Large-scale experimental validation of thermochemical water-splitting oxides discovered by defect graph neural networks

Tyra C. Douglas[1,†], Michael J. Dzara[2,†], Andrew J. E. Rowberg[3], Keith A. King[4], Maria Syrigou[4], Nicholas A. Strange[5], Robert T. Bell[2], Anuj Goyal[2,6], Pin-Wen Guan[4], Robert B. Wexler[7], Joel B. Varley[3], Tadashi Ogitsu[3], Stephan Lany[2], Anthony H. McDaniel[4], Sean R. Bishop[1,*], and Matthew D. Witman[4,*]

[1] Sandia National Laboratories Albuquerque, NM 87123, USA

[2] National Renewable Energy Laboratory, Golden, CO 80401, USA

[3] Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

[4] Sandia National Laboratories Livermore, CA 94550, USA

[5] Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA, 94025, USA

[6] Department of Materials Science and Metallurgical Engineering, IIT Hyderabad, Kandi, Sangareddy, Telangana 502284, India

[7] Department of Chemistry and Institute of Materials Science and Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA

[†] Equal contributions

[*] mwitman@sandia.gov ; srbisho@sandia.gov

Abstract

Thermochemical water-splitting (TCH) based on 2-step thermal redox cycles in metal oxides is a promising approach to generating H_2, but state-of-the-art (SOTA) CeO_2 has several practical limitations, which has motivated continued materials discovery efforts in this field. Here, we improve upon a SOTA defect graph neural network (dGNN) surrogate model's oxygen vacancy predictions and combine them with Materials Project phase diagrams to down-select and discover structurally diverse, experimentally known metal oxides whose TCH performance was previously unknown. Amongst twelve candidates selected based on our high-throughput screening and down-selection criteria, we achieved $\sim 80\%$ accuracy in identifying materials with stable redox cycling and hydrogen production in stagnation flow reactor water-splitting experiments. Closer to 100% accuracy can be achieved if higher-accuracy, hybrid DFT-predicted vacancy formation energies were computed and used in lieu of the most uncertain dGNN-based screening predictions, as they correct false positives to true negatives. Notably, two discovered candidates, $Sr_3PrMn_2O_8$ and $Ba_2Fe_2O_5$, display hydrogen yields greater than CeO_2 under specific redox conditions. These results demonstrate our ability to computationally predict and experimentally validate promising candidate TCH materials that have the potential to compete with CeO_2.

Introduction

Two-step redox of metal oxides via thermochemical (TCH) water-splitting[1,2] is widely considered a promising route for hydrogen production that, unlike electrolysis-based pathways,[3] does not depend primarily on redirecting electricity from the grid for fuel production. A specific subclass of these two-step cycles that use non-volatile, non-stoichiometric metal oxides as active material operate via a thermochemical process. High temperatures accompanied by low oxygen partial pressure reduce a metal oxide host by introducing oxygen vacancy defects, and re-oxidation of the material by water vapor at lower temperatures produces hydrogen.

$$\begin{aligned} \mathrm{MO}_x &\rightarrow \mathrm{MO}_{x-\delta} + \frac{\delta}{2}\mathrm{O}_2 \ \text{(red.)} \\ \mathrm{MO}_{x-\delta} + \delta\mathrm{H_2O} &\rightarrow \mathrm{MO}_x + \delta\mathrm{H}_2 \ \text{(ox.)} \end{aligned} \tag{1}$$

TCH water-splitting metal oxides have been discovered by experimental trial-and-error, intuition, or modeling-informed studies,[4–8] but CeO_2 remains the best material to date[9] and has reached the pilot-scale "large prototype" stage. This is primarily due to its exceptionally large reduction entropy and ability to re-oxidize in high H_2/H_2O ratios,[10] which simplifies practical reactor design considerations, i.e., by alleviating costly gas separations and by reducing the energy penalty associated with feeding a large excess of water vapor at high temperature. However, CeO_2 is not without its own drawbacks: it requires extremely high reduction temperatures ($\gtrsim$1500 °C) to achieve sufficiently high oxygen off-stoichiometry and hydrogen yields, which in turn complicates practical reactor design considerations such as material stability, heat source, etc.

This has motivated an intense push for computational materials discovery,[11–15] thermodynamic modeling,[16–19] and subsequent experimental synthesis[4–8] of alternative metal oxide phases with appropriate thermodynamics for TCH water-splitting. At the materials discovery stage, the key down-selection criteria for a promising TCH metal oxide relate to (1) its oxygen vacancy defect formation energies and (2) the overlap of the host structure's oxygen chemical stability window with typical TCH redox conditions. While additional material properties should in principle be considered and optimized (e.g., oxygen defect kinetics, thermal/melt stability, etc.), these two properties provide the most tractable down-selection criteria for computational screening that can be used to inform high-throughput TCH materials discovery.

In this work, we utilize a defect graph neural network (dGNN) surrogate model, first demonstrated in Ref. [15] and trained on an expanded set of density functional theory calculations (DFT) herein, for predicting oxygen vacancy formation energies. Materials Project (MP)[20] facilitated phase diagram construction was then used to predict the oxygen chemical potential stability range of candidate materials and to quantify its overlap with the oxygen chemical potential range (redox conditions) of a typical TCH process. This permits high-throughput screening of the 10,000s of known and hypothetical metal oxides in the MP for their oxygen vacancy defect formation energies and host stability, of which only ∼100 candidates are predicted to meet our strictest down-select criteria, representing a downsizing of roughly two orders of magnitude.

To test the robustness of our computational predictions for defect thermodynamics and host stability and their down-selection criteria, twelve metal oxides from the screening performed herein were chosen for synthesis and experimental validation through thermogravimetric analysis (TGA) and stagnation flow reactor (SFR) water-splitting experiments. Overall, the computational success at predicting TCH water-splitting activity is excellent with an $\sim$ 80% true positive rate. Here we define a material as a "positive" if it passes our experimental screening protocol by displaying reversible non-stoichiometry ($>$ 50% $\Delta\delta$ that of CeO_2 under similar conditions) in the TGA experiments for at least 3 redox cycles *and* stable hydrogen production in the SFR water-splitting experiments for at least 5 redox cycles. If we exclude materials that fail by mechanisms outside our down-selection criteria (i.e, insufficiently high melting temperature) and further validate uncertain dGNN predictions of vacancy formation energies with hybrid DFT, the true positive rate rises to 100%. Among true positive-predicted TCH water-splitters whose SFR re-oxidation was performed in pure steam, notably high performance is achieved by $Sr_3PrMn_2O_8$ and $Ba_2Fe_2O_5$, whose hydrogen yields are greater than that of CeO_2 when cycled under certain redox conditions. These materials should therefore be considered high-priority candidates for future experiments under more challenging and extended cycling conditions, as well as for further computationally-guided performance optimization via material modification and substitutions.

Results and Discussion

Our discovery workflow consists of four major steps. First, we (re)train an updated dGNN model for predicting vacancy formation energies. Second, we high-throughput screen vacancy formation energies across the known metal oxide space from MP and combine this information with first principles host stability predictions provided via MP. This permits down-selection of TCH candidates based on the primary criteria that indicate their predicted suitability for TCH. Third, as a preliminary experimental screen, TGA experiments determine whether selected oxides can reversibly cycle in thermochemical redox. Fourth, for those that pass the TGA screening, SFR experiments demonstrate whether a material can split water during thermochemical redox. Additional methodological details on DFT calculations, model architecture and training,

synthesis, and TGA/SFR testing are included in the Supplementary Information (SI).

dGNN training and validation

We use the dGNN modeling approach detailed in Ref. [15] for predicting neutral vacancy (V_X) formation energies,

$$\widehat{\Delta H}_{V_{X,i}} = f_{\text{dGNN}}(\mathcal{C}_h, i; \theta), \tag{2}$$

where $\mathcal{C}_h$ denotes the perfect (relaxed) host crystal structure, i the index of the crystallographic site hosting the vacancy, X the elemental identity of that site, and θ the learned model weights. Relying only on the host crystal structure as input, this surrogate model's inference predictions are of trivial computational cost (see Supplementary Section 4) compared to those of first-principles based predictions, $\Delta H_{V,i}$, based on DFT calculations that require large ($\gtrsim$100 atom) supercell relaxations for each unique site in the crystal structure. The neutral vacancy formation enthalpy $\Delta H_{V_{X,i}}$ for species X (i.e., cation or oxygen) is defined as follows:

$$\Delta H_{V_{X,i}} = E(V_{X,i}) - E_{\text{bulk}} + \mu_X^{\text{ref}}. \tag{3}$$

Here, $E(V_{X,i})$ is the total energy of a supercell containing the neutral vacancy $V_{X,i}$; E_{bulk} is the total energy of the pristine, defect-free supercell; and μ_X^{ref} is the reference chemical potential of the removed element, taken as the fitted elemental reference energies (FERE),[21,22] $\mu_X^{\text{ref}} = \mu_X^{\text{FERE}}$, which improve the description of thermochemical properties of oxides in DFT calculations.[23] See Supplementary Section 1 for additional details on DFT settings.[21,22,24–33]

Our previous training data set consisted of neutral oxygen and cation vacancies in binary and ternary metal oxides spanning a diverse chemical and structural space. In this work, we augment that training data with (1) the validation data from the same study, which again consists of neutral oxygen and cation vacancy formation energies, but for systems containing four or more elements and solid solution metal oxides;[5,7,19] (2) the ABO_3 perovskite dataset of Wexler et al.;[12] and (3) additional binary oxides and a handful of ternary oxides spanning an even wider chemical space computed in this work (see Supplementary Section 1).

For this augmented dataset, the number of structures containing a given cation is shown in Figure 1a. The training data's lack of uniform chemical coverage necessitates careful cross-validation (CV) to obtain a reasonable expectation estimate of model accuracy when high-throughput screening unseen structures for materials discovery.[34] We perform nested (K, L)-fold CV, with $K = 10$ and $L = 10$, described in detail in Supplementary Section 5. Here, $\langle\text{MAE}\rangle$ and $\langle R^2\rangle$ denote the model's expected mean absolute error (MAE) and coefficient of determination (R^2), which can be delineated by whether only oxygen vacancies, $\Delta H_{V_{X=O}}$, or cation vacancies, $\Delta H_{V_{X=\text{Cation}}}$, are considered in the expectation (Table 1). Note that our MAE for the neutral oxygen vacancy is comparable to that of Kumagai et al[35] (MAE $\sim$ 0.34 eV), who utilized a different surrogate modeling approach, DFT functional, and possible training elements (i.e., Mn, Fe, Ni, and Co were avoided but are of important considerations in TCH).

CV split criteria	Vac. type	$\langle\text{MAE}\rangle$ [eV]	$\langle R^2\rangle$
Structure	$V_{X=\text{All}}$	0.55	0.91
	$V_{X=O}$	0.38	0.84
	$V_{X=\text{Cation}}$	0.81	0.86
Element	$V_{X=\text{All}}$	1.09	0.84
	$V_{X=O}$	0.75	0.64
	$V_{X=\text{Cation}}$	1.64	0.74

Table 1: For structure-wise and element-wise nested CV splits, the expected MAE and R^2 of the model in predicting neutral vacancy formation energies, delineated by O-only, cation-only, or all vacancy types.

Two different train/test splitting criteria are considered: structure-wise (whereby all defects in a set of host training structures are held out for the test set) and element-wise (whereby all defects in any structure containing a specific cation type are held out for the test set). On unseen structures, our model achieves a sufficiently low error ($\langle\text{MAE}\rangle \lesssim 0.5$ eV for ΔH_{V_O}) needed for quantitatively reliable TCH materials discovery predictions. While still $\sim$ 2x greater than that of structure-wise splits, the models' expected $\langle\text{MAE}\rangle$ of $\sim$ 0.75 eV (and $\langle R^2\rangle \sim 0.64$) for $\widehat{\Delta H}_{V_O}$ and element-wise splits still indicates semi-quantitative predictive accuracy. In other words, the dGNN's generalization capabilities are sufficient that ΔH_{V_O} can be (roughly) approximated for compounds containing a cation that has little representation in the training data. Nonetheless, for poorly represented cations in our training data (deep purple elements in Figure 1a), we expect reduced accuracy, comparable in the best-case to the element-wise $\langle\text{MAE}\rangle$ in Table 1.

Parity plots of all test set predictions from the nested CV are shown in Figure 1(b,d), where (b) corresponds to models trained on structure-wise splits and (d) corresponds to models trained on element-wise splits. Each outer-fold test set prediction is derived from an ensemble average of inner-fold model predictions (see Supplementary Section 5), and the standard deviation of this ensemble of predictions, $\sigma(\widehat{\Delta H}_{V,i})$, is a heuristic metric for its uncertainty. Figure 1(c,e) shows that, while the absolute error (AE) of individual predictions is not well correlated with σ, the expectation value of AE *within a given bin of* σ is well approximated by σ, and especially so for large σ. Therefore, flagging vacancy predictions with high uncertainty (high σ) can be a useful strategy on average for avoiding experimental validation of materials that are poorly predicted by the dGNN.

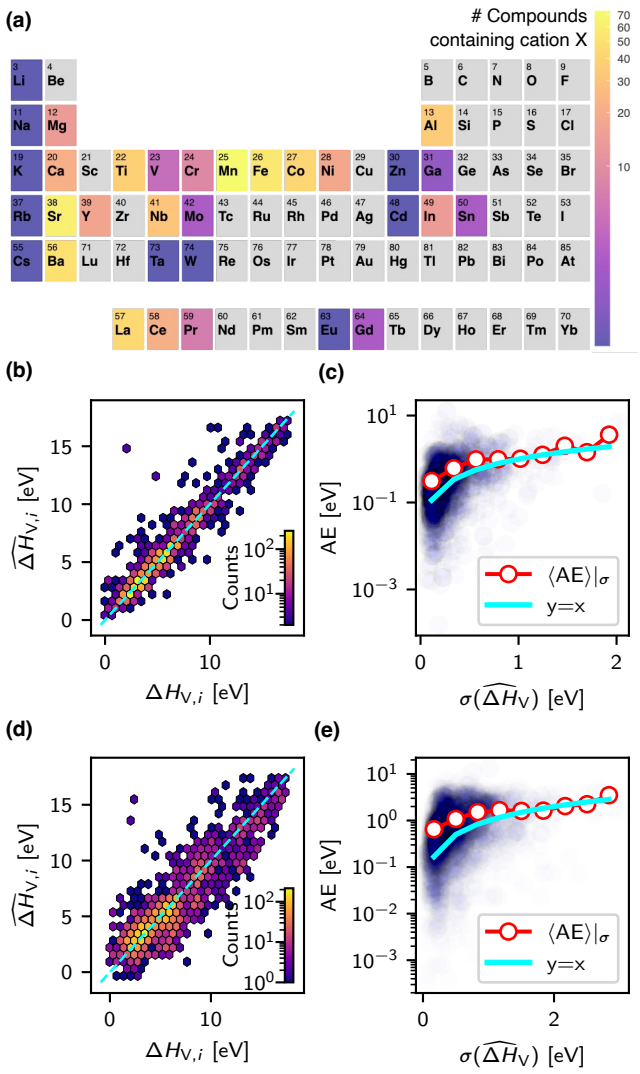

Figure 1: (a) Periodic table color-coded by the number of structures in the training data containing the designated cation. (b,d) Respective structure-wise vs. element-wise nested CV parity plots showing each test set example's ensemble averaged prediction. (c,e) Respective structure-wise vs. element-wise nested CV showing each test set example's prediction error vs. the uncertainty metric (blue circles), expectation of prediction error within a given σ bin (red line), and $y = x$ (cyan line).

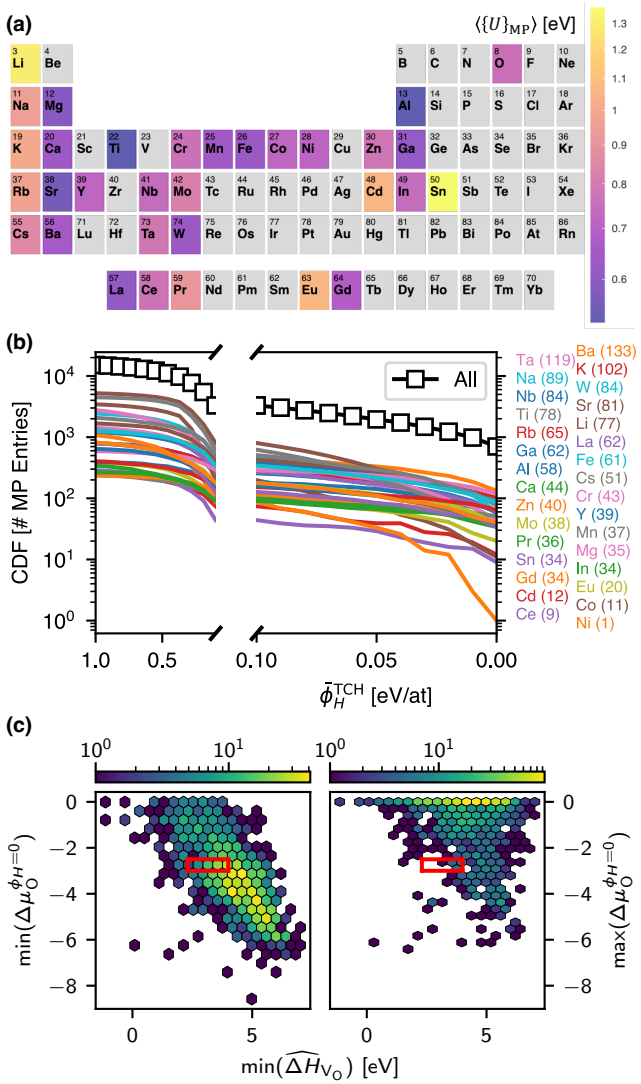

Figure 2: (a) Periodic table color-coded by $\langle\{\bar{U}\}_{\mathrm{MP}}\rangle$ for all MP screened structures containing the indicated cation. (b) Cumulative distribution function (CDF) for the # of MP metal oxides with $\bar{\phi}_H^{\mathrm{TCH}}$. The black line represents all queried MP20 compounds, while colored lines represent the CDF for the subset of compounds containing a given cation listed on the right (sorted from top to bottom by the number of candidates with $\bar{\phi}_H^{\mathrm{TCH}} = 0$, listed in parentheses). (c) The defect vs. host stability trade-off plotted as the minimum and maximum oxygen chemical potential (left and right subplots, respectively) at which a material is stable vs. the minimum oxygen vacancy formation energy. The color-bar denotes structure counts.

MP screening

We queried all metal oxides in the Materials Project 2020 (MP20) database,[20,36] excluding any structures with cations outside our training data (Figure 1a). Three key quantities are then computed for remaining candidates to aid down-selection for experiments.

First, oxygen vacancy formation energies of all oxygen sites in a given host are computed, $\{\widehat{\Delta H}_{V_O,i}\}_{i\in\mathcal{C}_h}$, and their range is extracted, $\mathrm{rng}(\widehat{\Delta H}_{V_O})$. The target vacancy defect formation energy range required for TCH operation is typically considered to be between $\Delta H_{V_O}^{\mathrm{TCH}} = [2.3, 4.0]$ eV. At the cost of increasing false positives, this range is judiciously selected to reduce false negatives arising from underlying dGNN prediction errors, systematic errors in DFT, and the omission (necessary for high-throughput DFT) of more complicated but possibly non-

negligible contributions to the free energy of vacancy formation.[10] Increased configurational entropy of the defect state under TCH conditions will occur for materials where $\mathrm{rng}(\widehat{\Delta H}_{V_O})$ is small and within $\Delta H_{V_O}^{\mathrm{TCH}}$. However, materials with only one defect in the TCH target are still considered, provided there are no defects below the lower bound, i.e., $\min(\widehat{\Delta H}_{V_O}) \in \Delta H_{V_O}^{\mathrm{TCH}}$.

Second, for a given oxygen chemical potential,

$$\mu_O = \mu_O^{\mathrm{ref}} + \Delta\mu_O, \tag{4}$$

we compute each structure's grand canonical energy above the hull, ϕ_H, using the MP20-provided formation energy mixing scheme. This yields the chemical potential *range*,

$$\Delta\mu_O^{\phi_H \leq \Theta} = [\Delta\mu_O | \phi_H(\Delta\mu_O) \leq \Theta], \tag{5}$$

for which the material is within some threshold, Θ, of the grand canonical hull energy. The material should be stable ($\phi_H = 0$) in a $\Delta\mu_O$ range that intersects or, preferably, spans the entire range of typical TCH operating conditions, taken here as $\Delta\mu_O^{\mathrm{TCH}} = [-3.0, -2.5]$ eV.[37,38] Nonetheless, MP20-predicted metastable materials can still be considered (especially those that are experimentally known). To facilitate down-selection of candidates for experiments, we therefore compute a scalar value, the average ϕ_H across $\Delta\mu_O^{\mathrm{TCH}}$,

$$\bar{\phi}_H^{\mathrm{TCH}} = \frac{\int_{\Delta\mu_O^{\mathrm{TCH}}} \phi_H(\Delta\mu_O) d\Delta\mu_O}{\int_{\Delta\mu_O^{\mathrm{TCH}}} d\Delta\mu_O}, \tag{6}$$

where materials with $\bar{\phi}_H^{\mathrm{TCH}}$ closer to 0 are more highly prioritized.

Requiring a lower $\bar{\phi}_H^{\mathrm{TCH}}$ reduces the number of potential candidates drastically (by several orders of magnitude when requiring $\bar{\phi}_H^{\mathrm{TCH}} = 0$), and this reduction is chemistry dependent (Figure 2b). For materials anywhere on the grand canonical hull, i.e., $\Delta\mu_O^{\phi_H \leq \Theta} \neq \emptyset$, Figure 2c shows the minimum $\Delta\mu_O$ at which they are stable vs. the minimum vacancy formation energy from the dGNN screening. Therefore the screening recovers, as one might expect, a reasonable correlation between defect stability and the minimum oxygen chemical potential at which a stable host structure becomes unstable, i.e., decomposes to more reduced metal oxide(s). This naturally occurring compensation effect, unfortunately for TCH discovery, imparts the large majority of screened metal oxides with $\min(\widehat{\Delta H}_{V_O})$ that is too high for the required TCH stability range. Specifically, for those materials with $\min(\widehat{\Delta H}_{V_O})$ in the appropriate TCH range, most are not stable enough under reducing conditions, i.e., $\min(\Delta\mu_O^{\phi_H=0}) > -2.5$ eV. This motivates future investigation of SOTA generative machine learning approaches to crystal structure prediction, such as conditioned diffusion models,[39] to expand the possible space of TCH oxide candidates beyond the limitations of currently known materials space.

Third, we average our heuristic uncertainty metric across all oxygen defects in a given host structure, $\bar{U} = \langle\{\sigma(\widehat{\Delta H}_{V_O,i})\}_{i \in \mathcal{C}_h}\rangle$. Since large values are well correlated with large prediction errors, we can additionally avoid experimental validation efforts on materials with large $\bar{U}$ to reduce the likelihood of false positives in the MP screening with the current dGNN models. Unsurprisingly, if we further average $\bar{U}$ over all MP structures containing a given cation, $\langle\{\bar{U}\}_{\mathrm{MP}}\rangle$, the largest uncertainties are broadly associated with the cations that are the least well represented in the training data (Figure 2a).

Final candidate selection

Table 2 lists materials that were selected from the MP screening for experimental validation and their predicted properties for our down-selection criteria. We additionally state the results of our experimental validation protocol: whether they successfully pass the TGA redox cycling test, whether they successfully pass the SFR water-splitting cycling test, and whether our down-selection criteria provided a true/false positive/negative classification (TP/FP/TN/FN) of the water-splitting capability. Specifically, a "positive" is assigned if a material displays reversible non-stoichiometry ($> 50\%$ $\Delta\delta$ that of CeO_2 under similar conditions) in the TGA experiments for at least 3 redox cycles *and* stable hydrogen production in the SFR water-splitting experiments for at least 5 redox cycles.

Only materials with $2.3 < \min(\widehat{\Delta H}_{V_O}) < 4.0$ eV, $\min(\Delta\mu_O^{\phi_H=0}) < -2.5$ eV, and $\bar{U} < 0.75$ eV were considered. The exceptions are $TiNb_2O_7$ and Ba_2CaMoO_6, which were selected as examples that do not meet our vacancy formation criteria (i.e., we tried to validate a true negative prediction for water-splitting ability). Since on the order of one hundred candidates still fulfill these criteria, further prioritization was performed based on experimental synthesizability and handling considerations, including demonstrated synthesis and stability in literature reports, non-hazardous precursors, etc. Note the abundance of Ba-containing oxides in our selected candidates is partly reflective of their over-representation in MP candidates satisfying our strictest host stability down-select criteria (Figure 2b), while additional selection criteria based on chemical intuition are provided in the SI. X-ray diffraction (XRD) characterization and verification of the intended as-synthesized phases is shown in the SI, as well as the post-TGA cycled XRD.

TGA redox cycle screening

First, each material was screened using TGA to assess its suitability for flow reactor experiments. The mass change of each material was recorded over three redox cycles. Oxidation conditions for each measurement were consistently set to 1000 °C (T_{ox}) and an oxygen partial pressure ($pO_{2,ox}$) of $10^{-0.76}$ atm. Reduction tempera-

Table 2: Down-selected, synthesized, and tested materials' MP identifier (MPID), chemical formula, and screening down-selection criteria of rng($\widehat{\Delta H}_{V_O}$), $\bar{U}$, $\bar{\phi}_H^{\text{TCH}}$, and $\Delta\mu_O^{\phi_H=0}$. TGA = {Yes, No} indicates whether reversible non-stoichiometry (> 50% $\Delta\delta$ that of CeO_2 under similar conditions) in the TGA experiments was observed for at least 3 redox cycles. SFR = {Yes, No} indicates whether the material exhibited stable redox cycling in SFR water-splitting experiments after five cycles. Rating = {TP, FP, TN, FN} indicates a true/false postive/negative classification of our down selection criteria's prediction of successful water-splitting capabilities in the SFR. DFT = {TP, FP, TN, FN} indicates the Rating when hybrid DFT predictions (see Table 4), rather than dGNN, for rng($\widehat{\Delta H}_{V_O}$) are used.

MPID	Formula	rng($\widehat{\Delta H}_{V_O}$) [eV]	$\bar{U}$ [eV]	$\bar{\phi}_H^{\text{TCH}}$ [eV/at]	$\Delta\mu_O^{\phi_H=0}$ [eV]	TGA	SFR	Rating	DFT
mp-19154	$BaFe_2O_4$	3.3 - 3.6	0.1	0.038	$[-2.6, -0.9]$	Y†,§	Y	TP	–
mp-1196071	$Ba_2Fe_2O_5$	2.9 - 3.8	0.1	0.0	$[-3.1, -0.9]$	Y†,§	Y	TP	–
mp-1228690	$Ba_6La_2Fe_4O_{15}$	3.6 - 4.5	0.3	0.0	$[-3.3, -0.8]$	Y	Y	TP	–
mp-1228552	$Ba_5SrLa_2Fe_4O_{15}$	3.7 - 4.4	0.3	0.0	$[-3.3, -0.6]$	Y	Y	TP	–
mp-1198058	$Ba_3YFe_2O_{7.5}$	3.8 - 4.5	0.4	0.0	$[-3.3, -0.5]$	Y	Y	TP	–
–	$Ba_{2.5}Sr_{0.5}YFe_2O_{7.5}$*	–	–	–	–	Y	Y	TP	–
mp-1218451	$Sr_3PrMn_2O_8$	2.4 - 3.3	0.5	0.011	$[-2.8, -1.5]$	Y	Y	TP	–
mp-19403	Ba_2CaMoO_6	4.5	0.6	0.0	$[-3.7, 0.0]$	N†	–	TN	–
mp-18967	Ba_2CoMoO_6	3.6	0.6	0.0	$[-3.1, -1.5]$	N†	–	FP	–
mp-769971	$BaGa_4O_7$	3.5 - 4.1	0.7	0.0	$[-3.9, 0.0]$	N	–	FP	TN
mp-17914	$CaGa_4O_7$	3.8 - 4.3	0.5	0.0	$[-3.9, 0.0]$	N	–	FP	TN
mp-759307	$TiNb_2O_7$	4.6 - 5.7	0.3	0.018	–	–	–	TN	–

* Not in MP, inferred by chemically intuited substitution analogous to mp-1228552.

† Indicates melting instability before the maximum reduction temperature of 1400 °C

§ Successful TGA redox cycling still achieved at a milder reduction temperature < 1400 °C

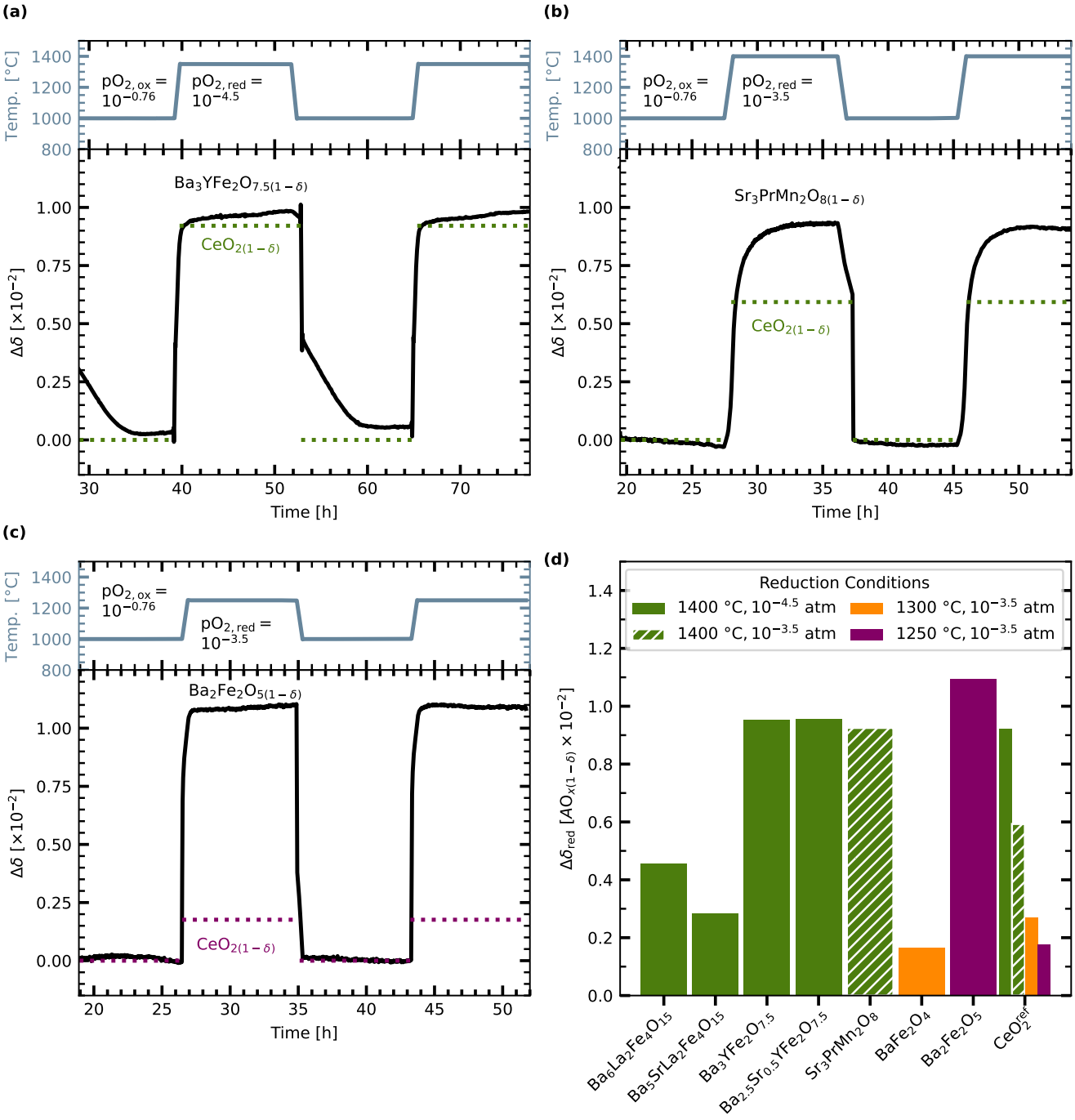

Figure 3: Reversible oxygen non-stoichiometry, expressed as $\Delta\delta$ vs. time in a TGA redox screening experiment for (a) $Ba_3YFe_2O_{7.5}$, (b) $Sr_3PrMn_2O_8$, and (c) $Ba_2Fe_2O_5$ respectively. Theoretical results for CeO_2 are also reported as colored dashed lines. The temperature profiles and pO_2s are indicated in the upper subplots. (d) The average maximum $\Delta\delta_{red}$ obtained from the last two redox cycles for screened materials under various reduction conditions. Theoretical $\Delta\delta_{red}$ for CeO_2[40] are also included.

tures (T_{red}) varied between 1250 and 1400 °C depending on the thermal stability of the material, which was evaluated in a furnace ex-situ. The oxygen partial pressure ($pO_{2,red}$) was set to either $10^{-4.5}$ or $10^{-3.5}$ atm, based on the capability of the instrument used. The first redox cycle was excluded from analysis due to the irreversible mass change resulting from "burn in effects" likely related to particle coarsening. Reversible mass changes in the remaining two cycles were attributed to changes in oxygen non-stoichiometry and can be expressed as the change (Δ)

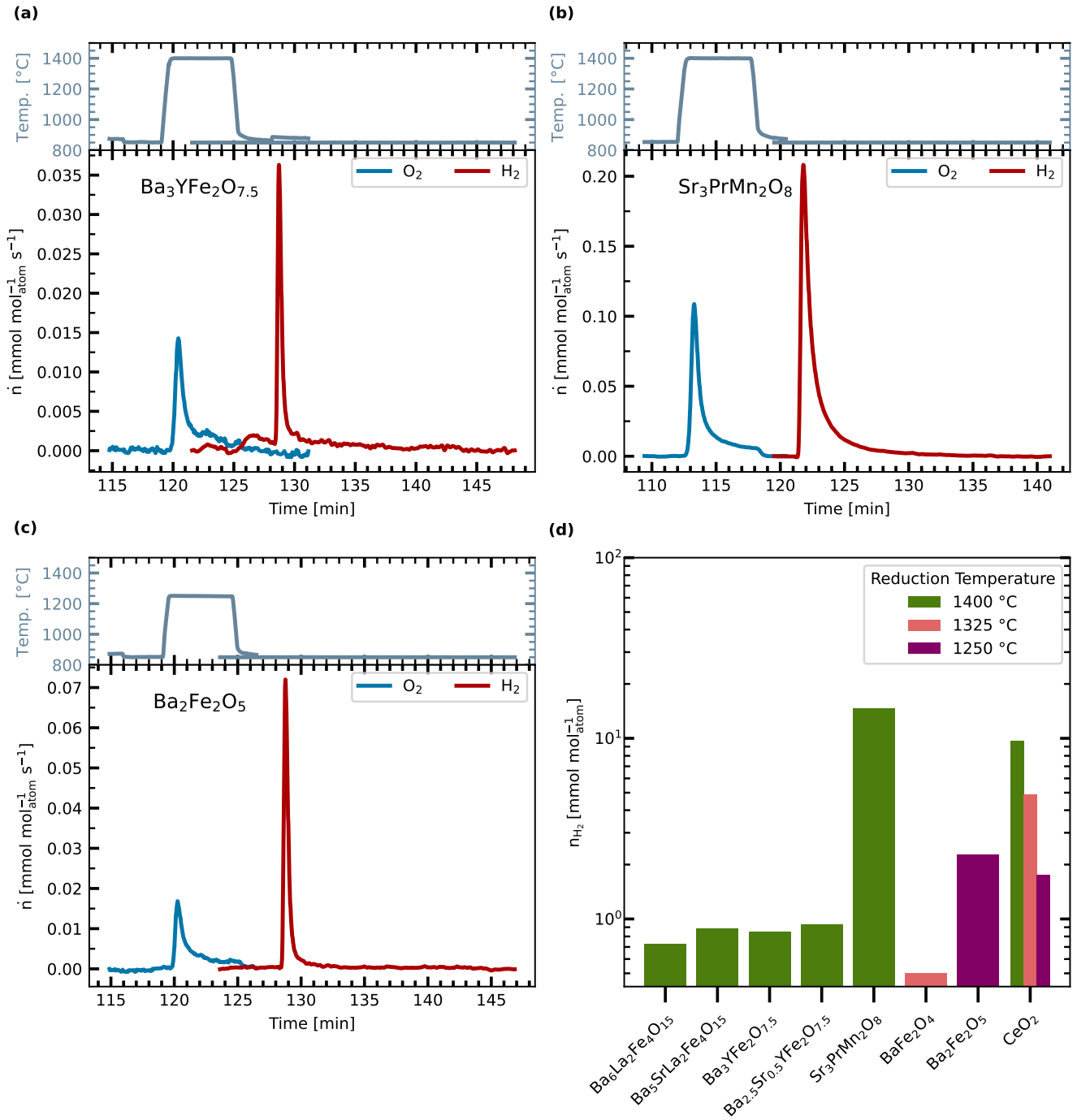

Figure 4: H_2 and O_2 production rates from the fifth and final water-splitting cycle of an SFR experiment for (a) $Ba_3YFe_2O_{7.5}$, (b) $Sr_3PrMn_2O_8$, and (c) $Ba_2Fe_2O_5$ respectively. (d) Average H_2 yield obtained from the last four water-splitting cycles for screened materials and CeO_2 under various T_{red}.

in δ by assuming $\Delta\delta$ reaches zero upon equilibration of the oxidation step of the second cycle to enable comparisons between materials with varying amounts of oxygen (see SI for details).

Figure 3a-c illustrates the oxygen non-stoichiometry results for three notable materials that passed the TGA screening (i.e. exhibited reversible mass changes and a $\Delta\delta$ >50% than that of CeO_2), $Ba_3YFe_2O_{7.5}$, $Sr_3PrMn_2O_8$, and $Ba_2Fe_2O_5$. Theoretically predicted oxygen non-stoichiometries for CeO_2,[40] using the identical redox conditions at which each material was measured, are shown as dashed colored lines for comparison in each subplot. The average maximum $\Delta\delta$ during the last two reduction cycles for each material passing the TGA screening

is reported in Figure 3d, along with the theoretical results for CeO_2. The reduction of $Ba_3YFe_2O_{7.5}$ resulted in a $\Delta\delta$ similar to that of CeO_2, while $Sr_3PrMn_2O_8$ and $Ba_2Fe_2O_5$ exhibited ∼50% and ∼500% larger $\Delta\delta$, respectively, than CeO_2 under the same conditions. Furthermore, visual inspection of the rate at which $\Delta\delta$ approaches zero indicates that the (re)oxidation kinetics of $Sr_3PrMn_2O_8$ and $Ba_2Fe_2O_5$ are notably faster than $Ba_3YFe_2O_{7.5}$. Within the 4 self-similar compounds based on monoclinic $Ba_3YFe_2O_{7.5}$ and hexagonal $Ba_6La_2Fe_4O_{15}$ structures, the two monoclinic materials yield $\Delta\delta$ values > 2x higher than the hexagonal counterparts. Sr substitution does not appear to impact $\Delta\delta$ within the monoclinic structure, while having a somewhat detrimental effect on hexagonal $Ba_5SrLa_2Fe_4O_{15}$. TGA cycling data for all materials are shown in the SI, including for materials with large irreversible mass changes (Ba_2CoMoO_6 and Ba_2CaMoO_6), which are likely due to cation volatility and/or melting instability. The materials that did not pass the TGA screen were not studied in the SFR.

SFR water-spitting screening

The TCH performance of any material passing the TGA screening was examined in a stagnation flow reactor (SFR), the details of which are given elsewhere.[5,41,42] Briefly, the SFR operates at sub-ambient pressure (0.1 atm) and is equipped with electromechanical systems for control of gas composition and mass flow, reactor back pressure, and process temperature. Material oxidation temperature (T_{ox}) is held fixed by placing the SFR into a tube furnace. Materials are rapidly heated at a rate of ∼15 °C/s from T_{ox} to the thermal reduction temperature (T_{red}) by focusing the radiant energy of a near-infrared diode laser normal to the sample surface. Gas composition (and thus hydrogen and oxygen production rate, $\dot{n}_{H_2}$ and $\dot{n}_{O_2}$) is measured downstream of the SFR by a mass spectrometer (with additional details presented in the SI). Prior to water-splitting experiments, each sample was first subjected to five redox cycles under a constant $P_{O_2} = 1.97 \cdot 10^{-4}$ atm to verify that it could in fact thermochemically cycle in a stable and consistent manner and remove any contributions from burn-in effects or sintering from the capacities measured during water-splitting redox.

Then, each sample underwent five water-splitting cycles according to conditions listed in Table 3. Under these conditions, samples typically reach stable cycling behavior within five cycles. The H_2 yield, n_{H_2}, is calculated by integrating $\dot{n}_{H_2}$ over t_{ox}. The $\dot{n}_{H_2}$ and $\dot{n}_{O_2}$ rates during the last SFR water-splitting cycle are shown in Figure 4a-c for $Ba_3YFe_2O_{7.5}$, $Sr_3PrMn_2O_8$, and $Ba_2Fe_2O_5$, while the same data for each material is presented in the SI. The average atom-normalized molar hydrogen yield measured over the last four cycles for all materials, at various T_{red}, is summarized in Figure 4d. Each material that passed the TGA screening test maintains a measurable hydrogen yield after five cycles under the conditions investigated, with the caveat that $T_{red} < 1400$ °C must be applied for $BaFe_2O_4$ and $Ba_2Fe_2O_5$ due to their melt stability issues at 1400 °C.

Table 3: Thermodynamic cycle (reduction and oxidation) conditions used to evaluate TCH water-splitting performance of screened materials in the SFR.

	Condition	Value
Reduction	Temperature (T_{red})	See Figure 4d
	O_2 Partial Pressure ($P_{O_2,red}$)	$9.87 \cdot 10^{-7}$ atm
	Time (t_{red})	5 min
Oxidation	Temperature (T_{ox})	850 °C
	H_2O Partial Pressure ($P_{H_2O,ox}$)	$3.95 \cdot 10^{-2}$ atm
	H_2 Partial Pressure ($P_{H_2,ox}$)	0 atm
	Time (t_{ox})	15 min

Among the screened materials, two are of particular interest, namely $Sr_3PrMn_2O_8$ and $Ba_2Fe_2O_5$. For $Sr_3PrMn_2O_8$, the H_2 yield is 30% greater than that of CeO_2 for the same $T_{red} = 1400$ °C. Meanwhile, $Ba_2Fe_2O_5$ displays a H_2 yield that is slightly greater than CeO_2 under the lowest T_{red} of 1250 °C examined in this work. The promising performance in these low T_{red} conditions motivates future work to modify the material via alloying to melt stabilize the material at higher temperatures and to improve material performance by enhancing defect concentration to increase H_2 production capacity. Furthermore, these materials exhibit promising kinetics with a significant amount of oxidation/reduction occurring between 15 and 5 mins. However, further kinetic studies are need to quantify rates under controlled sample particle size, surface area, and surface morphology. Finally, we note that CeO_2 is unique in its ability to split water at a high $P_{H_2,ox}/P_{H_2O,ox}$ ratio; while beyond the scope of this initial screening paper, re-examination of these materials under such conditions, especially for the promising materials like $Sr_3PrMn_2O_8$ and $Ba_2Fe_2O_5$, will be probed in future work.

DFT for uncertain screening predictions

Ga and Pr are among the least-sampled cations in our augmented training dataset. Due to their relatively higher $\bar{U}$ (Table 2) and the much larger expected MAE in element-wise CV (Table 1), we can reasonably expect the dGNN predictions for {Ca,Sr,Ba}Ga_4O_7 and $Sr_3PrMn_2O_8$ to exhibit higher error than other screened materials. In such cases, re-computing ΔH_{V_O} with DFT can provide a much higher-fidelity (albeit much higher-cost) prediction and possibly correct dGNN-predicted FPs to DFT-predicted TNs. We report min(ΔH_{V_O}) for {Ca, Sr,Ba}Ga_4O_7 (as well as some substituted variants)

and $Sr_3PrMn_2O_8$ in Table 4. Both the generalized gradient approximation of Perdew, Burke, and Ernzerhof (PBE-GGA)[27] and the higher fidelity hybrid exchange-correlation functional of Heyd, Scuseria, and Ernzerhof (HSE06) were used,[31,32] with additional details in Supplementary Section 1. As with the training data for the dGNN model, formation energies were calculated using Equation (3).

Table 4: DFT-computed $\min(\Delta H_{V_O})$ for selected screening candidates from this study, including hypothetical In→Ga substituted variants of {Ca,Sr,Ba}Ga_4O_7.

Material	DFT Method	$\min(\Delta H_{V_O})$ [eV]
$CaGa_4O_7$	HSE, $\alpha = 0.27$	4.17
	PBE-GGA	3.78
$SrGa_4O_7$	HSE, $\alpha = 0.32$	4.03
	PBE-GGA	3.59
$BaGa_4O_7$	HSE, $\alpha = 0.30$	3.91
	PBE-GGA	3.50
$CaIn_4O_7$	HSE, $\alpha = 0.27$	2.94
	PBE-GGA	2.60
$SrIn_4O_7$	HSE, $\alpha = 0.32$	3.02
	PBE-GGA	1.49
$BaIn_4O_7$	HSE, $\alpha = 0.30$	2.38
	PBE-GGA	1.93
$Sr_3PrMn_2O_8$	PBE-GGA	2.62

Beginning with {Ca,Sr,Ba}Ga_4O_7, we calculate ΔH_{V_O} with both PBE-GGA and HSE. Consistent with the dGNN model, the screening predictions (Table 2) are within the expected MAE of the PBE-GGA calculations. However, we find that the HSE-predicted formation energies are systematically larger than both dGNN and PBE-GGA (origins of this discrepancy are discussed in more details in the SI). The HSE-predicted $\min(\Delta H_{V_O})$ are approximately equal to or above our 4 eV threshold to be considered candidates for TCH applications, so recomputing with HSE (but not GGA) would have corrected the dGNN-predicted FP to a TN before attempted experimental validation.

Noting that the formation energy decreases with increasing *A*-site cation size (i.e., from Ca to Ba), we also considered whether substitution on the *B*-site might result in systems with more modest formation energies. Indeed, with further HSE calculations, we find that the In-containing analogues have formation energies that are significantly lower in energy. Notably, the $\min(\Delta H_{V_O})$ in $BaIn_4O_7$ is more than 1.5 eV lower in energy than that of $BaGa_4O_7$, thereby making vacancy formation likely much more favorable for TCH applications. $CaIn_4O_7$ and $SrIn_4O_7$, on the other hand, have formation energies well within the desired range. Unfortunately, while the Ga-containing compounds are thermodynamically stable with respect to limiting phases, the In-containing systems are not (see Supplementary Information); thus, these lower formation energies are not likely to be realized in practice. Nevertheless, these results suggest that In alloying in {Ca,Sr,Ba}Ga_4O_7 may be a promising strategy to reduce ΔH_{V_O}, providing that sufficient substitution of In for Ga can be realized.

For $Sr_3PrMn_2O_8$ we find $\min(\Delta H_{V_O}) = 2.62$ eV, which is slightly larger than the dGNN prediction but within both the expected model error and the range of formation energies targeted for TCH down-selection. As a result, $Sr_3PrMn_2O_8$ is a TP when using either DFT or dGNN predictions, despite the lack of data on Pr-containing compounds. Considering that $Sr_3PrMn_2O_8$ and related materials are candidates for electrodes in solid oxide cells,[43] it is unsurprising that formation energies are relatively modest.

As a final note, it is important to recognize that, in materials with a band gap, defects can be charged, depending on the position of the Fermi level relative to the band edges. V_O defects are commonly ionized in the +2 or +1 charge states (V_O^{+2} or V_O^{+1}, respectively), particularly for Fermi level positions close to the valence band maximum (VBM). Here, for the purposes of our screening approach, we focus solely on the neutral charge state, V_O^0; however, we have confirmed that, in each material in the {Ca,Sr,Ba}Ga_4O_7 family, V_O^0 is in fact the lowest energy V_O species at typical Fermi level positions. We include a full analysis to support this point in the SI. For $Sr_3PrMn_2O_8$, which does not have a band gap, all defects have neutral charge states. Therefore, our reported values here for neutral vacancies are representative of true ΔH_{V_O}.

Conclusions

A comprehensive TCH materials down-selection was performed, combining vacancy defect formation energy predictions from dGNN, thermochemical redox stability range calculated from MP20, and chemical intuition to target high priority phases for experimental validation. Building upon previous work, the dGNN was trained on additional oxide defect data (generated in this work) in less-conventional TCH chemistry spaces, expanding the diversity of compounds that could be targeted. Using the dGNN predicted vacancy formation energies, ~80% accuracy was achieved in predicting successful two-step TCH metal oxides, which we define as exhibiting reversible, stable, and cyclable hydrogen production in a stagnation flow reactor under at least five redox cycles.

For less certain dGNN screening predictions, additional DFT validation calculations were performed to predict vacancy formation energies. Corresponding to the level of theory at which the dGNN was trained, some PBE-GGA DFT calculations did not correct dGNN-predicted false positives; i.e., both the dGNN- and DFT-predicted

vacancy formation energy were within the target down-select range, but no oxygen off-stoichiometry could be observed in TGA screening. However, hybrid HSE calculations corrected these false positives to true negatives; i.e., the vacancy formation energies were shown to be too high, explaining the lack of off-stoichiometry observed in the TGA screening. Two materials from the experimental screening failed due to factors outside the scope of our computational screening capabilities, namely thermal stability issues like melting and (hypothesized) cation volatility leading to irreversible mass loss at TCH operating temperatures.

Beyond simply considering whether down-selected candidates could cycle and produce hydrogen in our SFR screening, two particularly interesting materials were identified based on their SFR performance in comparison with CeO_2, the SOTA TCH material. For $Sr_3PrMn_2O_8$, we observed ∼30% higher hydrogen yield relative to CeO_2 when cycled under a high reduction temperature of $T_{red} = 1400$ °C. Note that this behavior is reasonably common amongst perovskite oxides when reoxidizing in pure steam.[4,5] Perhaps more interesting and less typical was the increased hydrogen yield of $Ba_2Fe_2O_5$ relative to CeO_2 when cycled at substantially reduced $T_{red} = 1250$ °C. Thus, in addition to a high success rate of identifying possible water-splitters, this study identified a material exhibiting SOTA performance at milder T_{red}, which could potentially simplify future TCH reactor design and material requirements. Future work will involve more detailed explorations of these materials by probing their ability to reoxidize in higher H_2/H_2O ratios, more deeply investigating the origins of their promising water-splitting behavior, and optimizing their performance through substitution, alloying, or other modifications.

Acknowledgements

The authors gratefully acknowledge research support from the HydroGEN Advanced Water Splitting Materials Consortium, established as part of the Energy Materials Network under the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Fuel Cell Technologies Office. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration (DOE/NNSA) under contract DE-NA0003525. This written work is authored by an employee of NTESS. The employee, not NTESS, owns the right, title and interest in and to the written work and is responsible for its contents. Any subjective views or opinions that might be expressed in the written work do not necessarily represent the views of the U.S. Government. The publisher acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this written work or allow others to do so, for U.S. Government purposes. The DOE will provide public access to results of federally sponsored research in accordance with the DOE Public Access Plan. The National Renewable Energy Laboratory (NREL) is operated for the DOE under Contract No. DE-AC36-08GO28308. The work at the Lawrence Livermore National Laboratory was performed under the auspices of the U.S. Department of Energy (DOE) under Contract No. DE-AC52-07NA27344. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515 This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Fuel Cell Technologies Office (FCTO) under Award Number DE-EE0010733.

References

(1) Muhich, C. L.; Ehrhart, B. D.; Al-Shankiti, I.; Ward, B. J.; Musgrave, C. B.; Weimer, A. W. A review and perspective of efficient hydrogen generation via solar thermal water splitting. *Wiley Interdisciplinary Reviews: Energy and Environment* **2016**, *5*, 261–287.

(2) Gorman, B. T.; Lanzarini-Lopes, M.; Johnson, N. G.; Miller, J. E.; Stechel, E. B. Techno-Economic Analysis of a Concentrating Solar Power Plant Using Redox-Active Metal Oxides as Heat Transfer Fluid and Storage Media. *Front. Energy Res.* **2021**, *9*, 1–18.

(3) Worku, A. K.; Ayele, D. W.; Deepak, D. B.; Gebreyohannes, A. Y.; Agegnehu, S. D.; Kolhe, M. L. Recent Advances and Challenges of Hydrogen Production Technologies via Renewable Energy Sources. *Advanced Energy and Sustainability Research* **2024**, *5*, 2300273.

(4) McDaniel, A. H.; Miller, E. C.; Arifin, D.; Ambrosini, A.; Coker, E. N.; O'Hayre, R.; Chueh, W. C.; Tong, J. Sr- and Mn-doped $LaAlO_{3-\delta}$ for solar thermochemical H_2 and CO production. *Energy Environ. Sci.* **2013**, *6*, 2424.

(5) R. Barcellos, D.; Sanders, M. D.; Tong, J.; McDaniel, A. H.; O'Hayre, R. P. $BaCe_{0.25}Mn_{0.75}O_{3-\delta}$ - a promising perovskite-type oxide for solar thermochemical hydrogen production. *Energy Environ. Sci.* **2018**, *11*, 3256–3265.

(6) Qian, X.; He, J.; Mastronardo, E.; Baldassarri, B.; Wolverton, C.; Haile, S. M. Favorable Redox Thermodynamics of $SrTi_{0.5}Mn_{0.5}O_{3-\delta}$ in Solar Thermochemical Water Splitting. *Chem. Mater.* **2020**, *32*, 9335–9346.

(7) Zhang, D. et al. Compositionally Complex Perovskite Oxides for Solar Thermochemical Water Splitting. **2022**.

(8) Wexler, R. B.; Sai Gautam, G.; Bell, R. T.; Shulda, S.; Strange, N. A.; Trindell, J. A.; Sugar, J. D.; Nygren, E.; Sainio, S.; McDaniel, A. H., et al. Multiple and nonlocal cation redox in Ca–Ce–Ti–Mn oxide perovskites for solar thermochemical applications. *Energy Environ. Sci.* **2023**, *16*, 2550–2560.

(9) Chueh, W. C.; Falter, C.; Abbott, M.; Scipio, D.; Furler, P.; Haile, S. M.; Steinfeld, A. High-flux solar-driven thermochemical dissociation of CO_2 and H_2O using nonstoichiometric ceria. *Science* **2010**, *330*, 1797–1801.

(10) Lany, S. Chemical Potential Analysis as an Alternative to the van't Hoff Method: Hypothetical Limits of Solar Thermochemical Hydrogen. *Journal of the American Chemical Society* **2024**, *146*, 14114–14127.

(11) Deml, A. M.; Holder, A. M.; O'Hayre, R. P.; Musgrave, C. B.; Stevanović, V. Intrinsic Material Properties Dictating Oxygen Vacancy Formation Energetics in Metal Oxides. *J. Phys. Chem. Lett.* **2015**, *6*, 1948–1953.

(12) Wexler, R. B.; Gautam, G. S.; Stechel, E. B.; Carter, E. A. Factors Governing Oxygen Vacancy Formation in Oxide Perovskites. *J. Am. Chem. Soc.* **2021**, *143*, 13212–13227.

(13) Bare, Z. J. L.; Morelock, R. J.; Musgrave, C. B. A Computational Framework to Accelerate the Discovery of Perovskites for Solar Thermochemical Hydrogen Production: Identification of Gd Perovskite Oxide Redox Mediators. *Advanced Functional Materials* **2022**, *32*, 2200201.

(14) Baldassarri, B.; He, J.; Qian, X.; Mastronardo, E.; Griesemer, S.; Haile, S. M.; Wolverton, C. Accuracy of DFT computed oxygen-vacancy formation energies and high-throughput search of solar thermochemical water-splitting compounds. *Physical Review Materials* **2023**, *7*, 065403.

(15) Witman, M. D.; Goyal, A.; Ogitsu, T.; McDaniel, A. H.; Lany, S. Defect graph neural networks for materials discovery in high-temperature clean-energy applications. *Nat. Comput. Sci.* **2023**, *3*, 675–686.

(16) Wilson, S. A.; Muhich, C. L. A Bayesian method for selecting data points for thermodynamic modeling of off-stoichiometric metal oxides. *Journal of Materials Chemistry A* **2024**, *12*, 13328–13337.

(17) Wilson, S. A.; Stechel, E. B.; Muhich, C. L. Overcoming significant challenges in extracting off-stoichiometric thermodynamics using the compound energy formalism through complementary use of experimental and first principles data: A case study of $Ba_{1-x}Sr_xFeO_{3-\delta}$. *Solid State Ionics* **2023**, *390*, 116115.

(18) Guan, P.; Debusschere, B.; Bishop, S. R.; Witman, M.; McDaniel, A. H. LASSO for CALPHAD model selection enables data-efficient thermodynamic modeling: an application in thermochemical hydrogen production materials, 2024.

(19) Goyal, A.; Sanders, M. D.; O'Hayre, R. P.; Lany, S. Predicting Thermochemical Equilibria with Interacting Defects: $Sr_{1-x}Ce_xMnO_{3-\delta}$ Alloys for Water Splitting. *PRX Energy* **2024**, *3*, 013008.

(20) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013**, *1*, 011002.

(21) Stevanovi ć, V.; Lany, S.; Zhang, X.; Zunger, A. Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies. *Phys. Rev. B* **2012**, *85*, 115104.

(22) Sharan, A.; Lany, S. Computational discovery of stable and metastable ternary oxynitrides. *J. Chem. Phys.* **2021**, *154*, 234706.

(23) Lany, S. Semiconductor thermochemistry in density functional calculations. *Phys. Rev. B* **2008**, *78*, 245207.

(24) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **1964**, *136*, B864.

(25) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

(26) Kresse, G.; Furthmüller, J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169.

(27) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(28) Dudarev, S. L.; Botton, G. A.; Savrasov, S. Y.; Humphreys, C. J.; Sutton, A. P. Electron-energy-loss spectra and the structural stability of nickel oxide: An LSDA+U study. *Phys. Rev. B* **1998**, *57*, 1505–1509.

(29) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **1994**, *50*, 17953–17979.

(30) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, *59*, 1758–1775.

(31) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **2003**, *118*, 8207–8215.

(32) Heyd, G. E.; Scuseria, J.; Ernzerhof, M. Erratum: "Hybrid functionals based on a screened Coulomb potential"[J. Chem. Phys. 118, 8207 (2003)]. *J. Chem. Phys.* **2006**, *124*, 219906.

(33) Heyd, J.; Scuseria, G. E. Efficient hybrid density functional calculations in solids: Assessment of the Heyd–Scuseria–Ernzerhof screened Coulomb hybrid functional. *J. Chem. Phys.* **2004**, *121*, 1187–1192.

(34) Witman, M. D.; Schindler, P. MatFold: systematic insights into materials discovery models' performance through standardized cross-validation protocols. *Digital Discovery* **2025**, *4*, 625–635.

(35) Kumagai, Y.; Tsunoda, N.; Takahashi, A.; Oba, F. Insights into oxygen vacancies from high-throughput first-principles calculations. *Physical Review Materials* **2021**, *5*, 123803.

(36) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **2013**, *68*, 314–319.

(37) Lany, S. Communication: The electronic entropy of charged defect formation and its impact on thermochemical redox cycles. *J. Chem. Phys.* **2018**, *148*, 071101.

(38) Millican, S. L.; Clary, J. M.; Musgrave, C. B.; Lany, S. Redox Defect Thermochemistry of $FeAl_2O_4$ Hercynite in Water Splitting from First-Principles Methods. *Chem. Mater.* **2022**, *34*, 519–528.

(39) Zeni, C.; Pinsler, R.; Zügner, D.; Fowler, A.; Horton, M.; Fu, X.; Wang, Z.; Shysheya, A.; Crabbé, J.; Ueda, S., et al. A generative model for inorganic materials design. *Nature* **2025**, 1–3.

(40) Zinkevich, M.; Djurovic, D.; Aldinger, F. Thermodynamic modelling of the cerium–oxygen system. *Solid State Ionics* **2006**, *177*, 989–1001.

(41) Arifin, D.; Aston, V. J.; Liang, X.; McDaniel, A. H.; Weimer, A. W. CoFe2O4 on a porous Al2O3 nanostructure for solar thermochemical CO2 splitting. *Energy & Environmental Science* **2012**, *5*, 9438.

(42) Scheffe, J. R.; McDaniel, A. H.; Allendorf, M. D.; Weimer, A. W. Kinetics and mechanism of solar-thermochemical H2 production by oxidation of a cobalt ferrite–zirconia composite. *Energy & Environmental Science* **2013**, *6*, 963.

(43) Bahout, M.; Tonus, F.; Prestipino, C.; Pelloquin, D.; Hansen, T.; Fonda, E.; Battle, P. D. High-temperature redox chemistry of $Pr_{0.5}Sr_{1.5}Cr_{0.5}Mn_{0.5}O_{4-\delta}$ investigated in situ by neutron diffraction and X-ray absorption spectroscopy under reducing and oxidizing gas flows. *J. Mater. Chem.* **2012**, *22*, 10560–10570.