An Effective Machine Learning Frame for Materials Discovery Structured by a Chemical Concept

Yuanhui Sun[1,2], Austin Ellis[1], Samantha Scott[1], Xin Chen[2], Maosheng Miao[1]

[1] California State University Northridge, Northridge, CA, USA

[2] Suzhou Laboratory, Suzhou, China

DOI: 10.1021/jacs.5c11731

JACS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

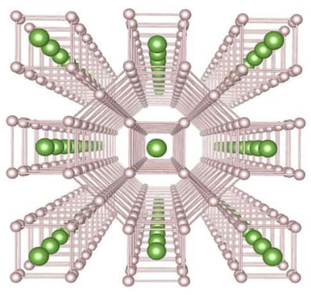
Abstract

Despite intensive work on hydrogen-rich binaries, large regions of composition space remain unexplored, especially systems with non-integer non-metal-to-metal ratios. We combine the chemical template concept with machine learning to build a targeted structure-discovery workflow that efficiently predicts new stable compounds. Applying this framework, we identify 13 new structural prototypes and 31 stable metal superhydrides, a 23% increase over previous datasets. Many of these phases feature 3D hydrogen clathrate cages and primitive cells with more than 50 atoms, a regime that is typically challenging to explore with conventional methods. Notably, 19 of the newly discovered superhydrides have predicted superconducting transition temperatures T_c above 100 K, underscoring the promise of chemical-template-guided machine learning for uncovering higher T_c materials in complex hydrogen-rich systems.

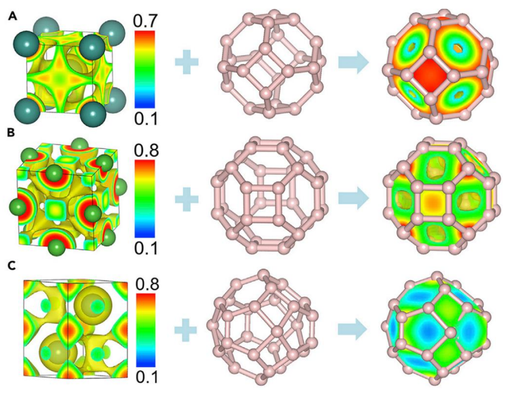
Background

High Pressure Chemistry

- High-pressure chemistry unveils unique material structures leading to discoveries like phase transitions and unusual stoichiometries
- It can induce superconductivity, transform the typical chemistry of elements, and stabilize compounds with distinctive electronic, mechanical, and optical characteristics

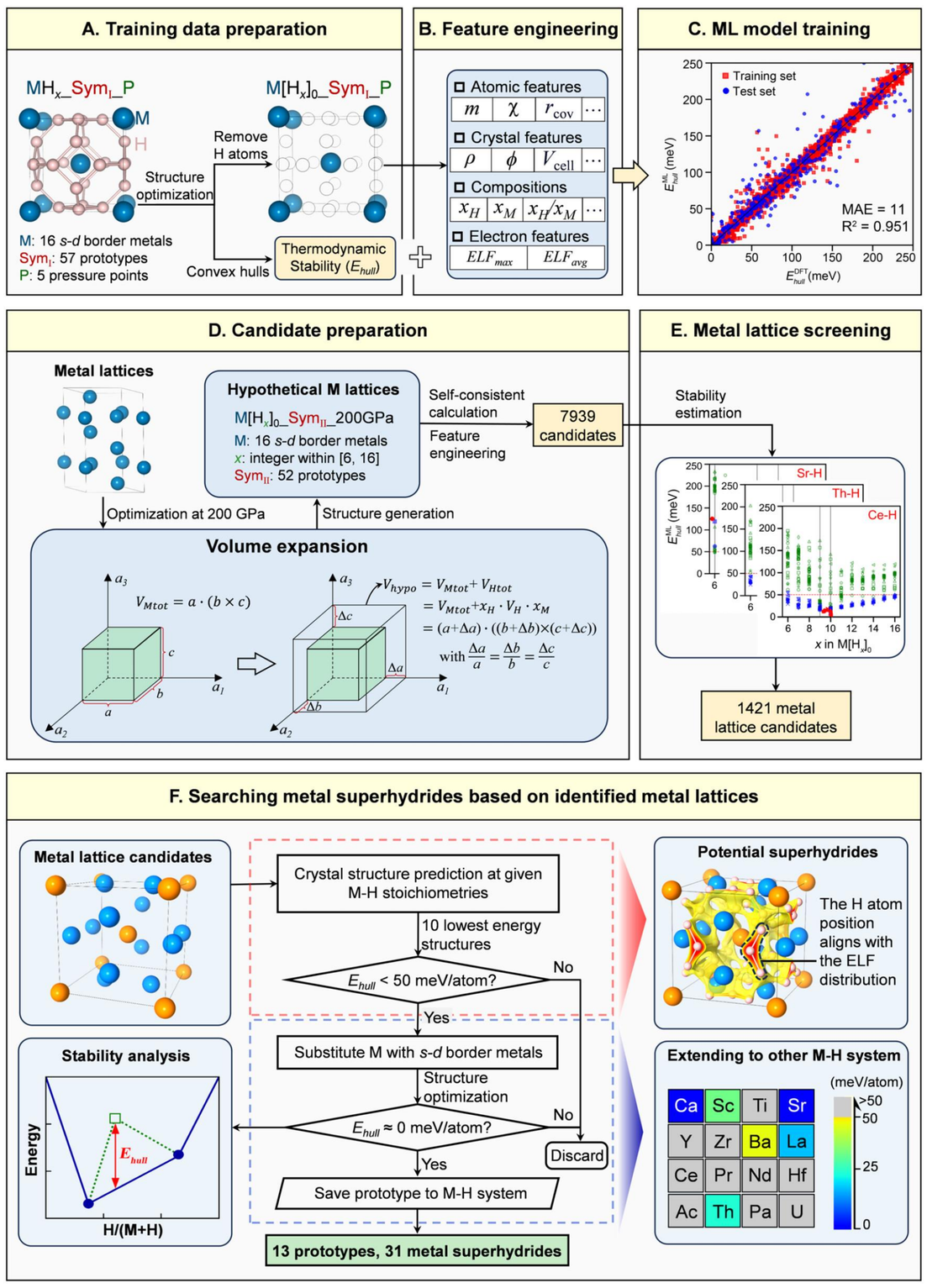


Chemical Template Effect



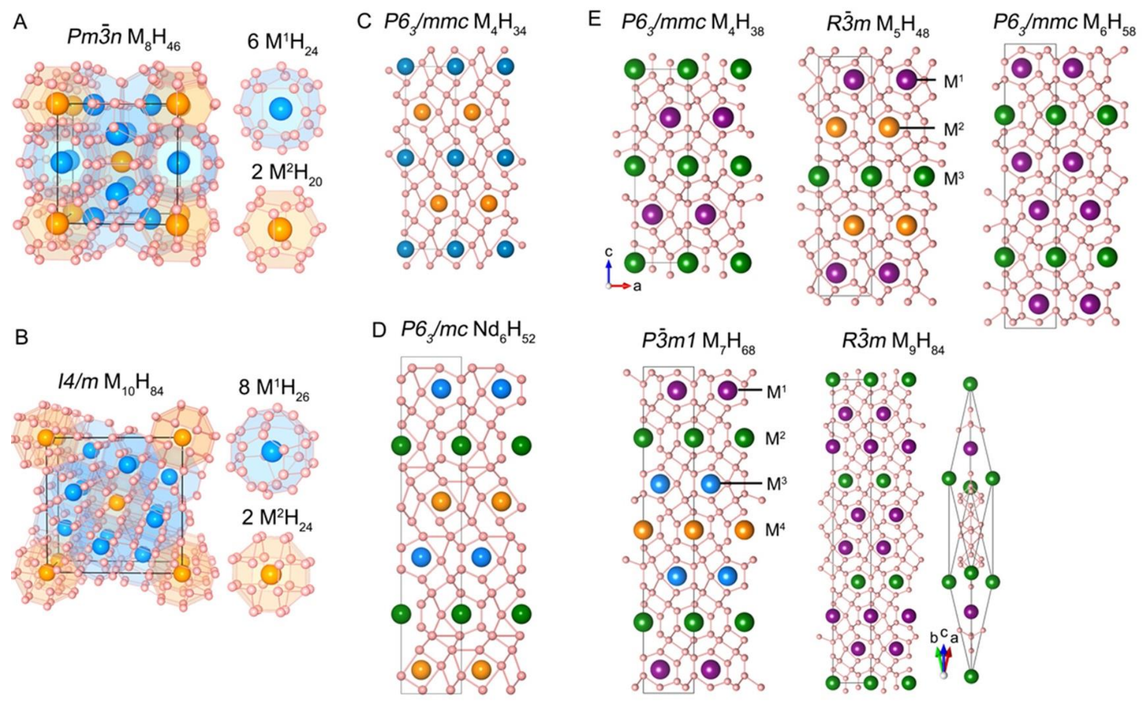
- Electrons occupying the quantum orbitals located at the interstitials of metal sublattices (quasi-atoms) form a template that assists the assembly of H lattices
- Utilizing a template can significantly improve the efficiency of searching for new complex compounds.

Methods and Workflow

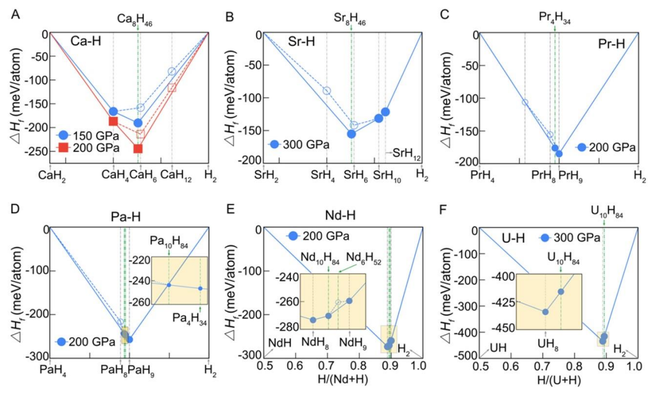


Results

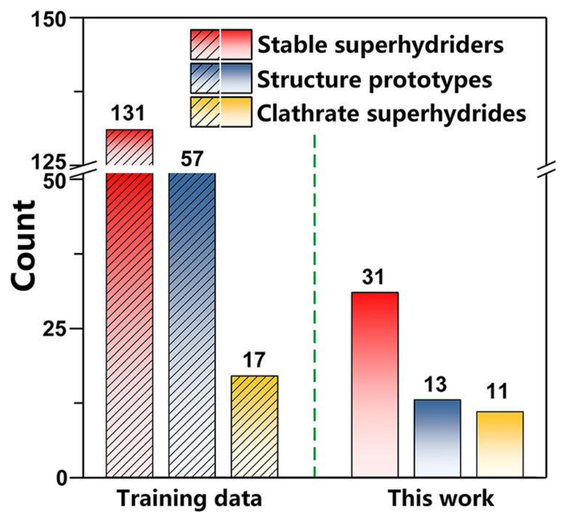
Crystal Structures of Newly Identified Metal Superhydrides



- Template guided search uncovers a family of previously unknown MH$_x$ structures, including M_8H_{46}, $M_{10}H_{84}$, $MH_{9+\delta}$, MH_{10}, and M_4H_{52} with distinct hydrogen cage topologies and multiple metal sublattices
- These structures show that the chemical template workflow can discover very complex superhydrides with non-integer ratios and complicated structures that are difficult to explore with conventional unconstrained crystal structure searches



- Convex hull calculations at 150 to 300 GPa show that the newly discovered MH$_x$ phases with non-integer ratios lie on or very near the ground state hull across several metal hydride systems
- The updated hulls are reshaped by the new phases and reveal previously missed stability windows at non-integer hydrogen ratios
- Template guided search adds 31 new stable superhydrides, 13 new structure prototypes, and 11 new clathrate superhydrides on top of the original training set, substantially enlarging the design space.
- Demonstrates that the new workflow based on chemical templates can discover complex compounds with modest training data.



Conclusion

- Template effect is a fundamental chemical mechanism that stabilizes many compounds with high non-metal content.
- Implementing a machine learning workflow based on chemical templates enables batch discovery of new compounds with complex stoichiometry and structure beyond the training set.
- Many of the predicted superhydrides lie on or very near the convex hull, and 19 have estimated superconducting transition temperatures above 100 K, so the workflow prioritizes experimentally relevant high T_c candidates.
- This concept driven ML strategy can be extended to ternary superhydrides and other materials where localized electrons act as templates, offering a scalable route to discover complex quantum materials.

Acknowledgments

M.M. and A.E. acknowledge the support of the Department of Defense under grant No.W911NF2310232. M.M. and Y.S. acknowledge the support of NSF under grant Nos. DMR 1848141 and OAC-2117956. M.S. also acknowledge the support of the Camille and Henry Dreyfus Foundation and the California State University RSCA awards. ***mmiao@csun.edu**