, *Nat. Commun.*, 2020, **11**(1), 6280, DOI: [10.1038/s41467-020-19964-7](https://doi.org/10.1038/s41467-020-19964-7).
- 52 Q. Zhou, P. Tang, S. Liu, J. Pan, Q. Yan and S.-C. Zhang, Learning Atoms for Materials Discovery, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**(28), E6411–E6417, DOI: [10.1073/pnas.1801181115](https://doi.org/10.1073/pnas.1801181115).
- 53 T. Xie and J. C. Grossman, Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties, *Phys. Rev. Lett.*, 2018, **120**(14), 145301, DOI: [10.1103/PhysRevLett.120.145301](https://doi.org/10.1103/PhysRevLett.120.145301).
- 54 K. Choudhary and B. DeCost, Atomistic Line Graph Neural Network for Improved Materials Property Predictions, *npj Comput. Mater.*, 2021, **7**(1), 185, DOI: [10.1038/s41524-021-00650-1](https://doi.org/10.1038/s41524-021-00650-1).
- 55 K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko and K.-R. Müller, SchNetPack: A Deep Learning Toolbox For Atomistic Systems, *J. Chem. Theory Comput.*, 2019, **15**(1), 448–455, DOI: [10.1021/acs.jctc.8b00908](https://doi.org/10.1021/acs.jctc.8b00908).