

# A graph-based network for predicting chemical reaction pathways in solid-state materials synthesis

Matthew J. McDermott<sup>1,2</sup>, Shyam S. Dwaraknath<sup>1</sup>  & Kristin A. Persson<sup>2,3</sup>  

Accelerated inorganic synthesis remains a significant challenge in the search for novel, functional materials. Many of the principles which enable “synthesis by design” in synthetic organic chemistry do not exist in solid-state chemistry, despite the availability of extensive computed/experimental thermochemistry data. In this work, we present a chemical reaction network model for solid-state synthesis constructed from available thermochemistry data and devise a computationally tractable approach for suggesting likely reaction pathways via the application of pathfinding algorithms and linear combination of lowest-cost paths in the network. We demonstrate initial success of the network in predicting complex reaction pathways comparable to those reported in the literature for  $\text{YMnO}_3$ ,  $\text{Y}_2\text{Mn}_2\text{O}_7$ ,  $\text{Fe}_2\text{SiS}_4$ , and  $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$ . The reaction network presents opportunities for enabling reaction pathway prediction, rapid iteration between experimental/theoretical results, and ultimately, control of the synthesis of solid-state materials.

<sup>1</sup>Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. <sup>2</sup>Department of Materials Science and Engineering, University of California, Berkeley, CA, USA. <sup>3</sup>Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ✉email: [kapersson@lbl.gov](mailto:kapersson@lbl.gov)

Dating back to 18<sup>th</sup> century mineralogy<sup>1</sup>, solid-state inorganic chemistry is a cornerstone in the design of novel, functional materials and continues to be driven by pressing technological demands. Consequently, the development of new techniques that accelerate materials synthesis/processing is vital for achieving multifunctional materials with complex properties. Solid materials with target functionality are often thermodynamically metastable, which can limit their accessibility via conventional solid-state synthesis routes, such as the classic “shake and bake” ceramic methods that typically require high temperatures to overcome diffusion barriers and often proceed to global thermodynamic equilibrium<sup>2</sup>. Indeed, solid-state chemistry itself has even been dubbed a black box that is best probed via systematic and extensive iteration, requiring significant experimental expertise akin to apprenticed artistry<sup>3</sup>. The optimization of synthesis procedures for new materials is hence both highly time-consuming and resource-consuming, demanding human-guided iteration over many combinations of precursors, processing steps, and environmental conditions.

A more efficient approach to synthesizing novel inorganic materials is “synthesis by design”, in which a set of guiding principles is used to quickly devise a synthesis method towards a target material, much like the paradigm central to synthetic organic chemistry<sup>4,5</sup>. Recent work, fueled by developments in solid-state *in situ* characterization techniques<sup>6,7</sup>, has advanced this direction by exploring reaction pathways in select case systems that provide insight into mechanistic relationships explaining how synthesis conditions (e.g., precursor selection, reaction environment) alter the reaction pathway and lead to selective formation of different target products. For example, Neilson and coworkers demonstrated the use of unconventional solid-state metathesis reactions to kinetically control the reaction pathway towards metastable polymorphs of CuSe<sub>2</sub><sup>8</sup> and YMnO<sub>3</sub><sup>9,10</sup>. Jiang et al. explored the use of iron silicide reactants to bypass kinetic limitations and achieve a low-temperature synthesis of Fe<sub>2</sub>SiS<sub>4</sub><sup>11</sup>. Miura et al. demonstrated the synthesis of MgCr<sub>2</sub>S<sub>4</sub> thiospinel via a metathesis route using novel precursors, which was shown to be thermodynamically favorable through computational phase diagram construction<sup>12</sup>. Bianchini et al. showed that the first phase formed in the synthesis of P2-type Na<sub>0.67</sub>MO<sub>2</sub> (M=Co, Mn) can be predicted by minimizing compositionally unconstrained reaction energies and that the initial phase formed may drastically alter both the kinetics of subsequent reactions, as well as final phase selectivity<sup>13</sup>. Each of these studies elucidates an important concept: chemical reaction pathways follow a complex thermodynamic free energy landscape that can be carefully manipulated and navigated via the thoughtful selection of precursors, processing, and environmental conditions.

Explicit modeling, as well as reaction network models derived from the atomistic potential energy surface (PES), have been successful in predicting chemical reaction pathways in molecular systems<sup>14,15</sup> but are much less developed for solid-state periodic systems, where monitoring each atom’s coordinates and interactions over the large time and spatial scales necessary rapidly becomes intractable. Despite these limitations, modeling of bounded solid-state reaction mechanisms at the atomistic level has been achieved in particular with molecular dynamics (MD)<sup>16</sup> and kinetic Monte Carlo (KMC)-based<sup>17</sup> approaches. Reactive force fields, such as ReaxFF<sup>18</sup> further permit the breaking of chemical bonds and can be used to study specific chemical reaction mechanisms and kinetic parameters<sup>19</sup>. KMC-based methods also explore parts of the PES, given reaction rate constants that can be approximated with quantum mechanical calculations. However, such methods are ultimately confined to an a priori selection of the relevant domains of the high dimensional solid-state PES. Recent work also suggests that the computational prediction of reaction

pathways in ceramic powder-based synthesis does not always require atomistic methods; significant predictive power can be derived from local thermodynamic equilibrium calculations of pairwise solid-solid interfaces<sup>20</sup>.

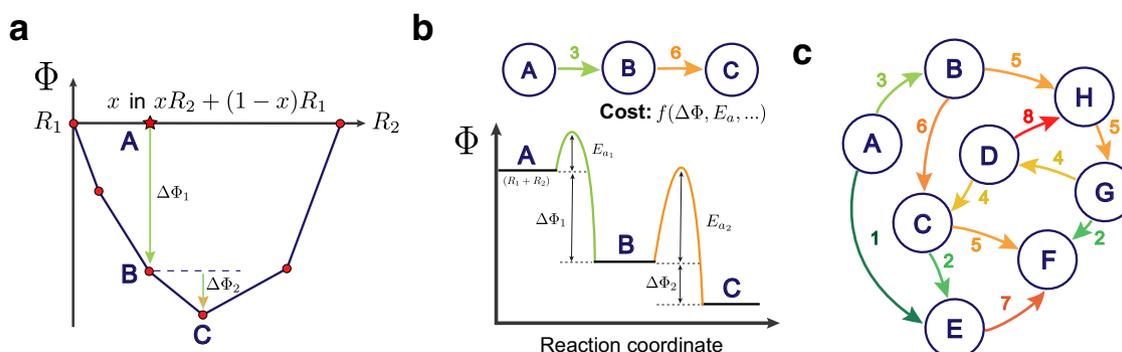
In this work, we describe a chemical reaction network framework for predicting and suggesting solid-state inorganic reaction pathways, which when combined with experimental efforts, aims to realize inorganic synthesis by design. We propose to leverage recent advances in data-driven methods that have resulted in computational/experimental thermochemistry databases<sup>21–24</sup> covering hundreds of thousands of materials and millions of associated reaction energies<sup>25</sup>. We employ a reaction network model that blends typical thermodynamic phase diagrams with the connectivity and kinetic heuristics derived from transition state theory. The network model serves as a convenient data structure for exploring the underlying free energy surface of thermodynamic phase space in solid-state chemistry via the power and efficiency of existing computational infrastructure for large graph networks. We outline the methodology used to create the chemical reaction network from thermochemistry databases and demonstrate its capacity for solid-state reaction pathway prediction by applying it to several reported experimental syntheses, as well as to recommend chemical routes to a novel battery cathode material that has not been previously synthesized.

## Results

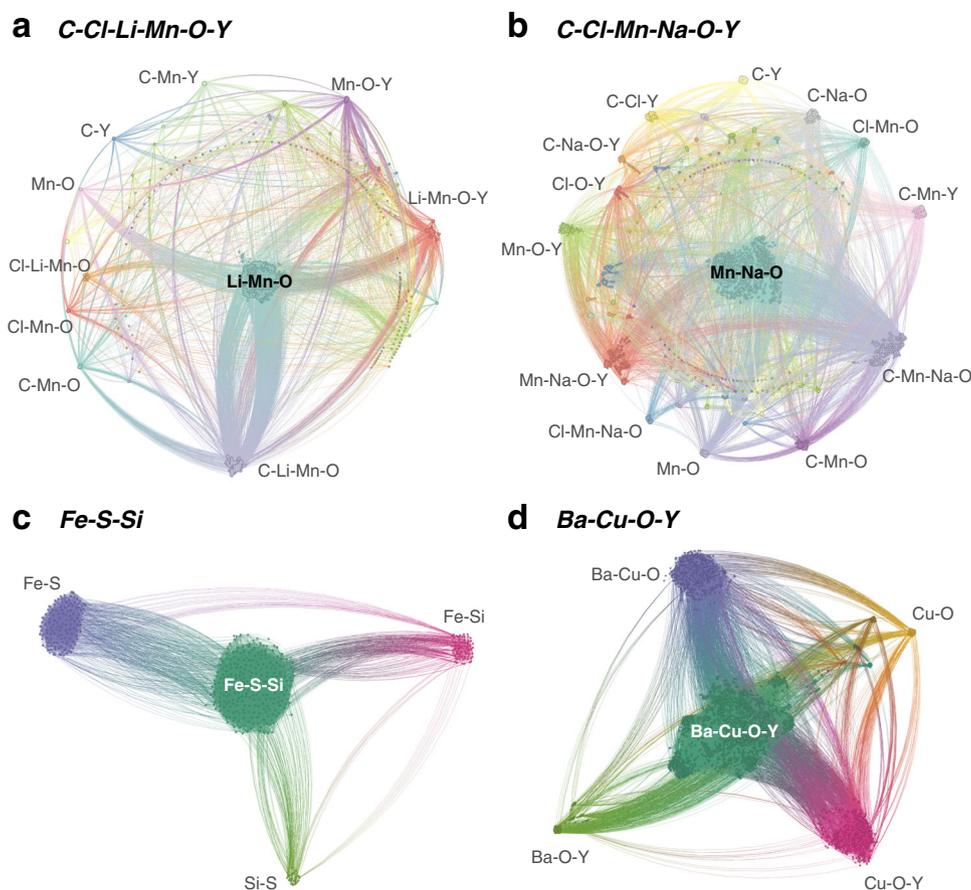
The solid-state reaction network is a model for thermodynamic phase space, which is represented by an energy landscape governed by a generalized thermodynamic potential or free energy,  $\Phi$ . The global minimum in this potential, which depends on the boundary conditions of a particular system, is the thermodynamic equilibrium state of the system. Figure 1 depicts three models of chemical reactions in the thermodynamic phase space, ordered by increasing the level of abstraction. The free energy convex hull construction of Fig. 1a is a purely thermodynamic model of a chemical reaction between two reactant phases,  $R_1$  and  $R_2$ <sup>26</sup>. The convex hull yields the set of products (and thus chemical reactions) that result in the largest decrease in free energy for a given mole ratio of the two reactants. Figure 1b abstracts the thermodynamic model further by incorporating the concept of activation energy,  $E_a$ , as defined by transition state theory<sup>27</sup>. This enables the inclusion of simple kinetic behavior of reactions, where the height of the activation energy barrier correlates with the rate of reaction.

Abstracting even further, we can consider these reaction coordinate diagrams as weighted directed graphs, like that shown in the upper portion of Fig. 1b. In these graphs, the cost/weight of a chemical reaction edge represents an a priori unknown function of synthesis parameters such as the thermodynamic driving force, activation energy, etc. Figure 1c shows the interlinking of many such graph representations within a set of phases, where each node represents a particular combination of phases (e.g.,  $R_1 + R_2$ ) and the edges represent chemical reactions with designated costs. This weighted directed graph, or chemical reaction network, is a densely connected model of thermodynamic phase space where thermodynamic/kinetic features can be combined and transformed into a unique cost representation for each reaction pathway.

In the following sections, the chemical reaction network method is applied to solid-state synthesis procedures discussed in the literature for YMnO<sub>3</sub>, Y<sub>2</sub>Mn<sub>2</sub>O<sub>7</sub>, Fe<sub>2</sub>SiS<sub>4</sub>, and YBa<sub>2</sub>Cu<sub>3</sub>O<sub>6.5</sub>. The reaction networks generated for each experimental system are illustrated in Fig. 2 and constructed primarily using thermochemistry data acquired from the Materials Project (MP)<sup>21</sup>. Via the application of pathfinding algorithms and several



**Fig. 1 Thermodynamic models of chemical reactions in increasing degree of abstraction.** The label, A, indicates a combination of arbitrary amounts of phases  $R_1$  and  $R_2$ . Similarly, the other labels (B–F) are combinations of arbitrary amounts of other phases that result from the chemical reactions. **a** The convex hull construction for consecutive reactions between two reactants,  $R_1$  and  $R_2$ . The points drawn indicate sets of chemical reaction products at different stoichiometric mixtures of the reactants, and the lines trace the convex hull indicating the thermodynamic equilibrium (minimum free energy) for all ratios of mixing,  $x$ . **b** A traditional reaction coordinate diagram used to represent both the free energies of reaction,  $\Delta\Phi$ , and activation energies,  $E_a$ . This is generalized by a weighted, directed graph connecting the three states (top). The cost/weight of the directed edges, shown as the colored number adjacent to each edge, is some function of the free energy change, activation energy, and other reaction features. **c** A chemical reaction network, which links together many such possible reaction pathways within a given set of phases.

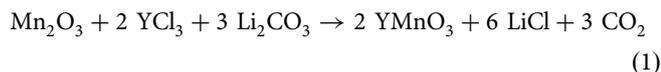


**Fig. 2 Visualizations of several computed reaction networks.** Chemical subsystems are illustrated by color and prominent subsystems are annotated with labels. Shown are reaction networks for the chemical systems: **a** C–Cl–Li–Mn–O–Y with 5855 nodes and 121,176 edges, **b** C–Cl–Mn–Na–O–Y with 4425 nodes and 46,427 edges, **c** Fe–S–Si with 509 nodes and 11,912 edges, and **d** Ba–Cu–O–Y with 2973 nodes and 33,957 edges.

post-processing steps (see “Methods”), we predict the top-ranked candidate reaction pathways to each of the targets and compare them with previous experimental results. Finally, we apply the network construction to predict synthesis routes towards a novel battery cathode material that has yet to be synthesized,  $\text{MgMo}_3(\text{PO}_4)_3\text{O}$ , further demonstrating the versatility of the method in advancing inorganic synthesis by design.

**Synthesis of  $\text{YMnO}_3$  using Li-based assisted metathesis.** First, we consider the synthesis of yttrium manganese oxide ( $\text{YMnO}_3$ ) through the solid-state Li-based assisted metathesis reaction first reported by Todd and Neilson<sup>9</sup>. This synthesis route has the advantage of yielding  $\text{YMnO}_3$  at temperatures significantly lower than the reaction between binary oxides (500 °C vs. 850 °C), enabling kinetic control and polymorph selectivity.

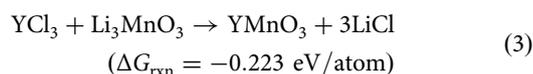
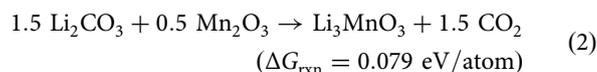
The overall reaction,



proceeds through several steps with distinct intermediate compounds, as determined experimentally through in situ temperature-dependent x-ray diffraction performed at a synchrotron beamline<sup>10</sup>.

Figure 2a shows the chemical reaction network generated for the C–Cl–Li–Mn–O–Y chemical system. The phase diagram for this system includes 853 entries; of these, 53 are predicted by density functional theory (DFT) to be stable at low temperatures. By incorporating vibrational entropic effects through a previously derived machine-learning methodology (see “Methods” section), we find the number of stable species to reduce to 41 at a temperature of 900 K. We include all of these stable entries, as well as metastable entries, not including polymorphs, up to a filter of +30 meV/atom above the hull. This cutoff is motivated by DFT calculations and statistics on experimentally available phases<sup>28</sup> showing that—while highly metastable compounds are by no means inaccessible—the distribution of energies above hull for all experimentally synthesized compounds peaks significantly below 30 meV/atom. This process results in a total of 76 phases considered (Supplementary Data 1) which yields a reaction network of 5855 nodes and 121,176 edges. Reaction edge costs were calculated using the softplus function applied to reaction free energies normalized by the number of reactant atoms. The 60 shortest paths (20 to each of the products  $\text{YMnO}_3$ ,  $\text{LiCl}$ , and  $\text{CO}_2$ ) were identified, and crossover reactions were generated considering oxygen as an open-element with a chemical potential of  $\mu_{\text{O}} = \mu_{\text{O}}^{\text{exp}}$  (900 K, 1 atm)<sup>22</sup>. The final reaction pathways were predicted via this candidate set of reactions by solving for all possible (mass-balanced) linear combinations of reactions up to a maximum size of five reaction steps. This resulted in 38 pathways, which are listed in full in Supplementary Data 1. This list further reduces to 20 pathways after removing pathways with interdependent reaction steps (see “Methods” section).

Of the 20 remaining total reaction pathways, 11 paths involve reaction steps that produce one or more of the following hypothetical intermediate compounds that, to our best knowledge, have never been experimentally synthesized:  $\text{Li}_3\text{MnO}_3$ ,  $\text{Li}_2\text{MnCO}_4$ , and  $\text{Li}_2\text{MnCO}_5$ . Furthermore, each of these phases is predicted to be metastable on the MP database, with energies above the hull of 17, 98, and 35 meV/atom, respectively. Interestingly, the lowest cost predicted pathway (before filtering) actually proceeds through the hypothetical  $\text{Li}_3\text{MnO}_3$  as follows:



This pathway through  $\text{Li}_3\text{MnO}_3$ , as well as similar ones that rely on the formation of hypothetical compounds, may provide useful insights in understanding the lack of accessibility of certain phases, but are deemed less likely to suggest experimentally viable reaction pathways. The five lowest-cost pathways, after removing the pathways which pass through hypothetical intermediates, are summarized in Table 1.

Paths 2 and 3 are identical to the experimentally observed low-temperature reaction pathway reported by Todd and Neilson<sup>10</sup>, involving the formation and reaction of ternary intermediates  $\text{YOCl} + \text{LiMnO}_2 \rightarrow \text{YMnO}_3 + \text{LiCl}$ , with the only difference being that Path 3 first includes the decomposition of  $\text{Li}_2\text{CO}_3$  to  $\text{Li}_2\text{O}$  before subsequent reaction with  $\text{YCl}_3$ . Note that Path 2 includes a reaction with three products (even though the network was restricted to a maximum combination size of  $n = 2$ ) due to the calculation of crossover reactions, which uses the compositional phase diagram to find the most thermodynamically favorable set of products (of any size). Paths 1, 4, and 5 also encompass a plausible reaction pathway and differ primarily in the manner in which  $\text{Y}_2\text{O}_3$  is formed and subsequently reacted with  $\text{Mn}_2\text{O}_3$  to produce the final target  $\text{YMnO}_3$ . Todd et al. report that the oxidation of  $\text{YCl}_3 \rightarrow \text{YOCl} \rightarrow \text{Y}_2\text{O}_3$  and reaction of

**Table 1** Top five predicted reaction pathways to  $\text{YMnO}_3$ .

Path	Total cost	Reactions	$\Delta G_{\text{rxn}}$ (eV/atom)	Cost
1	0.246	0.6667 $\text{Li}_2\text{O} + 0.3333 \text{YCl}_3 \rightarrow \text{LiCl} + 0.3333 \text{LiYO}_2$	−0.260	0.210
		0.5 $\text{YCl}_3 + 1.5 \text{LiYO}_2 \rightarrow \text{Y}_2\text{O}_3 + 1.5 \text{LiCl}$	−0.164	0.229
		0.5 $\text{Y}_2\text{O}_3 + 0.5 \text{Mn}_2\text{O}_3 \rightarrow \text{YMnO}_3$	−0.048	0.254
		$\text{Li}_2\text{CO}_3 \rightarrow \text{Li}_2\text{O} + \text{CO}_2$	0.145	0.300
2	0.249	$\text{Li}_2\text{CO}_3 + \text{YCl}_3 \rightarrow \text{YOCl} + 2 \text{LiCl} + \text{CO}_2$	−0.110	0.240
		$\text{YOCl} + \text{LiMnO}_2 \rightarrow \text{LiCl} + \text{YMnO}_3$	−0.069	0.249
		$\text{Li}_2\text{CO}_3 + \text{Mn}_2\text{O}_3 \rightarrow \text{CO}_2 + 2 \text{LiMnO}_2$	0.007	0.266
		$\text{Li}_2\text{O} + \text{YCl}_3 \rightarrow \text{YOCl} + 2 \text{LiCl}$	−0.281	0.206
3	0.254	$\text{YOCl} + \text{LiMnO}_2 \rightarrow \text{LiCl} + \text{YMnO}_3$	−0.069	0.249
		$\text{Li}_2\text{CO}_3 + \text{Mn}_2\text{O}_3 \rightarrow \text{CO}_2 + 2 \text{LiMnO}_2$	0.007	0.266
		$\text{Li}_2\text{CO}_3 \rightarrow \text{Li}_2\text{O} + \text{CO}_2$	0.145	0.300
		$4 \text{Li}_2\text{O} + 3 \text{YCl}_3 \rightarrow \text{Y}_3\text{O}_4\text{Cl} + 8 \text{LiCl}$	−0.297	0.203
4	0.260	0.6667 $\text{Li}_2\text{O} + 0.3333 \text{YCl}_3 \rightarrow \text{LiCl} + 0.3333 \text{LiYO}_2$	−0.260	0.210
		0.5 $\text{Y}_3\text{O}_4\text{Cl} + 0.5 \text{LiYO}_2 \rightarrow \text{Y}_2\text{O}_3 + 0.5 \text{LiCl}$	−0.057	0.252
		0.5 $\text{Y}_2\text{O}_3 + 0.5 \text{Mn}_2\text{O}_3 \rightarrow \text{YMnO}_3$	−0.048	0.254
		$\text{Li}_2\text{CO}_3 \rightarrow \text{Li}_2\text{O} + \text{CO}_2$	0.145	0.300
5	0.267	3 $\text{Li}_2\text{O} + 2 \text{YCl}_3 \rightarrow \text{Y}_2\text{O}_3 + 6 \text{LiCl}$	−0.306	0.202
		0.6667 $\text{Li}_2\text{O} + 0.3333 \text{YCl}_3 \rightarrow \text{LiCl} + 0.3333 \text{LiYO}_2$	−0.260	0.210
		0.5 $\text{YCl}_3 + 1.5 \text{LiYO}_2 \rightarrow \text{Y}_2\text{O}_3 + 1.5 \text{LiCl}$	−0.164	0.229
		0.5 $\text{Y}_2\text{O}_3 + 0.5 \text{Mn}_2\text{O}_3 \rightarrow \text{YMnO}_3$	−0.048	0.254
		$\text{Li}_2\text{CO}_3 \rightarrow \text{Li}_2\text{O} + \text{CO}_2$	0.145	0.300

These are the lowest total cost reaction pathways after filtering for interdependent reaction steps and previously unsynthesized compounds. Paths 2 and 3 match the experimentally reported pathway most closely, however, the oxidation of  $\text{YCl}_3 \rightarrow \text{YOCl} \rightarrow \text{Y}_2\text{O}_3$  and subsequent reaction with  $\text{Mn}_2\text{O}_3$  in Paths 1, 4, and 5 were experimentally suggested as a possible route, albeit much slower and only kinetically accessible at higher temperatures.

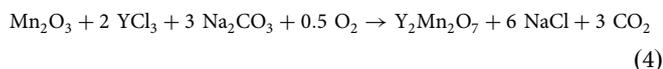
**Table 2 Selection of top predicted reaction pathways to Y<sub>2</sub>Mn<sub>2</sub>O<sub>7</sub>.**

Path	Total cost	Reactions	$\Delta G_{rxn}$ (eV/atom)	Cost	
1	0.238	Na <sub>2</sub> CO <sub>3</sub> + YCl <sub>3</sub> → CO <sub>2</sub> + YOCl + 2 NaCl	-0.178	0.226	
		0.5 O <sub>2</sub> + 2 YOCl + 2 NaMnO <sub>2</sub> → Y <sub>2</sub> Mn <sub>2</sub> O <sub>7</sub> + 2 NaCl	-0.164	0.229	
		Mn <sub>2</sub> O <sub>3</sub> + Na <sub>2</sub> CO <sub>3</sub> → CO <sub>2</sub> + 2 NaMnO <sub>2</sub>	0.024	0.271	
2	0.246	2.125 O <sub>2</sub> + 0.1667 YMn <sub>12</sub> + 0.9167 Y <sub>2</sub> O <sub>3</sub> → Y <sub>2</sub> Mn <sub>2</sub> O <sub>7</sub>	-0.926	0.113	
		0.5 O <sub>2</sub> + Na <sub>4</sub> Mn <sub>2</sub> O <sub>5</sub> + 4 YOCl → Y <sub>2</sub> Mn <sub>2</sub> O <sub>7</sub> + Y <sub>2</sub> O <sub>3</sub> + 4 NaCl	-0.222	0.217	
		Na <sub>2</sub> CO <sub>3</sub> + YCl <sub>3</sub> → CO <sub>2</sub> + YOCl + 2 NaCl	-0.178	0.226	
		0.5 Mn <sub>2</sub> O <sub>3</sub> + Na <sub>2</sub> CO <sub>3</sub> → CO <sub>2</sub> + 0.5 Na <sub>4</sub> Mn <sub>2</sub> O <sub>5</sub>	0.120	0.294	
		1.308 Mn <sub>2</sub> O <sub>3</sub> + 1.026 Y <sub>2</sub> O <sub>3</sub> → Y <sub>2</sub> Mn <sub>2</sub> O <sub>7</sub> + 0.05128 YMn <sub>12</sub>	0.144	0.300	
3	0.246	0.5 O <sub>2</sub> + Na <sub>4</sub> Mn <sub>2</sub> O <sub>5</sub> + 4 YOCl → Y <sub>2</sub> Mn <sub>2</sub> O <sub>7</sub> + Y <sub>2</sub> O <sub>3</sub> + 4 NaCl	-0.222	0.217	
		Na <sub>2</sub> CO <sub>3</sub> + YCl <sub>3</sub> → CO <sub>2</sub> + YOCl + 2 NaCl	-0.178	0.226	
		2 MnO <sub>2</sub> + Y <sub>2</sub> O <sub>3</sub> → Y <sub>2</sub> Mn <sub>2</sub> O <sub>7</sub>	-0.070	0.249	
		0.25 O <sub>2</sub> + 0.5 Mn <sub>2</sub> O <sub>3</sub> → MnO <sub>2</sub>	-0.058	0.252	
		0.5 Mn <sub>2</sub> O <sub>3</sub> + Na <sub>2</sub> CO <sub>3</sub> → CO <sub>2</sub> + 0.5 Na <sub>4</sub> Mn <sub>2</sub> O <sub>5</sub>	0.120	0.294	
4	0.248	4 Na <sub>2</sub> CO <sub>3</sub> + YCl <sub>3</sub> → Na <sub>5</sub> Y(CO <sub>3</sub> ) <sub>4</sub> + 3 NaCl	-0.193	0.223	
		0.5 O <sub>2</sub> + 2 YOCl + 2 NaMnO <sub>2</sub> → Y <sub>2</sub> Mn <sub>2</sub> O <sub>7</sub> + 2 NaCl	-0.164	0.229	
		0.25 Na <sub>5</sub> Y(CO <sub>3</sub> ) <sub>4</sub> + 0.75 YCl <sub>3</sub> → CO <sub>2</sub> + YOCl + 1.25 NaCl	-0.050	0.254	
		Mn <sub>2</sub> O <sub>3</sub> + Na <sub>2</sub> CO <sub>3</sub> → CO <sub>2</sub> + 2 NaMnO <sub>2</sub>	0.024	0.271	
		Na <sub>2</sub> O + YCl <sub>3</sub> → YOCl + 2 NaCl	-0.535	0.164	
11	0.259	2 YOCl + Na <sub>2</sub> O → Y <sub>2</sub> O <sub>3</sub> + 2 NaCl	-0.339	0.196	
		2 MnO <sub>2</sub> + Y <sub>2</sub> O <sub>3</sub> → Y <sub>2</sub> Mn <sub>2</sub> O <sub>7</sub>	-0.070	0.249	
		0.25 O <sub>2</sub> + 0.5 Mn <sub>2</sub> O <sub>3</sub> → MnO <sub>2</sub>	-0.058	0.252	
		Na <sub>2</sub> CO <sub>3</sub> → CO <sub>2</sub> + Na <sub>2</sub> O	0.328	0.351	

The complex defect cascade observed in the experiment is not directly present in the predicted pathways, although the reaction of 2MnO<sub>2</sub> + Y<sub>2</sub>O<sub>3</sub> → Y<sub>2</sub>Mn<sub>2</sub>O<sub>7</sub> present in Paths 3 and 11 represents the net effect of the defect cascade reaction, and Path 11 is overall the closest to the experimentally observed pathway.

binary oxides Y<sub>2</sub>O<sub>3</sub> + Mn<sub>2</sub>O<sub>3</sub> → 2 YMnO<sub>3</sub> is also a plausible reaction pathway that may simultaneously occur, although ex situ control reactions showed that it is much slower and only feasible at higher temperatures. While Li<sub>2</sub>O (Paths 1, 3–5) and LiYO<sub>2</sub> (Path 1, 4, 5) were not directly observed in the diffraction results, all of the other intermediate phases suggested by the top predicted paths (YOCl, Y<sub>3</sub>O<sub>4</sub>Cl, Y<sub>2</sub>O<sub>3</sub>, LiMnO<sub>2</sub>) were indeed observed in experimental data. Hence the reaction network predictions here seem to capture the form of both experimentally observed reaction pathways, suggesting that the model performs quite well with this particular system.

**Synthesis of Y<sub>2</sub>Mn<sub>2</sub>O<sub>7</sub> using Na-based assisted metathesis.** According to the original report by Todd and Neilson, substituting Na for Li in the aforementioned assisted metathesis synthesis of YMnO<sub>3</sub> changes the main product to the pyrochlore Y<sub>2</sub>Mn<sub>2</sub>O<sub>7</sub><sup>9</sup>, represented by the net reaction:



It was shown via in situ x-ray diffraction that, similar to the Li-based reaction, the Na-based reaction pathway similarly depends on the formation of an intermediate alkali manganese oxide phase—in this case, Na<sub>x</sub>MnO<sub>2</sub><sup>29</sup>. Through a cascade of defect reactions, Na<sub>x</sub>MnO<sub>2</sub> reacts with Y<sub>2</sub>O<sub>3</sub>, as formed through the previously described oxidation of YCl<sub>3</sub> → YOCl → Y<sub>2</sub>O<sub>3</sub>.

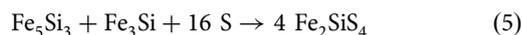
The reaction network for the C–Cl–Mn–Na–O–Y chemical system was constructed with the same parameters as the Li-based network: a temperature of  $T = 900$  K and a 30 meV/atom filter, resulting in 66 entries (Supplementary Data 2) mapped to 4425 nodes and 46,427 edges. Figure 2b illustrates the network for this system, which is similar in shape to the Li-based system, but with a higher number of prominent chemical subsystems. Pathfinding was also performed with identical parameters as the Li system, including  $k = 20$  paths to each target, a maximum reaction combination size of 5, and an open oxygen chemical potential of  $\mu_{\text{O}} = \mu_{\text{O}}^{\text{exp}}$  (900 K, 1 atm)<sup>22</sup>. The pathfinding process resulted in 44

potential reaction pathways (Supplementary Data 2), 35 of which do not contain interdependent reaction steps, and 24 of which further do not contain any reaction steps with the hypothetical phases Na<sub>3</sub>MnO<sub>3</sub>, Mn<sub>8</sub>Cl<sub>3</sub>O<sub>10</sub>, and Na<sub>2</sub>MnO<sub>3</sub>. A selection of the top predicted pathways are shown in Table 2.

The top predicted pathway (Path 1) is directly analogous to the observed kinetically favorable pathway in the Li-based synthesis, although this was not observed experimentally for the Na-based synthesis. Instead, Paths 3 and 11, which contain the final reaction step 2MnO<sub>2</sub> + Y<sub>2</sub>O<sub>3</sub> → Y<sub>2</sub>Mn<sub>2</sub>O<sub>7</sub>, more closely resemble the experimentally observed pathway, which involves a cascade of defect reactions where MnO<sub>2</sub> sub-units (anchored within Na<sub>x</sub>MnO<sub>2</sub>) react with Y<sub>2</sub>O<sub>3</sub>, resulting in a steady increase of the effective Na concentration,  $x$ , as Y<sub>2</sub>Mn<sub>2</sub>O<sub>7</sub> forms.

Interestingly, Na<sub>4</sub>Mn<sub>2</sub>O<sub>5</sub>, Na<sub>5</sub>Y(CO<sub>3</sub>)<sub>4</sub>, and YMn<sub>12</sub> appear as intermediate phases in several of the top predicted pathways but were not observed experimentally. While each of these phases is experimentally synthesizable and predicted by DFT as stable, their synthesis in the literature is reportedly difficult and involves long heating times<sup>30</sup>, hydrothermal methods<sup>31</sup>, or the use of a levitation furnace<sup>32</sup>, respectively. Hence these phases may be kinetically inaccessible using ceramic methods, including the assisted metathesis synthesis method discussed here.

**Synthesis of Fe<sub>2</sub>Si<sub>4</sub> using iron silicide precursors.** Jiang et al. previously showed that Fe<sub>2</sub>Si<sub>4</sub> can be synthesized at significantly lower temperatures by avoiding the kinetically limiting steps encountered when using elemental Fe and Si precursors<sup>11</sup>. They showed that pre-reacting iron with silicon to create iron silicide precursors greatly expedited the formation of Fe<sub>2</sub>Si<sub>4</sub> at temperatures as low as 550 °C, via the net reaction:



The reaction network for the Fe–S–Si chemical system was constructed at a temperature of  $T = 900$  K with much less stringent energy above hull filter of 0.5 eV/atom. This higher energy cutoff was chosen due to the relatively small size of the

**Table 3 Selection of top predicted reaction pathways to Fe<sub>2</sub>SiS<sub>4</sub>.**

Path	Total cost	Reactions	$\Delta G_{rxn}$ (eV/atom)	Cost
1	0.177	0.6667 FeSi <sub>2</sub> + 3.333 S → SiS <sub>2</sub> + 0.3333 Fe <sub>2</sub> SiS <sub>4</sub>	-0.787	0.129
		5 S + Fe <sub>3</sub> Si → Fe <sub>2</sub> SiS <sub>4</sub> + FeS	-0.586	0.156
		0.4286 Fe <sub>5</sub> Si <sub>3</sub> + 4 S → Fe <sub>2</sub> SiS <sub>4</sub> + 0.1429 FeSi <sub>2</sub>	-0.543	0.162
		SiS <sub>2</sub> + 2 FeS → Fe <sub>2</sub> SiS <sub>4</sub>	0.270	0.335
2	0.177	0.6667 FeSi <sub>2</sub> + 3.333 S → SiS <sub>2</sub> + 0.3333 Fe <sub>2</sub> SiS <sub>4</sub>	-0.787	0.129
		5 S + Fe <sub>3</sub> Si → Fe <sub>2</sub> SiS <sub>4</sub> + FeS	-0.586	0.156
		S + 0.4 Fe <sub>3</sub> Si → FeS + 0.2 FeSi <sub>2</sub>	-0.545	0.162
		0.4286 Fe <sub>5</sub> Si <sub>3</sub> + 4 S → Fe <sub>2</sub> SiS <sub>4</sub> + 0.1429 FeSi <sub>2</sub>	-0.543	0.162
3	0.179	SiS <sub>2</sub> + 2 FeS → Fe <sub>2</sub> SiS <sub>4</sub>	0.270	0.335
		0.6667 FeSi <sub>2</sub> + 3.333 S → SiS <sub>2</sub> + 0.3333 Fe <sub>2</sub> SiS <sub>4</sub>	-0.787	0.129
		2 FeSi + 6 S → SiS <sub>2</sub> + Fe <sub>2</sub> SiS <sub>4</sub>	-0.632	0.149
		5 S + Fe <sub>3</sub> Si → Fe <sub>2</sub> SiS <sub>4</sub> + FeS	-0.586	0.156
4	0.180	0.4286 Fe <sub>5</sub> Si <sub>3</sub> + 4 S → Fe <sub>2</sub> SiS <sub>4</sub> + 0.1429 FeSi <sub>2</sub>	-0.543	0.162
		S + 0.5 Fe <sub>3</sub> Si → FeS + 0.5 FeSi	-0.509	0.167
		SiS <sub>2</sub> + 2 FeS → Fe <sub>2</sub> SiS <sub>4</sub>	0.270	0.335
		0.6667 FeSi <sub>2</sub> + 3.333 S → SiS <sub>2</sub> + 0.3333 Fe <sub>2</sub> SiS <sub>4</sub>	-0.787	0.129
14	0.189	2 FeSi + 6 S → SiS <sub>2</sub> + Fe <sub>2</sub> SiS <sub>4</sub>	-0.632	0.149
		2 FeSi + 6 S → SiS <sub>2</sub> + Fe <sub>2</sub> SiS <sub>4</sub>	-0.632	0.149
		5 S + Fe <sub>3</sub> Si → Fe <sub>2</sub> SiS <sub>4</sub> + FeS	-0.586	0.156
		2 Fe <sub>5</sub> Si <sub>3</sub> + 20 S → Si + 5 Fe <sub>2</sub> SiS <sub>4</sub>	-0.541	0.163
		Si + 2 FeS <sub>2</sub> → Fe <sub>2</sub> SiS <sub>4</sub>	0.055	0.278
		SiS <sub>2</sub> + 2 FeS → Fe <sub>2</sub> SiS <sub>4</sub>	0.270	0.335
		0.5 Fe <sub>5</sub> Si <sub>3</sub> + 4 S → Fe <sub>2</sub> SiS <sub>4</sub> + 0.5 FeSi	-0.521	0.166
		2 FeSi + 6 S → SiS <sub>2</sub> + Fe <sub>2</sub> SiS <sub>4</sub>	-0.632	0.149

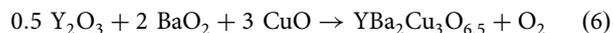
Each pathway successfully includes the experimentally observed reaction between intermediates SiS<sub>2</sub> and FeS. The observed FeSi intermediate only appears together with the other two intermediates in 4 of the 49 predicted paths. Path 14 is the lowest cost path which captures all experimentally observed intermediate phases.

chemical system (3 elements), which allowed for consideration of a wider range of metastability. The constructed network includes 22 entries (Supplementary Data 3) mapped to 509 nodes and 11,912 edges and its illustration is shown in Fig. 2c. Since there is only one target in the net reaction (Fe<sub>2</sub>SiS<sub>4</sub>), higher cutoffs were also chosen for the pathfinding process:  $k = 75$  shortest paths and a maximum reaction combination size of 6. The pathfinding process resulted in 340 suggested reaction pathways. However, as before, we demote pathways that include hypothetical compounds or compounds that only exist under conditions far from those considered here. While silicon monosulfide (SiS) has been experimentally synthesized, it primarily appears as a molecular gas and only exists naturally under astrophysical conditions, such as in massive star-forming<sup>33</sup>. In addition, the SiS<sub>4</sub> structure was experimentally reported<sup>34</sup> as a delithiated version of Li<sub>4</sub>SiS<sub>4</sub> but does not appear to have been synthesized on its own. Excluding these two compounds, as well as interdependent reaction steps (see “Methods” section), yields 49 possible pathways. A selection of the top predicted pathways after filtering is shown in Table 3.

The predicted pathways for this synthesis capture many of the intermediates that were experimentally observed. In the experiment, Jiang et al. observed the initial production of FeS<sub>2</sub>, which was quickly consumed to yield FeSi, Fe<sub>1-x</sub>S, and SiS<sub>2</sub>; the reaction of these phases yields the target Fe<sub>2</sub>SiS<sub>4</sub>, but not all intermediates were fully consumed in the process. Nearly every predicted path successfully identifies the experimentally observed reaction between SiS<sub>2</sub> and FeS, but only 4 of the 49 paths also include the formation of FeS<sub>2</sub> and FeSi with this final step. Path 14 is the lowest cost of such paths and involves elemental Si as an additional intermediate phase that reacts with FeS<sub>2</sub>. While Path 14 is the closest to the experimentally observed pathway, it differs slightly in two regards. In the predicted pathway, FeS<sub>2</sub> is consumed by reaction with elemental Si to produce Fe<sub>2</sub>SiS<sub>4</sub>,

and SiS<sub>2</sub> is formed from sulfidation of FeSi. In the experiment, FeSi actually appears after SiS<sub>2</sub> is formed, suggesting that there is a more complicated missing step (or several steps) involving the reaction of FeS<sub>2</sub> with precursors to form a mixture of Fe<sub>1-x</sub>S, FeSi, and SiS<sub>2</sub> intermediates.

**Synthesis of YBa<sub>2</sub>Cu<sub>3</sub>O<sub>6.5</sub> (YBCO) using barium peroxide.** Miura et al. recently investigated a significant improvement in the solid-state powder synthesis of the superconductor YBa<sub>2</sub>Cu<sub>3</sub>O<sub>6.5</sub> (YBCO) by substituting barium carbonate, BaCO<sub>3</sub>, with barium peroxide, BaO<sub>2</sub>, resulting in the following net reaction<sup>20</sup>:



The reaction network for the Ba–Cu–O–Y chemical system was constructed using a temperature of  $T = 1200$  K and 0.1 eV/atom filter. We again selected a less restrictive energy cutoff due to the smaller size of the system, resulting in a network of 54 entries (Supplementary Data 4) mapped to 2973 nodes and 33,957 edges. The network is illustrated in Fig. 2d. Similar to the assisted metathesis systems, the pathfinding was conducted with  $k = 20$  paths to each target, a maximum reaction combination size of 5, and an open oxygen chemical potential of  $\mu_{\text{O}} = \mu_{\text{O}}^{\text{exp}}$  (1200 K, 1 atm)<sup>22</sup>. The pathfinding process yielded 52 potential reaction pathways (Supplementary Data 4), only 22 of which do not contain interdependent reaction steps or reaction steps with the hypothetical phases Cu<sub>8</sub>O<sub>7</sub>, Ba<sub>8</sub>Cu<sub>8</sub>O<sub>19</sub>, and Y<sub>2</sub>Ba<sub>3</sub>O<sub>6</sub>. A selection of the top predicted pathways are shown in Table 4.

The predicted reaction pathways to YBCO capture many key aspects of the experimentally observed reaction pathway, but with some slight differences that can be attributed to the inability of the network to capture thermal decomposition/melting. In the experimental pathway, Miura et al. observed that Ba<sub>2</sub>Cu<sub>3</sub>O<sub>6</sub> was

**Table 4 Selection of top predicted reaction pathways to  $\text{YBa}_2\text{Cu}_3\text{O}_{6.5}$  (shown here as  $\text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$ ).**

Path	Total cost	Reactions	$\Delta G_{\text{rxn}}$ (eV/atom)	Cost
1	0.180	$\text{BaO}_2 + \text{CuO} \rightarrow \text{BaCuO}_2 + 0.5 \text{O}_2$	-0.160	0.177
		$\text{Y}_2\text{Ba}_4\text{O}_7 + 6 \text{CuO} \rightarrow \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	-0.137	0.181
		$3 \text{Y}_2\text{O}_3 + 12 \text{BaCuO}_2 \rightarrow \text{Y}_2\text{Ba}_4\text{O}_7 + 2 \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	0.039	0.212
2	0.182	$2 \text{BaO}_2 + 2 \text{CuO} \rightarrow \text{Ba}_2\text{Cu}_2\text{O}_5 + 0.5 \text{O}_2$	-0.225	0.167
		$\text{BaO}_2 + \text{CuO} \rightarrow \text{BaCuO}_2 + 0.5 \text{O}_2$	-0.160	0.177
		$0.8 \text{Y}_4\text{Ba}_3\text{O}_9 + 5.6 \text{CuO} \rightarrow \text{Y}_2\text{Cu}_2\text{O}_5 + 0.6 \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	-0.053	0.195
		$3.5 \text{Y}_2\text{O}_3 + 9 \text{BaCuO}_2 \rightarrow \text{Y}_4\text{Ba}_3\text{O}_9 + 1.5 \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	0.014	0.208
		$2 \text{Ba}_2\text{Cu}_2\text{O}_5 + \text{Y}_2\text{Cu}_2\text{O}_5 \rightarrow \text{O}_2 + \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	0.039	0.212
3	0.182	$\text{BaO}_2 + \text{CuO} \rightarrow \text{BaCuO}_2 + 0.5 \text{O}_2$	-0.160	0.177
		$0.8 \text{Y}_4\text{Ba}_3\text{O}_9 + 5.6 \text{CuO} \rightarrow \text{Y}_2\text{Cu}_2\text{O}_5 + 0.6 \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	-0.053	0.195
		$\text{Y}_2\text{Cu}_2\text{O}_5 + 4 \text{BaCuO}_2 \rightarrow \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	-0.010	0.203
		$3.5 \text{Y}_2\text{O}_3 + 9 \text{BaCuO}_2 \rightarrow \text{Y}_4\text{Ba}_3\text{O}_9 + 1.5 \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	0.014	0.208
4	0.185	$\text{BaO}_2 + \text{CuO} \rightarrow \text{BaCuO}_2 + 0.5 \text{O}_2$	-0.160	0.177
		$0.8 \text{Y}_4\text{Ba}_3\text{O}_9 + 5.6 \text{CuO} \rightarrow \text{Y}_2\text{Cu}_2\text{O}_5 + 0.6 \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	-0.053	0.195
		$\text{Y}_2\text{Cu}_2\text{O}_5 + 4 \text{BaCuO}_2 \rightarrow \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	-0.010	0.203
		$2.25 \text{Ba}_2\text{Cu}_2\text{O}_5 + 2.375 \text{Y}_2\text{O}_3 \rightarrow \text{Y}_4\text{Ba}_3\text{O}_9 + 0.375 \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	-0.007	0.204
		$0.5 \text{Y}_2\text{O}_3 + 4 \text{BaCuO}_2 \rightarrow \text{Ba}_2\text{Cu}_2\text{O}_5 + 0.5 \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	0.022	0.209
10	0.190	$4 \text{BaO}_2 + 6 \text{CuO} \rightarrow \text{O}_2 + 2 \text{Ba}_2\text{Cu}_3\text{O}_6$	-0.231	0.166
		$\text{BaO}_2 + \text{CuO} \rightarrow \text{BaCuO}_2 + 0.5 \text{O}_2$	-0.160	0.177
		$\text{Y}_2\text{Ba}_4\text{O}_7 + 6 \text{CuO} \rightarrow \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	-0.137	0.181
		$3 \text{Y}_2\text{O}_3 + 12 \text{BaCuO}_2 \rightarrow \text{Y}_2\text{Ba}_4\text{O}_7 + 2 \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	0.039	0.212
		$\text{Y}_2\text{O}_3 + 2 \text{Ba}_2\text{Cu}_3\text{O}_6 \rightarrow \text{O}_2 + \text{Y}_2\text{Ba}_4\text{Cu}_6\text{O}_{13}$	0.075	0.219

Seven of the top ten pathways capture the  $\text{BaCuO}_2$  intermediate, but  $\text{Ba}_2\text{Cu}_3\text{O}_6$  only shows up in Path 10 and beyond. Path 10 represents the closest match to the experimentally observed pathway, in which the predicted  $\text{Y}_2\text{Ba}_4\text{O}_7$  composition appears to facilitate representation of the observed reaction of  $\text{Y}_2\text{O}_3$  with Ba-Cu-O melt, which cannot be captured by the reaction network model.

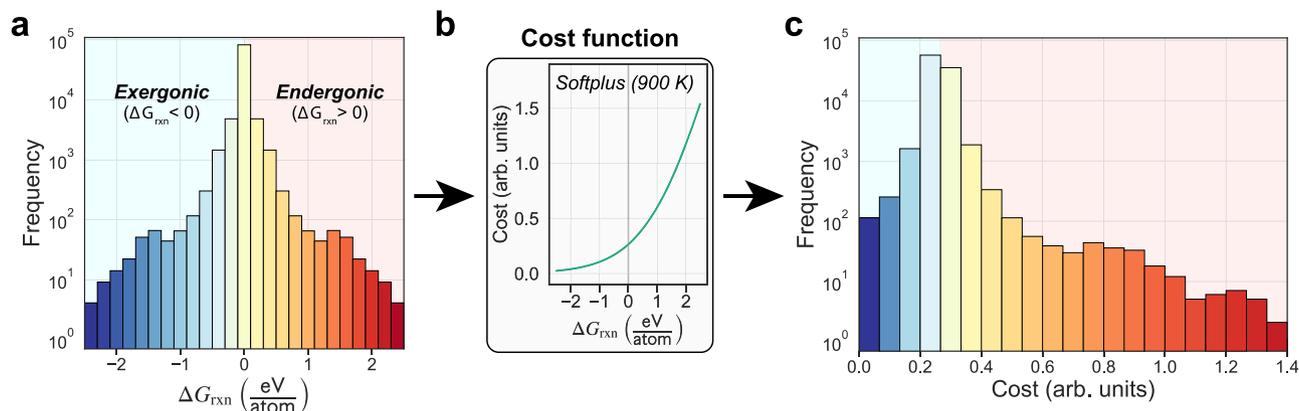
the first intermediate phase to form due to its high thermodynamic driving force reacting from the  $\text{BaO}_2|\text{CuO}_2$  interface, followed by peritectic decomposition of  $\text{Ba}_2\text{Cu}_3\text{O}_6$  into  $\text{BaCuO}_2$ , subsequent eutectic melting of the  $\text{BaCuO}_2|\text{CuO}$  interface, and finally rapid reaction of the Ba-Cu-O melt to form oxygen-deficient YBCO with a final  $\text{O}_2$  uptake step. In the predicted pathways, we see the formation of  $\text{BaCuO}_2$  in 7 of the top 10 predicted paths, although  $\text{Ba}_2\text{Cu}_3\text{O}_6$  only appears as an intermediate phase beginning with Path 10. Path 10 is the closest to the experimentally observed pathway, passing through both  $\text{Ba}_2\text{Cu}_3\text{O}_6$  and  $\text{BaCuO}_2$  intermediate phases. However, similar to Paths 1–4, the formation of YBCO from  $\text{BaCuO}_2$  involves the formation of a Y-Ba-O phase ( $\text{Y}_2\text{Ba}_4\text{O}_7$  or  $\text{Y}_4\text{Ba}_3\text{O}_9$ ) to balance the reaction; in each case, this intermediate is consumed by CuO to yield more of the YBCO product. Hence the Y-Ba-O intermediate effectively serves as an intermediate composition that allows the network to capture a more complex reaction appearing to involve three components, i.e., the reaction of  $\text{Y}_2\text{O}_3$  with a eutectic mixture of  $\text{BaCuO}_2$ -CuO. So while the Y-Ba-O intermediates were not observed in the experiment, their presence in the predicted pathways clearly serves as a convenient mechanism for splitting the  $\text{Y}_2\text{O}_3|\text{Ba-Cu-O(liquid)}$  reaction into two smaller steps that can be captured by the network. Interestingly, two other routes to form YBCO also appear in Paths 2–4, involving the reaction of  $\text{Ba}_2\text{Cu}_2\text{O}_5$  (or  $\text{BaCuO}_2$ ) with  $\text{Y}_2\text{Cu}_2\text{O}_5$  to yield YBCO directly. These may serve as alternative routes towards YBCO, which should be explored in future studies.

**Design of synthesis route for novel Mg-ion battery cathode  $\text{MgMo}_3(\text{PO}_4)_3\text{O}$ .** Finally, we apply the same reaction network code to demonstrate a different use case: designing synthesis routes to the desired target material. Previously, Rong et al. identified a novel Mg battery cathode material,  $\text{MgMo}_3(\text{PO}_4)_3\text{O}$ , which was predicted by DFT calculations to possess an unprecedented, fast  $\text{Mg}^{2+}$  mobility in the dilute Mg concentration limit<sup>35</sup>. To our best knowledge, since the publishing of their report in 2017, no experimental studies on the synthesis of

$\text{MgMo}_3(\text{PO}_4)_3\text{O}$  have been published. According to MP, this material is metastable with significantly large energy above the hull of 0.103 eV/atom. Since approximately 20% of all known oxides exhibit energy above hull higher than 100 meV/atom<sup>28</sup>, this metric does not directly imply that  $\text{MgMo}_3(\text{PO}_4)_3\text{O}$  is impossible to synthesize, but energy above the hull of this magnitude is typically a sign that the material is at least challenging to synthesize.

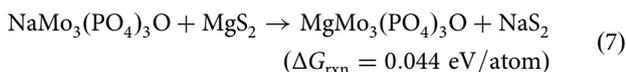
To assist in the development of solid-state synthesis routes, we search for possible reactions within a very large reaction network encompassing both the Mg-Mo-P-O system and 39 additional elements that frequently appear in solid-state syntheses, including nearly all the major anions, alkali, and alkaline earth cations, and several transition metals. The total chemical system includes 43 elements: Ag, Al, B, Ba, Be, Bi, Br, C, Ca, Cd, Ce, Cl, Co, Cr, Cs, Cu, F, Fe, I, K, La, Li, Mg, Mn, Mo, N, Na, Ni, O, P, Pb, Rb, S, Sc, Se, Si, Sr, Te, Ti, V, Y, Zn, and Zr. We calculated the thermodynamic stability of all phases in this system at  $T = 800 \text{ K}$  (a relatively low synthesis temperature) and used an energy above hull filter of 0.11 eV/atom, as this cutoff is just large enough to include the target phase. The final reaction network, which includes only reaction edges that can be successfully balanced to form the target phase, contains 21,564 entries mapped to 660,909 nodes and 663,176 edges. We found a total of 2,270 unique reactions yielding  $\text{MgMo}_3(\text{PO}_4)_3\text{O}$  as a product.

The full list of 2270 discovered reactions (Supplementary Data 5) includes many reactions that would be challenging to achieve in an experiment for reasons other than the metastability of the target phase, such as reactions that also yield side products that are metastable or highly reactive (e.g., Li metal). One method for distilling down the list of candidate reactions is to use a “metathesis-like” approach and search for reactions that involve the formation of a byproduct that does not include any of the elements in the target, i.e., a second product that does not contain Mg, Mo, P, or O. This is a similar principle as that of metathesis reactions in the sense that it includes the formation of an easily separable compound or highly stable phase which acts as a thermodynamic “sink” that may increase the driving force of the



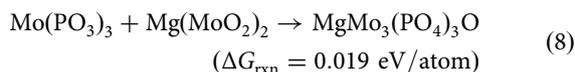
**Fig. 3** Effect of the cost function transformation on reaction energies. **a** Distribution of reaction Gibbs free energies (normalized per reactant atom) in the C-Cl-Li-Mn-O-Y chemical system. **b** The softplus cost function, described in Eq. (12), is used to monotonically transform reaction energies. **c** Final distribution of reaction costs (in arb. units) after transformation. Reaction free energy of zero corresponds to a cost of -0.265.

reaction. Filtering by this constraint reduces the list of candidate reactions down to a more manageable set of 186. Interestingly, nearly every reaction in this candidate set has a similar form: ion exchange from a closely related phase, such as  $\text{NaMo}_3\text{P}_3\text{O}_{13}$ ,  $\text{MgFe}_3\text{P}_3\text{O}_{13}$ , or  $\text{MgV}_3\text{P}_3\text{O}_{13}$ , as demonstrated by the example:



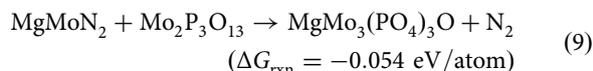
While this reaction may be feasible, it still requires first developing a synthesis route to make the  $\text{NaMo}_3(\text{PO}_4)_3\text{O}$  precursor, which is also significantly metastable and not known to be experimentally synthesizable.

Besides ion-exchange routes, there are only two other reactions that emerge. The first is the direct formation of  $\text{MgMo}_3(\text{PO}_4)_3\text{O}$  (with no byproduct) from simple ternary phases:



While both of these reactants are predicted to be stable by DFT,  $\text{Mg}(\text{MoO}_2)_2$  does not appear to have been previously synthesized in the literature. Assuming this phase is synthesizable (which is likely considering its predicted stability), it is still possible that this reaction might not be spontaneous due to its predicted positive  $\Delta G_{\text{rxn}}$ . However, the small magnitude of this reaction energy is close enough to zero that it may truly be negative, considering both the uncertainties of DFT and the applied Gibbs free energy descriptor.

Finally, the second reaction is an interesting pathway using the ternary nitride,  $\text{MgMoN}_2$ , and  $\text{Mo}_2\text{P}_3\text{O}_{13}$ :



Both precursors of this reaction are predicted to be stable via DFT calculations on the MP database, but we could not find any literature discussing the synthesis of  $\text{Mo}_2\text{P}_3\text{O}_{13}$ .  $\text{MgMoN}_2$ , on the other hand, has been reported to be synthesized via solid-state reaction<sup>36</sup>. This reaction between ternary nitrides may be a promising synthesis route to  $\text{MgMo}_3(\text{PO}_4)_3\text{O}$ , assuming that  $\text{Mo}_2\text{P}_3\text{O}_{13}$  can first be synthesized—which is a strong possibility given its predicted stability in DFT calculations.

## Discussion

The cost function approach that is central to our model is a necessary transformation for applying pathfinding methods that

not only allows for the combination of various reaction metrics (e.g.,  $\Delta G_{\text{rxn}}$ ,  $E_a$ ) but also serves as a particularly powerful way to navigate uncertainties in thermochemistry data. By transforming reaction free energies to positive costs, we no longer restrict possible reactions to only those with negative free energies (Fig. 3). This seems to be a crucial reason for the model's success since reaction steps that occur experimentally are not always predicted to be thermodynamically favorable using computed—or even experimental—data. For example, the well-studied thermal decomposition of  $\text{Li}_2\text{CO}_3 \rightarrow \text{Li}_2\text{O} + \text{CO}_2$  which appears frequently throughout Table 1 is highly endergonic ( $\Delta G_{\text{rxn}} > 0$ ) when modeled with MP data (+0.333 eV/atom at  $T = 0$  K). This is somewhat expected, as MP often reports much more negative formation energies for carbonate compounds due to the elemental energy corrections used, which are not always applicable to polyatomic ions. However, this reaction is still highly endergonic using NIST-JANAF experimental data (+0.145 eV/atom at  $T = 900$  K), despite the fact that the decomposition of lithium carbonate has been observed to occur spontaneously just above temperatures in this range ( $T \sim 900$  K)<sup>37</sup>. We hypothesize that the specific local environment conditions during synthesis, which can differ from the average or global conditions, play a major role in governing the extent of decomposition. Hence it is desirable to retain some degree of flexibility in deciding what is thermodynamically feasible.

It is worth noting that the final ranking of the candidate set of predicted reaction pathways is very sensitive to the choice and form of the cost function. Here, we apply the softplus function to approximate reaction Gibbs free energies derived via a machine-learned model of the vibrational energy that does not explicitly include configurational degrees of freedom. Indeed, in the examples presented, it is not always the lowest total cost path that corresponds to the experimentally observed reaction sequence. The set of suggested pathways also typically includes one or more predictions that are very close to the observed pathways. Since the differences between the highest-ranked (lowest-cost) pathways are small, we strongly recommend users of the model consider several highly ranked pathways. The candidate sets are limited enough in size (often fewer than 100 paths) such that they can be manually inspected by the user.

The primary challenge that is addressed by our work is the high degree of complexity inherent to thermodynamic phase space, which ordinarily leads to a combinatorial explosion during both the creation of the network and subsequent pathfinding steps. For example, consider a reaction network with  $N$  phases and a

maximum phase combination size,  $n$ . If during the graph generation every possible chemical reaction between any two nodes is considered, the number of reactions calculated,  $R$ , would be:

$$R = \left[ \sum_{i=1}^n \binom{N}{i} \right]^2 \quad (10)$$

$$= \left[ \binom{N}{1} + \binom{N}{2} + \dots + \binom{N}{n} \right]^2$$

In the C–Cl–Li–Mn–O–Y reaction network shown in Fig. 2a, which contains  $N = 76$  distinct phases, the maximum number of possible reactions described by Eq. (10) are  $R \approx 5.78 \times 10^3$  ( $n = 1$ ),  $8.56 \times 10^6$  ( $n = 2$ ), and  $5.36 \times 10^9$  ( $n = 3$ ). This equation scales quickly as constraints are relaxed, often leading to the consideration of millions—or even billions—of possible reactions.

Our implementation reduces the complexity and degrees of freedom of the phase space by introducing a series of filters, including (1) restricting the number of phases considered via thermodynamic stability arguments (energy above the hull), (2) limiting the maximum number of phases present on each side of the reaction to a small number ( $n = 2$ ), (3) using a cost function to prioritize reactions which are more likely to occur, (4) enforcing mass conservation via stoichiometric constraints, and (5) removing interdependent reaction steps. The first two filters work together during graph generation to limit the combinatorial size/complexity of the network. This number can be reduced by decreasing either  $N$ ,  $n$ , or both. Since it is typically optimal to consider as many phases as possible, and because we see that the complexity scales especially quickly with increasing  $n$ , it is more favorable to maintain a large  $N$  and enforce a constraint of  $n = 2$ . This choice greatly minimizes the combinatorics of the network but does not inherently sacrifice the complexity of reaction pathways that can occur. In fact, the choice of  $n = 2$  may more realistically capture the behavior of reacting solids by dividing reaction pathways into pseudo-elementary steps that more closely follow the free energy surface. This decision is further justified by recent work suggesting that the most thermodynamically favorable pairwise reactions direct the reaction pathway in typical solid-state synthesis procedures<sup>20</sup>.

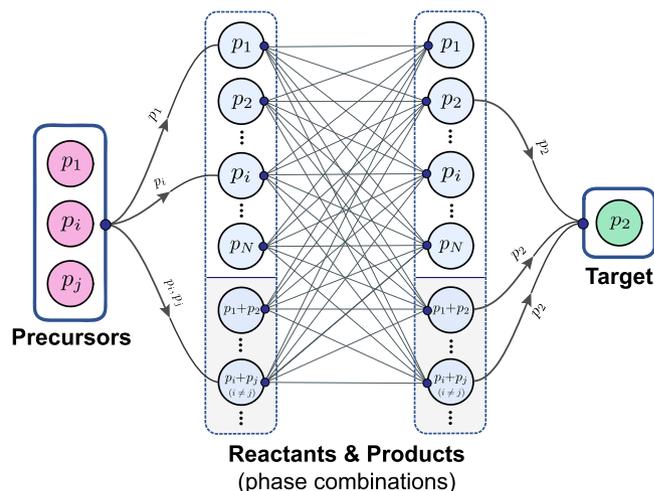
Addressing the combinatorial challenges associated with reaction pathway prediction in large chemical spaces can result in approximations of complex behavior that do not always capture specifics. This complex behavior includes: (1) changes in reaction kinetics due to melting, (2) amorphous intermediates, and (3) defect reactions involving non-stoichiometric compounds. Each of these challenges massively expands the configurational degrees of freedom and necessitates the collection of significantly more thermodynamic data. However, a complete predictive understanding of solid-state synthesis can not be attained without first developing models that incorporate these effects. In future investigations, we expect to address some aspects of these challenges, for example, by including kinetic features leading to more complex reaction pathway selection. While we did not include any kinetic features in this work, we anticipate that several metrics from recent data-oriented studies in chemistry and materials science may be included as extra parameters in the cost function. Such parameters may include the structural (dis)similarity between phases<sup>38,39</sup>, the average number of bonds broken/created in a reaction, the change in the information entropy description of atomic configurations, the change in atomic density, etc. The exact weighting of these parameters within the cost function model is unknown as of now, however, and would best be investigated via high-throughput, automated experiments and detailed studies that systematically probe the impact of precursor

composition/morphology, reaction conditions, etc. on the resulting reaction pathways.

In conclusion, we designed a solid-state chemical reaction network model constructed from available thermochemistry data and demonstrated the model as a predictor of reaction pathways in solid-state materials synthesis. The framework effectively reduces the large, complex thermodynamic landscape to a computationally tractable structure through (i) creation of a weighted directed graph representation of the available thermodynamic phase space, (ii) mapping of rigorous thermodynamic data and possible heuristics into a versatile cost function, and (iii) application of graph pathfinding algorithms to identify probable reaction routes. While the framework explores reaction trajectories in the most general way possible, allowing for parallel combined pathways, the combinatorial complexity is reduced by chemically motivated filters such as: (1) restricting the number of phases considered via thermodynamic stability arguments, (2) limiting the maximum number of simultaneously reacting phases, and (3) enforcing mass conservation via stoichiometric constraints. As a demonstration, the framework was shown to identify complex reaction pathways comparable to those observed in several experimental solid-state synthesis studies. We envision our methodology to be used to suggest possible synthesis precursors/routes that create efficient thermodynamic conditions for targeting desired phases, as well as identification of byproducts and possible intermediates along synthesis routes. Future work will benefit tremendously by combining the framework ‘live’ with automated data collection, in situ phase identification, rapid analysis techniques, and automated feedback loops, moving towards active control of solid-state synthesis.

## Methods

**Network model construction.** Figure 4 illustrates the generalized graph structure of a reaction network for any chemical system. Here, the chemical system refers to the set of all  $N$  phases  $p_i$  ( $i = 1, 2, \dots, N$ ) that can be produced from a designated set of chemical elements. Each reactant/product node on the graph is created by



**Fig. 4 Generalized graph architecture of a solid-state chemical reaction network.** The reaction network is constructed for a hypothetical chemical system containing  $N$  phases  $p_i$  ( $i = 1, 2, \dots, N$ ), with nodes made of distinct phase combinations up to maximum size,  $n$ . The precursors and target nodes link into and out of the densely connected network of reactions, respectively. The edge directions between the reactant and product nodes have been omitted for clarity; an alternate depiction of the network with edge directions is shown in Supplementary Fig. 1. These central edges include chemical reactions weighted by their cost value, as well as zero-weight edges which create loops from product nodes back to reactant nodes so that multiple-step reaction pathways can be captured.

considering combinations of distinct phases up to a maximum size,  $n$ . This yields the set of all nodes,  $P$ , given by:

$$P = \{p_i | i \leq N\} \cup \{p_i + p_j | i, j \leq N; i \neq j\} \cup \dots \cup \{p_i + p_j + \dots + p_n | i, j, \dots, n \leq N; i \neq j \neq \dots \neq n\} \quad (11)$$

In the graph, each of these phase combinations is added twice: first as a reactants node and again as a products node. While higher values of  $n$  enable more complex reactions, in general, it suffices to choose  $n = 2$  among three or more reactants are less likely due to kinetic and steric constraints in a solid composite.

To create the dense set of directed edges at the center of the network, we algorithmically iterate through every possible chemical reaction between all pairs of reactants and product nodes. Using a reaction balancing algorithm, we then solve for the stoichiometric coefficients and add a weighted, directed edge from the reactant node to the corresponding product node for every chemical reaction that is successfully balanced. Note that a vast majority of generated trial reactions cannot be stoichiometrically balanced and hence are excluded from the graph; for example, there are no  $x$ ,  $y$ , or  $z$  that satisfy  $xY_2O_3 + yMnO_2 \rightarrow zYMnO_3$ , so this reaction edge would not appear in the graph. We also exclude trivial identity-like reactions between identical reactants and products, e.g.,  $Y_2O_3 \rightarrow Y_2O_3$ . The weight of the reaction edge is determined by a “cost function” that maps features of the chemical reaction (e.g.,  $\Delta\Phi_{rxn}$ ) to a single cost value. To facilitate product phases being capable of reacting again, zero-weight edges are added which connect each product node to all reactant nodes that contain, as a subset, at least one of the product phases and/or starting reactant phases (regardless of consideration of stoichiometric coefficients). This creates a large degree of cycles in the network that enable the network to capture multiple-step reaction pathways.

Finally, two more nodes are added: one for the synthesis precursors and one for the selected target. These two external nodes act as single-source and destination nodes linking into and out of the dense network of reactions, defining a net (overall) synthesis reaction. The precursors node connects into the network via zero-weight edges directed towards all reactants nodes that contain, as a subset, at least one of the precursor phases. The target node is connected via a set of zero-weight edges directed from all product nodes which contain the target phase.

**Cost function derivation.** The cost function determines the weighting of edges in the network and its nature is critical to the generation of probable reaction pathways. The simplest, and possibly most intuitive, cost function is a one-to-one mapping onto the thermodynamic landscape, such as the measured or calculated Gibbs free energy of reaction,  $\Delta G_{rxn}$ . However, using unprocessed reaction energies alone poses several problems: (1) negative reaction energies result in infinite cycles during pathfinding, which precludes the use of Dijkstra’s algorithm and many other pathfinding methods, (2) kinetic effects and other known heuristics about the reaction are necessarily excluded, and (3) reaction costs are affected by stoichiometric scaling. Instead, here we choose a single, positive cost function that maps the Gibbs free energy of reaction, normalized per reactant atom, to a positive value for each reaction. The choice of a functional mapping also provides the opportunity to create different cost functions for chemical reactions where additional information is known, such as experimental data, kinetic factors, and/or other heuristics. One example of a cost function that captures reaction thermodynamics is the softplus function, which was originally developed for use as an activation function in neural networks<sup>40</sup>. This function maps the Gibbs free energy of a chemical reaction,  $\Delta G_{rxn}$ , to a positive cost value,  $C$ , via:

$$C = \ln \left( 1 + \frac{273 \text{ K}}{T} e^{\Delta G_{rxn}} \right) \quad (12)$$

where  $T$  is the absolute temperature in Kelvin and  $\Delta G_{rxn}$  is the Gibbs free energy of a reaction in units of eV per reactant atom, divided by unity to be dimensionless. Since molar reaction energies scale with the stoichiometric balancing of the reaction,  $\Delta G_{rxn}$  must be normalized on a per-atom basis independent of the stoichiometric coefficients. The softplus cost function transforms highly exergonic ( $\Delta G_{rxn} \ll 0$ ) reactions into low (near-zero) cost events, whereas endergonic ( $\Delta G_{rxn} > 0$ ) reactions exhibit a finite cost that smoothly approaches a linear scaling as  $\Delta G_{rxn} \rightarrow \infty$ . Note that different environmental boundary conditions, such as open elements, can be modeled by replacing  $\Delta G_{rxn}$  with  $\Delta\Phi_{rxn}$ , where  $\Phi$  represents a customized thermodynamic potential.

Other monotonically increasing functions, such as those in Fig. 5, were briefly tested. The Rectified Linear Unit (ReLU) function was not appropriate as it did not discriminate between exergonic reactions, which significantly alters the reaction energy distribution and affects the final ranking of pathways. The piecewise linear function preserved the distribution but requires more unknown parameters, and the discontinuity of the first derivative at zero unnecessarily penalized reactions with marginally positive energies. Hence the softplus function was chosen due to its simplicity, smoothness, and differentiable form.

**Application of graph pathfinding methods.** Ideally, predicting a reaction pathway using the network would be fully equivalent to solving the single-source

shortest path problem from graph theory using existing algorithms. However, inorganic chemical reactions rarely trace a set of linear steps even in simple syntheses; instead, the precursor phases often undergo reactions concurrently in parallel or react again. Within the network, parallel reactions can be modeled as simultaneous travel along multiple reaction edges. These reactions must obey mass conservation, and phases produced in one reaction may react with phases in another. These so-called “crossover” reactions, along with the possibility of parallel reactions, prohibit the direct application of shortest path algorithms.

To accommodate parallel paths, we identify not only the single shortest path from precursors to target phase, but a set of  $k$ -shortest paths to the target that are present in the network. To computationally generate the  $k$ -shortest paths, we utilize Yen’s algorithm<sup>41</sup> which iteratively produces the next  $k - 1$  shortest path via deviations from the first shortest path, as calculated with Dijkstra’s algorithm<sup>42</sup>. The purpose of using Yen’s algorithm is two-fold: (1) to identify a candidate set of reactions that may occur in parallel, and (2) to account for uncertainties in the knowledge of the local synthesis environment, as well as the thermochemistry data used to create the network. For syntheses that involve multiple targets or byproducts (e.g.,  $CO_2$ ), pathfinding is performed towards each target phase separately to ensure that all targets appear in the generated set of shortest paths.

To allow for crossover reactions, we first identify all intermediate phases that appear in the  $k$ -shortest paths to each target and then compute all possible reactions between these intermediates which result in formation of at least one of the targets. These reactions are computed via two approaches: (1) a simple  $n$ -combinatorial approach analogous to the one used in the network generation, and (2) a compositional phase diagram approach whereby reaction products are predicted to be the set of phases that yield the minimum thermodynamic potential,  $\Phi_{min}$ , along a compositional tie-line, as previously developed by Richards et al.<sup>26</sup>. This second approach allows crossover reactions to exceed the constraints of the value of  $n$  chosen, due to both the inclusion of open elements via a grand potential, as well as the construction of phase diagram simplexes including several product phases in equilibrium. Since it is recommended to select  $n = 2$ , this second approach makes it possible to additionally capture the reaction of two solids in a flowing gas (e.g.,  $O_2$ ), which is a commonly encountered experimental scenario in solid-state synthesis.

**Combining chemical reactions via mass conservation.** When a net reaction is known a priori, reaction steps identified during pathfinding can be linearly combined to satisfy the stoichiometric mass constraints of the overall reaction. These constraints correspond to numerically solving the linear system of equations given by

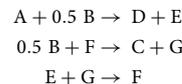
$$A\mathbf{m} = \mathbf{c} \quad (13)$$

where  $\mathbf{m}$  is a vector containing the multiplicity of each reaction (i.e., the factor by which the entire reaction is multiplied),  $A$  is the matrix containing the stoichiometric coefficients of all phases present in all reactions where reactants/products have negative/positive coefficients, respectively, and  $\mathbf{c}$  is a vector containing the stoichiometric coefficients of the net synthesis reaction. We solve this system of equations for the multiplicity vector,  $\mathbf{m}$ , via application of the Moore-Penrose matrix pseudoinverse as implemented within the SciPy package<sup>43</sup>.

The total cost,  $C_{tot}$ , for a balanced reaction pathway with  $l$  reaction steps is the weighted mean of individual reaction costs,  $s_i$ , where the weights are the multiplicities,  $m_i$ , for each reaction:

$$C_{tot} = \frac{\sum_i m_i s_i}{\sum_i m_i} \quad (14)$$

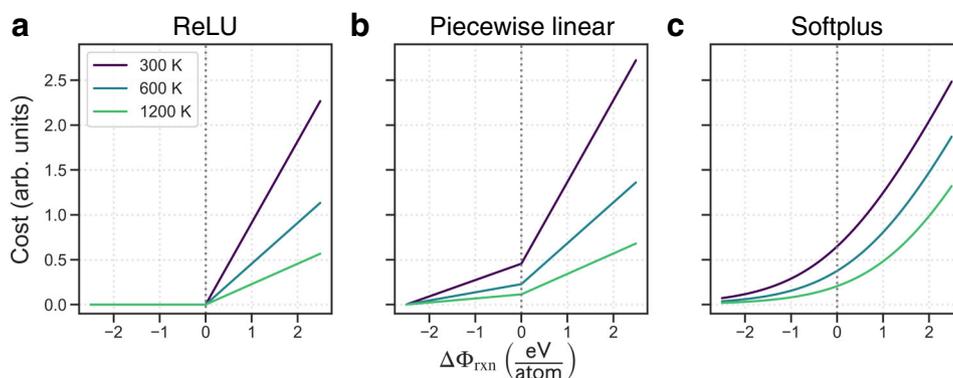
**Identifying interdependent reaction steps.** It is often possible to encounter reaction steps in a reaction pathway that are interdependent, which occurs when two or more reactions exclusively share intermediate phases in such a way that no reaction can proceed without the other(s) happening first. To demonstrate this situation, consider a predicted reaction pathway for the hypothetical net reaction  $A + B \rightarrow C + D$ :



The net stoichiometry of this pathway indeed balances exactly to  $A + B \rightarrow C + D$ , but the second and third reactions are an interdependent pair. In other words, Phase F is a reactant in the second reaction but must be produced by the third reaction, while phase G is a reactant in the third reaction but must be produced in the second reaction. It is often possible to combine interdependent reactions to form a more simple single reaction; in this case, the second and third reactions are combined as  $0.5 B + E \rightarrow C$ .

At the end of the reaction pathway search, we consolidate suggested pathways that contain interdependent reactions. This produces more realistic outputs and tends to greatly reduce the number of predicted pathways.

**Thermochemistry data and graph software.** While the chemical reaction network can be created from any thermochemistry data—computed, experimental, or



**Fig. 5 Example cost functions.** Each cost function is a monotonically increasing function of reaction free energy,  $\Delta\Phi_{\text{rxn}}$ , and scaled according to temperature,  $T$ . The functions include (a) Rectified Linear Unit (ReLU), b) piecewise linear, and c) softplus. Each has the result that the thermodynamic free energy of the reaction is rectified to positive values and scaled such that lower reaction temperatures result in higher costs.

a combination of both— in this work, we primarily employ the Materials Project database, which contains well-benchmarked ab initio calculated formation enthalpies for over one hundred thousand different materials as calculated with density functional theory (DFT)<sup>21,44</sup>. To capture the temperature dependence of vibrational degrees of freedom, we employ the machine-learned Gibbs free energy descriptor reported by Bartel et al.<sup>45</sup>, which estimates the finite temperature contribution to the Gibbs free energy of formation of solids,  $\Delta G_f(T)$ . This contribution incorporates both temperature-dependent enthalpic and entropic effects, although the entropic contribution ( $TS$ ) typically dominates. The elemental Gibbs free energies used for these formation energy calculations are acquired from FactSage<sup>22</sup>. The Gibbs free energies of formation for non-elemental gases (e.g.,  $\text{CO}_2$ ), as well as a small set of solid compounds ( $\text{A}_2\text{O}$  and  $\text{A}_2\text{CO}_3$ ;  $\text{A}=\text{Li}, \text{Na}$ ) are acquired from NIST-JANAF experimental thermochemical tables<sup>23</sup>.

Thermodynamic phase diagram calculations and reaction balancing procedures are performed using algorithms implemented in the pymatgen package<sup>46</sup>. Graphs are implemented using the graph-tool package<sup>47</sup> and visualized in Fig. 2 using Graphistry Hub<sup>48</sup>.

### Data availability

All results data supporting the findings of this paper are included within the article or supplied as Supplementary Data. Ab initio thermochemistry data used in this work are available via the open-access Materials Project database<sup>21</sup> (version 2020.09.08). Supplemental experimental data were acquired from the open-access NIST-JANAF database<sup>23</sup> and are also included as part of the provided software. Source data for Fig. 3 are provided with the paper. Source data are provided with this paper.

### Code availability

A Python implementation of the solid-state reaction network method is available at <https://github.com/GENESIS-EFRC/reaction-network>, archived in ref.<sup>49</sup>, and provided as Supplementary Software 1.

Received: 24 June 2020; Accepted: 20 April 2021;

Published online: 25 May 2021

### References

- Marshall, J. L. & Marshall, V. R. Rediscovery of the elements: cronstedt and nickel. *Hexagon Alpha Chi Sigma* **105**, 24–29 (2014).
- Disalvo, F. J. Solid-state chemistry: a rediscovered chemical frontier. *Science* **247**, 649–655 (1990).
- Kohlmann, H. Looking into the black box of solid-state synthesis. *Eur. J. Inorganic Chem.* **2019**, 4174–4180 (2019).
- Soderholm, L. & Mitchell, J. F. Perspective: Toward “synthesis by design”: exploring atomic correlations during inorganic materials synthesis. *APL Mater.* **4**, 053212 (2016).
- Stein, A., Keller, S. W. & Mallouk, T. E. Turning down the heat: Design and mechanism in solid-state synthesis. *Science* **259**, 1558–1564 (1993).
- Shoemaker, D. P. et al. In situ studies of a platform for metastable inorganic crystal growth and materials discovery. *Proc. Natl Acad. Sci. USA* **111**, 10922–10927 (2014).
- O’Nolan, D. et al. A thermal-gradient approach to variable-temperature measurements resolved in space. *J. Appl. Crystallogr.* **53**, <https://doi.org/10.1107/S160057672000415X>. (2020).
- Martinolich, A. J. & Neilson, J. R. Toward reaction-by-design: achieving kinetic control of solid state chemistry with metathesis. *Chem. Mater.* **29**, 479–489 (2017).
- Todd, P. K. & Neilson, J. R. Selective formation of yttrium manganese oxides through kinetically competent assisted metathesis reactions. *J. Am. Chem. Soc.* **141**, 1191–1195 (2019).
- Todd, P. K., Smith, A. M. M. & Neilson, J. R. Yttrium manganese oxide phase stability and selectivity using lithium carbonate assisted metathesis reactions. *Inorganic Chem.* **58**, 15166–15174 (2019).
- Jiang, Z., Ramanathan, A. & Shoemaker, D. P. In situ identification of kinetic factors that expedite inorganic crystal formation and discovery. *J. Mater. Chem. C* **5**, 5709 (2017).
- Miura, A. et al. Selective metathesis synthesis of  $\text{MgCr}_2\text{S}_4$  by control of thermodynamic driving forces. *Mater. Horiz.* <https://doi.org/10.1039/C9MH01999E>. (2020).
- Bianchini, M. et al. The interplay between thermodynamics and kinetics in the solid-state synthesis of layered oxides. *Nat. Mater.* <https://doi.org/10.1038/s41563-020-0688-6>. (2020).
- Steinfeld, J. I., Francisco, J. S. & Hase, W. L. *Chemical Kinetics and Dynamics* (Prentice Hall, Upper Saddle River, 1999).
- Dewyer, A. L., Argüelles, A. J. & Zimmerman, P. M. Methods for exploring reaction space in molecular systems. *WIREs Comput. Mol. Sci.* **8**, e1354 (2018).
- Allen, M. P. *Introduction to Molecular Dynamics Simulation* (Computational soft matter: from synthetic polymers to proteins (NIC Series), Julich, 2004).
- Voter, A. F. Introduction to the Kinetic Monte Carlo Method. (eds. Sickafus, K. E., Kotomin, E. A. & Uberuaga, B. P.) In *Radiation Effects in Solids*. 1–23 (Springer Netherlands, Dordrecht, 2007).
- van Duin, A. C. T., Dasgupta, S., Lorant, F. & Goddard, W. A. ReaxFF: a reactive force field for hydrocarbons. *J. Phys. Chem. A* **105**, 9396–9409 (2001).
- Ilyin, D. V., Goddard, W. A., Oppenheim, J. J. & Cheng, T. First-principles-based reaction kinetics from reactive molecular dynamics simulations: application to hydrogen peroxide decomposition. *Proc. Natl Acad. Sci. USA* **116**, 18202–18208 (2019).
- Miura, A. et al. Observing and Modeling the Sequential Pairwise Reactions that Drive Solid-State Ceramic Synthesis. *Adv. Mater.* 2100312. <https://doi.org/10.1002/adma.202100312> (2021).
- Jain, A. et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
- Bale, C. et al. Factsage thermochemical software and databases, 2010–2016. *Calphad* **54**, 35–53 (2016).
- Malcolm, W. & Chase, J. *NIST-JANAF thermochemical tables* (Fourth edition. American Chemical Society; American Institute of Physics for the National Institute of Standards and Technology, 1998).
- Bale, C. W. & Eriksson, G. Metallurgical thermochemical databases—a review. *Can. Metallur. Quart.* **29**, 105–132 (1990).
- Jain, A., Shin, Y. & Persson, K. A. Computational predictions of energy materials using density functional theory. *Nat. Rev. Mater.* **1**, 15004 (2016).
- Richards, W. D., Miara, L. J., Wang, Y., Kim, J. C. & Ceder, G. Interface stability in solid-state batteries. *Chem. Mater.* **28**, 266–273 (2016).
- Eyring, H. The activated complex in chemical reactions. *J. Chem. Phys.* **3**, 107–115 (1935).
- Sun, W. et al. The thermodynamic scale of inorganic crystalline metastability. *Sci. Adv.* **2**, <https://advances.sciencemag.org/content/2/11/e1600225.full.pdf> (2016).
- Todd, P. et al. Selectivity in materials synthesis via local chemical potentials in hyperdimensional phase space. Preprint at <https://arxiv.org/abs/2104.05986> (2021).

30. Brachtel, G. & Hoppe, R. Die Koordinationszahl 5 bei Mn(III): Na<sub>4</sub>Mn<sub>2</sub>O<sub>5</sub>. *Zeitschrift für anorganische und allgemeine Chemie* **468**, 130–136 (1980).
31. Awaleh, M., Ben Ali, A., Maisonneuve, V. & Leblanc, M. Microwave-assisted synthesis, crystal structures and thermal behaviour of Na<sub>3</sub>Y(CO<sub>3</sub>)<sub>4</sub> and Na<sub>5</sub>Yb(CO<sub>3</sub>)<sub>4</sub> · 2H<sub>2</sub>O. *J. Alloys Compounds* **349**, 114–120 (2003).
32. Deportes, J. & Givord, D. Magnetic structure of YMn<sub>12</sub>. *Solid State Commun.* **19**, 845–851 (1976).
33. Morris, M., Gilmore, W., Palmer, P., Turner, B. E. & Zuckerman, B. Detection of interstellar SiS and a study of the IRC +10216 molecular envelope. *Astrophys. J.* **199**, L47–L51 (1975).
34. Murayama, M. et al. Synthesis of new lithium ionic conductor thio-lisicon-lithium silicon sulfides system. *J. Solid State Chem.* **168**, 140–148 (2002).
35. Rong, Z. et al. Fast Mg<sup>2+</sup> diffusion in Mo<sub>3</sub>(PO<sub>4</sub>)<sub>3</sub>O for Mg batteries. *Chem. Commun.* **53**, 7998–8001 (2017).
36. Wang, L. et al. Solid state synthesis of a new ternary nitride MgMoN<sub>2</sub> nanosheets and micromeshes. *J. Mater. Chem.* **22**, 14559–14564 (2012).
37. Timoshevskii, A. N., Ktalkherman, M. G., Emel'kin, V. A., Pozdnyakov, B. A. & Zamyatin, A. P. High-temperature decomposition of lithium carbonate at atmospheric pressure. *High Temp.* **46**, 414–421 (2008).
38. Zimmermann, N. E. R. & Jain, A. Local structure order parameters and site fingerprints for quantification of coordination environment and crystal structure similarity. *RSC Adv.* **10**, 6063–6081 (2020).
39. De, S., Bartók, A. P., Csányi, G. & Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **18**, 13754–13769 (2016).
40. Dugas, C., Bengio, Y., Bélisle, F., Nadeau, C. & Garcia, R. Incorporating second-order functional knowledge for better option pricing. In *Proceedings of the 13th International Conference on Neural Information Processing Systems*, NIPS'00, 451–457 (MIT Press, Cambridge, 2000).
41. Yen, J. Y. Finding the K shortest loopless paths in a network. *Manag. Sci.* **17**, 712–716 (1971).
42. Dijkstra, E. W. A note on two problems in connexion with graphs. *Numerische Mathematik* **1**, 269–271 (1959).
43. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
44. Jain, A. et al. Formation enthalpies by mixing GGA and GGA + U calculations. *Phys. Rev. B* **84**, 45115 (2011).
45. Bartel, C. J. et al. Physical descriptor for the Gibbs energy of inorganic crystalline solids and temperature-dependent materials chemistry. *Nat. Commun.* **9**, 4168 (2018).
46. Ong, S. P. et al. Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
47. Peixoto, T. P. The graph-tool python library. *figshare*. [http://figshare.com/articles/graph\\_tool/1164194](http://figshare.com/articles/graph_tool/1164194), <https://doi.org/10.6084/m9.figshare.1164194.v14> (2014).
48. Graphistry Inc. *Graphistry Hub*. <https://hub.graphistry.com> (Accessed 27 Dec 2020) (2020).
49. McDermott, M., Dwaraknath, S. & Persson, K. A graph-based network for predicting chemical reaction pathways in solid-state materials synthesis. <https://doi.org/10.5281/zenodo.4690495>. (2021).

## Acknowledgements

This work was supported as part of GENESIS: A Next-Generation Synthesis Center, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award Number DE-SC0019212. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. The authors would like to thank J. Neilson, P. Todd, O. Kononova, and C. Bartel for their helpful discussion regarding the reaction network model, as well as E. Persson for math skills, and L. Meyerovich for assistance with graph visualization.

## Author contributions

M.J.M. and S.S.D. conceived the idea of the presented work. M.J.M. designed and developed the code with feedback from S.S.D. and K.A.P. M.J.M. performed all calculations and wrote the manuscript with the guidance of S.S.D. and K.A.P.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-23339-x>.

**Correspondence** and requests for materials should be addressed to K.A.P.

**Peer review information** *Nature Communications* thanks Ankit Agrawal, Caleb Phillips and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021