Machine Learning-Empowered Study of Metastable γ-CsPbI₃ under Pressure and Strain

Minkyung Han[1,2,*], Cheng Peng[3], Ruyi Song[3], Feng Ke[3,#], Youssef S. G. Nashed[4], Wendy L. Mao[1,3], Chunjing Jia[5,*], and Yu Lin[3,*]

[1]Department of Earth and Planetary Sciences, Stanford University, Stanford CA, 94305, United States
[2]Department of Materials Science and Engineering, Stanford University, Stanford CA, 94305, United States
[3]Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, Menlo Park CA, 94025, United States
[4]Machine Learning Initiative, SLAC National Accelerator Laboratory, Menlo Park CA, 94025, United States
[5]Department of Physics, University of Florida, Gainesville FL, 32611, United States
[*]mhan8@stanford.edu, chunjing@phys.ufl.edu, lyforest@slac.stanford.edu

Supplementary Information

Accuracy of DFT-calculated γ-CsPbI$_3$ under pressure and strain

Accurate ML predictions compared to experimental values rely on the precision of the underlying DFT calculations. Thus, it is crucial to assess the error between calculations and experiments and identify optimal DFT settings to minimize inaccuracies. Here, we evaluated the impact of different exchange-correlation functionals by comparing zero-pressure band gap values calculated using PBE and PBEsol pseudopotential libraries in Quantum Espresso (QE). Ultrasoft pseudopotentials with scalar relativistic effects were used, and a 4×3×4 K-point grid centered on the Γ point was employed for the relax-calculation. The PBE functional yielded a band gap of 1.83 eV, while PBEsol resulted in 1.64 eV. It is noteworthy that the experimental band gap value of γ-CsPbI$_3$ at ambient pressure is approximately 1.7 eV[1].

It is widely recognized that the commonly used GGA exchange-correlation functionals, including the PBE functional utilized in our study, tend to underestimate the band gap of materials[2]. To evaluate the effects of more accurate approximations on the band gap calculation, we performed calculations using the Vienna Ab initio Simulation Package (VASP) code[3, 4]. We employed VASP 6.4.1 with PBE projector-augmented wave (PAW) pseudopotentials and various approximations for band structure calculations. The plane-wave basis has been set with a cut-off energy of 240 eV. The experimental lattice constants in a primitive cell are a = 9.06 Å, b = 12.62 Å, c = 8.75 Å, α = 90°, β = 90°, and γ = 90°. The band structures, which further consider the GW0 quasiparticle method[5] and spin-orbit coupling (SOC), are shown in Figure S11. After the self-consistent field (scf) calculation using the PBE functional at zero pressure, VASP and QE yielded band gap values of 1.80 eV and 1.83 eV, respectively. The GW0 quasiparticle method increased the band gap by 1.28 eV, whereas considering SOC significantly decreased the band gap by 0.95 eV. After applying both GW0 and SOC, the resulting band gap value was 2.04 eV, slightly higher than the original band gap of 1.80 eV calculated using the PBE without any corrections. These results suggest that using PBE solely provides band gap values closer to the experimental band gap of CsPbI$_3$ while maintaining a relatively low computational cost. Additionally, the ML prediction performance is independent of the DFT accuracy, which is the primary focus of this work. Thus, we concentrate on the ML prediction results based on PBE functionals.

Comparison between DFT-calculated results and experimental data

We further compared our high-pressure DFT results with limited experimental observations and the comparison is overall consistent. The experimental optical band gap values in $CsPbI_3$ perovskite nanocrystals, based on high-pressure photoluminescence (PL) results reported by Beimborn II *et al.*, displayed a similar trend and range of band gap change in the order of tens of meV, compared to our DFT calculations[6]. The observed band gap exhibited a redshift towards lower energy up to 0.3 GPa, attributed to the contraction of Pb-I bond lengths within the octahedra, followed by a blueshift towards higher energy up to 2.5 GPa due to the dominance of octahedral tilting over octahedral compression. The calculated range of band gap change is also consistent with our own PL measurements on pressure-preserved bulk γ-$CsPbI_3$ at further compression, although no redshift was observed within 0-2 GPa[7]. Distinctions of the experimental materials, e.g., different particle sizes, different structures that arise from the material's intrinsic instability and complex synthesis routes, led to discrepancies within experimental reports. In addition, temperature differences between experiments and DFT simulations remain an issue that can lead to varying trends and values in band gap evaluations especially for this class of soft and dynamic materials.

In our examination of the pressure-tuned system, we observed discontinuities in structural parameters including lattice parameters and unit cell volume at approximately 0.4 GPa and 1.5 GPa, suggesting potential first-order phase transitions as depicted in Figure S1. The observed phase transitions within these halide perovskites are consistent with expectations, given that these materials are inherently soft, possessing low bulk moduli, making them particularly susceptible to compression. There is substantial evidence in the literature of pressure-induced phase transitions in analogous materials[8]. Similarly, in the strain-tuned system, we identified discontinuous changes in some structural features at specific strain levels, approximately ±1~2%. Figure S3 illustrates this discontinuity. These observed discontinuities in structural parameters may significantly contribute to the enhanced stability of the γ-$CsPbI_3$ perovskite phases, corroborating findings from previous studies. For instance, Steele *et al.* enhanced the stability of γ-$CsPbI_3$ thin films at room temperature by applying approximately 1% of biaxial strain[9]. Complementary to this, Li *et al.* presented computational evidence that suggests the free energy of the γ-phase is lower than that of the δ-phase at strain levels around 3%, which also coincides with a notable decrease in the band gap[10]. They determined that strain along the <012> direction optimally stabilizes the γ-phase. In the

context of our study, the equivalent direction in our unit cell system corresponds to the <120> direction, which involves the *a* and *b* lattice parameters. Hence, we infer that strain levels ranging from approximately ±1~3% are pivotal for inducing the stabilization of the perovskite γ-phase over the non-perovskite δ-phase.

For further comparisons between the calculations and the experimental measurements from a comparable system, we turned to $CsPbBr_3$, which shares a similar perovskite structure with γ-$CsPbI_3$. $CsPbBr_3$ demonstrates sufficient stability to allow for experimental investigation of its structural responses under high pressure. We conducted high-pressure X-ray diffraction (XRD) measurements on powdered $CsPbBr_3$ at beamline 11-ID-B at the Advanced Photon Source, Argonne National Laboratory, utilizing X-ray photon energy at 86.7 keV (0.143 Å)[11]. Concurrently, we performed DFT calculation to computationally apply hydrostatic pressure to the 0 GPa $CsPbBr_3$ structure[12] up to 2.6 GPa, following the same procedure as γ-$CsPbI_3$. The DFT-calculated lattice parameters and unit cell volumes exhibit reasonable agreement with the XRD results at the corresponding pressure points (Figure S12). There is some discrepancy in lattice parameters between XRD and DFT, with *a* and *c* showing differences of below 2.5%, while lattice parameter *b* exhibits a smaller gap of below 1.2% as the pressure increases. Consequently, the unit cell volume can differ by up to 3.2% up to 2.6 GPa between the computational and experimental values. We assume a similar experiment-calculation error for the γ-$CsPbI_3$ system in this study.

Feature importance evaluation

In our analysis of feature importance, we observed a significant correlation among structural features (Figure S8). This high degree of correlation can lead to unreliable feature selection results and misleading interpretation[13, 14]. To demonstrate the influence of multicollinearity on the feature selection, we employed lasso regression for the pressure-tuned system and permutation feature importance for the strain-tuned system, as illustrated in Figure S9. Utilizing the features identified through these methods, we developed new models for predicting band gap and enthalpy. The performance of these models, along with the selected input features, is elaborated in Figure S10. Notably, the new linear regression model for predicting the band gap of the pressure-tuned system, reconstructed with features chosen by lasso regression, demonstrated a significant reduction in performance compared to the original model that incorporated all eight features. Although numerous feature selection methods exist to address the issues of multicollinearity among input features, they bear the risk of excluding crucial features[14]. Given that

our models incorporate only eight structural features, we deduced that retaining all eight is imperative for achieving high accuracy in the band gap predictions of the pressure-tuned system. Conversely, the pressure-tuned enthalpy predictions with the new model based on selected features showed comparable performance with that of the original model. This consistency can be attributed to the *PV* term's direct linear relationship with pressure, which ensures a linear correlation with each structural feature, allowing for accurate predictions with a limited set of features. Interestingly, the new models for both band gap and enthalpy predictions in the strain-tuned system, developed based on features selected through permutation feature importance, exhibited comparable performance to the original models, despite the existing multicollinearity. This can likely be ascribed to the inherent randomness of the random forest model in selecting input subsets, which is more adept at mitigating the effects of multicollinearity compared to linear regression models.

R^2		Linear Regression	Random Forest	CGCNN	ALIGNN
Band gap	Pressure	**0.9980±0.0008**	0.9622±0.0061	0.9757±0.0113	0.9924±0.0032
	Strain	0.7495±0.0673	0.9905±0.0030	0.9846±0.0050	**0.9984±0.0007**
Enthalpy	Pressure	0.9995±0.0001	0.9995±0.0002	0.9995±0.0001	**0.9996±0.0001**
	Strain	0.9386±0.0120	0.9665±0.0158	0.9960±0.0013	**0.9988±0.0006**

Table S1. Summary of ML prediction performance. Average R^2 values and corresponding standard deviations from the 5-fold cross-validation are presented. The best prediction results for each target are indicated in bold.

References

1. Wang, Y. *et al.* Thermodynamically stabilized β-CsPbI$_3$ –based perovskite solar cells with efficiencies > 18%. *Science* **365**, 591–595 (2019).

2. Rajan, A. C. *et al.* Machine-learning-assisted accurate band gap predictions of functionalized mxene. *Chem. Mater.* **30**, 4031–4038 (2018).

3. Kresse, G. & Hafner, J. Ab initio molecular dynamics for open-shell transition metals. *Phys. Rev. B* **48**, 13115–13118, DOI: 10.1103/PhysRevB.48.13115 (1993).

4. Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251–14269, DOI: 10.1103/PhysRevB.49.14251 (1994).

5. Aryasetiawan, F. & Gunnarsson, O. The gw method. *Reports on Prog. Phys.* **61**, 237 (1998).

6. Beimborn, J. C., Hall, L. M., Tongying, P., Dukovic, G. & Weber, J. M. Pressure response of photoluminescence in cesium lead iodide perovskite nanocrystals. *J. Phys. Chem. C* **122**, 11024–11030 (2018).

7. Ke, F. *et al.* Preserving a robust $CsPbI_3$ perovskite phase via pressure-directed octahedral tilt. *Nat. Commun.* **12**, 461 (2021).

8. Jaffe, A. *et al.* High-pressure single-crystal structures of 3D lead-halide hybrid perovskites and pressure effects on their electronic and optical properties. *ACS Cent. Sci.* **2**, 201–209 (2016).

9. Steele, J. A. *et al.* Thermal unequilibrium of strained black $CsPbI_3$ thin films. *Science* **365**, 679–684 (2019).

10. Li, Z. *et al.* Elastic and electronic origins of strain stabilized photovoltaic γ-CsPbI3. *Phys. Chem. Chem. Phys.* **22**, 12706–12712 (2020).

11. Girdzis, S. P. P. *Local and Average Structural Evolution of Halide Perovskites Upon Compression* (Stanford University, 2021).

12. Szafrański, M., Katrusiak, A. & Ståhl, K. Time-dependent transformation routes of perovskites $CsPbBr_3$ and $CsPbCl_3$ under high pressure. *J. Mater. Chem. A* **9**, 10769–10779 (2021).

13. Biau, G. & Scornet, E. A random forest guided tour. *Test* **25**, 197–227 (2016).

14. Chan, J. Y.-L. *et al.* Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics* **10**, 1283 (2022).

15. Hinuma, Y., Pizzi, G., Kumagai, Y., Oba, F. & Tanaka, I. Band structure diagram paths based on crystallography. *Comput. Mater. Sci.* **128**, 140–184 (2017).

16. Hubbard, C. R. & Calvert, L. D. The pearson symbol. *Bull. Alloy. Phase Diagrams* **2**, 153–157 (1981).
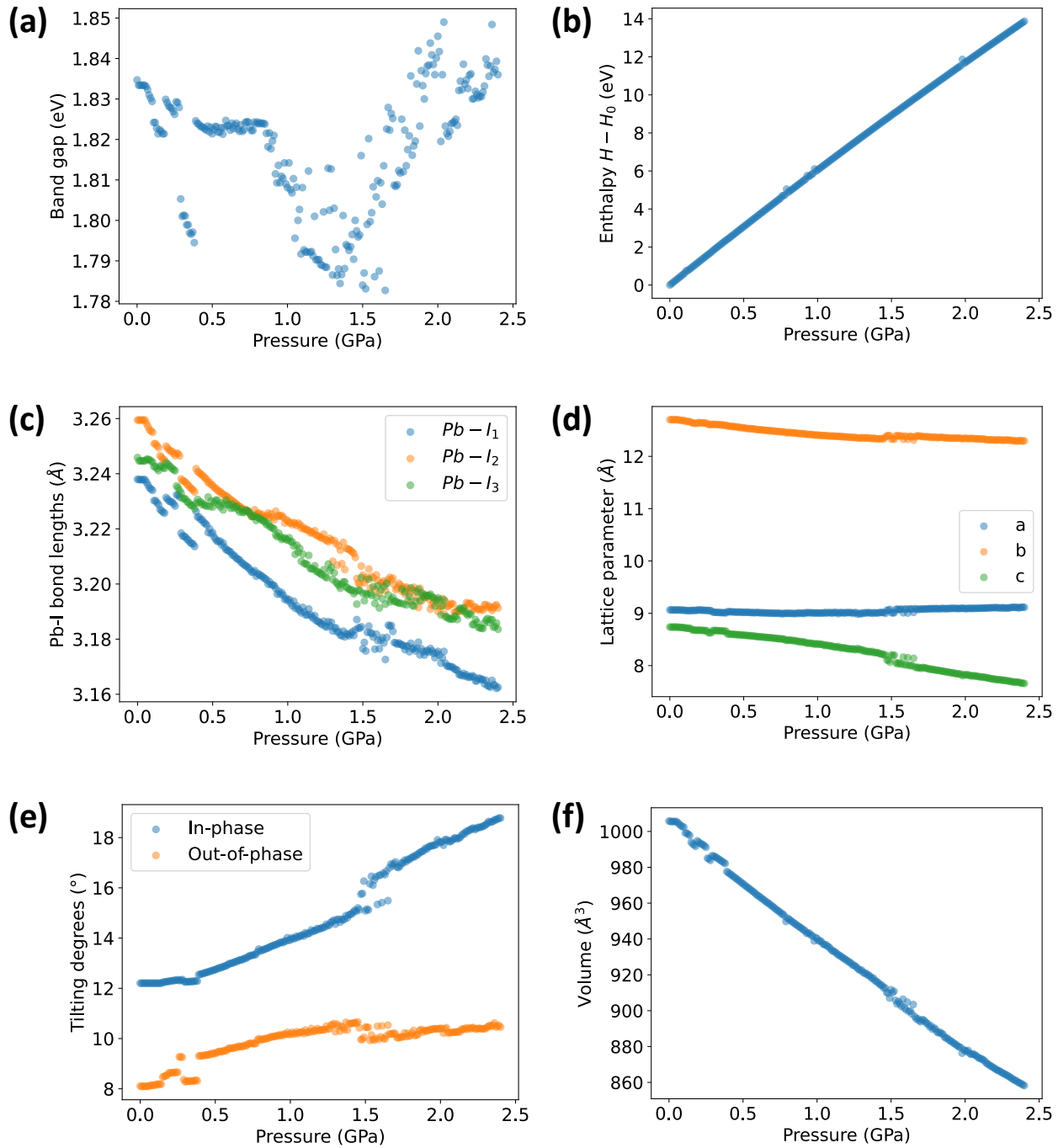
Figure S1. DFT relax-calculation results of pressure-tuned γ-CsPbI$_3$. (a) Band gap (b) enthalpy (c) Pb-I bond lengths (d) lattice parameters (e) octahedral tilting angles and (f) volume.
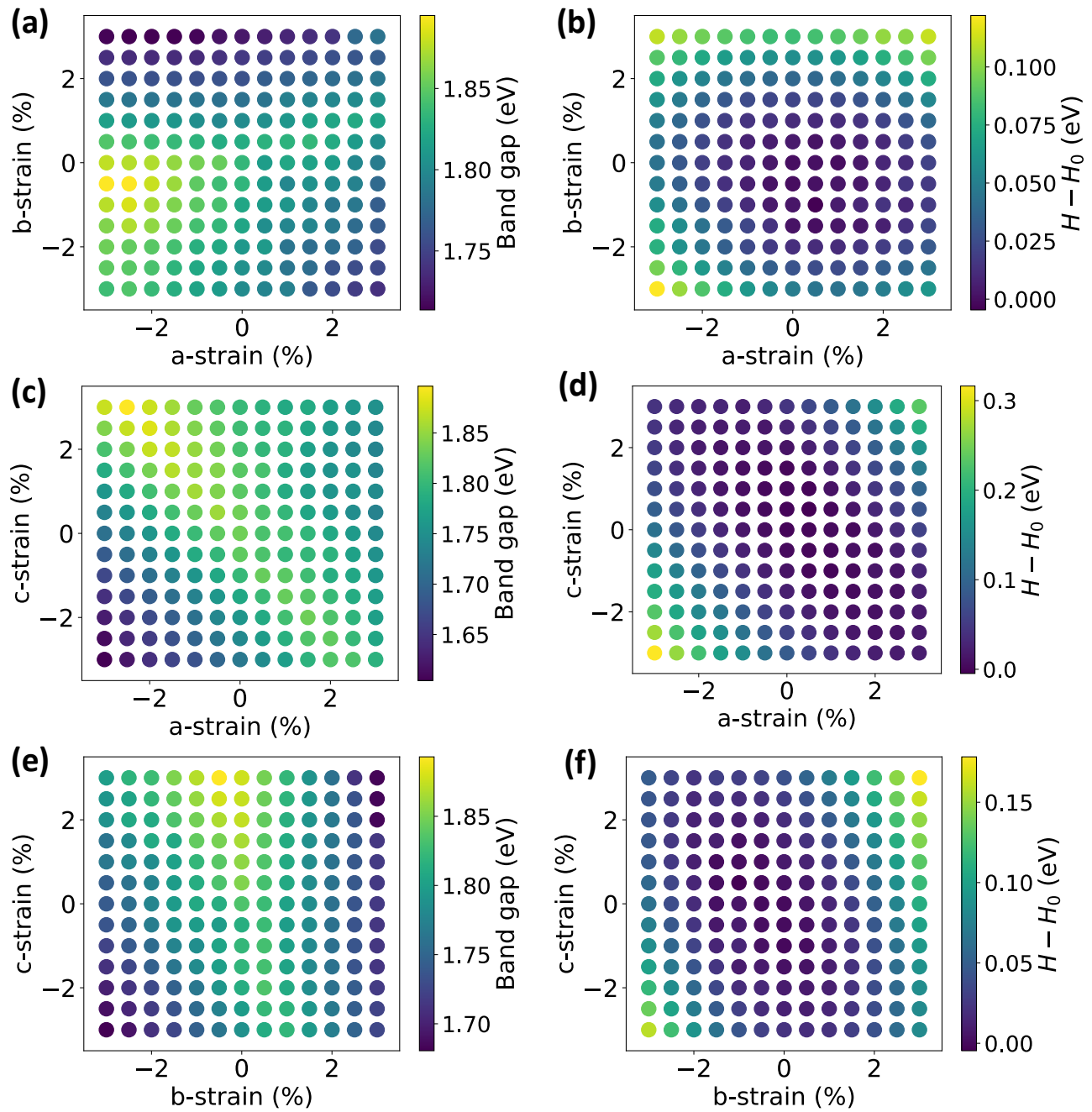
Figure S2. DFT relax-calculation results of strain-tuned γ-CsPbI$_3$. Band gap and enthalpy values from strain-tuning along (a-b) a and b directions, (c-d) a and c directions, and (e-f) b and c directions. Note that positive and negative strains represent tensile and compressive strains, respectively.
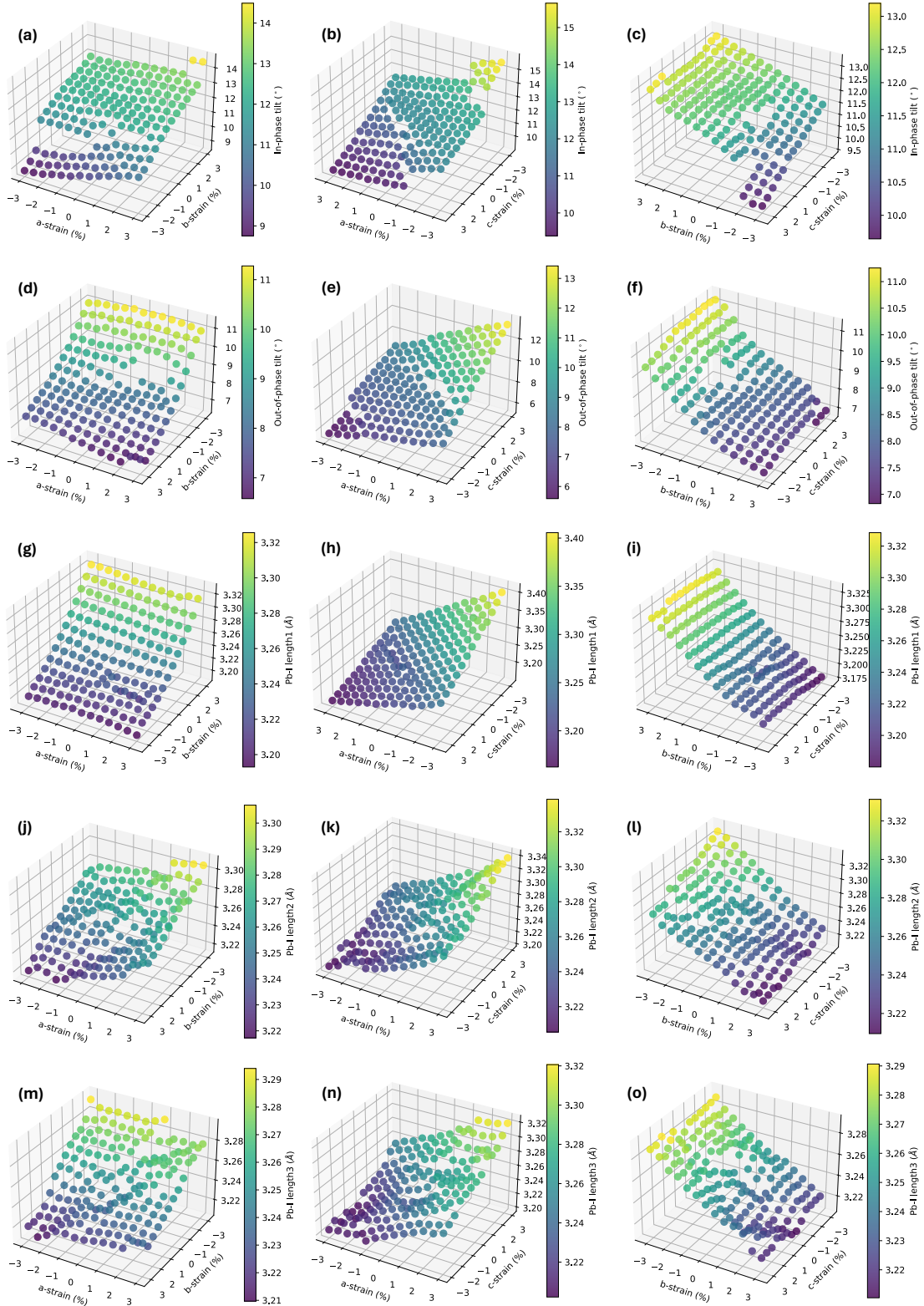
Figure S3. Discontinuous changes in structural features of the DFT-calculated, strain-tuned γ-CsPbI$_3$ system. In-phase tilts (a-c) out-of-phase tilts (d-f) Pb-I bond length$_1$ (g-i) Pb-I bond length$_2$ (j-l) and Pb-I bond length$_3$ (m-o) across the *ab*, *ac*, and *bc* directions, respectively. Axes in some plots have been inverted for better visibility.
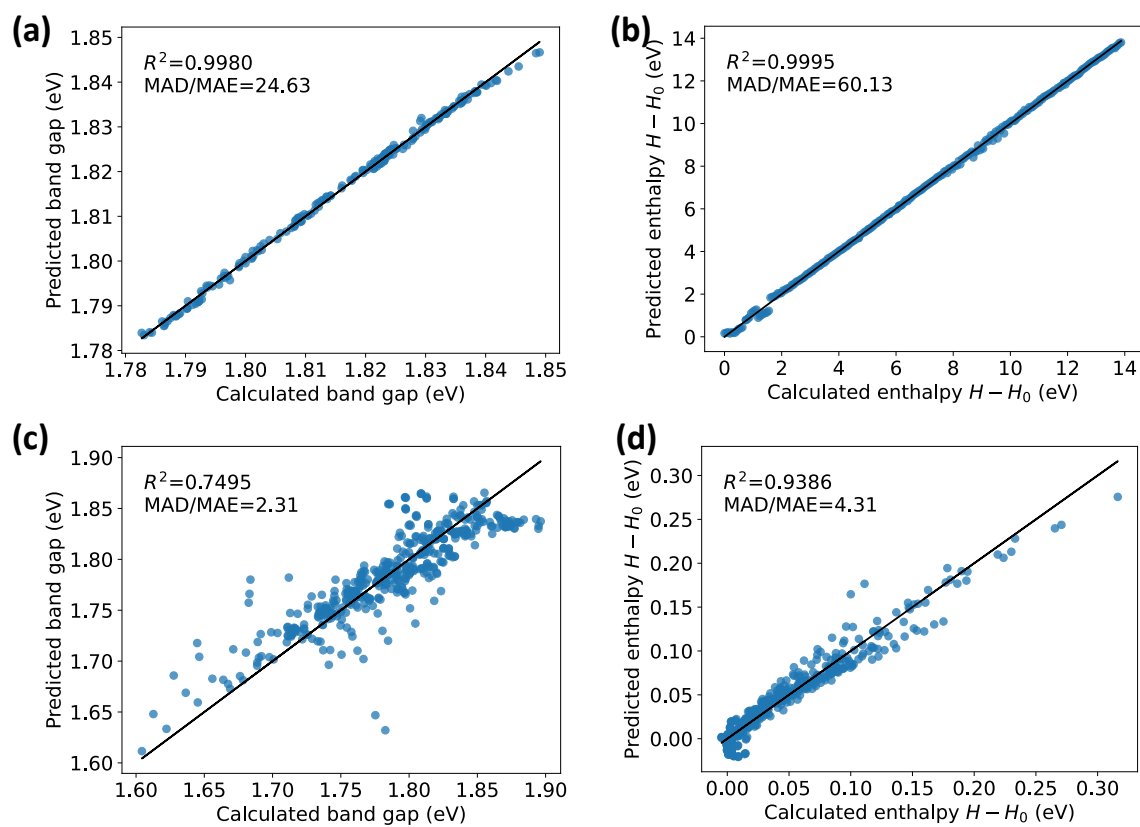
Figure S4. Multiple linear regression prediction results of γ-CsPbI$_3$. (a) Band gap and (b) enthalpy of the pressure-tuned system. (c) Band gap and (d) enthalpy from strain-tuning.
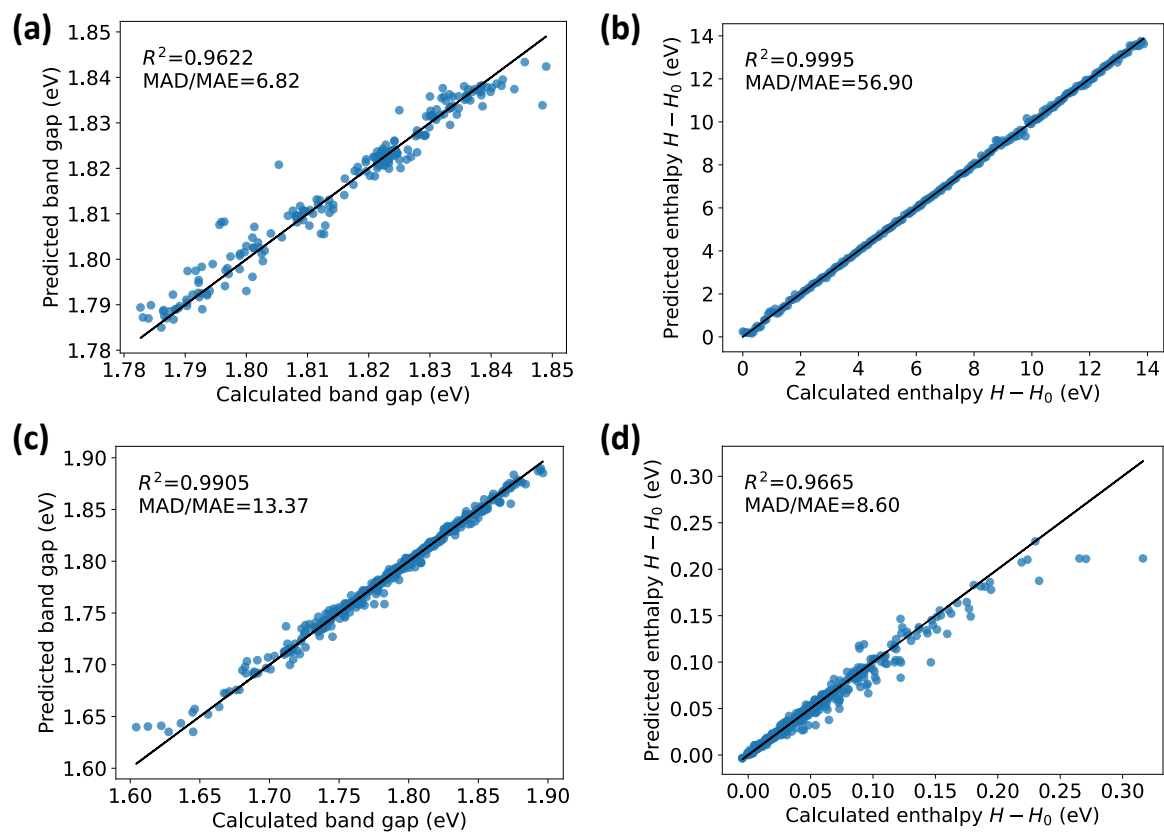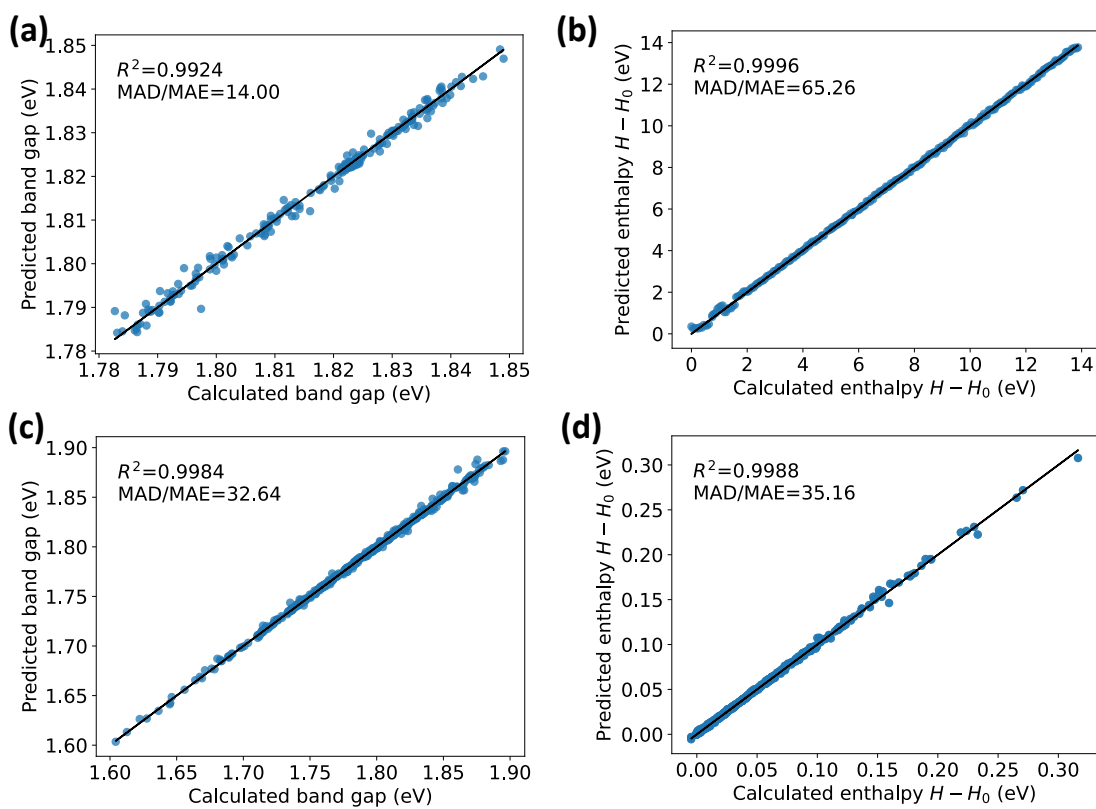
Figure S5. Random forest prediction results of γ-CsPbI$_3$. (a) Band gap and (b) enthalpy of the pressure-tuned system. (c) Band gap and (d) enthalpy from strain-tuning.
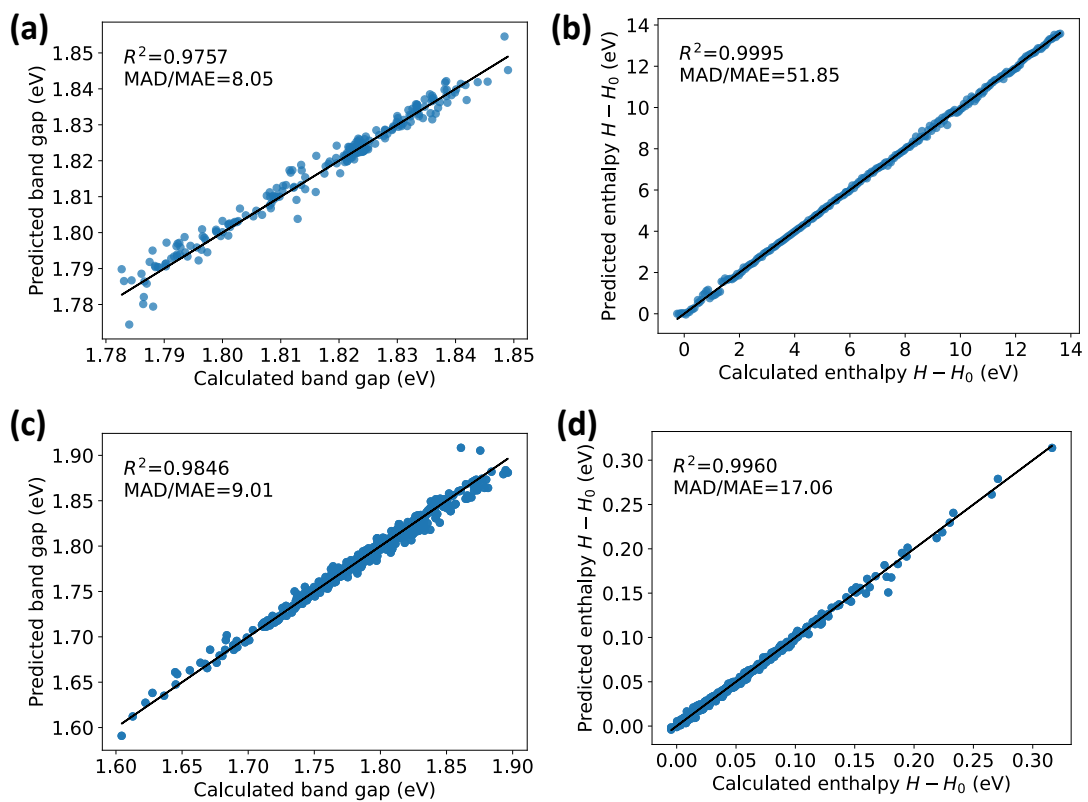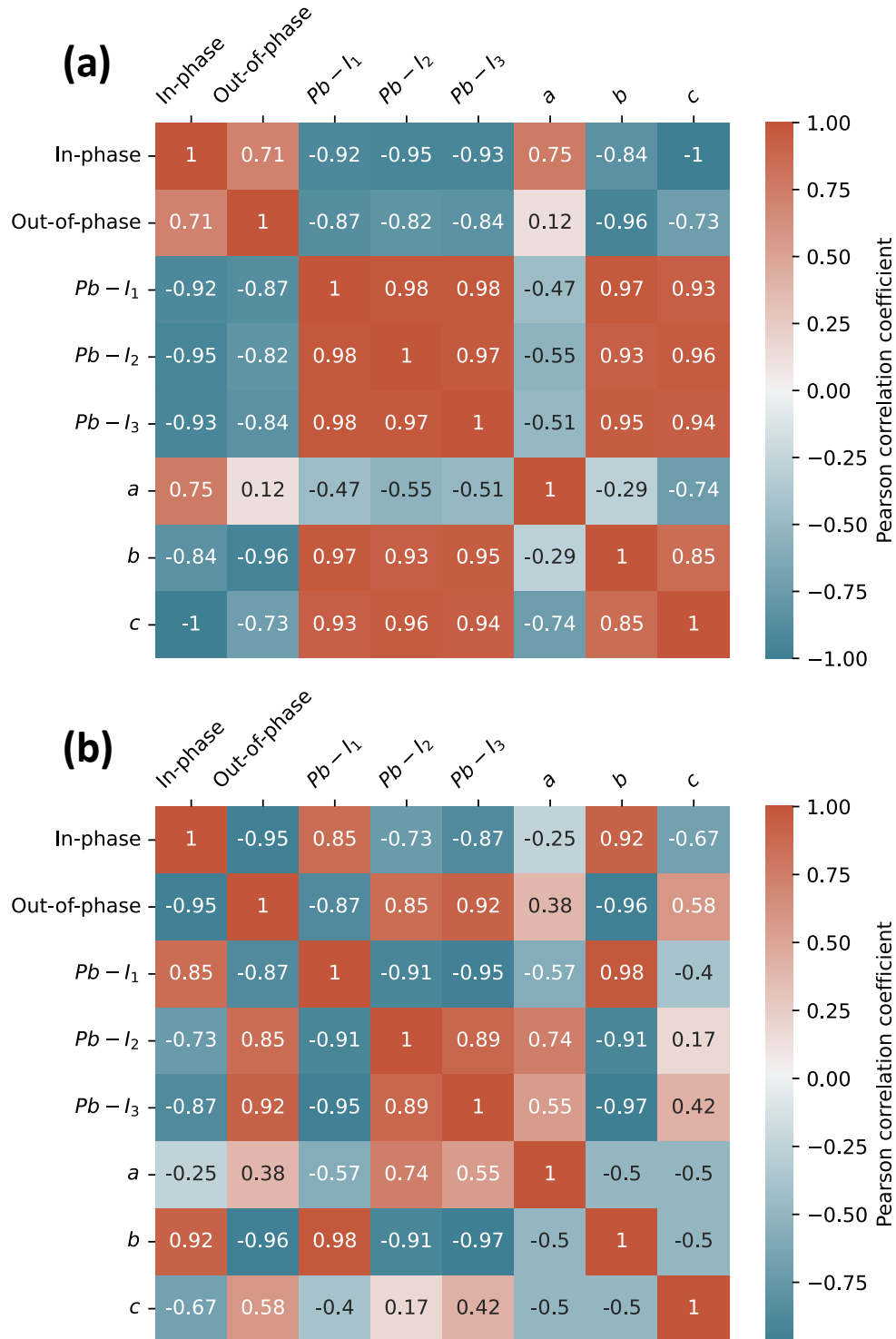
Figure S6. CGCNN prediction results of γ-CsPbI$_3$. (a) Band gap and (b) enthalpy of the pressure-tuned system. (c) Band gap and (d) enthalpy from strain-tuning.

Figure S7. ALIGNN prediction results of γ-CsPbI$_3$. (a) Band gap and (b) enthalpy of the pressure-tuned system. (c) Band gap and (d) enthalpy from strain-tuning.

Figure S8. Pearson correlation coefficients among eight structural input features of γ-CsPbI$_3$ systems. Calculations from (a) pressure-tuned and (b) strain-tuned systems are presented.
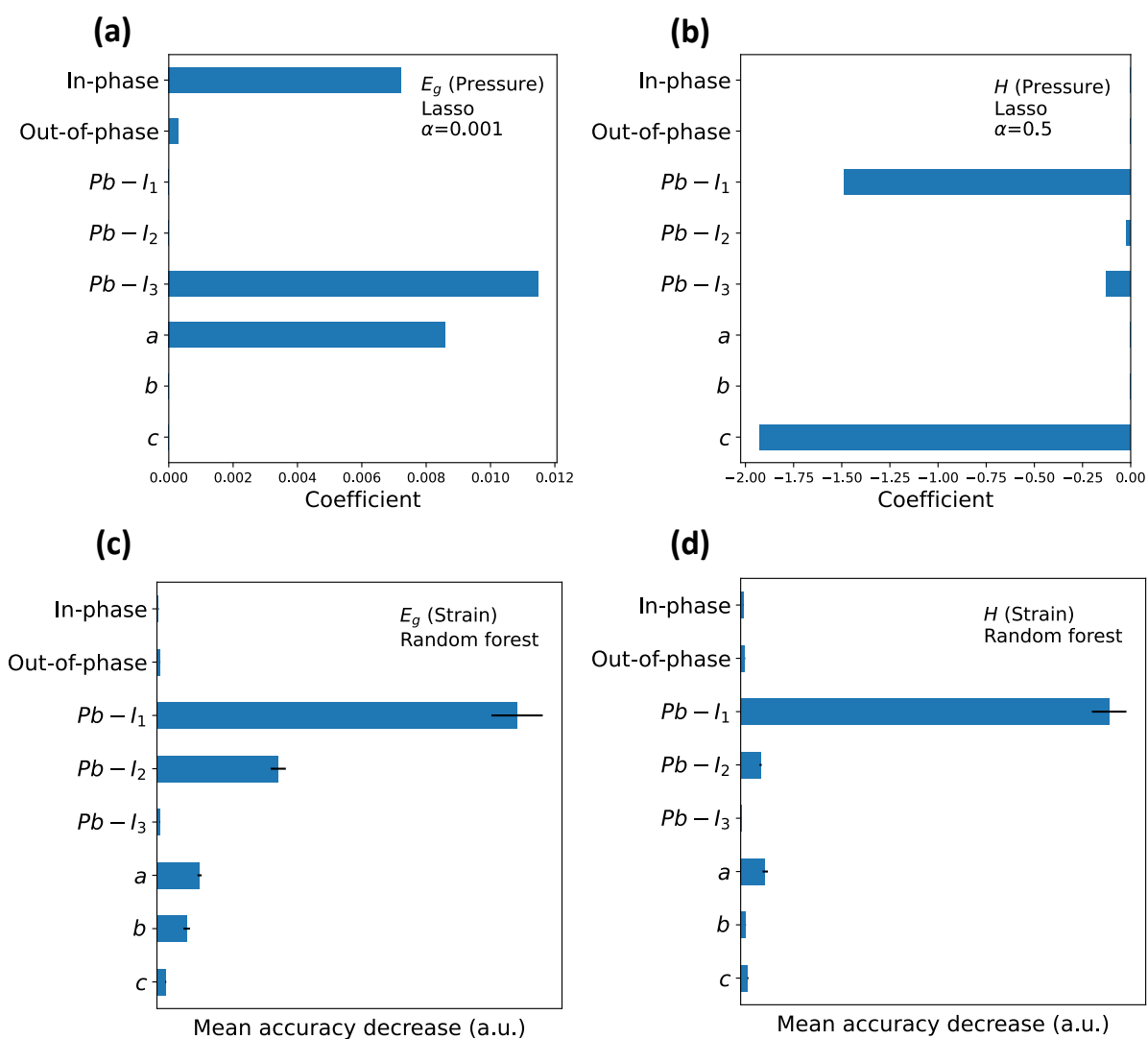
Figure S9. The feature selection results of inputs having high multicollinearity using lasso regression and permutation feature importance. The feature importance of pressure-tuned (a) band gap (E_g) and (b) enthalpy (H) is evaluated from the coefficients of lasso regression with different regularization strength α. Here, α is chosen based on the smallest value that simplifies the model by eliminating less important features, leaving only a few features. Strain-tuned (c) band gap and (d) enthalpy used random forest-based permutation feature importance. The higher mean accuracy decrease indicates the larger importance of the corresponding features. The choice of each tool, lasso and permutation feature importance technique, is driven by their respective prediction performances, with linear regression demonstrating better prediction in the pressure-tuned system, and random forest with non-linearity yielding better results in the strain-tuned system.
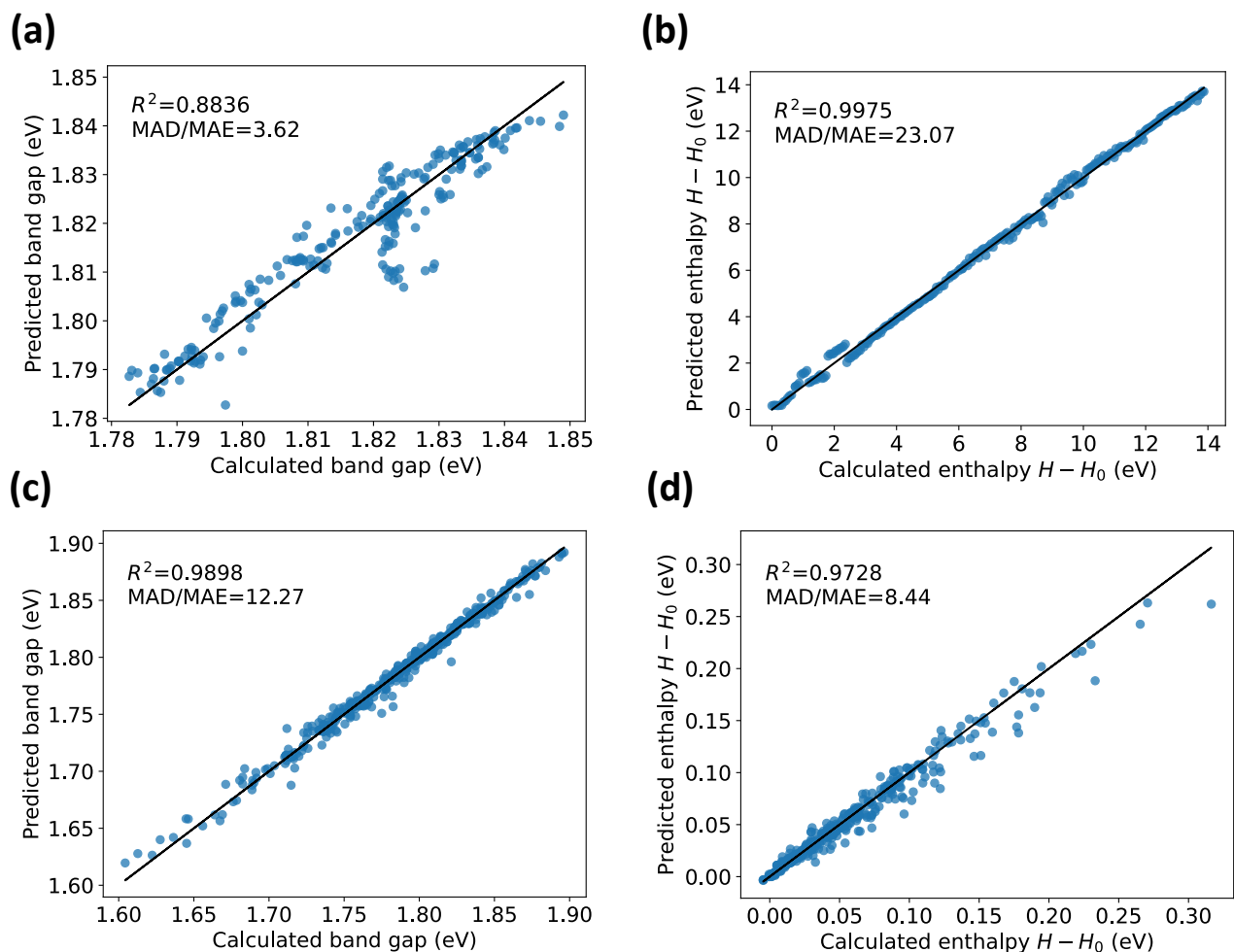
Figure S10. Prediction accuracy of new models built on further selected leading features. (a) Band gap and (b) enthalpy predictions for the pressure-tuned system using multiple linear regression models with input features selected from lasso regression. The band gap model consists of Pb-I$_3$, lattice parameters a, and in-phase tilting angle, whereas the enthalpy model is built with lattice parameters c and Pb-I$_1$. Conversely, for the strain-tuned system, (c) band gap and (d) enthalpy predictions employed random forest models with inputs selected from permutation feature importance comprising Pb-I$_1$, Pb-I$_2$, and lattice parameter a.
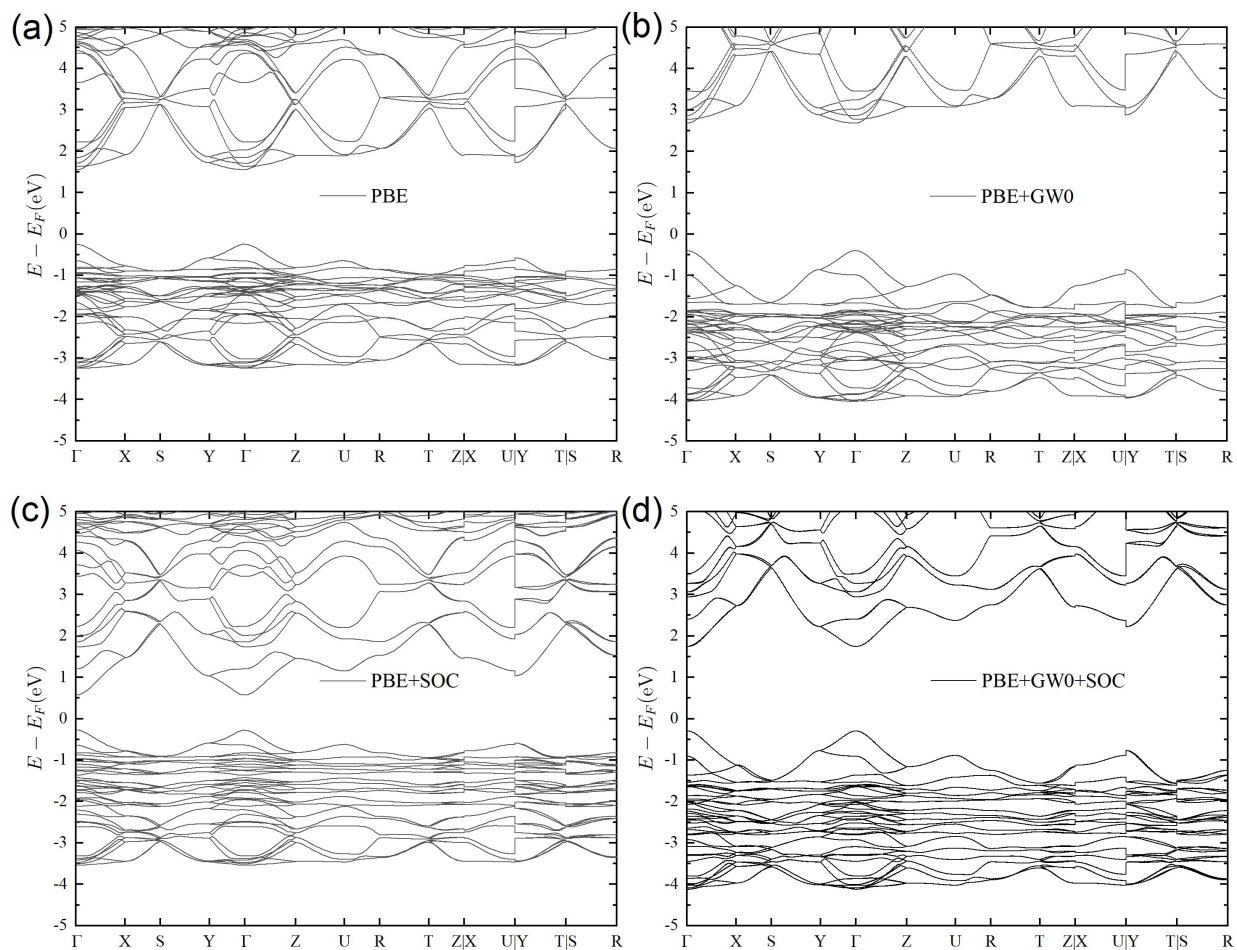
Figure S11. Band structures calculated from VASP 6.4.1 with (a) PBE pseudopotential and (b-d) PBE plus various approximations. (b) PBE+GW0, (c) PBE+SOC (spin orbit coupling), and (d) PBE+GW0+SOC. Here we use the high symmetry K path for oP lattices[15]. Here, oP lattice represents the Pearson symbol denoting the orthorhombic primitive unit cell[16].
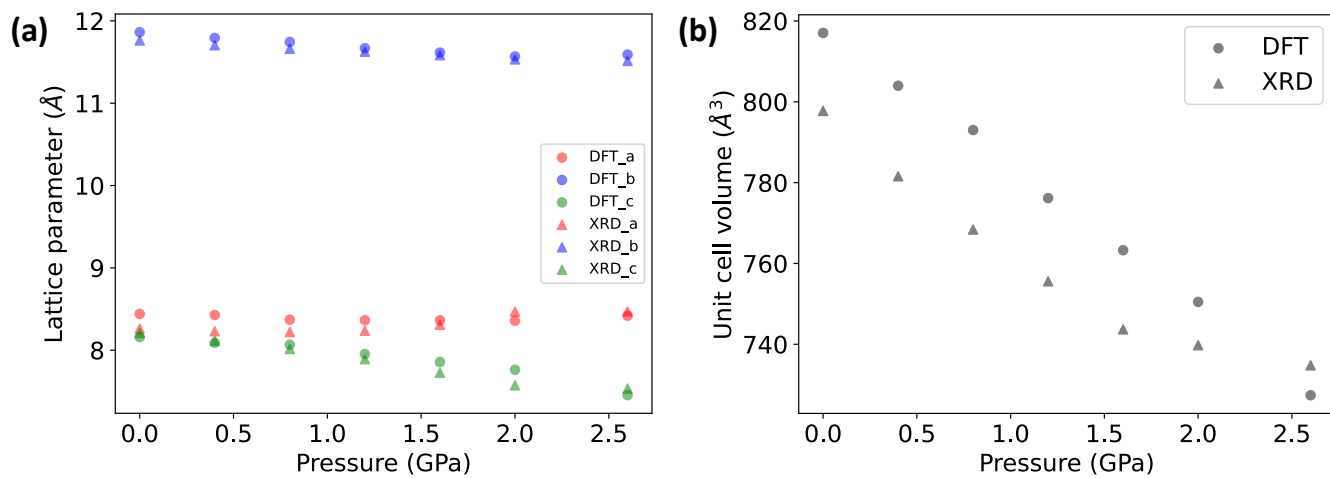
Figure S12. Comparison between DFT and XRD values from pressurized CsPbBr$_3$ structures. (a) Lattice parameters and (b) unit cell volume changes are presented.