Supplementary Information

Design of Lightweight BCC Multi-Principal Element Alloys with Enhanced Hydrogen Storage Using a Machine Learning-driven Genetic Algorithm

Kevin Ji,[1] Tanumoy Banerjee,[1,2] Matthew D. Witman,[3] Mark D. Allendorf,[3]
Vitalie Stavila,[3] and Prashant Singh[1,*]

[1]Ames National Laboratory, US Department of Energy, Iowa State University, Ames, Iowa, USA
[2]Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
[3]Sandia National Laboratories, Livermore, California 94551-0969, United States

***Keywords**: Hydrogen storage, High-Entropy Alloys, Genetic algorithm, ML, DFT, Reaction Enthalpy*

*Corresponding author Email: psingh84@ameslab.gov/prashant40179@gmail.com

Density-functional theory (DFT) calculations: The electronic-structure, local-lattice distortion, and hydrogen diffusion analysis was done employing first-principles DFT as implemented in the Vienna Ab-initio Simulation Package (VASP) [**S1,S2**]. The generalized gradient approximation of Perdew, Burke, and Ernzerhof (PBE) was employed in all calculations [**S3**] with a plane-wave cut-off energy of 520 eV. The choice of PBE over LDA or meta-GGA [**S4,S5**] functionals is based on the work of Söderling et al. [**S6**] and Giese et al. [**S7**] that establishes the effectiveness of GGA functionals. Large Supercell Random Approximates (SCRAPs), i.e., 60 atoms per cell, with the optimized disorder (zero-correlation) were generated (a single, optimized configurational representation) for DFT calculations [**S8**]. The energy and force convergence criterion of 10^{-8} eV and 10^{-6} eV/Å, respectively, were used for full (volume and atomic) relaxation of SCRAPs. The Monkhorst-Pack *k*-mesh was used for Brillouin zone integration during structural optimization and charge self-consistency calculations [**S9**].

Feature engineering: The experimental dataset selected in this study includes only the thermodynamic features like pressure, temperature, reaction enthalpy alongside atomic compositions. To handle the scarcity of features, we added different statistical-based atomic and electronic features like Pauling electronegativity ($\mathcal{H}_{pauling}$), Mulliken electronegativity ($\mathcal{H}_{mulliken}$), covalent, atomic radii, atomic volume, weight, density, and local-lattice distortion or atomic-mismatch. We approached our model development including extended feature list to achieve higher accuracy without bias and overdependence on specific features. The mean, variance, minimum and maximum values of each of those atomic, electronic and material features was added, and model was trained using only these feature sets. The mean and variance were both weighted according to the proportion of the element within each alloy structures. Mathematically, the mean (μ), variance (σ^2), minimum (min), and maximum (max) values of a feature *X* = *{x₁, x₂,….,xₙ}* are calculated as, $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$; $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2$; $\min(X) = min\{x_1,\ x_2,\ x_3,\ldots,x_n\}$; and $\max(X) = max\{x_1,\ x_2,\ x_3,\ldots,x_n\}$.

Additionally, atomic packing fraction (APF), a measure of how densely the atoms is closely packed in each structure, is added as a feature which is calculated by the fraction of volume in a crystal structure that is occupied by the constituent atoms. APF can be a critical feature in hydrogen storage for alloys, as it is strongly correlated with the availability of interstitial sites where hydrogen atoms can reside within

$$APF = \frac{\frac{4}{3}\pi\rho N_a \sum (r_i^3 c_i)}{\sum M_i c_i}$$

lattice structures. The APF is calculated by, , where ρ is the density of the material, r_i, c_i, and M_i are the atomic radii, compositional fraction, and molar mass of i-th element respectively, and N_a is Avogadro's number. By addition of all these atomic, electronic, and structural features; calculated from elemental properties using the alloy compositions of the dataset; we finally reached a total of 44 features for each composition (see Metho section in the main text). This statistical approach helps in reducing bias and prevents over-reliance on all thermodynamic features, thus improving the model's accuracy without overfitting, which is the primary objective of this work.

Table S1: Features used in machine learning.

Feature Name	Models Used In	Notes	Dimensions
Temperature	H/M, Reaction Enthalpy		1
Pressure	H/M, Reaction Enthalpy		1
Atomic Radius	All		4
Covalent Radius	All		4
Density	All		4
Electron Affinity	All		4
Heat of Formation	All		4
Lattice Constant	All		4
Melting Point	All		4
Specific Heat Capacity	All		4
Electronegativity Pauling	All		4
Electronegativity Mulliken	All		4
Atomic Number	All		4
Enthalpy of Mixing	H/M, Bulk Moduli, Phase Prediction	Calculated as $\Sigma c_i \log c_i$, where c_i is the concentration of the i^{th} element	1
Bulk Modulus	H/M	Predicted using bulk modulus ML model	1
FCC probability	H/M	Predicted using phase prediction ML model	1
BCC probability	H/M		1

Atomic Packing Factor	All	$\frac{\frac{4}{3}\pi\rho N_a \sum (r_i^3 c_i)}{\sum M_i c_i}$, Calculated as where is the density of the material, r_i and c_i are the atomic radii and proportion of element i respectively, M_i is the molar mass of element i, and N_a is Avagadro's number.	1

Model evaluation (see the discussion in the main text): Evaluating the performance of machine learning models is a critical step in assessing their effectiveness. Here we have considered 4 parameters- R-squared (R^2) score, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), for evaluating our models' performance (**Eq. 1-4**). The R-squared score, also known as the coefficient of determination, first introduced by Wright [**S10**] measures the proportion of the variance in the target variable that is explained by the model i.e., it shows how well the dependent variable is evaluated by all the independent variables [**S11**]. It varies from 0 to 1 where a higher R^2 score means a better fit of the dataset. MAE is straightforward and measures the average absolute deviation between the model's predicted values and the actual target values. RMSE penalizes heavily to outliers thus making it more sensitive to outliers in a dataset, whereas MAPE focuses on percentage error and becomes effective in quantifying the relative variations of the predicted data and actual data, but it is ineffective when large errors are determined [**S11**].

$$R2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \quad (1)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y})^2} \quad (3)$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N} \frac{|y_i - \hat{y}|}{|y_i|} \times 100\% \quad (4)$$

where $\bar{y}$ is the mean value of y, $\hat{y}$ is the predicted value of y, and y_i is the i-th data point.

Although it is almost impossible to predict the performance of a machine-learning regression task just by one single evaluation parameter, studies show that the R^2 score (i.e., the coefficient of

determination) becomes the most informative and truthful than any other performance matrices like MAE, MAPE or RMSE [**S11**]

Additionally, the atomic size mismatch adopted to measure the LLD effect here is calculated as [**10**],

$$\delta = 100\sqrt{\sum_{i=1}^{n} c_i\left(1-\frac{r_i}{\bar{r}}\right)^2} \quad (5)$$

Where c_i and r_i are the atomic percentage and radius of individual constituent elements of the alloy respectively.

Extended Table of MPEAs high wt.% H_2 (see main text Table 1):

Table S2. The Cr-Mg-Nb-Mo-Fe-Ti-V based top eight HEAs with improved hydrogen storage properties predicted by ML-GA optimizer, displaying estimated wt% H, formation enthalpy, BCC phase probability, and applied constraints.

HEAs	Estimated H_2 wt %	ΔH [meV/atom]	% probability of BCC phase formation (pBCC)	Constraints
$Cr_{0.16}Mg_{0.09}Nb_{0.08}Ti_{0.35}V_{0.32}$	3.37	-54.53	85%	|ΔH| > 40 p_{BCC} > 0.8 c_{min} = 0.05; c_{max} = 0.35 5 elements
$Cr_{0.11}Fe_{0.05}Mg_{0.32}T_{i0.35}V_{0.17}$	3.18	-61.64	96%	|ΔH| > 60 p_{BCC} > 0.8 c_{min} = 0.05; c_{max} = 0.35 5 elements
$Cr_{0.11}Mg_{0.15}Mo_{0.15}Ti_{0.29}V_{0.3}$	3.01	-50.4	98%	|ΔH| > 40 p_{BCC} > 0.97 c_{min} = 0.05; c_{max} = 0.35 5 elements
$Cr_{0.25}Mg_{0.35}Ti_{0.35}V_{0.05}$	3.24	-59.7	95%	|ΔH| > 40 p_{BCC} > 0.8 c_{min} = 0.05; c_{max} = 0.35 4 elements
$Cr_{0.21}Nb_{0.11}Ti_{0.35}V_{0.33}$	3.45	-38.3	91%	|ΔH| > 30 p_{BCC} > 0.7 c_{min} = 0.05; c_{max} = 0.35 4 elements
$Cr_{0.09}Mg_{0.73}Ti_{0.18}$	4.22	-60.7	91%	|ΔH| > 40 p_{BCC} > 0.8 c_{min} = 1.00; c_{max} = 0.0 3 elements
$Cr_{0.24}Fe_{0.17}Mg_{0.22}Ti_{0.2}V_{0.17}$	3.05	-51.3	85%	|ΔH| > 40 p_{BCC} > 0.8 c_{min} = 0.2; c_{max} = 0.25

				5 elements
$Cr_{0.21}Mg_{0.3}Ti_{0.27}V_{0.22}$	3.09	-57	90%	$\|\Delta H\| > 40$ $p_{BCC} > 0.8$ $c_{min} = 0.2$; $c_{max} = 0.25$ 4 elements

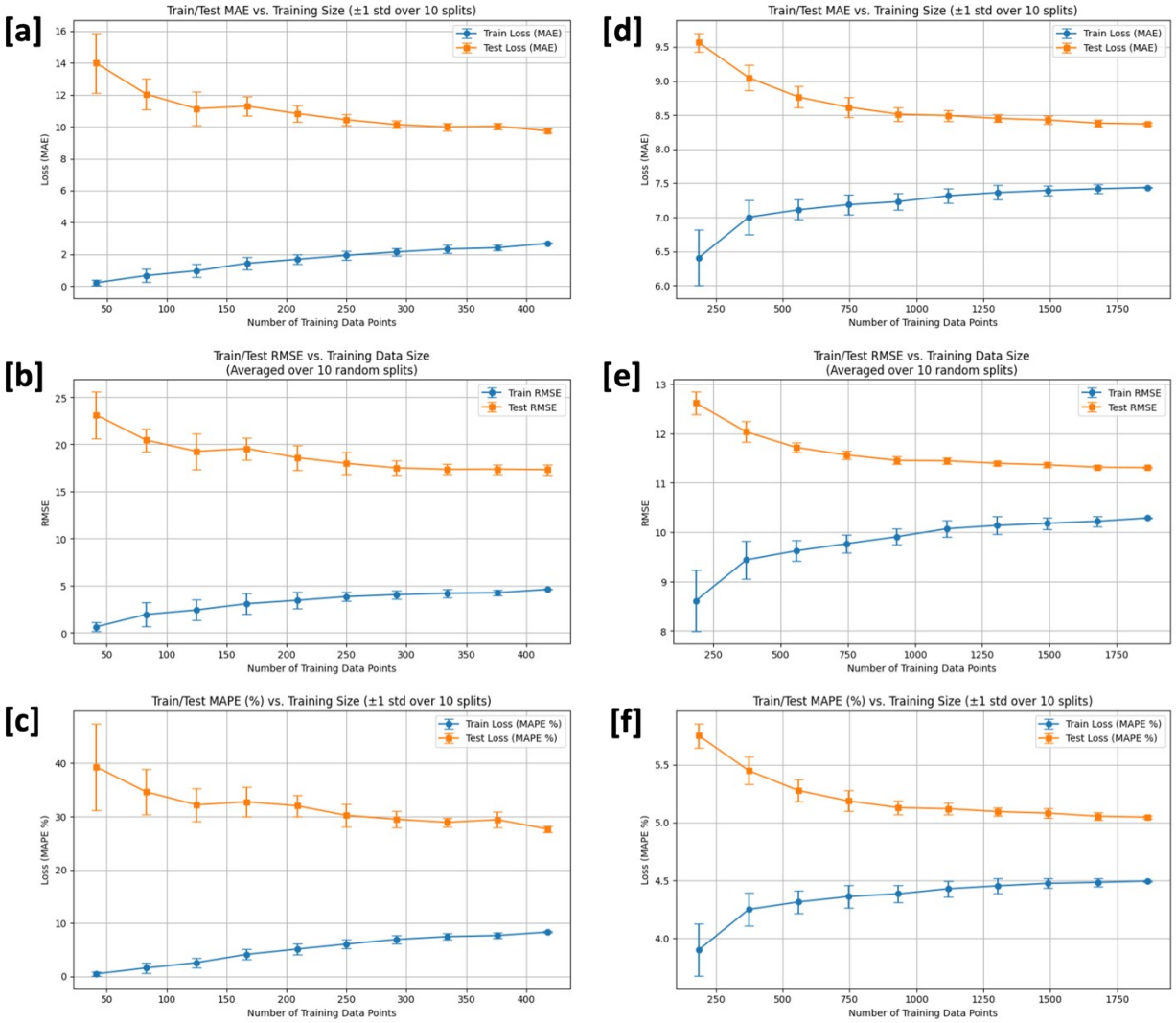

Figure S1. [a-c] Train and test loss on MAE, RMSE, MAPE with respect to data points for reaction enthalpy model; [d-f] Train and test loss on MAE, RMSE, MAPE with respect to data points for bulk modulus model. For both cases the figures show that the loss function in test dataset (also the error bars) goes down with increasing the number of datapoints, which suggest the model did not overfit. Also, this shows the bulk modulus model performs much better (with a MAPE of ~5%) compared to the reaction enthalpy model (MAPE>20%) as also predicted in Figure 3 of main manuscript.

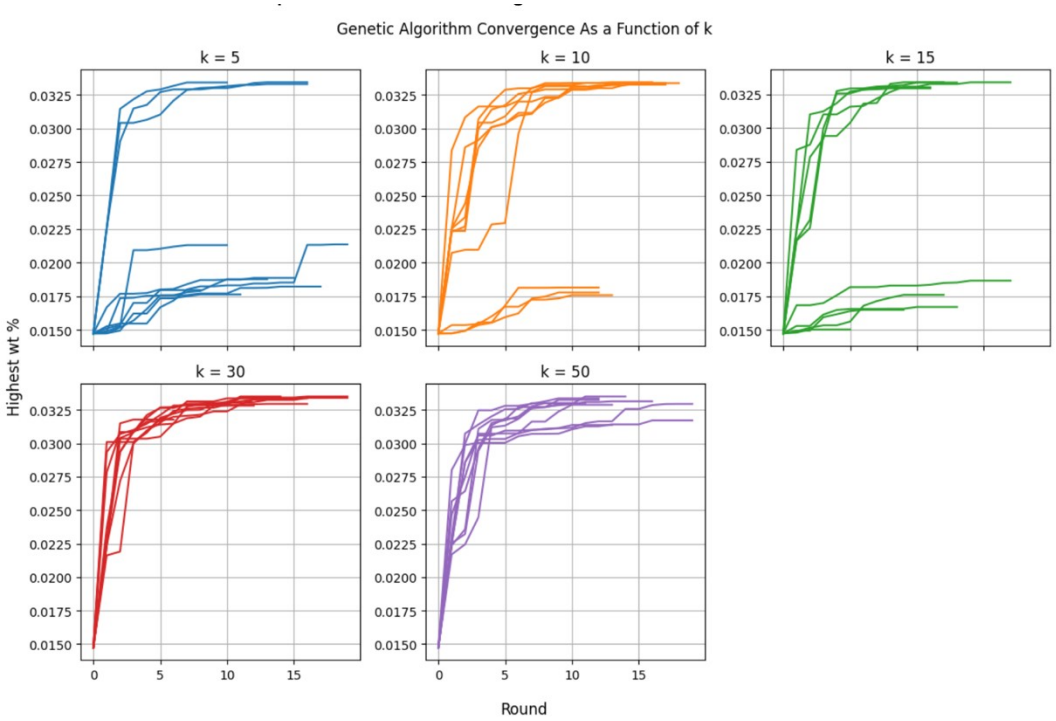

Figure S2. GA convergence vs. penalty strength *k*. Each panel shows the highest predicted wt% (fitness = predicted hydrogen wt% in the current population) as a function of GA round for independent ML-GA replicates (different random seeds), with panels for *k* = 5, 10, 15, 30 and 50 (top left → bottom right). The x-axis is GA round (evaluation generation) and the y-axis is the best predicted wt% observed in that round. The parameter *k* is the multiplicative penalty applied inside the fitness function (the "*k*" penalty term discussed in the text).

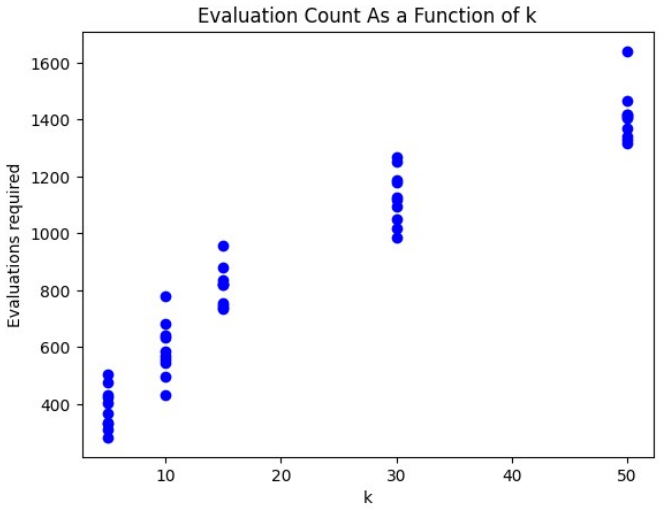

Figure S3. Evaluation count required as a function of parameter *k*. Each point represents the number of evaluations needed for a given value of *k*. The plot shows that the evaluation count increases with *k*, indicating higher computational cost for larger *k*.

Feature selection and hyperparameter optimization: In addition to the features described in main text, we incorporated the featurization and feature selection strategies reported by Zhang et al. [**S21**]. Hyperparameter optimization and feature selection were simultaneously performed using a genetic algorithm applied to the 145 default features available in Matminer (**Table S3**). The developed framework is fully generalizable and applicable to both prediction and optimization tasks. For demonstration, we re-ran our ΔH and phase classification models entirely within this new framework. The updated results are shown below. A modest improvement in performance was observed when using LightGBM compared to Random Forest, though the enhancement remained marginal even after hyperparameter tuning. Importantly, this minor gain did not lead to any qualitative change in the outcomes of the genetic algorithm. Detailed results and comparisons of these new techniques are provided in the Supplementary Information.

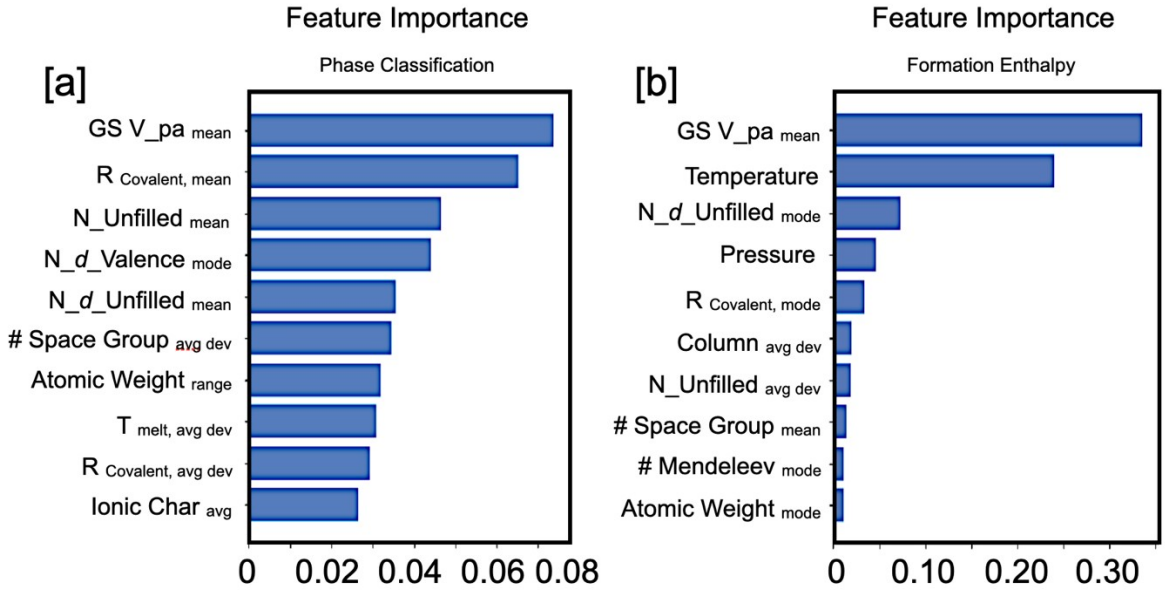

Figure S4. Feature importance for (a) phase classification and (b) formation enthalpy prediction using the updated framework. In both models, GS_V_pa (mean) emerges as the dominant descriptor, highlighting the strong role of atomic volume in governing phase stability and thermodynamics. Other key contributors include covalent radius, unfilled d-states, and space group variability, reflecting the interplay between atomic size, electronic structure, and crystallographic symmetry in determining alloy behavior.

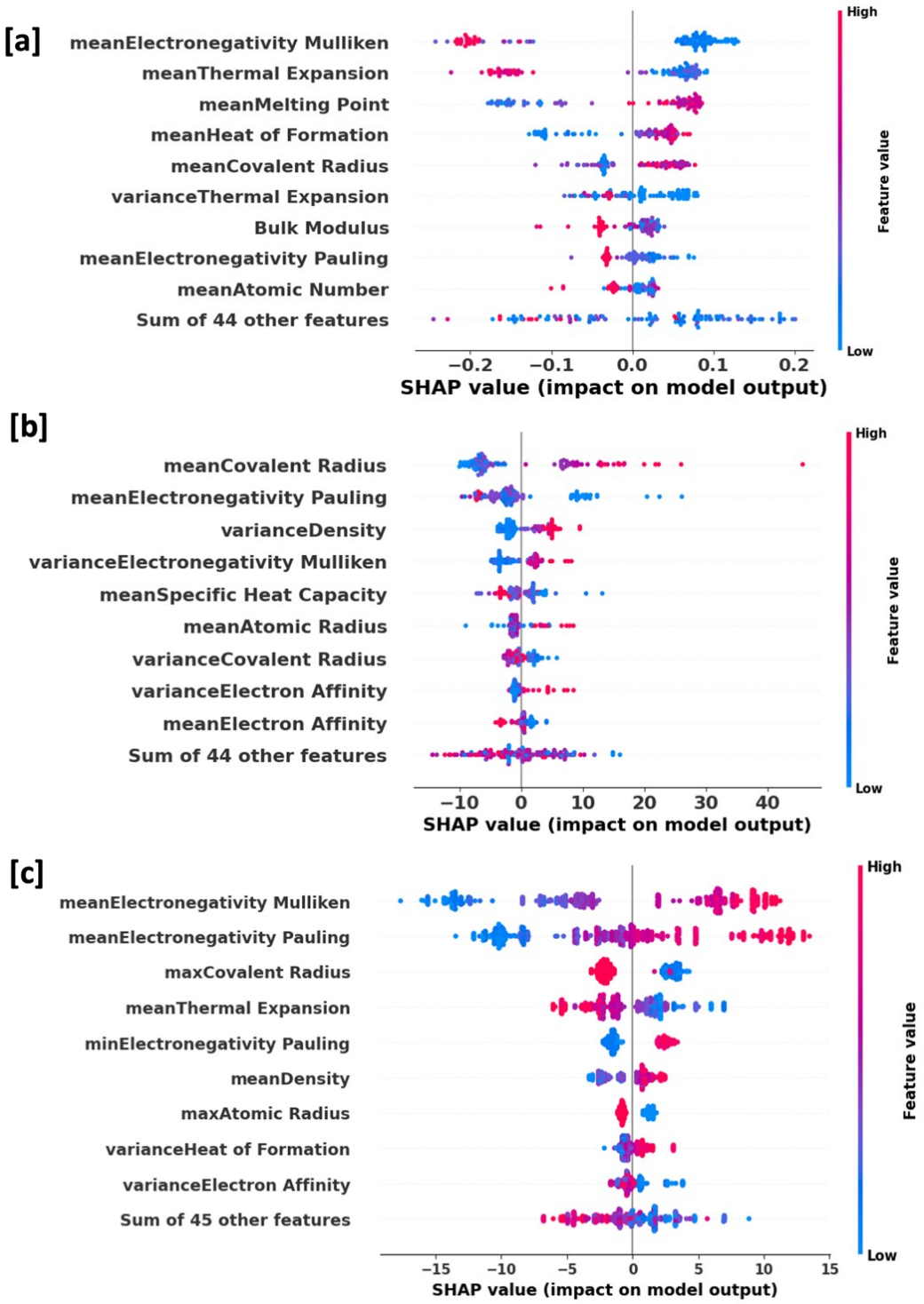

Figure S5. SHAP analysis on (a) H/M, (b) Reaction enthalpy, (c) Bulk modulus models. These results agree well with the global feature importance plots shown in Figure 3(c, f, i). For all the 3 models we find the top 3-5 global features responsible for the ML prediction (Figure 3) also came out in the SHAP analysis. For example, higher values of *meanElectronegativity Mulliken* (blue points on the right of Figure S2 'a') reduce predicted H/M, whereas lower values (pink/red points) increase H/M.

Table S3: Feature selection and hyperparameter optimization for phase classification and formation enthalpy models for hydrogen storage materials.

	Phase Classification	**Enthalpy of Formation**
Optimal model type	Random Forest	Random Forest
Old 5-fold CV accuracy/R^2	0.77	0.71
New 5-fold CV accuracy/R^2	0.80	0.72
Optimized Hyperparameters	"n_estimators": 360, "max_depth": 16, "min_samples_split": 4, "min_samples_leaf": 2, "max_features": 0.52, "max_samples": 0.85	"n_estimators": 119, "min_samples_split": 3, "min_samples_leaf": 2, "max_samples": 1.0

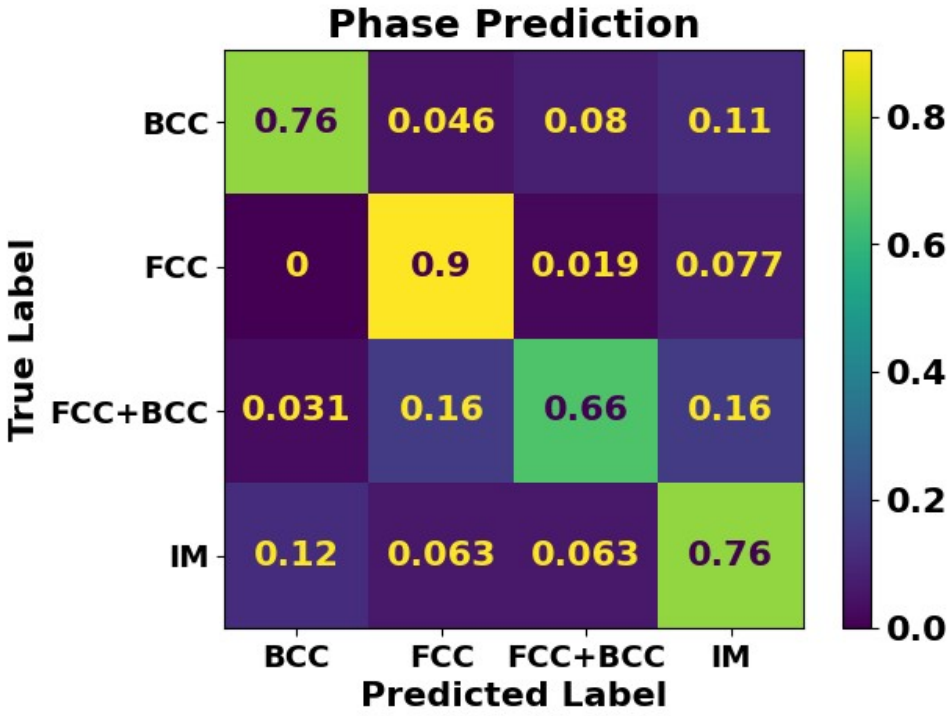

Figure S6. Random Forest (RF) based confusion matrix showing lower phase classification accuracy across different phases compared to gradient boost classifier (see **Fig. 4a**).

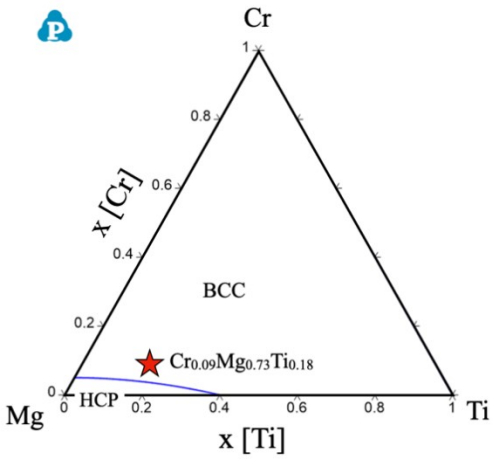

Figure S7: Thermodynamic phase diagram of Cr-Mg-Ti MPEA at 800 K, showing BCC phase stability over much of the Mg composition range. HCP phase appears only for Cr < 0.05 at.-frac. And Ti < 0.4 at.-frac.

Pair-distribution function: The molecular dynamic simulations were performed within LAMMPS using Lennard Jones potentials [**S12-S20**]. In **Figure S8**, the radial distribution functions for all five BCC MPEAs show a very similar, well-defined first-neighbor peak centered at roughly 2.7 to 3.0 Å with a peak height of order g(r) $\approx$ 4 to 5 at 300 K, followed by a pronounced second peak near 4.6 to 5.0 Å with g(r) $\approx$ 1.5 to 2.0 and weaker, damped oscillations beyond 6 Å. These features are exactly what one expects for a BCC-derived short-range order: a sharp, high-intensity first peak (nearest-neighbor shell) and subsequent broader peaks corresponding to second and higher coordination shells. Small alloy-to-alloy differences are visible as slight peak shifts (on the order of $\leq$ 0.1–0.2 Å) and modest variations in first-peak height: for example one composition (red/MPEA1) shows a marginally higher and slightly left-shifted first peak compared with others, consistent with either a slightly smaller average nearest-neighbor distance or stronger local ordering around certain species.

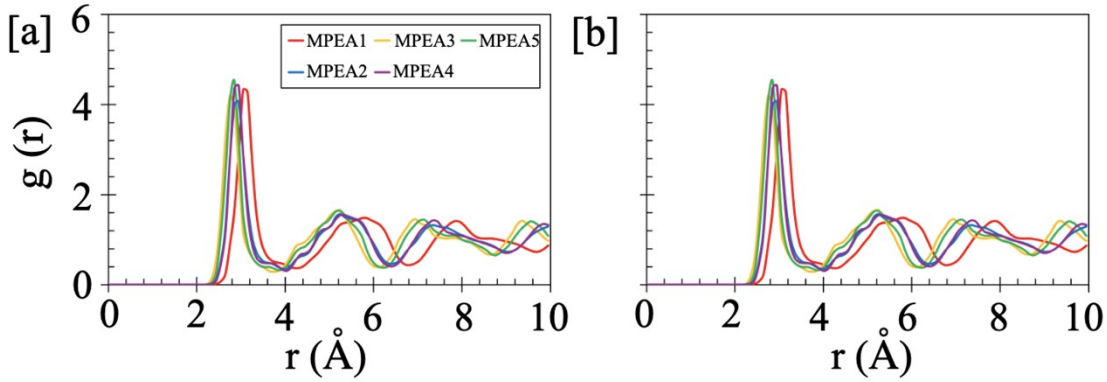

Figure S8. Pair distribution function for the five different BCC high entropy alloys (MPEAs) in presence of hydrogen as shown in **Table 2** at (a) 300 K, and (b) 600 K, showing structural integrity at BCC phase and are in good agreement with DFT and ML-GA study.

Mean squared displacement (MSD) profiles of individual atoms calculated with *molecular dynamic simulation:*

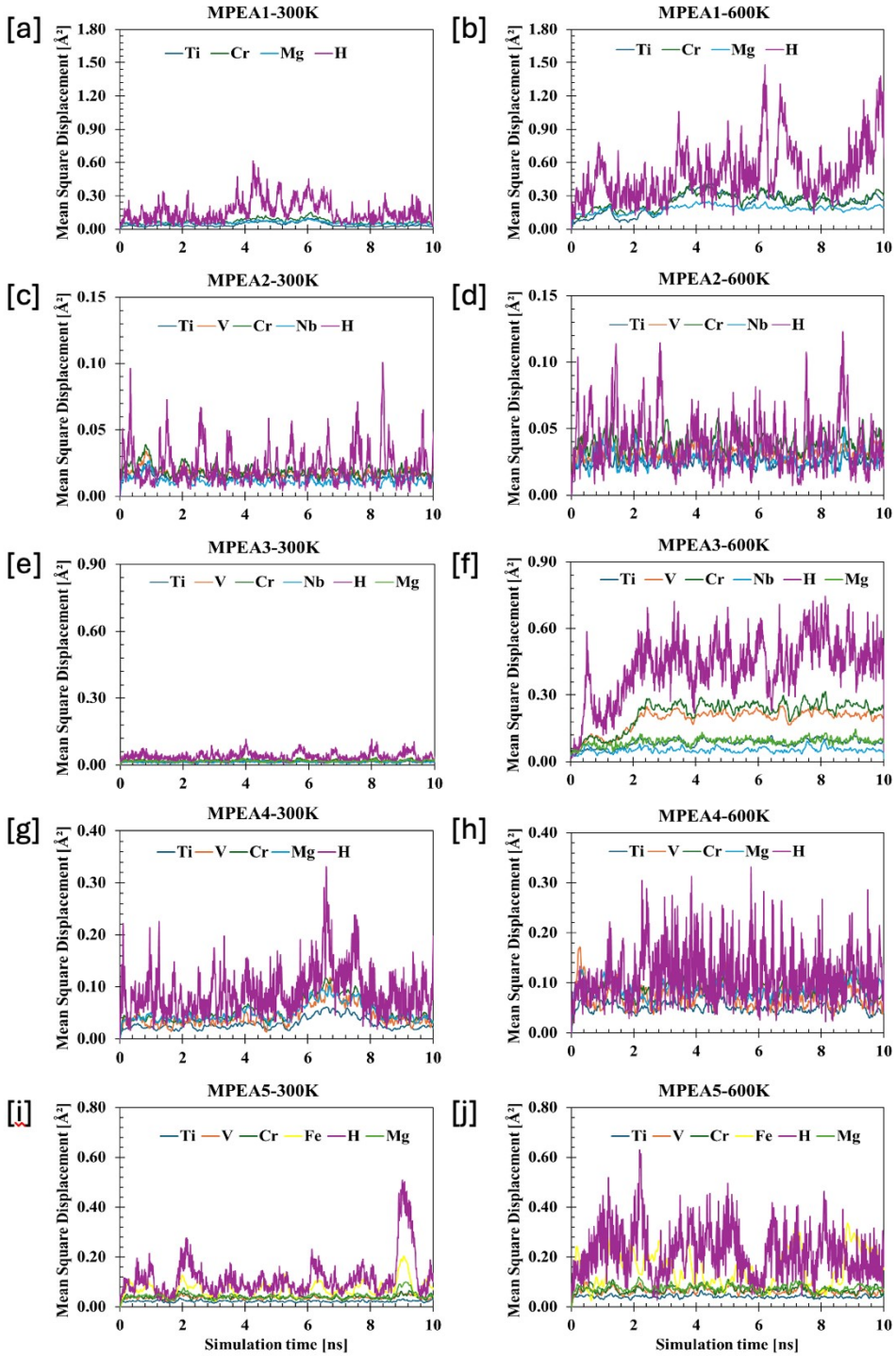

Figure S9. Mean squared displacement (MSD) profiles of individual atoms in) in MPEA1–MPEA5 at 300 K and 600 K (Ti, V, Cr, Nb, Fe, and Mg) in presence of hydrogen as a function of simulation time. The MSD trajectories exhibit stable and converged behavior over 10 ns, confirming adequate statistical sampling. Hydrogen shows consistently higher MSD values relative to the metallic constituents especially in MPEA1 and MPEA3 at higher temperature, indicating its dependence on temperature and alloy compositions.

Electronic-structure analysis of BCC $Cr_{0.09}Mg_{0.73}Ti_{0.18}$ MEA: In the pristine **BCC $Cr_{0.09}Mg_{0.73}Ti_{0.18}$** alloy [**Fig. S10a**], the projected density of states (PDOS) reveals a metallic character with finite states at the Fermi level, primarily arising from Ti and Cr 3*d*-orbitals, while Mg contributes a broad s-like distribution extending below the Fermi level. When hydrogen occupies the octahedral site [**Fig. S10a**], a distinct H-1*s* peak emerges around -5 eV, well below the Fermi level, signifying a deeply bound and localized hydrogen state with minimal hybridization to the surrounding metal orbitals. This deep, isolated H state reflects strong ionic-type bonding and limited perturbation of the alloy's metallic backbone. Conversely, hydrogen at the tetrahedral site [**Fig. S10a**] introduces an H-1*s* feature that shifts upward to approximately -2 eV and overlaps noticeably with Ti and Cr 3*d*-states, indicating enhanced orbital mixing and stronger covalent hybridization. This interaction slightly modifies the metallic DOS near E_{Fermi}, suggesting more active electronic coupling between H and the transition-metal sublattice. Overall, the octahedral site provides deeper stabilization of hydrogen, while the tetrahedral site promotes stronger hybridization without disrupting the metallic nature of the alloy.

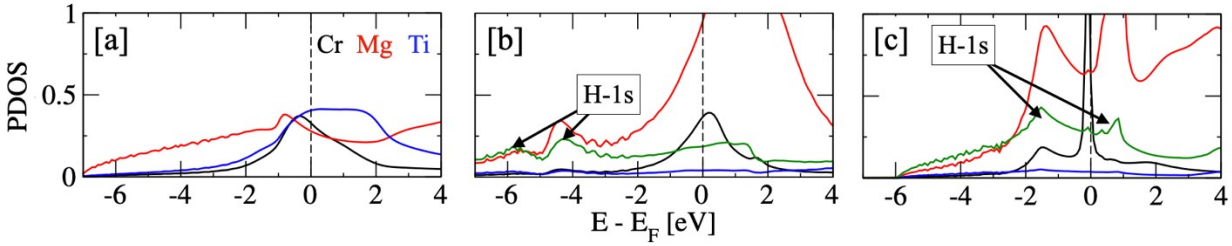

Figure S10. Projected density of states (PDOS) of BCC $Cr_{0.09}Mg_{0.73}Ti_{0.18}$ without hydrogen (a), with hydrogen at the octahedral site (b), and at the tetrahedral site (c). Black, red, blue, and green lines represent Cr, Mg, Ti, and H-*1s* contributions, respectively. The Fermi level is set to zero energy.

References:

S1. G. Kresse, and J. Hafner, Ab initio molecular dynamics for liquid metals, Phys. Rev. B, 1993, **47**, 558-561.

S2. G. Kresse, and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B, 1999, **59**,1758-1775.

S3. J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett., 1997, **77**, 3865.

S4. P Singh, MK Harbola, B Sanyal, A Mookerjee, Accurate determination of band gaps within density functional formalism, Phys. Rev. B, 2014, **87** (23), 235110.

S5. P Singh, MK Harbola, M Hemanadhan, A Mookerjee, DD Johnson, Better band gaps with asymptotically corrected local exchange potentials, Phys. Rev. B, 2016, **93** (8), 085204.

S6. P. Söderlind, PEA Turchi PEA, A. Landa, and V. Lordi, Ground-state properties of rare-earth metals: an evaluation of density-functional theory, Journal of Physics: Condensed Matter, 2014, **26**, 416001.

S7. T.J. Giese, and D.M. York, A GPU-Accelerated Parameter Interpolation Thermodynamic Integration Free Energy Method, J Chem Theory Comput, 2018, **14**, 1564-82.

S8. R. Singh et al. Accelerating computational modeling and design of high-entropy alloys. Nat. Comput Sci, 2021, **1**, 54-61.
S9. H. J. Monkhorst and J. D. Pack, Special points for Brillouin-zone integrations, Phys. Rev. B, 1976, **13**, 5188.
S10. T. Chiba et al. Correlation and causation, J Agric Res, 1921, **20**, 557.
S11. D. Chicco et al. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci, 2021, **7**, 1–24.
S12. J. Delhommelle et al. Inadequacy of the Lorentz-Berthelot combining rules for accurate predictions of equilibrium properties by molecular simulation, Mol Phys, 2001, **99**, 619–25.
S13. V.P. Filippova et al. Calculation of the parameters of the Lennard-Jones potential for pairs of identical atoms based on the properties of solid substances, Inorganic Materials: Applied Research, 2015, **6**, 1–4.
S14. D.W. Jacobson et al. Lennard Jones, Morse, and N-M potentials for metals, Comput Mater Sci, 2022, **205**, 111206.
S15. V. Varshney et al. MD simulations of molybdenum disulphide (MoS2): Force-field parameterization and thermal transport behavior, Comput Mater Sci, 2010, **48**, 101–8.
S16. S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, Sandia Report, Sand91-1144, May 1993. https://doi.org/10.2172/10176421.
S17. W.B. Paul, Molecular dynamics simulation, elementary methods. By J. M. Haile, Wiley, Chichester 1992, 489 pp.; Advanced Materials, 1993, **5**, 223–4.
S18. W.C. Swope et al. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. J Chem Phys ,1982, **76**, 637–49.
S19. S. Nosé, A unified formulation of the constant temperature molecular dynamics methods. J Chem Phys, 1984, **81**, 511–9.
S20. W.G. Hoover, Canonical dynamics: Equilibrium phase-space distributions. Phys Rev A, 1985, **31**, 1695.
S21. Y. Zhang et al. A New Phase Classifier with an Optimized Feature Set in ML-Based Phase Prediction of High-Entropy Alloys. *Applied Sciences*, 2023, ***13***(20), 11327.