















Cite this: DOI: 10.1039/d6sc03089k

 All publication charges for this article have been paid for by the Royal Society of Chemistry

A unified descriptor framework for hydrogen storage capacity and equilibrium pressure in interstitial hydrides

Seong-Hoon Jang, ^{*ab} Di Zhang, ^{ac} Xue Jia, ^a Hung Ba Tran, ^a Linda Zhang, ^{ac} Ryuhei Sato, ^d Yusuke Hashimoto, ^c Yusuke Ohashi, ^e Toyoto Sato, ^e Kiyoe Konno, ^f Shin-ichi Orimo ^{*ae} and Hao Li ^{*a}

Hydrogen is a promising energy carrier, yet its practical deployment is limited by the lack of storage materials that simultaneously achieve high storage capacity (w) and practical equilibrium pressure at room temperature ($P_{\text{eq,RT}}$). Interstitial metal hydrides offer fast kinetics and favorable thermodynamics (high $P_{\text{eq,RT}}$) but suffer from intrinsically low w . Here, we establish a physically interpretable, data-driven framework to uncover descriptor–property relationships in interstitial hydrides using a curated database of pressure-composition-temperature measurements (Digital Hydrogen Platform, *DigHyd*) and white-box symbolic regression. Strikingly, the analysis reveals a clear separation of governing mechanisms, in which w is governed by geometric and lattice conditions, captured by the average atomic radius ($\langle r_M \rangle$) and average thermal conductivity ($\langle \kappa \rangle$), with an optimal regime of $\langle r_M \rangle \sim 1.47 \text{ \AA}$ and relatively low $\langle \kappa \rangle$. In contrast, $P_{\text{eq,RT}}$ is governed by elastic properties, captured by the average shear modulus ($\langle G \rangle$) and average Poisson's ratio ($\langle \nu \rangle$), reflecting the role of lattice rigidity and mechanical compliance. These relationships are translated into compositional optimization pathways that follow the descriptor trends above, enabling the design of candidate materials with enhanced w under practical equilibrium conditions ($P_{\text{eq,RT}} \sim 0.1 \text{ MPa}$). This work establishes a general, interpretable strategy for physics-informed design of energy materials systems.

Received 14th April 2026
Accepted 24th May 2026

DOI: 10.1039/d6sc03089k

rsc.li/chemical-science

Introduction

Hydrogen is widely regarded as a key enabler for carbon-neutral energy systems, primarily due to its high energy density by mass and the absence of carbon emissions upon use.^{1,2} Despite these advantages, its widespread application in fuel cells and related technologies remains constrained by the lack of storage solutions that are simultaneously compact, safe, and reversible.³ Among the various approaches proposed to address this challenge, solid-state hydrogen storage in metal hydrides has attracted considerable attention because of its high volumetric density, cyclability, and compatibility with engineered systems.^{4–6} Metal hydrides, including representative systems such as MgH_2 , Mg_2NiH_4 , FeTiH_2 , $\text{PdH}_{0.6}$, and LaNi_5H_6 , span

a wide thermodynamic range depending on their bonding characteristics.^{7–15} Saline-type hydrides composed of light elements (*e.g.*, MgH_2) can achieve high gravimetric capacities but typically require elevated temperatures for hydrogen release,^{13,16} whereas interstitial hydrides based on transition or heavier metals (*e.g.*, LaNi_5H_6) exhibit fast kinetics and favorable equilibrium pressures, albeit with intrinsically limited hydrogen storage capacity.¹¹

Despite decades of extensive investigation, the compositional landscape of hydride-forming alloys remains far from fully explored. While a vast number of binary and multicomponent systems are theoretically accessible, only a limited fraction has been experimentally synthesized and evaluated. This challenge is further exacerbated by the absence of predictive frameworks that are both quantitatively accurate and physically interpretable, which hinders rational materials design. Although recent machine learning approaches have shown promise in accelerating property prediction,^{17–19} they often rely on relatively small or inconsistently curated datasets and employ black-box models that provide limited insight into the underlying physicochemical mechanisms. To overcome these limitations, we previously developed the Digital Hydrogen Platform (*DigHyd*: <https://www.dighyd.org>), a curated database constructed through large-scale extraction of experimental pressure-composition-temperature (PCT) data from the

^aAdvanced Institute for Materials Research (WPI-AIMR), Tohoku University, Sendai 980-8577, Japan. E-mail: jang.seonghoon.b4@tohoku.ac.jp; shin-ichi.orimo.a6@tohoku.ac.jp; li.hao.b8@tohoku.ac.jp

^bUnprecedented-scale Data Analytics Center, Tohoku University, Sendai 980-8578, Japan

^cFrontier Research Institute for Interdisciplinary Sciences (FRIS), Tohoku University, Sendai 980-8577, Japan

^dDepartment of Materials Engineering, The University of Tokyo, Tokyo 113-8656, Japan

^eInstitute for Materials Research, Tohoku University, Sendai, 980-8577, Japan

^fInstitute of Fluid Science, Tohoku University, Sendai, 980-8577, Japan



literature.^{20,21} Building on this dataset, symbolic regression was performed using a white-box modeling approach, enabling the construction of explicit relationships between materials descriptors and key hydrogen storage metrics, namely gravimetric capacity (w) and equilibrium pressure at room temperature ($P_{\text{eq,RT}}$).^{22,23} This approach identified a compact set of physically meaningful descriptors, including contributions from atomic mass, electronic structure, and packing characteristics, which govern hydrogen storage behavior. In particular, systems containing light elements were found to favor higher hydrogen capacity, whereas electronic and structural descriptors play a dominant role in determining equilibrium pressure. Within this descriptor space, beryllium (Be)-containing alloys emerged as promising candidates; however, their practical applicability is severely limited by toxicity and associated handling constraints.²⁴

These considerations motivate a shift toward more practical materials systems, as schematically illustrated in Fig. 1. Fig. 1a shows the construction of a curated database of PCT measurements, *DigHyd*, which serves as the foundation for this study. Building on this dataset, Fig. 1b presents the development of interpretable, white-box symbolic regression models that link materials descriptors to key hydrogen storage properties, w and $P_{\text{eq,RT}}$. Focusing on interstitial hydrides, we aim to identify the key descriptors governing both w and $P_{\text{eq,RT}}$ by extracting physically meaningful descriptors from the symbolic models (see Fig. 1c). Finally, as shown in Fig. 1d, these descriptor–property relationships are translated into materials design guidelines that enable enhanced w under practical operating conditions. In particular, we target equilibrium pressures near ambient conditions, $P_{\text{eq,RT}} \sim 0.1$ MPa, corresponding in this study to the window of $-1.5 \log_{10}[\text{MPa}] < \log_{10} P_{\text{eq,RT}} < -0.5 \log_{10}[\text{MPa}]$.

Results

Data curation and feature construction for interstitial hydrides

Often, PCT experiments involve multi-phase materials, which can introduce unintended noise in regression modeling. To mitigate this issue, we filtered the dataset and retained only single-phase or near-single-phase cases. Here, “near-single-phase” refers to materials in which a single phase accounts for more than 80 wt% of the total. As a result, the total number of entries is 706. However, not all entries include multi-temperature measurements required to determine $P_{\text{eq,RT}}$. Consequently, the number of data points (n_{data}) used for modeling w and $P_{\text{eq,RT}}$ are 706 and 299, respectively. This reduction primarily reflects the limited availability of consistent multi-temperature PCT measurements in the literature, despite the broader abundance of hydrogen capacity data. Within this context, the present framework is designed to extract robust and physically meaningful insights from limited but high-quality datasets. For BCC-type alloys exhibiting double plateaus, the plateau corresponding to practical operating conditions near ambient pressure was consistently selected, as the lower plateau is often located far below atmospheric pressure and is not practically accessible.

Fig. 2a shows the distribution of data points in the two-dimensional $w - P_{\text{eq,RT}}$ materials map for cases where multi-temperature PCT data are available. Most structural classes with different parent structures, such as Laves (C14), LaNi_5 , LaMgNi_4 , and TiFe , are located in the low- w region ($w < 2.5\%$), whereas the BCC class extends into the higher-capacity region ($2 < w < 5\%$). However, BCC metals and alloys generally exhibit two distinct plateaus in their PCT curves, with the lower plateau typically located far below atmospheric pressure at room

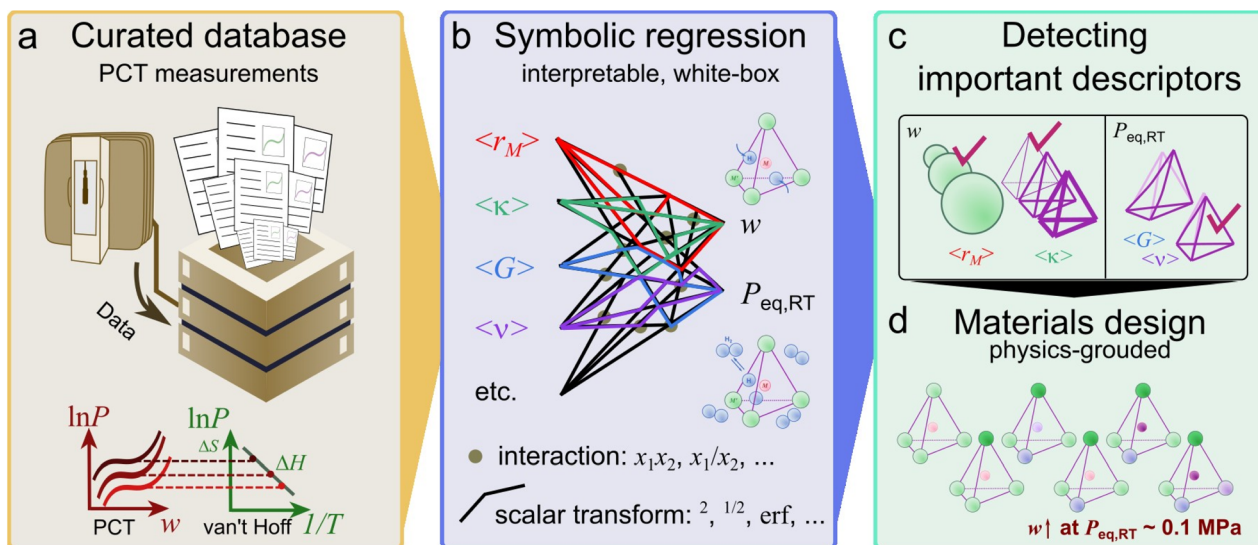


Fig. 1 Workflow of the present study for physics-informed materials design of interstitial hydrides. (a) Construction of a curated database of pressure-composition-temperature (PCT) measurements (*DigHyd*). (b) Development of interpretable, white-box symbolic regression models linking materials descriptors to hydrogen storage properties, including gravimetric capacity (w) and equilibrium pressure at room temperature ($P_{\text{eq,RT}}$). (c) Identification of key descriptors governing w and $P_{\text{eq,RT}}$ within the interstitial hydride space. (d) Translation of descriptor–property relationships into materials design guidelines, enabling the optimization of w under practical equilibrium pressure conditions ($P_{\text{eq,RT}} \sim 0.1$ MPa).



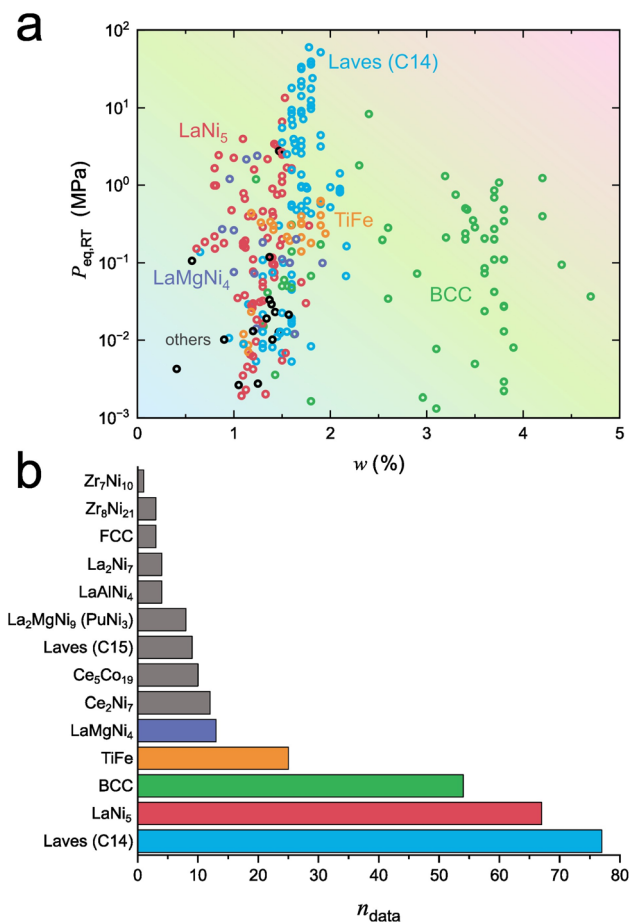


Fig. 2 Data distribution and feature construction for interstitial hydrides. (a) Distribution of data points in the $w - P_{\text{eq,RT}}$ map for entries with multi-temperature PCT data. (b) Number of data points (n_{data}) for each structural class.

temperature, thereby limiting the practical accessibility of their full w .²⁵ Fig. 2b presents n_{data} for each reported class. Notably, three classes, Laves (C14), LaNi_5 , and BCC, out of 14 classes account for $n_{\text{data}} = 198$, corresponding to 66.2% of the dataset with multi-temperature PCT measurements ($n_{\text{data}} = 299$), which is close to the Pareto principle.

For symbolic regression modeling, features were extracted as candidate descriptors for w and $P_{\text{eq,RT}}$. The full list of features and their denotations is provided in Table 1; hereafter, the denotations are omitted for simplicity. A total of 57 features were constructed, comprising 18 chemical, physical, and structural features applied to their averaged values $\langle \dots \rangle$, standard deviations $\sigma(\dots)$, and skewness $r(\dots)$, along with three additional compositional features (R_{Tr} , $R_{\text{Tr(IV)}}$, and $R_{\text{Tr+RE}}$). For example, for a compound $A_aB_bC_c$, the averaged atomic mass is defined as $M = (aM_a + bM_b + cM_c)/(a + b + c)$, where M_x denotes the atomic mass of element x . Structural descriptors such as Ω , Ω_σ , and V_p require the definition of coordination polyhedra XY_n ($X, Y = A, B$, and C crystallographic sites; X -centered). The maximum cutoff distance for metal-metal pairs was set to 3.5 Å, corresponding to the shoulder of the first peak in the radial distribution function averaged across crystal structures (see the

section “Averaged Radial Distribution Function of Metal-alloy Parent Structures and the Construction of XY_n Polyhedra” in the SI). In addition, the heatmap of Pearson correlation coefficients (r_{col}) among all feature pairs is provided in the section “Pearson Correlation Heatmap” in the SI.

Key descriptors governing w and $P_{\text{eq,RT}}$

Given the strong predictive performance of the white-box modeling approach we adopted (see the section “Benchmarking of Regression Models” in the SI),¹⁹ further symbolic regression models were reconstructed for the target metrics $\log_{10}(w/M)$ and $\log_{10}P_{\text{eq,RT}}$, using the full dataset without reserving a separate test set. Fig. 3a demonstrates the strong predictive performance of the model for $\log_{10}(w/M)$, achieving $R^2 = 0.754$, $\text{RMSE} = 0.151 \log_{10}[\% \text{ mol g}^{-1}]$, and $\text{MAE} = 0.0868 \log_{10}[\% \text{ mol g}^{-1}]$ over the entire dataset ($n_{\text{data}} = 706$). Fig. 3b and c present the partial dependence plots (PDPs) for two most important descriptors for w . In each case, the PDP curve is obtained by fixing all other features to their average values. As a result, the experimental and regressed data points, which reflect the full compositional effects and multicollinearity among features (see the section “Pearson Correlation Heatmap” in the SI), do not necessarily follow the PDP curve. Importantly, PDPs isolate the effect of each descriptor by fixing all others, thereby revealing intrinsic relationships beyond correlations in the dataset. Nevertheless, comparing the PDP trends with the corresponding data points provides insight into the underlying physics governing the target metrics.

In Fig. 3b, the PDP alone suggests that smaller $\langle r_{\text{M}} \rangle$, which is often associated with lower atomic mass, favors higher w . However, in real materials, excessively small $\langle r_{\text{M}} \rangle$ can instead induce steric constraints, requiring lattice expansion to accommodate interstitial hydrogens. As a result, an optimal value emerges around $\langle r_{\text{M}} \rangle \sim 1.47$ Å, whereas larger values lead to expanded interstitial sites that are unable to effectively stabilize hydrogen, likely due to geometric mismatch between the interstitial cage size and the effective size required for hydrogen accommodation. Notably, in the BCC structure, $\langle r_{\text{M}} \rangle = 1.47$ Å yields a geometric maximum hard-sphere radius of 0.43 Å for interstitial hydrogen in tetrahedral sites. In Fig. 3c, the PDP suggests that lower thermal conductivity $\langle \kappa \rangle$, which is linked to electronic structure near the Fermi level and the strength of metallic bonding, and typically associated with softer lattice structures, correlates with higher w . Here, $\langle \kappa \rangle$ is interpreted not as an isolated causal factor, but as an effective descriptor reflecting coupled electronic-structure and lattice-related characteristics of the host metal framework. In particular, softer lattice environments can more readily accommodate lattice expansion upon hydrogen insertion, thereby facilitating higher hydrogen uptake. In practice, however, κ exhibits a more nuanced influence. In contrast, for $\langle \kappa \rangle > 110 \text{ W m}^{-1} \text{ K}^{-1}$, w shows little variation, indicating that the effect of κ becomes saturated in this regime. Taken together, these results indicate that w is maximized when $\langle r_{\text{M}} \rangle$ is tuned to approximately 1.47 Å in conjunction with relatively low $\langle \kappa \rangle$.



Table 1 Chemical, physical, structural, and compositional properties of constituent elements of interstitial metal hydrides for regression models, given as features for symbolic regression modeling

Properties	Description	Unit
$\chi - \chi_{\text{H}}$	Difference in electronegativity between metal atom and hydrogen	—
M	Atomic mass	g mol^{-1}
n_{ve}	The number of valence electrons. For s/p-block and d/f-block, electrons in the shells of $n_{\text{p}s}$ and $n_{\text{p}p}$, and in those of $n_{\text{p}s}$ and $(n_{\text{p}} - 1)d$ are counted, respectively; n_{p} is the highest principal quantum number	—
r_{M}	Metallic radius	\AA
ρ	Density	g cm^{-3}
ρ_{mol}	Molar density, defined as ρ/M	mol cm^{-3}
η_{IM}	Metallic filling rate (per unit volume), defined as $\eta_{\text{IM}} = N_{\text{avo}}\rho_{\text{mol}}\left(\frac{4}{3}\pi r_{\text{M}}^3\right)$ where N_{avo} is the Avogadro's constant	—
B	Bulk modulus	GPa
G	Shear modulus	GPa
ν	Poisson's ratio	—
κ	Thermal conductivity	$\text{W m}^{-1} \text{K}^{-1}$
α	Thermal expansion coefficient (linear not volumetric)	K^{-1}
θ_{D}	Debye temperature	K
χ_{m}	Molar susceptibility	$\text{m}^3 \text{mol}^{-1}$
n_{c}	Coordination number to neighbor metal atoms within 3.5 \AA for each crystallographic site (A , B , and C). For example, in the compound LaMgNi_4 , A , B , and C correspond to La, Mg, and Ni, respectively	—
Ω	Solid angle per face constructed in polyhedra of XY_n (X , $Y = A$, B , and C crystallographic sites; X -centered; and X - Y distance shorter than 3.5 \AA). Estimated from the parent structures listed in Fig. 2b	Radian
Ω_{σ}	Standard deviation across Ω for each crystallographic site (A , B , and C). Estimated from the parent structures listed in Fig. 2b	Radian
V_{p}	Volume of polyhedra of XY_n for each crystallographic site (A , B , and C). Estimated from the parent structures listed in Fig. 2b	\AA^3
...	Average over the constituent metal ions	Same with the unit of ...
$\sigma(\dots)$	Standard deviation over the constituent metal ions	Same with the unit of ...
$r(\dots)$	Skewness over the constituent metal ions	—
R_{Tr}	Compositional fraction of transition metal elements	—
$R_{\text{Tr(IV)}}$	Compositional fraction of the fourth-row transition metal elements (Sc, ..., Zn)	—
$R_{\text{Tr+RE}}$	Compositional fraction of transition and rare-earth metal elements (La, ..., Lu)	—

Fig. 3d demonstrates the strong predictive performance of the model for $\log_{10}P_{\text{eq,RT}}$, achieving $R^2 = 0.889$, $\text{RMSE} = 0.527 \log_{10}[\text{MPa}]$, and $\text{MAE} = 0.400 \log_{10}[\text{MPa}]$ over the entire dataset ($n_{\text{data}} = 299$). Fig. 3e and f present the PDPs for two most important descriptors for $\log_{10}P_{\text{eq,RT}}$. In Fig. 3e, both the PDP and experimental and regressed data points show good agreement. A higher $\langle G \rangle$ indicates a more rigid lattice, which increases the elastic energy penalty associated with interstitial hydrogen insertion and thereby destabilizes hydride formation, leading to higher $P_{\text{eq,RT}}$. In Fig. 3f, the PDP suggests that the influence of $\langle \nu \rangle$ on $P_{\text{eq,RT}}$ is relatively weak. In real materials, however, $P_{\text{eq,RT}}$ decreases rapidly when ν is below approximately 0.3 and becomes nearly constant above this threshold. This behavior suggests that materials with higher $\langle \nu \rangle$ are more mechanically compliant and can more readily accommodate lattice deformation during hydride formation, consistent with the negative correlation between $\langle G \rangle$ and $\langle \nu \rangle$ shown in the section “Pearson Correlation Heatmap” in the SI. Taken together, these results indicate that lattice rigidity plays a central role in governing equilibrium pressure, with stiffer

crystal structures leading to higher $P_{\text{eq,RT}}$. To further assess the robustness of the model across different structural classes, a class-resolved sensitivity analysis was performed (see the section “Class-resolved Model Performance and Sensitivity Analysis” in the SI), confirming that the predictive performance is not dominated by any single class.

Materials design for high w at practical $P_{\text{eq,RT}}$

All *DigHyd* entries with multi-temperature PCT data ($n_{\text{data}} = 299$) were subjected to optimization toward higher w under the target condition of $P_{\text{eq,RT}} \sim 0.1 \text{ MPa}$, guided by the symbolic models. For each parent structure class, the composition with the highest predicted w within the window $-1.5 \log_{10}[\text{MPa}] < \log_{10}P_{\text{eq,RT}} < -0.5 \log_{10}[\text{MPa}]$ was selected, resulting in 8 representative cases. Fig. 4a–h are arranged in descending order of the optimized w , corresponding to the following structure types: BCC > Laves (C14) > LaMgNi_4 > La_2MgNi_9 (PuNi_3) > TiFe > LaNi_5 > Laves (C15) > Ce_2Ni_7 .

As an illustrative example, Fig. 4a shows the optimization of the BCC-type alloy TiVnBcCrNi_2 , initially characterized by



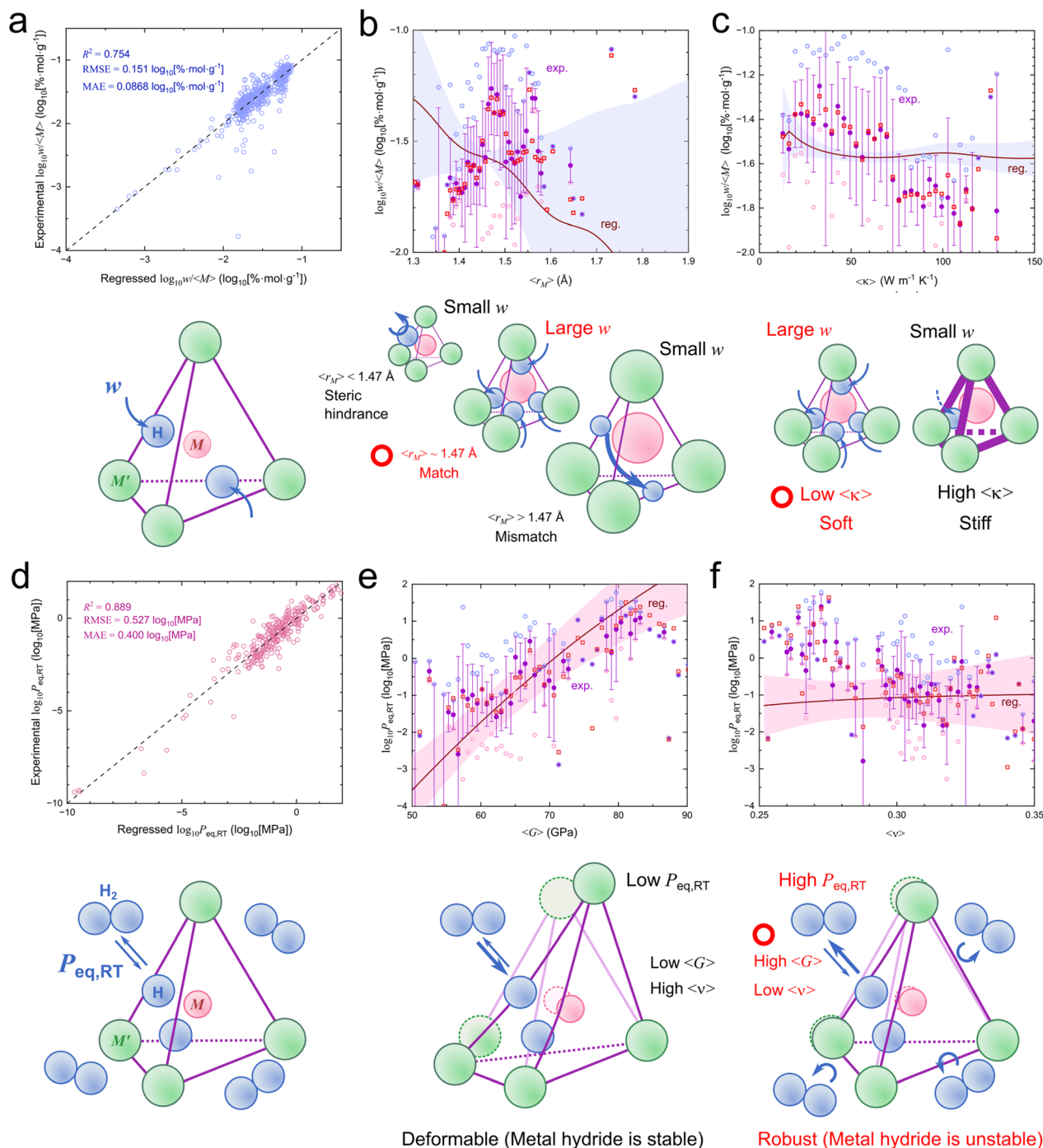


Fig. 3 Predictive performance and descriptor analysis for hydrogen capacity and equilibrium pressure. (a) Parity plot for $\log_{10}(w/M)$ using the symbolic regression model. (b) and (c) Partial dependence plots (PDPs) for the two most important descriptors governing $\log_{10}(w/M)$. (d) Parity plot for $\log_{10}P_{\text{eq,RT}}$. (e) and (f) PDPs for key descriptors governing $\log_{10}P_{\text{eq,RT}}$. In the PDPs of (b), (c), (e), and (f), the regression curve ("reg.") is obtained by fixing all other features to their average values. Experimental data ("exp.") are partitioned into 100 bins along the horizontal axis, where markers denote the minimum (represented by a magenta circle), average (a purple circle), with standard deviation represented by a purple bar, and maximum values (a cyan circle) within each bin, as well as the averaged regressed value (a red rectangle). The shaded region represents the variation across individual symbolic models within the ensemble. Schematic illustrations are presented below each panel to define w and $P_{\text{eq,RT}}$ and to illustrate the relationships between the descriptors and the corresponding properties. As materials design principles, the ideal conditions for achieving high w and high $P_{\text{eq,RT}}$ are highlighted in red and indicated by red circles.

a descriptor-performance vector $p = (\langle r_M \rangle / \text{Å}, \langle \kappa \rangle / (\text{W m}^{-1} \text{K}^{-1}), w / \%) = (1.42, 63.8, 3.5)$.²⁶ The optimization proceeds through a multi-step compositional pathway: first, 90% of Ni is replaced

with V to yield TiVNbCrV_{1.8}Ni_{0.2}; next, 10% of Cr is substituted with Tm, resulting in TiVNbCr_{0.9}Tm_{0.1}V_{1.8}Ni_{0.2}; finally, 10% of Ti is replaced with Nb, leading to the composition



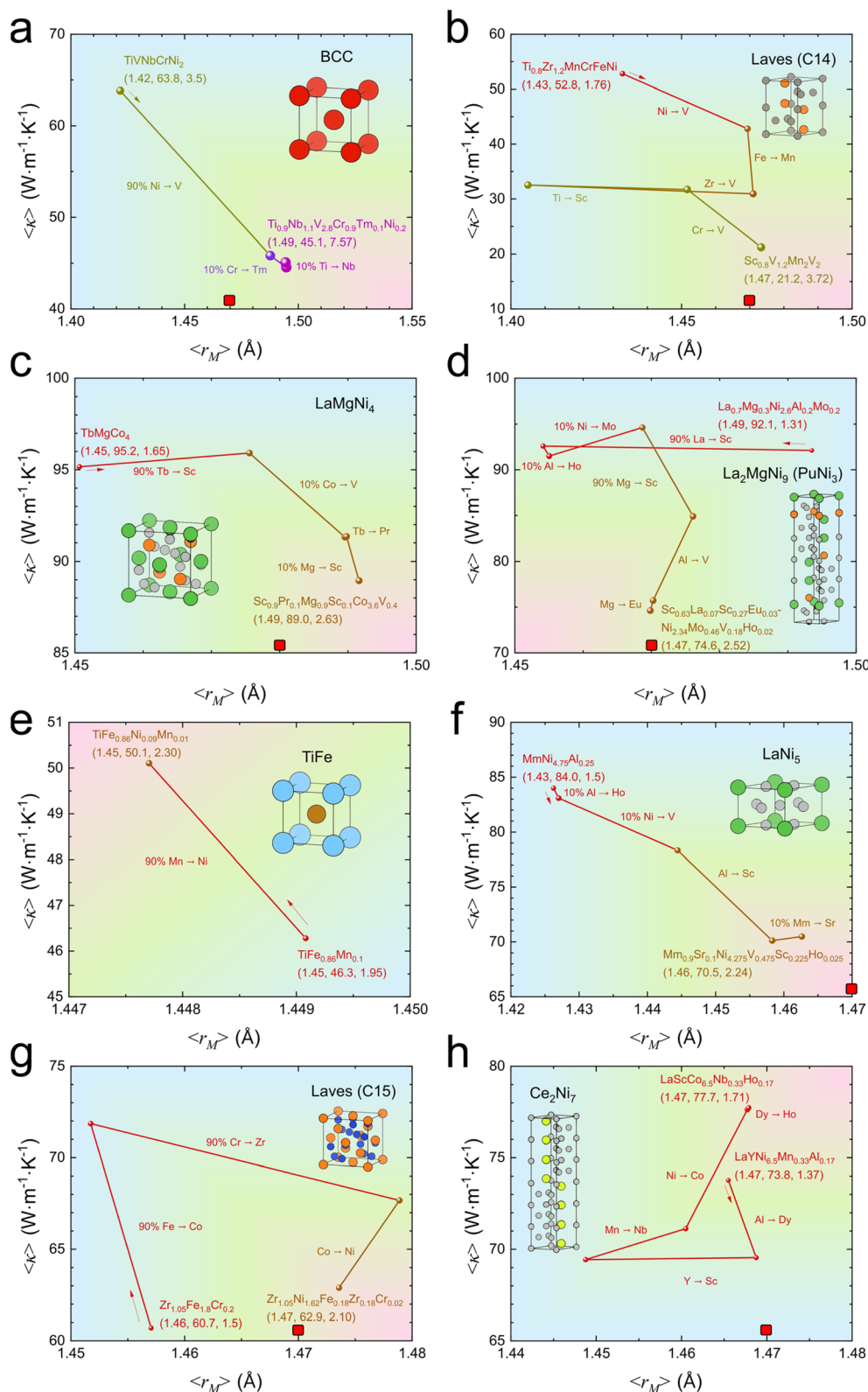


Fig. 4 Materials design pathways toward high hydrogen capacity under practical equilibrium pressure, mapped onto the $\langle \kappa \rangle - \langle r_M \rangle$ plot. Optimization trajectories for representative parent structure classes: (a) BCC,²⁶ (b) Laves (C14),²⁷ (c) LaMgNi_4 ,²⁸ (d) La_2MgNi_9 (PuNi₃),²⁹ (e) TiFe,³⁰ (f) LaNi_5 ,³¹ (g) Laves (C15),³² and (h) Ce_2Ni_7 .³³ Each panel shows the compositional optimization pathway guided by the symbolic models within the target window $-1.5 \log_{10}[\text{MPa}] < \log_{10} P_{\text{eq,RT}} < -0.5 \log_{10}[\text{MPa}]$. The marker size and color, as well as the line color at each step, are scaled by w . Initial and optimized compositions are represented by descriptor-performance vectors $p = (\langle r_M \rangle / \text{\AA}, \langle \kappa \rangle / (\text{W m}^{-1} \text{K}^{-1}), w / \%)$, together with intermediate substitution steps. The red rectangle in each panel indicates a shared design target identified by symbolic models, where $\langle r_M \rangle$ converges to approximately 1.47 Å with reduced $\langle \kappa \rangle$. In most cases, $\langle \kappa \rangle$ either decreases or remains nearly constant, consistent with adherence to



$\text{Ti}_{0.9}\text{Nb}_{0.1}\text{VNbCr}_{0.9}\text{TM}_{0.1}\text{V}_{1.8}\text{Ni}_{0.2}$. This sequence yields the optimized composition $\text{Ti}_{0.9}\text{Nb}_{1.1}\text{V}_{2.8}\text{Cr}_{0.9}\text{TM}_{0.1}\text{Ni}_{0.2}$, with final $p = (1.49, 45.1, 7.57)$. Similar symbolic-model-guided optimization systematically shifts compositions toward enhanced w across other structure types^{27–33} by modulating $\langle r_{\text{M}} \rangle$ and $\langle \kappa \rangle$ (Fig. 4b–h). Here, all descriptors are updated self-consistently with composition and are not fixed to their average values, in contrast to the PDP analysis in Fig. 3.

Across all optimized cases, two consistent design principles emerge in most cases, represented by red rectangles in Fig. 4a–h. First, $\langle r_{\text{M}} \rangle$ converges toward an optimal value of approximately 1.47 Å, represented by dashed magenta lines. Second, $\langle \kappa \rangle$ either decreases or remains nearly unchanged during optimization. These trends directly reflect the descriptor–property relationships identified in Fig. 3b and c, demonstrating that the symbolic models capture transferable design rules for achieving high w under practical $P_{\text{eq,RT}}$ conditions. Although the predicted values of w themselves may be subject to uncertainty due to their extrapolative nature, the proposed optimization pathways provide physically grounded and clear design directions, based on the white-box symbolic models. Also, this design pathway can be further enhanced through integration with a closed-loop discovery framework,^{34,35} in which symbolic models are iteratively refined using synthesis and measurement data guided by the optimization trajectories, thereby enabling more accurate, real-world-relevant predictions.

Discussion

The present study reveals that hydrogen storage behavior in interstitial hydrides can be described by a small number of physically interpretable descriptors that govern distinct aspects of performance. In particular, w is governed by the interplay between geometric and electronic-lattice factors, captured by $\langle r_{\text{M}} \rangle$ and $\langle \kappa \rangle$. The existence of an optimal $\langle r_{\text{M}} \rangle \sim 1.47$ Å reflects a geometric constraint on hydrogen accommodation, while lower $\langle \kappa \rangle$, associated with softer lattices, facilitates hydrogen incorporation. These results indicate that maximizing w requires simultaneous optimization of both interstitial geometry and lattice adaptability. Atomic size effects have long been recognized as an important factor in hydrogen storage alloys, particularly through their influence on interstitial site size and lattice geometry; it was shown that variations in atomic radius modulate the available void space for hydrogen accommodation, thereby affecting hydrogen storage capacity w .³⁶ In this context, the emergence of $\langle r_{\text{M}} \rangle$ as a key descriptor for w is consistent with established physical understanding. In contrast, the identification of $\langle \kappa \rangle$ for w is to the best of our knowledge, less explored in the literature.

In contrast, $P_{\text{eq,RT}}$ is primarily governed by elastic properties, described by $\langle G \rangle$ and $\langle \nu \rangle$. A higher $\langle G \rangle$ increases the elastic penalty for hydrogen insertion, leading to higher equilibrium pressure, whereas larger ν reflects greater mechanical compliance and lowers $P_{\text{eq,RT}}$. Thus, $P_{\text{eq,RT}}$ is fundamentally controlled by the elastic response of the host lattice. These descriptors provide complementary representations of lattice rigidity and deformability rather than independent contributions. This also can be understood in the context of the van't Hoff equation, $\ln P_{\text{eq},T} = \frac{\Delta H}{RT} - \frac{\Delta S}{R}$, where $P_{\text{eq},T}$ is the equilibrium pressure at temperature T , and ΔH and ΔS are the enthalpy and entropy changes associated with hydride formation. For many metal-hydrogen systems, ΔS is largely dominated by the loss of gas-phase entropy when molecular hydrogen is incorporated as atomic hydrogen into the lattice. Consequently, variations in $P_{\text{eq,RT}}$ are primarily governed by changes in ΔH . In this context, $\langle G \rangle$ and $\langle \nu \rangle$ reflecting the elastic energy penalty is consistent with a prior study demonstrating that lattice strain has been shown to modify the effective ΔH and shift $P_{\text{eq},T}$ over several orders of magnitude by altering the elastic energy associated with hydrogen insertion.³⁷ Furthermore, $\langle G \rangle$ has recently been reported as a key descriptor governing ΔH in the formation of BCC high-entropy alloy monohydrides, based on a machine learning study.³⁸

Thus, we emphasize that these relationships are consistent mainly with established physical understanding, and the present analysis provides a quantitative framework to support and organize these insights. Here, it should be noted that the present descriptor framework is most applicable to systems undergoing isostructural hydrogenation. In materials with non-isostructural phase transitions, additional thermodynamic and structural factors may become dominant.

Taken together, these results establish a clear separation of roles between descriptors governing capacity and thermodynamics. While w is controlled by geometric-electronic factors, $P_{\text{eq,RT}}$ is dictated by elastic constraints. This separation provides a physically transparent and actionable framework for materials design under practical operating conditions. Importantly, these descriptor relationships are consistently observed across different structure types, indicating that they define a transferable design space. Compositions across multiple classes converge toward $\langle r_{\text{M}} \rangle \sim 1.47$ Å with reduced $\langle \kappa \rangle$, demonstrating the robustness of the identified design principles. The present framework is expected to be most reliable within the probed compositional and structural domains of interstitial hydrides. Application beyond this space may require incorporation of additional descriptors to capture different bonding regimes, phase transformations, or thermodynamic contributions not present in interstitial hydrides.

the design rule of $\langle r_{\text{M}} \rangle \rightarrow \sim 1.47$ Å. The parent crystal structures are also visually represented. The color gradient is used as a visual guide to indicate the overall direction of materials design trends, rather than representing a quantitative colormap. The reliability of the optimized candidate compositions was further evaluated using ensemble uncertainty metrics (see the section "Uncertainty Assessment of Optimized Candidate Materials" in the SI), indicating that the proposed candidates lie within a low-uncertainty prediction regime. These compositions should be interpreted as physically guided design candidates rather than directly synthesizable materials, and further experimental validation will be required.



These findings should also be understood in the broader context of descriptor discovery across hydrogen storage materials. In our previous study, we showed that across a global materials space, including both interstitial and saline hydrides, w and $P_{\text{eq,RT}}$ are strongly governed by elemental electronegativity, and can therefore impose an intrinsic trade-off between capacity and equilibrium pressure.¹⁹ This indicates that, at the global scale, electronegativity acts as a unifying descriptor controlling hydrogen storage behavior across fundamentally different bonding types. In contrast, the present work focuses on the local materials space of interstitial hydrides, where more specific descriptors emerge, reflecting finer structural and mechanical effects. Thus, while global electronegativity-driven trends may introduce trade-offs, the local descriptor framework identified here provides a route to navigate and potentially mitigate these trade-offs for simultaneous optimization of high w and practical $P_{\text{eq,RT}}$. Together, these results highlight a “hierarchical” descriptor framework, in which global trends are governed by simple chemical parameters, while local optimization requires more detailed, physically specific descriptors.

More broadly, the present results demonstrate that symbolic regression can resolve coupled physical mechanisms into interpretable descriptor relationships, enabling a transition from empirical exploration toward mechanism-informed materials design. This approach provides a generalizable strategy for descriptor-based materials design in hydrogen-related systems. Furthermore, such a framework is expected to be directly applicable to saline hydrides and, more broadly, to systems beyond hydrogen storage, including hydride-based solid-state electrolytes for batteries.³⁹

Conclusion

In this work, we established a physically interpretable, data-driven framework to uncover descriptor–property relationships governing hydrogen storage behavior in interstitial hydrides using curated PCT data and symbolic regression. A small set of physically meaningful descriptors was identified, revealing that w and $P_{\text{eq,RT}}$ are governed by fundamentally distinct mechanisms. In particular, w is controlled by geometric and electronic-lattice factors, whereas $P_{\text{eq,RT}}$ is dictated by elastic properties of the host lattice. These insights were further translated into actionable materials design guidelines, enabling the identification of compositions with enhanced hydrogen capacity under practical operating conditions. The resulting design principles are transferable across different structure types and compositional spaces, demonstrating the robustness of the descriptor-based framework. Beyond hydrogen storage, the present approach provides a general strategy for integrating interpretable machine learning with physically grounded materials design, with potential applications extending to other energy materials.

Methods

DigHyd database

The Digital Hydrogen Platform (*DigHyd*) is a rigorously curated database of hydrogen storage materials constructed through an

AI-assisted, human-in-the-loop literature mining framework. As described in ref. 21, *DigHyd* integrates experimentally reported data from more than 4000 literature sources, comprising over 30 000 curated data entries on hydrogen storage properties and thermodynamic parameters. Importantly, the database spans a wide range of material classes beyond conventional metal hydrides, including interstitial hydrides, complex hydrides, ionic (saline) hydrides, multi-component and destabilized systems, as well as porous materials such as metal–organic frameworks. This breadth enables systematic analysis across fundamentally different hydrogen storage mechanisms. The curated data include gravimetric hydrogen capacity (w), as well as the enthalpy (ΔH) and entropy (ΔS) changes associated with hydrogenation reactions, primarily defined as $M + \frac{1}{2}\text{H}_2 \rightleftharpoons MH$.

Symbolic modeling

While the algorithmic details of the white-box symbolic regression modeling package (*GoodRegressor*) are described in ref. 23, a brief summary of its use in this study is provided here. For each symbolic model constructed using the *Regressor* module described in ref. 20, the dataset was randomly split into training and validation sets with a ratio of 8 : 2. Given the set of candidate descriptors listed in Table 1, the number of possible model combinations including interaction terms is given as 1.95×10^{463} for the 109 default scalar transforms provided in the *Regressor* module.

Model construction was therefore performed in a staged manner. Starting from models with 20 active independent variables, the number of active variables was iteratively reduced one by one ($20 \rightarrow 19 \rightarrow 18 \rightarrow \dots \rightarrow 1$), while progressively allowing more complex interaction terms (*i.e.*, increasing interaction depth). At each stage, candidate models were evaluated, and the model with the highest R^2 over the entire dataset was selected.

All models were generated under the so-called *full Fisher condition*, whereby the p -values of both the overall model (F -test) and all individual coefficients (t -tests) were required to be less than 0.05. For each target metric ($\log_{10}(w/M)$ and $\log_{10}P_{\text{eq,RT}}$), 10 symbolic models were constructed and subsequently combined into a single “stacking-ensembled” model. Key descriptors were then identified by examining features that consistently appeared across the 10 symbolic models, as well as by evaluating the magnitude of their contributions through z -scored coefficients. In particular, descriptors associated with terms exhibiting large average z -scored coefficient magnitudes across the ensemble were considered to play dominant roles in determining the target properties.

To ensure model reliability, the standard deviation across the regressed values in the ensemble was constrained to be below $0.5 \log_{10}[\% \text{ mol g}^{-1}]$ and $0.7 \log_{10}[\text{MPa}]$ for $\log_{10}(w/M)$ and $\log_{10}P_{\text{eq,RT}}$, respectively. Predictions exceeding these thresholds were discarded. In the 5-fold benchmark tests, the discard ratios were approximately 1% and 10% for $\log_{10}(w/M)$ and $\log_{10}P_{\text{eq,RT}}$, respectively. Notably, when models were constructed using the entire dataset, no data points were discarded,



as all predictions satisfied the ensemble consistency criteria. Representative examples of symbolic model ensembles, their error distributions, and comparisons across different target metrics ($\log_{10}w$ and $\log_{10}(wM)$) are provided in the section “Symbolic Models: Formulation and Error Analysis” in the SI.

Materials design

For materials optimization, the *Designer* module described in ref. 23 was employed, using the *DigHyd* dataset as input. During the search for optimization pathways from existing experimental compositions, only candidate compositions yielding ensemble standard deviations below $0.5 \log_{10}[\% \text{ mol g}^{-1}]$ and $0.7 \log_{10}[\text{MPa}]$ for $\log_{10}(w/M)$ and $\log_{10}P_{\text{eq,RT}}$, respectively, were considered. For the final optimized compositions, a stricter criterion was applied: only those with ensemble standard deviations below $0.1 \log_{10}[\% \text{ mol g}^{-1}]$ were retained for $\log_{10}(w/M)$.

Compositional modifications were performed through controlled substitution operations while preserving the original stoichiometry. Specifically, full substitution, 90% substitution, and 10% substitution of constituent metal elements were allowed. Candidate substitution elements were required to satisfy two criteria relative to the original element: the differences in electronegativity and metallic radius must not exceed 0.5 and 0.5 Å, respectively.

Crystal structure visualization

Crystal structures were visualized using VESTA.⁴⁰

Code availability

The source code supporting materials prediction and design in this study is openly available at <https://github.com/JerryGarcia1995/OxygenIonConductor>. The repository contains the general symbolic regression framework used in this study.

Author contributions

H. Li and S.-i. Orimo conceived and supervised the project. S.-H. Jang carried out conceptualization, data curation, formal analysis, methodology development, software implementation, validation, and visualization, and wrote the original draft of the manuscript. D. Zhang contributed to data curation, investigation, methodology, and software development. X. Jia, H. B. Tran, and K. Konno contributed to data curation and investigation. L. Zhang, R. Sato, and Y. Hashimoto reviewed and edited the manuscript. Y. Ohashi and T. Sato contributed to conceptualization and manuscript review and editing. All authors discussed the results and approved the final manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data can be accessed in the Digital Hydrogen Platform (*DigHyd*: <https://www.dighyd.org>).

Supplementary information (SI): Radial distribution function averaged over metal-alloy parent structures for hydrogen storage and the construction of XY_n polyhedra, Pearson correlation heatmap, benchmarking of regression models, class-resolved model performance and sensitivity analysis, uncertainty assessment of optimized candidate materials, and symbolic models: formulation and error analysis. See DOI: <https://doi.org/10.1039/d6sc03089k>.

Acknowledgements

This work was supported by The Green Technologies of Excellence (GteX) Program, Japan (Grant No. JPMJGX23H1).

References

- 1 N. Johnson, M. Liebreich, D. M. Kammen, P. Ekins, R. McKenna and I. Staffell, Realistic roles for hydrogen in the future energy transition, *Nat. Rev. Clean Technol.*, 2025, **1**, 351–371, DOI: [10.1038/s44359-025-00050-4](https://doi.org/10.1038/s44359-025-00050-4).
- 2 A. P. Zhao, S. Li, D. Xie, Y. Wang, Z. Li, P. J.-H. Hu and Q. Zhang, Hydrogen as the nexus of future sustainable transport and energy systems, *Nat. Rev. Electr. Eng.*, 2025, **2**, 447–466, DOI: [10.1038/s44287-025-00178-2](https://doi.org/10.1038/s44287-025-00178-2).
- 3 A. G. Gebretatios, F. Banat and C. K. Cheng, A critical review of hydrogen storage: Toward the nanoconfinement of complex hydrides from the synthesis and characterization perspectives, *Sustain. Energ. Fuels.*, 2024, **8**(22), 5091–5130, DOI: [10.1039/d4se00353e](https://doi.org/10.1039/d4se00353e).
- 4 A. Züttel, Materials for hydrogen storage, *Mater. Today*, 2003, **6**(9), 24–33, DOI: [10.1016/S1369-7021\(03\)00922-2](https://doi.org/10.1016/S1369-7021(03)00922-2).
- 5 J. Bellosta von Colbe, J.-R. Ares, J. Barale, M. Baricco, C. Buckley, G. Capurso, N. Gallandat, D. M. Grant, M. N. Guzik, I. Jacob, E. H. Jensen, T. Jensen, J. Jepsen, T. Klassen, M. V. Lototskyy, K. Manickam, A. Montone, J. Puszkil, S. Sartori, D. A. Sheppard, A. Stuart, G. Walker, C. J. Webb, H. Yang, V. Yartys, A. Züttel and M. Dornheim, Application of hydrides in hydrogen storage and compression: Achievements, outlook and perspectives, *Int. J. Hydrogen Energy*, 2019, **44**(15), 7780–7808, DOI: [10.1016/j.ijhydene.2019.01.104](https://doi.org/10.1016/j.ijhydene.2019.01.104).
- 6 M. Hirscher, V. A. Yartys, M. Baricco, J. Bellosta von Colbe, D. Blanchard, R. C. Bowman, D. P. Broom, C. E. Buckley, F. Chang, P. Chen, Y. W. Cho, J.-C. Crivello, F. Cuevas, W. I. F. David, P. E. de Jongh, R. V. Denys, M. Dornheim, M. Felderhoff, Y. Filinchuk and G. E. Froudakis, Materials for hydrogen-based energy storage - Past, recent progress and future outlook, *J. Alloys Compd.*, 2020, **827**, 153548, DOI: [10.1016/j.jallcom.2019.153548](https://doi.org/10.1016/j.jallcom.2019.153548).
- 7 E. Wicke, H. Brodowsky and H. Züchner, Hydrogen in palladium and palladium alloys. Hydrogen in Metals II. *Topics in Applied Physics*, ed. G. Alefeld and J. Völkl,



- Springer, 1978, vol. 29, pp 73–155, DOI: [10.1007/3-540-08883-0_19](https://doi.org/10.1007/3-540-08883-0_19).
- 8 G. Sandrock, A panoramic overview of hydrogen storage alloys from a gas reaction point of view, *J. Alloys Compd.*, 1999, **293–295**, 877–888, DOI: [10.1016/S0925-8388\(99\)00384-9](https://doi.org/10.1016/S0925-8388(99)00384-9).
- 9 R. C. Bowman Jr and B. Fultz, Metallic hydrides I: Hydrogen storage and other gas-phase applications, *MRS Bull.*, 2002, **27**, 688–693, DOI: [10.1557/mrs2002.223](https://doi.org/10.1557/mrs2002.223).
- 10 S. Orimo, Y. Nakamori, J. R. Eliseo, A. Züttel and C. M. Jensen, Complex hydrides for hydrogen storage, *Chem. Rev.*, 2007, **107**, 4111–4132, DOI: [10.1021/cr0501846](https://doi.org/10.1021/cr0501846).
- 11 B. Sakintuna, F. Lamari-Darkrim and M. Hirscher, Metal hydride materials for solid hydrogen storage: A review, *Int. J. Hydrogen Energy*, 2007, **32(9)**, 1121–1140, DOI: [10.1016/j.ijhydene.2006.11.022](https://doi.org/10.1016/j.ijhydene.2006.11.022).
- 12 I. P. Jain, P. Jain and A. Jain, Novel hydrogen storage materials: A review of lightweight complex hydrides, *J. Alloys Compd.*, 2010, **503(2)**, 303–339, DOI: [10.1016/j.jallcom.2010.04.250](https://doi.org/10.1016/j.jallcom.2010.04.250).
- 13 I. P. Jain, C. Lal and A. Jain, Hydrogen storage in Mg: A most promising material, *Int. J. Hydrogen Energy*, 2010, **35(10)**, 5133–5144, DOI: [10.1016/j.ijhydene.2009.08.088](https://doi.org/10.1016/j.ijhydene.2009.08.088).
- 14 G. Scarpati, E. Frasci, G. Di Ilio and E. Jannelli, A comprehensive review on metal hydrides-based hydrogen storage systems for mobile applications, *J. Energy Storage*, 2024, **102**, 113934, DOI: [10.1016/j.est.2024.113934](https://doi.org/10.1016/j.est.2024.113934).
- 15 E. Nemukula, C. B. Mtshali and F. Nemangwele, Metal hydrides for sustainable hydrogen storage: A review, *Int. J. Energy Res.*, 2025, 630225, DOI: [10.1155/er/6300225](https://doi.org/10.1155/er/6300225).
- 16 Z. Gao, X. Yang, Z. Zhuang, Y. Zhang, J. Cai, Y. Li, W. Fu, H. Li and W. Yang, Catalytic strategies and mechanisms for enhancing MgH₂ solid-state hydrogen storage, *Chem Catal.*, 2026, **6**, 101692, DOI: [10.1016/j.checat.2026.101692](https://doi.org/10.1016/j.checat.2026.101692).
- 17 P. Zhou, X. Xiao, X. Zhu, Y. Chen, W. Lu, M. Piao, Z. Cao, M. Lu, F. Fang, Z. Li, L. Jiang and L. Chen, *Energy Storage Mater.*, 2023, **63**, 102964, DOI: [10.1016/j.ensm.2023.102964](https://doi.org/10.1016/j.ensm.2023.102964).
- 18 P. Zhou, Q. Zhou, X. Xiao, X. Fan, Y. Zou, L. Sun, J. Jiang, D. Song and L. Chen, Machine learning in solid-state hydrogen storage materials: Challenges and perspectives, *Adv. Mater.*, 2025, **37(6)**, 2413430, DOI: [10.1002/adma.202413430](https://doi.org/10.1002/adma.202413430).
- 19 P. Zhou, H. Shen, N. Xu, Q. Zhou, C. Yan, X. Li, N. Lei, J. Jiang, D. Song, J. Zheng, Y. Zou, L. Sun, Z. Han, X. Fan, X. Xiao and L. Chen, Ambient-condition hydrogen storage: Accelerating the development of high-capacity hydrides via a quantitative interpretable machine learning framework, *Energy Storage Mater.*, 2026, **84**, 104789, DOI: [10.1016/j.ensm.2025.104789](https://doi.org/10.1016/j.ensm.2025.104789).
- 20 D. Zhang, X. Jia, H. B. Tran, S. H. Jang, L. Zhang, R. Sato, Y. Hashimoto, T. Sato, K. Konno, S. Orimo and H. Li, “DIVE” into hydrogen storage materials discovery with AI agents, *Chem. Sci.*, 2026, **17**, 3031–3042, DOI: [10.1039/D5SC09921H](https://doi.org/10.1039/D5SC09921H).
- 21 S.-H. Jang, D. Zhang, X. Jia, H. B. Tran, L. Zhang, R. Sato, Y. Hashimoto, T. Sato, K. Konno, S. Orimo and H. Li, Digital hydrogen platform (DigHyd): A rigorously curated database for hydrogen storage materials empowered by AI-assisted literature mining, *arXiv*, 2026, preprint, arXiv.2603.14139, DOI: [10.48550/arXiv.2603.14139](https://doi.org/10.48550/arXiv.2603.14139).
- 22 S.-H. Jang, D. Zhang, H. B. Tran, X. Jia, K. Konno, R. Sato, S. Orimo and H. Li, Physically interpretable descriptors drive the materials design of metal hydrides for hydrogen storage, *Chem. Sci.*, 2025, **16**, 23111–23120, DOI: [10.1039/D5SC07296D](https://doi.org/10.1039/D5SC07296D).
- 23 S.-H. Jang, GoodRegressor: A hierarchical inductive bias for navigating high-dimensional compositional space, *arXiv*, 2026, preprint, arXiv.2510.1832, DOI: [10.48550/arXiv.2510.1832](https://doi.org/10.48550/arXiv.2510.1832).
- 24 World Health Organization & International Programme on Chemical Safety, *Beryllium: Health and safety guide*, World Health Organization, 1990, <https://iris.who.int/handle/10665/40004>.
- 25 E. Akiba and M. Okada, Metallic hydrides III: Body-centered-cubic solid-solution alloys, *MRS Bull.*, 2002, **27(9)**, 699–703, DOI: [10.1557/mrs2002.225](https://doi.org/10.1557/mrs2002.225).
- 26 B. Cheng, L. Kong, H. Cai, Y. Li, Y. Zhao, D. Wan and Y. Xue, Exploring microstructure variations and hydrogen storage characteristics in TiVnBnCrnNi high-entropy alloys with different Ni incorporation, *Int. J. Hydrog. Energy*, 2024, **72**, 29–40, DOI: [10.1016/j.ijhydene.2024.05.317](https://doi.org/10.1016/j.ijhydene.2024.05.317).
- 27 V. Enblom, R. Clulow, T.-J. Ha, M. D. Witman, L. E. Way, S. J. Han, P. H. B. Brant Carvalho, V. Stavila, J.-Y. Suh, M. Sahlberg and J. O. Fadonougbo, A combined experimental and machine learning exploration of Ti_{2-x}Zr_xMnCrFeNi high entropy Laves hydrides, *Mater.*, 2025, **40**, 102414, DOI: [10.1016/j.mtla.2025.102414](https://doi.org/10.1016/j.mtla.2025.102414).
- 28 V. V. Shtender, V. Paul-Boncour, R. V. Denys, J.-C. Crivello and I. Y. Zavalij, TbMgNi_{4-x}Co_x(H,D)₂ System. I: Synthesis, hydrogenation properties, and crystal and electronic structures, *J. Phys. Chem. C*, 2020, **124(1)**, 196–204, DOI: [10.1021/acs.jpcc.9b10252](https://doi.org/10.1021/acs.jpcc.9b10252).
- 29 X. B. Zhang, D. Z. Sun, W. Y. Yin, Y. J. Chai and M. S. Zhao, Crystallographic and electrochemical characteristics of La_{0.7}Mg_{0.3}Ni_{3-x}(Al_{0.5}Mo_{0.5})_x ($x = 0-0.4$) hydrogen storage alloys, *Electrochim. Acta*, 2005, **50(16–17)**, 3407–3413, DOI: [10.1016/j.electacta.2004.12.020](https://doi.org/10.1016/j.electacta.2004.12.020).
- 30 H. Qu, J. Du, C. Pu, Y. Niu, T. Huang, Z. Li, Y. Lou and Z. Wu, Effects of Co introduction on hydrogen storage properties of Ti-Fe-Mn alloys, *Int. J. Hydrog. Energy*, 2015, **40(6)**, 2729–2735, DOI: [10.1016/j.ijhydene.2014.12.089](https://doi.org/10.1016/j.ijhydene.2014.12.089).
- 31 B. Molinas, A. Pontarollo, M. Scapin, H. Peretti, M. Melnichuk, H. Corso, A. Aurora, D. M. Gattia and A. Montone, The optimization of MmNi_{5-x}Al_x hydrogen storage alloy for sea or lagoon navigation and transportation, *Int. J. Hydrog. Energy*, 2016, **41(32)**, 14484–14490, DOI: [10.1016/j.ijhydene.2016.05.222](https://doi.org/10.1016/j.ijhydene.2016.05.222).
- 32 C. Zhou, H. Wang, L. Z. Ouyang, J. W. Liu and M. Zhu, Achieving high equilibrium pressure and low hysteresis of Zr-Fe based hydrogen storage alloy by Cr/V substitution, *J. Alloys Compd.*, 2019, **806**, 1436–1444, DOI: [10.1016/j.jallcom.2019.07.170](https://doi.org/10.1016/j.jallcom.2019.07.170).
- 33 E. H. Jensen, L. Lombardo, A. Girella, M. N. Guzik, A. Züttel, C. Milanese, P. Whitfield, D. Noréus and S. Sartori, The effect



- of Y content on structural and sorption properties of A_2B_7 -type phase in the La-Y-Ni-Al-Mn system, *Molecules*, 2023, **28**(9), 3749, DOI: [10.3390/molecules28093749](https://doi.org/10.3390/molecules28093749).
- 34 D. Zhang, X. Jia, Y. Wang, H. Liu, Q. Wang, S.-H. Jang, D. Shah, S. Ye, H. B. Tran and H. Li, Digital materials ecosystem: from databases to AI agents for autonomous discovery, *Chem. Sci.*, 2026, **17**, 5782–5804, DOI: [10.1039/D5SC09229A](https://doi.org/10.1039/D5SC09229A).
- 35 D. Zhang, Y. Chen, C. Liu, Y. Liu, H. Xin, J. Peng, P. Ou and H. Li, Accelerating catalyst materials discovery with large artificial intelligence models, *Angew. Chem., Int. Ed.*, 2026, e26150, DOI: [10.1002/anie.202526150](https://doi.org/10.1002/anie.202526150).
- 36 K. Panwar and S. Srivastava, On structural model of AB_5 -type multi-element hydrogen storage alloy, *Int. J. Hydrogen Energy*, 2019, **44**(57), 30208–30217, DOI: [10.1016/j.ijhydene.2019.09.138](https://doi.org/10.1016/j.ijhydene.2019.09.138).
- 37 P. Ngene, A. Longo, L. Mooij, W. Bras and B. Dam, Metal-hydrogen systems with an exceptionally large and tunable thermodynamic destabilization, *Nat. Commun.*, 2017, **8**, 1846, DOI: [10.1038/s41467-017-02043-9](https://doi.org/10.1038/s41467-017-02043-9).
- 38 E. Halpren, X. Yao, Z. W. Chen and C. V. Singh, Machine learning assisted design of BCC high entropy alloys for room temperature hydrogen storage, *Acta Mater.*, 2024, **270**, 119841, DOI: [10.1016/j.actamat.2024.119841](https://doi.org/10.1016/j.actamat.2024.119841).
- 39 Q. Wang, F. Yang, Y. Wang, D. Zhang, R. Sato, L. Zhang, J. Cheng, Y. Yan, Y. Chen, K. Kisu, S. Orimo and H. Li, Unraveling the complexity of divalent hydride electrolytes in solid-state batteries *via* a data-driven framework with large language model, *Angew. Chem., Int. Ed.*, 2025, **64**(5), e202506573, DOI: [10.1002/anie.202506573](https://doi.org/10.1002/anie.202506573).
- 40 K. Momma and F. Izumi, VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data, *J. Appl. Cryst.*, 2011, **44**, 1272–1276, DOI: [10.1107/S0021889811038970](https://doi.org/10.1107/S0021889811038970).

