Wu *et al. J Mater Inf* 2024;4:7
DOI: 10.20517/jmi.2024.10

Journal of Materials Informatics

Research Article

Open Access

Check for updates

Data-driven strategy for bandgap database construction of perovskites and the potential segregation study

Bobin Wu[1,2,3], Xinyu Zhang[1,2], Zixuan Wang[1,2], Zijian Chen[1,2], Shaohui Liu[1,2,3], Jie Liu[4], Zhenming Xu[5,*], Qingde Sun[6,7,*], Haitao Zhao[1,2,*]

[1]Center for Intelligent and Biomimetic Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, Guangdong, China.
[2]Digital Intelligent Manufacturing Research Center, Wenzhou Institute of Technology of the Chinese Academy of Sciences, Wenzhou, 325000, Zhejiang, China.
[3]Nano Science and Technology Institute, University of Science and Technology of China, Suzhou 215000, Jiangsu, China.
[4]Department of Chemistry, The University of Hong Kong, Hong Kong 999077, China.
[5]Jiangsu Key Laboratory of Electrochemical Energy Storage Technologies, College of Materials Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, Jiangsu, China.
[6]School of Physics and Electronic Science, Changsha University of Science and Technology, Changsha 410114, Hunan, China.
[7]School of Materials Science & Engineering, Nanyang Technological University, Singapore 639798, Singapore.

[*]**Correspondence to:** Dr. Haitao Zhao, Center for Intelligent and Biomimetic Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Avenue, Shenzhen University Town, Shenzhen 518055, Guangdong, China. E-mail: haitaozhao@outlook.com; Dr. Zhenming Xu, Jiangsu Key Laboratory of Electrochemical Energy Storage Technologies, College of Materials Science and Technology, Nanjing University of Aeronautics and Astronautics, No. 29 Yudao Street, Nanjing 210016, Jiangsu, China. E-mail: xuzhenming@nuaa.edu.cn; Dr. Qingde Sun, School of Materials Science & Engineering, Nanyang Technological University, 50 Nanyang Ave, Singapore 639798, Singapore. E-mail: qingde.sun@ntu.edu.sg

How to cite this article: Wu B, Zhang X, Wang Z, Chen Z, Liu S, Liu J, Xu Z, Sun Q, Zhao H. Data-driven strategy for bandgap database construction of perovskites and the potential segregation study. *J Mater Inf* 2024;4:7. https://dx.doi.org/10.20517/jmi.2024.10

Received: 13 Mar 2024 **First Decision:** 25 Apr 2024 **Revised:** 4 May 2024 **Accepted:** 16 May 2024 **Published:** 28 May 2024

Academic Editor: Xiang-Dong Ding **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Light-induced segregation limits the practical application of mixed halide perovskites in solar cells. Herein, halide segregation is evaluated by a data-driven approach with constructing a bandgap database of 53,361 mixed ABX_3 [where A = Cs, formamidinium (FA) or methylammonium (MA); B = Pb or Sn; X = Br, Cl, or I] perovskites. A transfer learning strategy was employed to fine-tune the parameters of a Graph Neural Network model using experimental and density functional theory (DFT)-calculated bandgaps. This approach accelerated the construction of a unique database, distinguishing it from others primarily focused on ABX_3 perovskite element

substitution. The database is characterized by continuously varying compositions and accurate bandgaps. It was utilized to calculate the free energy of 20,688 mixed iodine-bromine perovskites and generate corresponding phase diagrams for predicting their light-induced segregation behavior. It is found that the bandgap increases with decreasing ionic radii at the A-site and X-site. This composition-dependent bandgap difference drives halide segregation. Moreover, using a higher Cs content at the A-site, rather than MA, reduces this bandgap difference, enhancing photostability. The proposed data-driven strategy can facilitate the targeted design of novel perovskites with mixed compositions and the investigation of halide perovskite segregation.

Keywords: Mixed halide perovskites, bandgap database, machine learning, halide segregation

INTRODUCTION

Metal halide perovskites have rapidly gained attention due to their extraordinary optoelectronic properties in solar energy harvesting[1-5]. Up to now, experimental implementations of tandem cells based on metal halide perovskite/silicon combinations have improved quickly[6]. These cells have achieved an efficiency exceeding 33%, surpassing the Shockley-Queisser limit (SQ-limit) of single-junction solar cells[7-9]. This remarkable progress is primarily attributed to the tunable bandgaps of metal halide perovskites[10]. The ABX_3-type perovskites [where A = Cs, formamidinium (FA) or methylammonium (MA); B = Pb or Sn; X = Br, Cl, or I] allow for continuous tuning of their bandgaps through compositional engineering and adjustment of proportions on different lattice sites[5,11,12], ranging from 1.2 eV for $MAPb_{0.5}Sn_{0.5}I_3$[13], to > 3 eV for $MAPbCl_3$[14]. While adjusting the ratios of anions and cations in mixed halide perovskites achieves a tunable bandgap, making them an ideal choice for tandem solar cells[15-19], this modification of compositional ratios can lead to halide segregation, ultimately causing photo-induced instability[20-24].

Halide segregation in mixed halide perovskites is important for engineering their compositions[12,24-27]. density functional theory (DFT) calculations are widely applied to validate and explain experimental observations of halide perovskites[28-30]. Based on experimental observations, several microscopic models have been developed using DFT methods to learn light-induced halide segregation in mixed halide perovskites[31-35]. Previous studies suggested that photo-induced halide segregation was caused by factors such as charge carrier gradients[34,36,37], polaron-induced strain gradients under illumination[36,38-40], or thermodynamic origins[41-43]. These theories indicate that the bandgap differences between mixed-halide perovskites are a crucial factor contributing to halide segregation under illumination, which means that the bandgaps significantly influence the segregation behavior[33,34,44,45].

Given the significance of bandgaps in determining the properties and behavior of mixed halide perovskites, it becomes important to rapidly identify materials with ideal bandgaps. Traditional trial-and-error experimentation, although effective in obtaining true bandgap values, is time-consuming, labor-intensive, and limited by discontinuous perovskite compositions, resulting in a restricted dataset[46]. Due to the vast compositional space of mixed cation-anion perovskites, high-throughput DFT methods have been applied for bandgap prediction to overcome the challenges. While standard DFT calculations employing the semi-local generalized gradient approximation (GGA) functionals such as Perdew-Burke-Ernzerhof (PBE) can cover a wide range of perovskite compositions, they often underestimate bandgaps[47]. On the other hand, more advanced functionals such as Heyd-Scuseria-Ernzerhof (HSE) or Green's function-based GW approximation, combined with spin-orbit coupling (SOC), can precisely predict bandgaps[47,48]. However, the high computational resource consumption of these advanced DFT methods renders it impractical to perform high-throughput calculations across the vast chemical composition space of mixed perovskites[49,50]. Given this limitation, is it feasible to accurately predict bandgaps for the numerous mixed perovskite materials resulting from continuous variations in anions and cations?

In recent years, with the advancement of artificial intelligence (AI), data-driven approaches have become a valuable tool for rapidly discovering materials and their properties[51-55]. Notably, AI has significantly progressed in predicting bandgaps[56-58]. For example, Priyanga *et al.* predicted the bandgaps of perovskites using elemental composition, ionic radius, ionic character, and electronegativity as inputs for a random forest (RF) algorithm, and the prediction results were comparable to the mean absolute error (MAE) of DFT computation[59]. However, they did not consider the structural information of the perovskites. Chen *et al.* established a universal MatErials Graph Network (MEGNet) model trained on ~60,000 crystals in the Materials Project for accurate property prediction in both molecules and crystals, achieving better than DFT accuracy over a much larger dataset[60]. In the MEGNet model, each crystal was represented by a crystal graph, which not only considered constituent elements and the basic physicochemical properties of chemical components but also incorporated crystal structural features as model inputs. However, despite its comprehensive approach, the model pre-trained by the PBE bandgaps still exhibited significant errors when predicting experimental bandgap values. Chen *et al.* further developed the Atomsets framework encoded by MEGNet by employing a transfer learning strategy to significantly improve the prediction accuracy of experimental bandgaps[61]. To the best of our knowledge, no work has been carried out on utilizing data-driven approaches to construct a mixed perovskite bandgap database and then explore the light-induced segregation in halide perovskites.

In this study, we proposed a data-driven technique to construct a bandgap database of mixed perovskites and study their segregation behaviors. Considering the limited amount of experimental bandgap data with discontinuous components but high accuracy, abundant continuous components yet less precise PBE bandgap data, and precise but time-consuming and expensive HSE bandgap data, we employed a transfer learning strategy to construct machine learning models that can address these challenges. Furthermore, we established a comprehensive mixed perovskite database with continuous components, a rich variety of elements, and high precision. Database analysis shows a crucial relationship between bandgap variations and ionic radius changes at the A-site and X-site. Based on the phase segregation theory model, we further explored this relationship by generating phase diagrams. These diagrams revealed how compositional variations led to bandgap differences, which drove halide phase segregation.

MATERIALS AND METHODS

Dataset preparation

The ABX_3 perovskite experimental bandgap dataset was derived from three sources: the perovskite solar cell (PSC) database established by Jacobsson *et al.*[62], 227 experimental data collected by Liu *et al.*[63], and 610 experimental data collected by Yang *et al.*[64]. After excluding duplicate data, for perovskites with the same composition but different bandgaps, the average value was taken. Finally, a total of 645 experimental bandgap data were obtained. The experimental perovskite bandgap formulas for cubic $FAPb(I_{1-x}Br_x)_3$, $MAPb(I_{1-x}Br_x)_3$, $CsPb(I_{1-x}Br_x)_3$, tetragonal $MAPb(I_{1-x}Br_x)_3$, and orthorhombic $CsPb(I_{1-x}Br_x)_3$ are given in Supplementary Table 1.

To further extend the bandgap dataset, we calculated the bandgaps of the cubic mixed halide perovskites $FAPb(I_{1-x}Br_x)_3$, $MAPb(I_{1-x}Br_x)_3$, and $CsPb(I_{1-x}Br_x)_3$. Starting from the unit cell of the pure iodides containing three formula units, 3 × 1 × 1 supercells of a cubic perovskite were built (using the structure module within the Pymatgen toolkit[65]). I anions were randomly replaced by Br anions at different concentrations: x = 0, 1/9, 2/9, …, 1. To obtain tetragonal $MAPb(I_{1-x}Br_x)_3$ and orthorhombic $CsPb(I_{1-x}Br_x)_3$, we considered a supercell with 2 × 2 × 1 expansion of a pseudo cubic perovskite building block. Br anions randomly replaced I anions at varying concentrations: x = 0, 1/12, 2/12, …, 1. Furthermore, the perovskite crystal structures containing four formula units were constructed with 2 × 2 × 1 cubic supercells to calculate the bandgaps of

halide perovskites such as $FA_{0.75}Cs_{0.25}PbI_3$, $FA_{0.5}MA_{0.5}PbBr_3$, $FA_{0.5}MA_{0.5}Pb(I_{0.33}Br_{0.67})_3$, $FA_{0.5}MA_{0.5}Pb(I_{0.5}Br_{0.5})_3$, $FA_{0.5}MA_{0.5}Pb(I_{0.67}Br_{0.33})_3$ and $FA_{0.5}MA_{0.5}Pb(I_{0.83}Br_{0.17})_3$. In total, 108 perovskite crystal structures were considered for the bandgap calculations.

All the calculations were carried out using the Vienna Ab initio Simulation Package (VASP) with the PBE functional for DFT. For structure relaxation calculations, the plane wave energy cutoff was set at 520 eV. All atomic structures were fully relaxed until forces on all atoms were less than 0.03 (eV/Å). The Brillouin zone was sampled with a gamma-centered grid, the K-spacing value was set to 0.03 (2π/Å), and then set to 0.015 (2π/Å) for band structure calculations. The DFT calculated electronic band structures for six halide perovskites are shown in Supplementary Figure 1.

Machine learning model construction

Machine learning algorithms, including MEGNet[60], Materials Graph Library (MatGL)[66], Atomsets[61], Support Vector Regression (SVR), RF, and multi-layer perceptron (MLP), were applied to predict the bandgaps of perovskites. Additionally, cross-validation was implemented to evaluate the accuracy and generalization ability of the trained models for all the machine learning algorithms, with the DFT and experimental dataset being randomly partitioned into a 90% training set and a 10% test set.

The DFT bandgap dataset was combined with the corresponding experimental data to create a new dataset. Then, the pre-trained MEGNet model was fine-tuned on the new dataset to predict the DFT bandgap values, the experimental bandgap values, and the difference between these two. After converting the perovskite crystal structure into a crystal graph using MEGNet, it was fitted with the corresponding DFT bandgap, experimental bandgap, and the difference between them. The experiments were run for 100 epochs, and the R^2 and MAE were recorded. The main training parameters were bond attributes (nfeat_bond) and cutoff radius (r_cutoff).

In MEGNet, the composition features are stored by element embedding vectors[67]. The element embedding vectors themselves do not directly encode specific chemical information, but they are designed to provide useful feature inputs for MEGNet[60]. These embedding vectors capture certain chemical or physical relationships between elements, learned from a vast amount of chemical data through machine learning algorithms[60]. Since MEGNet is a black box model, we cannot determine which composition features are important for bandgap prediction. Nevertheless, the literature reports that the orbital energy of atoms is crucial for bandgap prediction[68]. Furthermore, bond length and coordination number in structure features are important for bandgap prediction because they influence the interaction strength between atoms and the overlap of charge density[56].

For the three datasets, the bond attributes and cutoff radius with the best correlation coefficient (R^2) and MAE performance in the training sets were selected. As shown in Supplementary Figure 2, the performance of predicting the difference between the actual and DFT bandgaps was better. Therefore, in this work, the DFT bandgap and the gap between the DFT and the experimental bandgap of different perovskites were first predicted through MEGNet, and then the experimental bandgap of a perovskite was obtained by adding the two together.

The 645 experimental bandgap data were encoded using the Atomsets framework to represent the components of the mixed perovskites. After converting the components into element embedding vectors using Atomsets, three machine learning algorithms, MLP, RF, and SVR, were used to fit these vectors with their corresponding experimental bandgap data to construct a machine learning prediction model. Once the

optimal parameters were selected, the experimental bandgap values were predicted on the test set.

Enthalpy calculation

To calculate the total energy of each structure of $FAPb(I_{1-x}Br_x)_3$, specifically, the graph neural networks with three-body interactions (M3GNet)[69] and Crystal Hamiltonian graph neural network (CHGNet)[70] were employed. In detail, these models predicted the total energy of each perovskite in the DFT bandgap dataset. The prediction results are shown in Supplementary Figure 3. M3GNet was chosen for energy prediction due to its lower MAE between the predicted energy values and those obtained from DFT, indicating better predictive performance.

The four types of energies for $FAPb(I_{1-x}Br_x)_3$ were calculated using the multicomponent phase diagram module within the Pymatgen toolkit. Decomposition energy represents the energy required to decompose $FAPb(I_{1-x}Br_x)_3$ into FAI, FABr, PbI_2, and $PbBr_2$. Mixing energy is the energy needed to decompose $FAPb(I_{1-x}Br_x)_3$ into $FAPbI_3$ and $FAPbBr_3$. Formation energy was calculated using the energy values of the most stable phases for each element in the Materials Project database[71]. Energy above the hull (E_hull) was determined by comparing the energies of competing phases in the Materials Project database.

Helmholtz free energy calculation

The total number of structures for each system of mixed halide perovskites is 2^{12} = 4,096 (12 halide atoms in 2 × 2 × 1 supercells) for $APb(I_{1-x}Br_x)_3$ in cubic structures (A = $FA_{0.75}MA_{0.25}$, $FA_{0.5}MA_{0.5}$, $FA_{0.25}MA_{0.75}$, $FA_{0.75}Cs_{0.25}$, $FA_{0.5}Cs_{0.5}$, $FA_{0.25}Cs_{0.75}$, $MA_{0.75}Cs_{0.25}$, $MA_{0.5}Cs_{0.5}$, $MA_{0.25}Cs_{0.75}$), $MAPb(I_{1-x}Br_x)_3$ (in tetragonal structures) and $CsPb(I_{1-x}Br_x)_3$ (in orthorhombic structures), and 2^9 = 512 (9 halide atoms in 3 × 1 × 1 supercells) for $APb(I_{1-x}Br_x)_3$ in cubic structures (A = FA, MA, Cs). In these structures, I anions were randomly substituted with Br anions at varied concentrations. Considering symmetry, there are 20,688 independent structures. The total energy E of each independent structure was predicted by M3GNet.

The thermodynamic properties of the alloy were determined using the generalized quasi-chemical approximation (GQCA)[72]. This method has been successfully employed in the thermodynamic analysis of semiconductor alloys[73,74]. By considering the total energy and the degeneracy of each configuration, the method provides simple expressions for the mixing contribution to the alloy internal energy ΔU and the configurational entropy ΔS, as functions of the composition x and temperature T according to a Boltzmann distribution. The Helmholtz free energy in dark conditions was calculated by:

$$\Delta F_{dark}(x,T) = \Delta U(x,T) - T\Delta S(x,T) \quad (1)$$

In this work, we used the hypothesis put forward by Zhu *et al.* that the energy resulting from illumination could be addressed within the thermodynamic framework by considering the above-bandgap photon energy deposited in the lattice via thermalization through phonon relaxation[75]. Then, for each structure J, this light contribution term can be defined as follows:

$$E_{light}(J) = \hbar\nu_{exc} - E_g(J) \quad (2)$$

where $\hbar$ is Planck constant, the ν_{exc} is the frequency of the absorbed photon, and $E_g(J)$ is the bandgap of each structure J predicted by Atomsets-MLP.

By additionally considering the mixing contribution of this energy term, $\Delta E_{light}(J)$ is given by:

$$\Delta E_{light}(J) = E_{light}(J) - \left[(1-x)E_{light}(A) + xE_{light}(B)\right] = \Delta E_g(J) \quad (3)$$

which only negatively depends on the mixing contribution of the bandgap energy, $\Delta E_g(J)$, defined as the bandgap difference between each structure, J, and the fractions of the parent compound A (pure I, $x = 0$) and B (pure Br, $x = 1$).

By further utilizing the probability $\delta(J, T)$ given by the GQCA, we obtained $\Delta E_{light}(J)$, which was coupled with the composition effect and temperature effect using the Boltzmann distribution. The combination of $\Delta E_{light}(J)$ and the $\Delta F_{dark}(x, T)$ led to the excitation-intensity-dependent expression. Light intensity in this model was described through the photon deposited energy density, η: the ratio of absorbed photon energy units (E_{light}) to unit cells under illumination. Thus, the excitation-intensity dependent system free energy is given in:

$$\Delta F_{light}(\eta, x, T) = \Delta F_{dark}(x, T) + \eta \Delta E_{light}(x, T) \quad (4)$$

The phase diagram of the alloy can be built by calculating the free energy at different temperatures (T) and light intensity (η). For each temperature and light intensity, the binodal points are determined by collecting the compositions for which ΔF has a common tangent. The spinodal points are those in which the second derivative of ΔF vanishes.

RESULTS AND DISCUSSION

Framework of data-driven strategy for perovskite bandgap study

Herein, we developed a data-driven framework for predicting perovskite bandgaps to study halide segregation of perovskite. The general outline of this framework is shown in Figure 1. In the first step, we used a transfer learning approach to construct four machine learning algorithms based on the pre-trained MEGNet and MatGL models. These models were trained on datasets from the Materials Project to predict perovskite bandgaps. The perovskite bandgap prediction model, Atomsets-MLP, was selected after evaluation [Figure 1A]. Secondly, as can be seen from Figure 1B, we established a database containing 53,361 mixed perovskite compounds with continuous compositions. Their bandgaps were predicted using Atomsets-MLP and then compared with experimental values collected from the literature. Subsequently, we generated 20,688 mixed halide perovskites to study the halide segregation. Finally, as illustrated in Figure 1C, using the bandgaps predicted by Atomsets-MLP, we obtained the light contribution of mixed halide perovskites. By further utilizing the free energy variation under illumination conditions, the corresponding phase diagrams were generated to predict halide perovskite segregation.

Evaluation of machine learning methods for bandgap prediction

The predictive performance of pre-trained MEGNet and MatGL in the DFT bandgap dataset is shown in Supplementary Figure 4A and B. It can be clearly seen that these models show poor predictive performance. This is because MEGNet and MatGL were trained using DFT (PBE) calculated bandgaps of 69,640 crystals in the Materials Project and were used directly without changing any model parameters to predict the experimental bandgaps of perovskites. As can be seen from Supplementary Figure 4C, most data points are located below the y = x line, which indicates that the DFT (PBE) method tends to underestimate the bandgap of perovskite materials. The results are consistent with reports from other literature[76]. In conclusion, a pre-trained model learned from one large dataset is hard to use to directly predict a different small dataset.

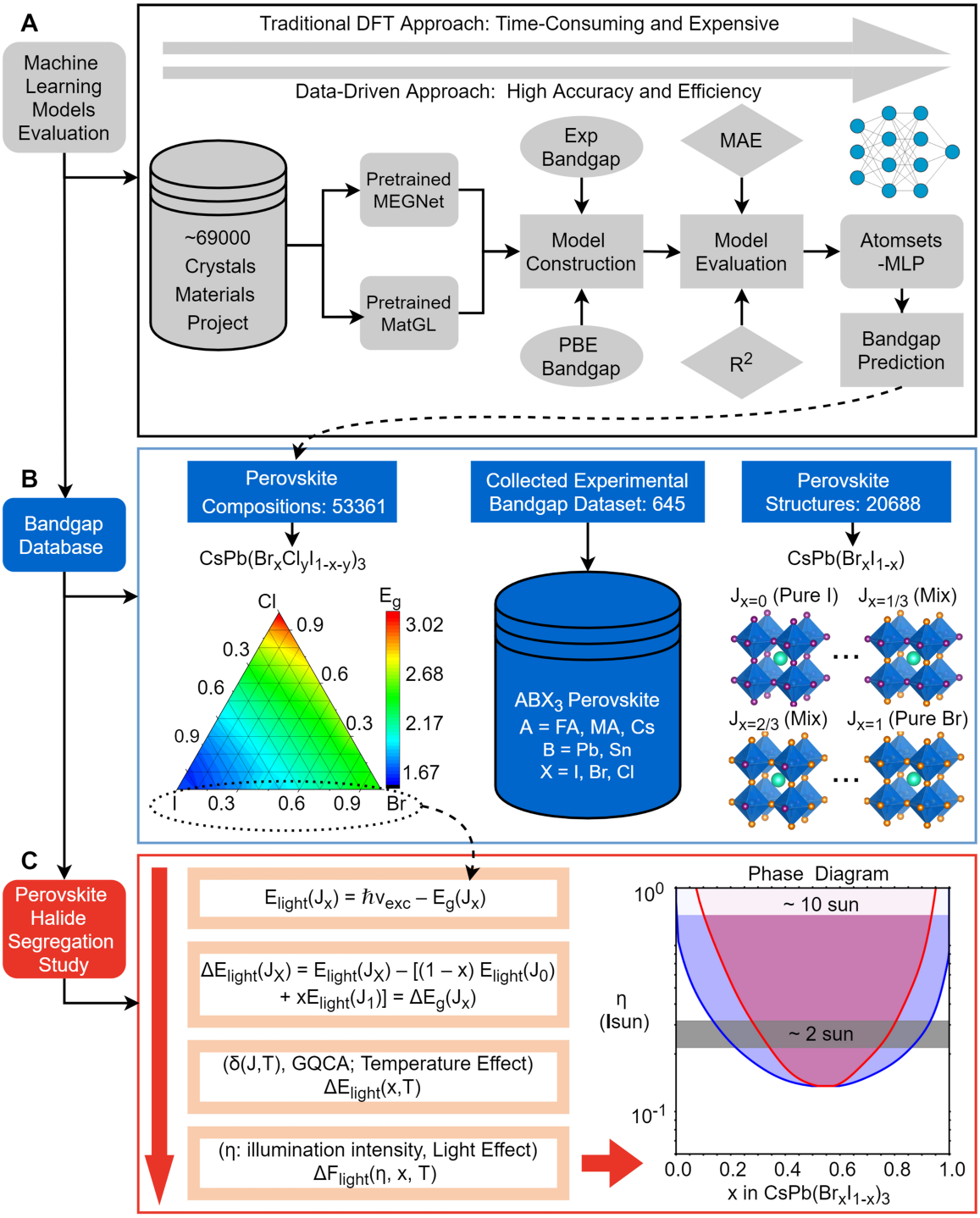

Figure 1. The framework of data-driven strategy for bandgap database construction of perovskites and the potential segregation study. (A) The comparison of traditional and data-driven approaches and evaluation of machine learning models; (B) The construction of perovskite bandgap database; (C) The application of database for perovskite halide segregation study.

To address this challenge, we further developed the Fine-Tuned MEGNet by fine-tuning the model parameters. As shown in Figure 2, the Fine-Tuned MEGNet exhibits low MAE values, indicating that the application of transfer learning has greatly improved prediction accuracy. Notably, the Atomsets

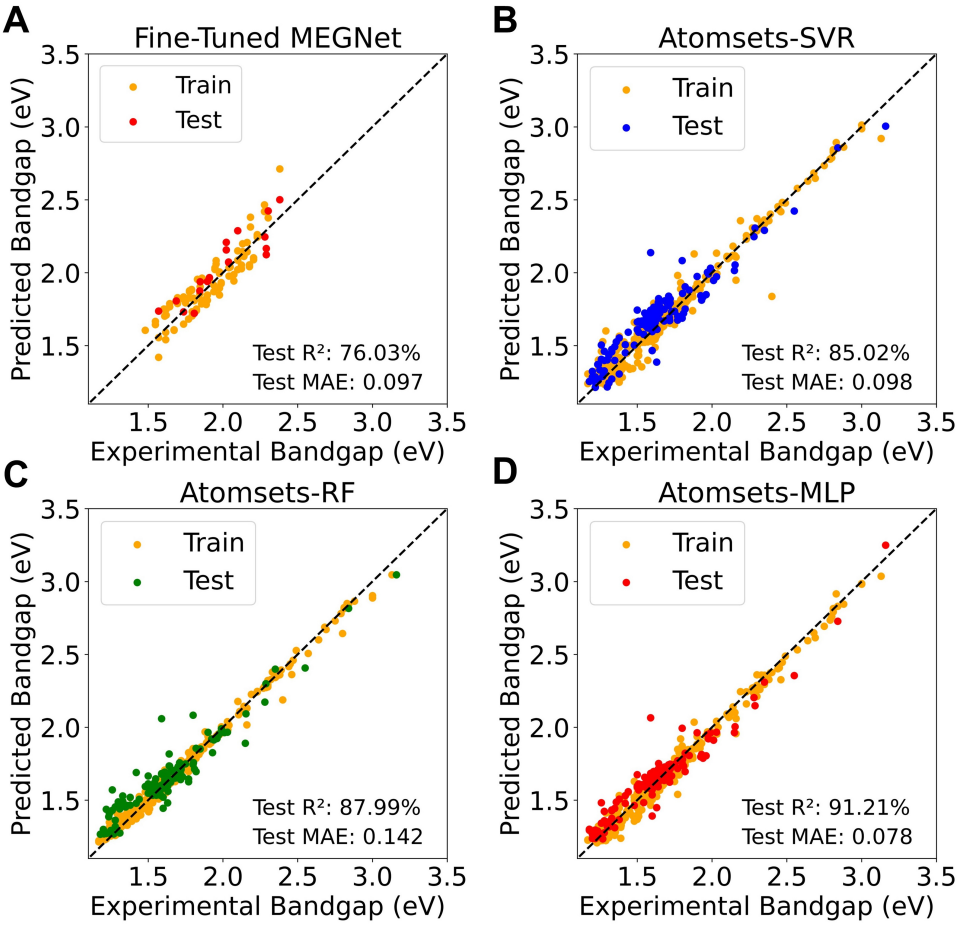

Figure 2. The bandgap prediction results for the training set and test set of (A) Fine-Tuned MEGNet; (B) Atomsets-SVR; (C) Atomsets-RF; (D) Atomsets-MLP. MEGNet: MatErials Graph Network; SVR: Support Vector Regression; MLP: multi-layer perceptron.

framework, which combines supervised machine learning algorithms such as Support Vector Machine (SVM), RF, and MLP, outperforms the Fine-Tuned MEGNet. This enhanced performance results from the fact that the Atomsets framework is a hierarchical transfer learning framework derived from pre-trained MEGNet, which has better transferability and shows lower errors than the MEGNet at a different type of dataset. In summary, the transfer learning method effectively overcomes reduced accuracy encountered when Machine Learning (ML) models trained on one type of large dataset are applied to another distinct small dataset.

Although the MAE of the Atomsets-MLP model has reached 0.078, indicating relatively high prediction accuracy, its R^2 is only 91.21%, implying that there is still potential for enhancing the model's predictive performance. This limitation can be partially attributed to various sources for the training data. In this study, the experimental data on the bandgap of perovskite primarily originates from diverse literature. Multiple bandgap measurement techniques and preparation methods of perovskite thin films can influence these data[77]. To further enhance the model's predictive accuracy, adopting a high-throughput experimental approach in the future could be beneficial. This would facilitate the acquisition of a substantial amount of precise and consistent bandgap data from a sole, dependable source, and these data can then train machine learning models to elevate their predictive performance.

Machine learning accelerated the construction of perovskite bandgap database

To overcome the discontinuity of perovskite compositions, a database was constructed containing 53,361 perovskite compounds with the component $FA_{1-a-b}MA_aCs_bPb(I_{1-x-y}Br_xCl_y)_3$ (a, b, x, y = 0.05, 0.1, 0.15, 0.2, …, 1). The bandgap of each compound was predicted using the Atomsets-MLP model. The probability density distribution is shown in Supplementary Figure 5. Notably, a large portion of the predicted bandgaps is distributed in the 1.75 to 2.25 eV range, which is precisely the bandgap range prone to halide segregation[78,79]. It is worthwhile mentioning that it took only 320 s to predict 53,361 compounds on a graphics processing unit (GPU). This means that Atomsets-MLP required only 0.006 s to predict the bandgap for a single compound. On the other hand, DFT methods would require three hours to compute the bandgap of a perovskite. Thus, Atomsets-MLP greatly accelerated the construction of the experimental bandgap database of perovskites.

To explore the correlations between the compositions and bandgaps of perovskites, the heat maps of the predicted bandgaps of $FA_{1-y}Cs_yPb(I_{1-x}Br_x)_3$, $MA_{1-y}Cs_yPb(I_{1-x}Br_x)_3$, and $FA_{1-y}MA_yPb(I_{1-x}Br_x)_3$ are shown in Figure 3A. It is evident that as the cation radius at the A site decreases, the bandgap rises correspondingly, indicating an inverse relationship between them. Likewise, decreased ion radii at the X site increase bandgaps. Moreover, changes in the ionic radius at the X site exert a more pronounced influence on the bandgap than the A site. In summary, the bandgap widens as the ion radius diminishes at both the A and X sites. This is consistent with the findings of Liu *et al.*, who combined machine learning models with SHapley Additive exPlanations (SHAP) to calculate the impact of each chemical composition in ABX_3-type perovskites on the bandgaps[63]. In addition, the influence of such compositional changes on bandgaps also agrees with the research conclusion of Yang *et al.* who combined feature engineering, the gradient-boosted regression tree (GBRT) algorithm, and the genetic algorithm-based symbolic regression (GASR) algorithm to identify the influencing factors of bandgaps[64].

Generally, the A-site cation does not directly contribute to the states governing the bandgap of ABX_3 perovskite, whereas the s-p antibonding of B and X contributes to the valence band maximum (VBM), and the p-p antibonding contributes to the conduction band minimum (CBM)[63,80,81]. Therefore, a decrease in the A cation radius leads to lattice compression, resulting in a simultaneous increase in both VBM and CBM[82]. However, the CBM increases more significantly, leading to a slight increase in the bandgap as the A cation radius decreases[83]. On the other hand, at the X site, in addition to the bandgap narrowing caused by lattice compression, orbital energy also decreases following the trend of Br 4p < I 5p[84]. This results in decreased VBM and increased CBM when transiting from I to Br, ultimately leading to a wider bandgap[80,82]. The combined effects of lattice compression and orbital contributions lead to a more rapid increase in the bandgap as the X anion radius decreases[63,85]. To summarize, the theory of physics explains why A-site cation and X-halogen ions influence the change of the perovskite bandgap.

As shown in Figure 3B, the deviations between the experimental and the predicted results are extremely small, revealing that Atomsets-MLP model can accurately predict the controlling effect of the A-site cation and X-site halide ions on the bandgap. Considering that difference in bandgaps of perovskites with mixed compositions is a crucial factor causing halide segregation, these systems are selected to generate 20,688 mixed halide perovskites for halide segregation study [Figure 3C].

Application of the database in perovskite segregation study

To predict halide segregation in $FAPb(I_{1-x}Br_x)_3$, various energies were calculated, including decomposition energy, mixing energy, formation energy, and E_hull. The results are shown in Supplementary Figure 6. Calculations of decomposition energy and mixing energy suggested that a bromide ratio of 0.66 would yield the most stable perovskite in $FAPb(I_{1-x}Br_x)_3$. However, experiments have revealed decreased stability due to

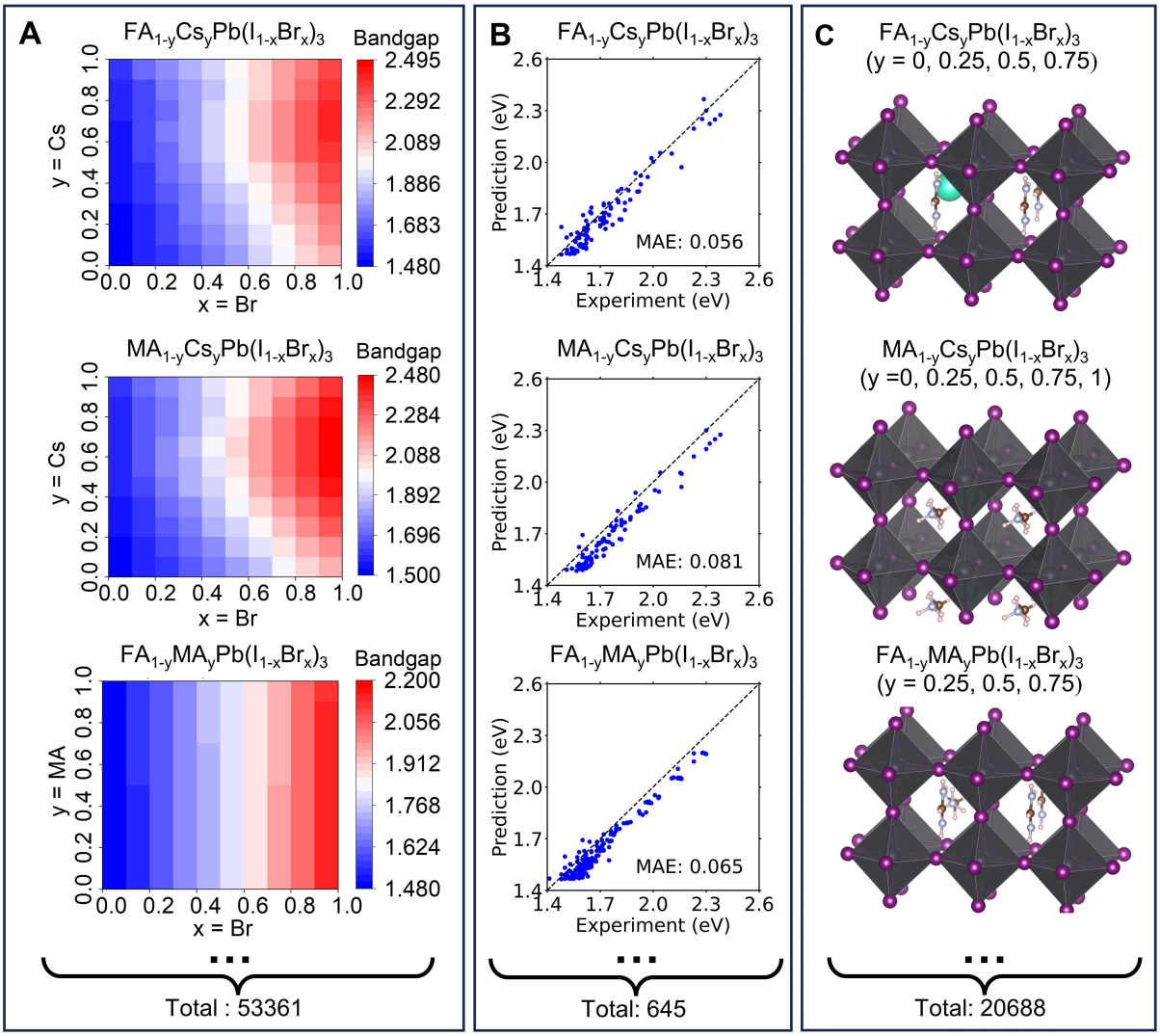

Figure 3. The construction of perovskite bandgap database. (A) Heat maps of predicted bandgap of $FA_{1-y}Cs_yPb(I_{1-x}Br_x)_3$, $MA_{1-y}Cs_yPb(I_{1-x}Br_x)_3$, and $FA_{1-y}MA_yPb(I_{1-x}Br_x)_3$; (B) The comparison between predicted bandgaps and experimental bandgaps; (C) The mixed halide perovskites in database.

halide segregation at this ratio, and after exposure to light, the segregated domains remix in dark conditions[44,86-88]. Likewise, neither the E_hull nor the formation energy accurately predicted these segregation and reversal behaviors under illumination and in the dark. This mismatch results from limitations in the calculation methods used to determine these energies, as they exclude critical factors such as temperature and illumination conditions, which significantly influence perovskite stability. Thus, these approaches are insufficient for predicting complex behaviors without considering these external factors.

In order to accurately predict the experimental behavior, a thermodynamic model that Zhu *et al.* provided was applied in $FAPb(I_{1-x}Br_x)_3$[75]. As shown in Figure 4, the curve shape undergoes significant changes with excitation intensity, exhibiting a more asymmetric profile at higher intensities. In dark conditions or at lower excitation intensities ($\eta = 1/9$), the convex profile of $\Delta F(\eta, x, T)$ at 300 K means that a miscibility gap is not apparent (no points have the same tangent line), suggesting that under these conditions, the perovskite will not phase segregate [Figure 4A]. However, at higher excitation intensities, $\Delta F(\eta, x, T)$ begins to show a

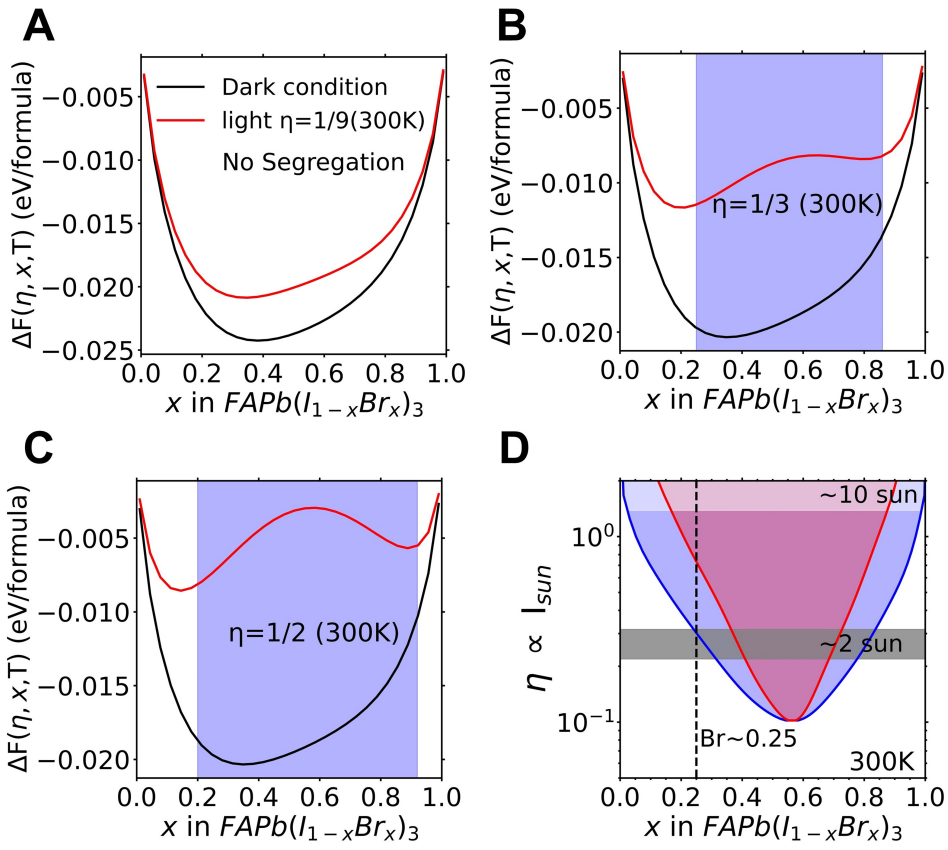

Figure 4. The free energy calculations for perovskite phase segregation. (A-C) The excitation intensity-dependent free energy variation at 300 K with different light intensities (η = 1/9, 1/3, 1/2); (D) The predicted phase diagram along with different light intensities at 300 K. The binodal and spinodal lines and regions are shown in blue and red, respectively. The light intensities are illustrated by greyscale shaded areas (~2-sun, 10-sun).

concave shape for higher excitation intensities (η = 1/3, 1/2), indicating the perovskites will segregate as illumination increases [Figure 4B and C].

Based on the Helmholtz free energy variation, we built the phase diagram of $FAPb(I_{1-x}Br_x)_3$ [Figure 4D]. The red lines are the spinodals separating the metastable (blue) and unstable (red) regions. The blue lines are the binodals separating the stable (white) and metastable regions. The phase diagram, which reflects the asymmetry of the free energy, reveals the emergence of miscibility gaps below a critical Br concentration. Specifically, at a temperature of 300 K and an intensity of ~2-sun, the above 25% Br concentration makes the mixed compounds thermodynamically unstable. The transition from a stable to an unstable state is consistent with experimental evidence[34,87,89,90]. In conclusion, the model can help construct the phase diagrams for each compound, both in the dark and under illumination, to distinguish between stable, metastable, and unstable regions.

To identify candidates with the potential to suppress light-induced phase segregation in mixed halide perovskites, a screening study was conducted across different A-cations in compounds of the form $APb(I_{1-x}Br_x)_3$, where A = $FA_{0.75}Cs_{0.25}$, $FA_{0.5}Cs_{0.5}$, $FA_{0.25}Cs_{0.75}$, $MA_{0.75}Cs_{0.25}$, $MA_{0.5}Cs_{0.5}$, $MA_{0.25}Cs_{0.75}$, $FA_{0.75}MA_{0.25}$, $FA_{0.5}MA_{0.5}$, $FA_{0.25}MA_{0.75}$. As shown in Figure 5, these phase diagrams indicate that the threshold below which Br content must be kept to prevent phase segregation was a component-dependent value. Figure 5A-C

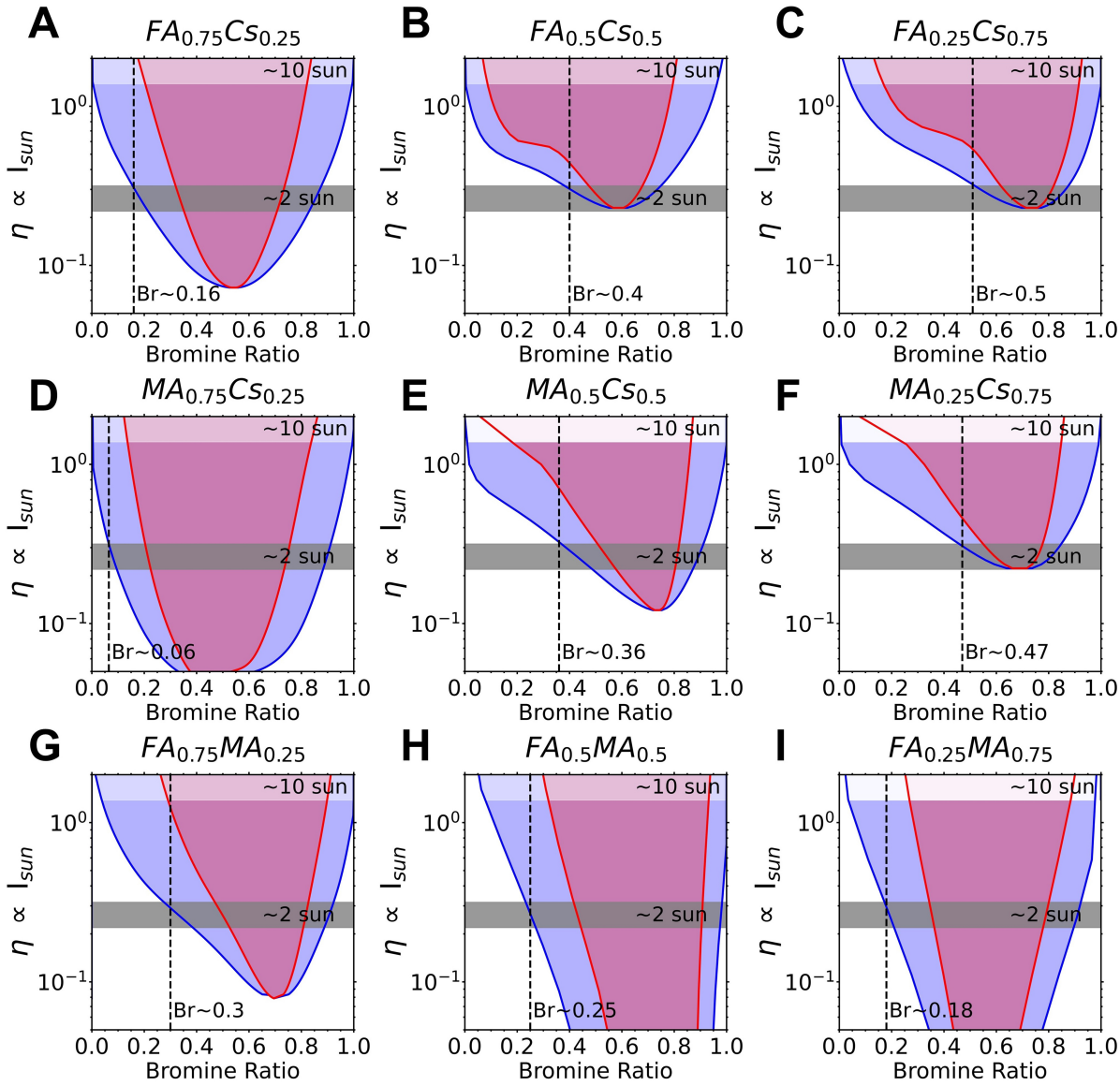

Figure 5. The predicted phase diagrams of mixed halide perovskites [$APb(I_{1-x}Br_x)_3$], A = (A) $FA_{0.75}Cs_{0.25}$; (B) $FA_{0.5}Cs_{0.5}$; (C) $FA_{0.25}Cs_{0.75}$; (D) $MA_{0.75}Cs_{0.25}$; (E) $MA_{0.5}Cs_{0.5}$; (F) $MA_{0.25}Cs_{0.75}$; (G) $FA_{0.75}MA_{0.25}$; (H) $FA_{0.5}MA_{0.5}$; (I) $FA_{0.25}MA_{0.75}$. The binodal and spinodal lines and regions are shown in blue and red, respectively. The light intensities are illustrated by greyscale shaded areas (~2-sun, 10-sun).

shows that as the Br content threshold increases, the Cs content correspondingly rises. Similarly, for the $MAPb(I_{1-x}Br_x)_3$, the Br content threshold increases for different cation compositions in the order $MA_{0.75}Cs_{0.25}$-$MA_{0.5}Cs_{0.5}$-$MA_{0.25}Cs_{0.75}$ [Figure 5D-F]. The results indicate that mixing in Cs in the MA and FA compounds has a stabilizing effect. However, Figure 5G-I presents an inverse relationship between the Br content threshold and the MA content, indicating that partial MA alloying does not improve the photostability of $FAPb(I_{1-x}Br_x)_3$. In conclusion, rational alloying of A-site cations, such as alloying Cs into both MA and FA compounds, can effectively combat halide segregation.

As mentioned above, the bandgap variation for different halide compositions primarily results from differences in the energy of the valence band maxima[84]. Specifically, an increase in Br concentration lowers

the energy of the VBM[85]. This energetic variation subsequently leads to bandgap differences between the parent, iodine-rich, and bromine-rich phases. When photon energy exceeds the bandgap of perovskite, it converts to phonons and deposits in the crystal lattice, facilitating ion migration[75]. This leads to uneven component distribution, forming regions with diverse bandgaps. Regions with smaller bandgaps more easily absorb photons, becoming rich in electrons and holes, while regions with larger bandgaps absorb less light. This disparity causes uneven charge distribution, promoting the nucleation of I- and Br-rich phases, ultimately leading to phase separation. Therefore, these bandgap differences are the thermodynamic drivers of free energy differences that favor halide segregation. As shown in Supplementary Figure 7A and B, incorporating Cs into both MA and FA compounds reduces the bandgap differences (ΔE_g), hindering the formation of the iodine-rich phase and thus preventing phase segregation. However, excessive doping with MA yields an opposing effect, widening the bandgap differences [Supplementary Figure 7C].

Based on the kinetics of phase segregation, photo-induced phase segregation is due to the migration of halide ions[91,92]. If ion migration is suppressed, phase segregation becomes difficult[93]. Recently, it has been observed that lattice distortion plays a crucial role in this process[94,95]. Lattice distortion, which can be modulated by incorporating differently sized ions, such as small A-site cations or large X-site anions, effectively raises the energy barrier for ion migration[33]. This increase in the migration barrier is primarily due to the decreased average interatomic distance, especially between the A-site cation and iodide, caused by the lattice distortion[27]. Consequently, this hindrance to ion movement serves as a significant factor in inhibiting the onset of phase segregation[96,97]. Additionally, strategies such as incorporating small A-site ions can lead to lattice contraction, further enhancing the halide ion migration barrier and contributing to suppressing phase segregation[94,98]. To summarize, these findings from the kinetic perspective are consistent with the discoveries in this paper from the thermodynamic perspective. Specifically, small-sized A-site ions prove advantageous in preventing phase segregation.

The consequence of light-induced halide segregation studied here is that the attractive bandgap tunability of mixed halide perovskites simultaneously leads to photostability problems. Nevertheless, our study serves as a guide for further experimental research on light-induced halide segregation. The challenge can be overcome using informed alloying strategies. For example, using relatively high Cs content rather than MA at the A-site is preferable in improving photostability of mixed perovskites, which agrees with the findings of Chen *et al.* that screened the components of mixed halide perovskites capable of inhibiting phase segregation through a high-throughput robotic system[99]. Additionally, Cao *et al.* discovered that by adjusting and increasing the Cs content in wide-bandgap perovskites, the crystallinity of the perovskite films can be improved, thereby aiding in the suppression of phase segregation[100]. Appropriate component selection can also avoid light-induced segregation. As shown in Figure 5B, at 2 Sun illumination and room temperature, $FA_{0.5}Cs_{0.5}Pb(I_{1-x}Br_x)_3$ should be photostable up to 40% Br concentration. This allows reaching a bandgap of 1.82 eV, which is sufficient for the monolithic tandem solar cell with a narrow bandgap perovskite of ~1.2 eV.

CONCLUSIONS

To summarize, photo-induced halide segregation in mixed halide perovskites limits the practical application of many alloy compositions. To address this challenge, we developed a data-driven framework to predict the bandgaps and study halide segregation. Firstly, by adopting a transfer learning strategy, we fine-tuned the parameters of the Graph Neural Network model using collected experimental bandgap data and DFT-calculated bandgap data. The transfer learning model, Atomsets-MLP, improved the prediction accuracy of the perovskite bandgap to a MAE of 0.078 eV, better than the original Graph Neural Network without transfer learning, significantly improving transferability and accuracy. Secondly, applying Atomsets-MLP,

we predicted the bandgaps of 53,361 mixed perovskites with continuous compositions and built a comprehensive and high-accuracy mixed perovskite database with a rich variety of elements and compositions. Finally, utilizing this database and based on the light-induced segregation theoretical model, we calculated the free energy of 20,688 mixed iodine-bromine perovskites and generated corresponding phase diagrams to study their light-induced segregation behavior.

We found that the bandgap increases as the ionic radius at the A-site and X-site decreases. Reaching a desired bandgap is achieved by adjusting the ratio of I/Br. However, this adjustment will lead to the bandgap differences between the parent phase, the iodine-rich phase, and the bromine-rich phase, which are the driving force of phase segregation. To avoid the segregation behavior, the Br content must be kept below a certain threshold, which is related to the A-site composition. Furthermore, using a relatively high Cs content at the A-site rather than MA can reduce the bandgap difference in mixed perovskites. Therefore, it is preferable for improving the photostability of mixed perovskites. By accurately predicting bandgaps and studying halide segregation behavior, we can significantly reduce the need for costly and time-consuming experimental trials. The approach paves the way for more efficient and targeted material design, enabling the rapid identification of promising mixed perovskite compositions with enhanced photostability for practical applications in solar cells and other optoelectronic devices.

DECLARATIONS

Acknowledgments

We thank Shaogang Hao and Ming Shao of the Tencent Quantum Laboratory for their help in this work.

Authors' contributions

Conceived the idea and designed the project: Wu B, Zhao H
Performed data analysis and interpretation: Wu B, Xu Z, Sun Q
Supervised the project: Xu Z, Sun Q
Drafted the manuscript: Wu B
Revised and finalized the manuscript: Zhang X, Wang Z, Chen Z, Liu S, Liu J
All authors read and approved the final manuscript.

Availability of data and materials

Supplementary Materials are available from the *Journal of Materials Informatics* or the authors. Perovskite experimental bandgap dataset and predicted bandgap dataset are provided with this paper. Mixed ABX_3 perovskites for this work can be found at http://43.138.168.107/.

Financial support and sponsorship

This work was supported by the National Natural Science Foundation of China (Grant Nos. 52002016, 52173234, and 22209074), Shenzhen Science and Technology Program (Grant Nos. RCJC20200714114435061, JCY20210324102008023, ZDSYS20220606100606013 and JSGG20210802153408024), Shenzhen-Hong Kong-Macau Technology Research Program (Type C, SGDX2020110309300301), Natural Science Foundation of Guangdong Province (2022A1515010554 and 2022A1515011959), Open Project of Zhengzhou JiShu Institute of AI Science (ZZJSB2023001), Hong Kong Quantum AI Lab, AIR@InnoHK of the Hong Kong Government, Natural Science Foundation of Hunan Province (Grant No. 2023JJ40041), High Performance Computing Platform of Nanjing University of Aeronautics and Astronautics.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

REFERENCES

1. Yu W, Zou Y, Wang H, et al. Breaking the bottleneck of lead-free perovskite solar cells through dimensionality modulation. *Chem Soc Rev* 2024;53:1769-88. DOI
2. Raza H, Imran T, Gao Y, et al. Potential-induced degradation: a challenge in the commercialization of perovskite solar cells. *Energy Environ Sci* 2024;17:1819-53. DOI
3. Liang Z, Zhang Y, Xu H, et al. Homogenizing out-of-plane cation composition in perovskite solar cells. *Nature* 2023;624:557-63. DOI PubMed PMC
4. Park SM, Wei M, Lempesis N, et al. Low-loss contacts on textured substrates for inverted perovskite solar cells. *Nature* 2023;624:289-94. DOI PubMed
5. Brinkmann KO, Wang P, Lang F, et al. Perovskite-organic tandem solar cells. *Nat Rev Mater* 2024;9:202-17. DOI
6. Suchan K, Jacobsson TJ, Rehermann C, Unger EL, Kirchartz T, Wolff CM. Rationalizing performance losses of wide bandgap perovskite solar cells evident in data from the *Perovskite Database*. *Adv Energy Mater* 2024;14:2303420. DOI
7. Wang YK, Singh K, Li JY, et al. In situ inorganic ligand replenishment enables bandgap stability in mixed-halide perovskite quantum dot solids. *Adv Mater* 2022;34:e2200854. DOI PubMed
8. Yao Y, Cheng C, Zhang C, Hu H, Wang K, De Wolf S. Organic hole-transport layers for efficient, stable, and scalable inverted perovskite solar cells. *Adv Mater* 2022;34:e2203794. DOI PubMed
9. Green MA, Dunlop ED, Yoshita M, et al. Solar cell efficiency tables (version 62). *Prog Photovoltaics* 2023;31:651-63. DOI
10. Aydin E, Allen TG, De Bastiani M, et al. Pathways toward commercial perovskite/silicon tandem photovoltaics. *Science* 2024;383:eadh3849. DOI PubMed
11. Xie L, Yin L, Liu Y, et al. Interface engineering for efficient raindrop solar cell. *ACS Nano* 2022;16:5292-302. DOI PubMed
12. Grandhi GK, Hardy D, Krishnaiah M, et al. Wide-bandgap perovskite-inspired materials: defect-driven challenges for high-performance optoelectronics. *Adv Funct Mater* 2023:2307441. DOI
13. Feng HJ, Paudel TR, Tsymbal EY, Zeng XC. Tunable optical properties and charge separation in $CH_3NH_3Sn_xPb_{1-x}I_3/TiO_2$-based planar perovskites cells. *J Am Chem Soc* 2015;137:8227-36. DOI PubMed
14. Liu Z, Krückemeier L, Krogmeier B, et al. Open-circuit voltages exceeding 1.26 V in planar methylammonium lead iodide perovskite solar cells. *ACS Energy Lett* 2019;4:110-7. DOI
15. Brinkmann KO, Becker T, Zimmermann F, et al. Perovskite-organic tandem solar cells with indium oxide interconnect. *Nature* 2022;604:280-6. DOI PubMed
16. Jiang Q, Tong J, Xian Y, et al. Surface reaction for efficient and stable inverted perovskite solar cells. *Nature* 2022;611:278-83. DOI PubMed
17. Lin R, Xu J, Wei M, et al. All-perovskite tandem solar cells with improved grain surface passivation. *Nature* 2022;603:73-8. DOI PubMed
18. Chen W, Zhu Y, Xiu J, et al. Monolithic perovskite/organic tandem solar cells with 23.6% efficiency enabled by reduced voltage losses and optimized interconnecting layer. *Nat Energy* 2022;7:229-37. DOI
19. Li L, Wang Y, Wang X, et al. Flexible all-perovskite tandem solar cells approaching 25% efficiency with molecule-bridged hole-selective contact. *Nat Energy* 2022;7:708-17. DOI
20. Datta K, Caiazzo A, Hope MA, et al. Light-induced halide segregation in 2D and quasi-2D mixed-halide perovskites. *ACS Energy Lett* 2023;8:1662-70. DOI PubMed PMC
21. Zhang J, Duan J, Guo Q, et al. A universal grain "cage" to suppress halide segregation of mixed-halide inorganic perovskite solar cells. *ACS Energy Lett* 2022;7:3467-75. DOI
22. Wen J, Zhao Y, Liu Z, et al. Steric engineering enables efficient and photostable wide-bandgap perovskites for all-perovskite tandem solar cells. *Adv Mater* 2022;34:e2110356. DOI PubMed
23. Yang G, Ni Z, Yu ZJ, et al. Defect engineering in wide-bandgap perovskites for efficient perovskite-silicon tandem solar cells. *Nat Photon* 2022;16:588-94. DOI

24. Wang C, Shao W, Liang J, et al. Suppressing phase segregation in wide bandgap perovskites for monolithic perovskite/organic tandem solar cells with reduced voltage loss. *Small* 2022;18:e2204081. DOI PubMed
25. Wright AD, Patel JB, Johnston MB, Herz LM. Temperature-dependent reversal of phase segregation in mixed-halide perovskites. *Adv Mater* 2023;35:e2210834. DOI PubMed
26. Shirzadi E, Tappy N, Ansari F, Nazeeruddin MK, Hagfeldt A, Dyson PJ. Deconvolution of light-induced ion migration phenomena by statistical analysis of cathodoluminescence in lead halide-based perovskites. *Adv Sci* 2022;9:e2103729. DOI PubMed PMC
27. Wang Z, Zeng L, Zhu T, et al. Suppressed phase segregation for triple-junction perovskite solar cells. *Nature* 2023;618:74-9. DOI PubMed
28. Cai R, Bektas H, Wang X, et al. Accelerated perovskite oxide development for thermochemical energy storage by a high-throughput combinatorial approach. *Adv Energy Mater* 2023;13:2203833. DOI
29. Sabino FP, Dalpian GM, Zunger A. Light-induced frenkel defect pair formation can lead to phase-segregation of otherwise miscible halide perovskite alloys. *Adv Energy Mater* 2023;13:2301539. DOI
30. Shin Y, Galli G. Tunable ferroelectricity in oxygen-deficient perovskites with Grenier structure. *npj Comput Mater* 2023;9:218. DOI
31. Barker AJ, Sadhanala A, Deschler F, et al. Defect-assisted photoinduced halide segregation in mixed-halide perovskite thin films. *ACS Energy Lett* 2017;2:1416-24. DOI
32. Mao W, Hall CR, Bernardi S, et al. Light-induced reversal of ion segregation in mixed-halide perovskites. *Nat Mater* 2021;20:55-61. DOI PubMed
33. Brennan MC, Ruth A, Kamat PV, Kuno M. Photoinduced anion segregation in mixed halide perovskites. *Trends Chem* 2020;2:282-301. DOI
34. Chen Z, Brocks G, Tao S, Bobbert PA. Unified theory for light-induced halide segregation in mixed halide perovskites. *Nat Commun* 2021;12:2687. DOI PubMed PMC
35. Bechtel JS, Van der Ven A. First-principles thermodynamics study of phase stability in inorganic halide perovskite solid solutions. *Phys Rev Mater* 2018;2:045401. DOI
36. Frohna K, Anaya M, Macpherson S, et al. Nanoscale chemical heterogeneity dominates the optoelectronic response of alloyed perovskite solar cells. *Nat Nanotechnol* 2022;17:190-6. DOI PubMed
37. Ghasemi M, Zhang Y, Zhou C, et al. Controllable acceleration and deceleration of charge carrier transport in metal-halide perovskite single-crystal by Cs-cation induced bandgap engineering. *Small* 2022;18:e2107680. DOI PubMed
38. Mussakhanuly N, Soufiani AM, Bernardi S, et al. Thermal disorder-induced strain and carrier localization activate reverse halide segregation. *Adv Mater* 2024;36:e2311458. DOI PubMed
39. Zhu C, Niu X, Fu Y, et al. Strain engineering in perovskite solar cells and its impacts on carrier dynamics. *Nat Commun* 2019;10:815. DOI PubMed PMC
40. Yang F. Effects of size mismatch of halide ions on the phase stability of mixed halide perovskites. *Phys Scr* 2024;99:025937. DOI
41. Slotcavage DJ, Karunadasa HI, Mcgehee MD. Light-induced phase segregation in halide-perovskite absorbers. *ACS Energy Lett* 2016;1:1199-205. DOI
42. Wang P, Wu Y, Cai B, Ma Q, Zheng X, Zhang W. Solution-processable perovskite solar cells toward commercialization: progress and challenges. *Adv Funct Mater* 2019;29:1807661. DOI
43. Li N, Luo Y, Chen Z, et al. Microscopic degradation in formamidinium-cesium lead iodide perovskite solar cells under operational stressors. *Joule* 2020;4:1743-58. DOI
44. Gautam SK, Kim M, Miquita DR, Bourée J, Geffroy B, Plantevin O. Reversible photoinduced phase segregation and origin of long carrier lifetime in mixed-halide perovskite films. *Adv Funct Mater* 2020;30:2002622. DOI
45. Das B, Aguilera I, Rau U, Kirchartz T. Effect of doping, photodoping, and bandgap variation on the performance of perovskite solar cells. *Adv Opt Mater* 2022;10:2101947. DOI
46. Wang T, Li R, Ardekani H, et al. Sustainable materials acceleration platform reveals stable and efficient wide-bandgap metal halide perovskite alloys. *Matter* 2023;6:2963-86. DOI
47. Mannodi-kanakkithodi A, Chan MKY. Data-driven design of novel halide perovskite alloys. *Energy Environ Sci* 2022;15:1930-49. DOI
48. Al-qaisi S, Ali MA, Alrebdi TA, et al. First-principles investigations of Ba_2NaIO_6 double perovskite semiconductor: material for low-cost energy technologies. *Mater Chem Phys* 2022;275:125237. DOI
49. Liu Y, Tan X, Liang J, Han H, Xiang P, Yan W. Machine learning for perovskite solar cells and component materials: key technologies and prospects. *Adv Funct Mater* 2023;33:2214271. DOI
50. Li C, Zheng K. Methods, progresses, and opportunities of materials informatics. *InfoMat* 2023;5:e12425. DOI
51. Zhao Y, Heumueller T, Zhang J, et al. A bilayer conducting polymer structure for planar perovskite solar cells with over 1,400 hours operational stability at elevated temperatures. *Nat Energy* 2022;7:144-52. DOI
52. Zhao H, Chen W, Huang H, et al. A robotic platform for the synthesis of colloidal nanocrystals. *Nat Synth* 2023;2:505-14. DOI
53. Liu S, Chen Z, Liu Y, et al. Data-driven controlled synthesis of oriented quasi-spherical $CsPbBr_3$ perovskite materials. *Angew Chem Int Ed Engl* 2024;63:e202319480. DOI PubMed
54. Wu L, Chen Z, Yuan Z, et al. Data-driven fine element tuning of halide double perovskite for enhanced photoluminescence. *Adv Opt Mater* 2024;12:2301245. DOI
55. Wang Z, Sun Z, Yin H, et al. Data-driven materials innovation and applications. *Adv Mater* 2022;34:e2104113. DOI PubMed

56. Wang Z, Yang M, Xie X, et al. Applications of machine learning in perovskite materials. *Adv Compos Hybrid Mater* 2022;5:2700-20. DOI
57. Liu Y, Yan W, Han S, et al. How machine learning predicts and explains the performance of perovskite solar cells. *Solar RRL* 2022;6:2101100. DOI
58. Hu Y, Hu X, Zhang L, et al. Machine-learning modeling for ultra-stable high-efficiency perovskite solar cells. *Adv Energy Mater* 2022;12:2201463. DOI
59. Priyange GS, Mattur MN, Nagappan N, Rath S, Thomas T. Prediction of nature of band gap of perovskite oxides (ABO_3) using a machine learning approach. *J Materiomics* 2022;8:937-48. DOI
60. Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. *Chem Mater* 2019;31:3564-72. DOI
61. Chen C, Ong SP. AtomSets as a hierarchical transfer learning framework for small and large materials datasets. *npj Comput Mater* 2021;7:173. DOI
62. Jacobsson TJ, Hultqvist A, García-fernández A, et al. An open-access database and analysis tool for perovskite solar cells based on the FAIR data principles. *Nat Energy* 2022;7:107-15. DOI
63. Liu Y, Yan W, Zhu H, Tu Y, Guan L, Tan X. Study on bandgap predications of ABX_3-type perovskites by machine learning. *Org Electron* 2022;101:106426. DOI
64. Yang C, Chong X, Hu M, et al. Accelerating the discovery of hybrid perovskites with targeted band gaps via interpretable machine learning. *ACS Appl Mater Interfaces* 2023;15:40419-27. DOI PubMed
65. Ong SP, Richards WD, Jain A, et al. Python materials genomics (pymatgen): a robust, open-source python library for materials analysis. *Comp Mater Sci* 2013;68:314-9. DOI
66. Chen C, Zuo Y, Ye W, Li X, Ong SP. Learning properties of ordered and disordered materials from multi-fidelity data. *Nat Comput Sci* 2021;1:46-53. DOI PubMed
67. Coley CW, Barzilay R, Green WH, Jaakkola TS, Jensen KF. Convolutional embedding of attributed molecular graphs for physical property prediction. *J Chem Inf Model* 2017;57:1757-72. DOI PubMed
68. Wang J, Xu P, Ji X, Li M, Lu W. Feature selection in machine learning for perovskite materials design and discovery. *Materials* 2023;16:3134. DOI PubMed PMC
69. Chen C, Ong SP. A universal graph deep learning interatomic potential for the periodic table. *Nat Comput Sci* 2022;2:718-28. DOI PubMed
70. Deng B, Zhong P, Jun K, et al. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat Mach Intell* 2023;5:1031-41. DOI
71. Jain A, Ong SP, Hautier G, et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Materials* 2013;1:011002. DOI
72. Sher A, van Schilfgaarde M, Chen AB, Chen W. Quasichemical approximation in binary alloys. *Phys Rev B Condens Matter* 1987;36:4279-95. DOI PubMed
73. Phan AT, Paek M, Kang Y. Phase equilibria and thermodynamics of the Fe–Al–C system: critical evaluation, experiment and thermodynamic optimization. *Acta Mater* 2014;79:1-15. DOI
74. Wasiur-rahman S, Medraj M. Critical assessment and thermodynamic modeling of the binary Mg–Zn, Ca–Zn and ternary Mg–Ca–Zn systems. *Intermetallics* 2009;17:847-64. DOI
75. Zhu T, Teale S, Grater L, et al. Coupling photogeneration with thermodynamic modeling of light-induced alloy segregation enables the discovery of stabilizing dopants. arXiv. [Preprint.] Jan 30, 2023. [accessed on 22 May 2024]. Available from: https://doi.org/10.48550/arXiv.2301.12627.
76. Abedi S, Tarighi Ahmadpour M, Baninajarian S, et al. Statistical analysis of the performance of a variety of first-principles schemes for accurate prediction of binary semiconductor band gaps. *J Chem Phys* 2023;158:184109. DOI PubMed
77. Ramadan AJ, Oliver RDJ, Johnston MB, Snaith HJ. Methylammonium-free wide-bandgap metal halide perovskites for tandem photovoltaics. *Nat Rev Mater* 2023;8:822-38. DOI
78. Peng S, Wang Y, Braun M, et al. Kinetics and mechanism of light-induced phase separation in a mixed-halide perovskite. *Matter* 2023;6:2052-65. DOI
79. Wu P, Thrithamarassery Gangadharan D, Saidaminov MI, Tan H. A Roadmap for efficient and stable all-perovskite tandem solar cells from a chemistry perspective. *ACS Cent Sci* 2023;9:14-26. DOI PubMed PMC
80. Eperon GE, Stranks SD, Menelaou C, Johnston MB, Herz LM, Snaith HJ. Formamidinium lead trihalide: a broadly tunable perovskite for efficient planar heterojunction solar cells. *Energy Environ Sci* 2014;7:982. DOI
81. Brivio F, Walker AB, Walsh A. Structural and electronic properties of hybrid perovskites for high-efficiency thin-film photovoltaics from first-principles. *APL Mater* 2013;1:042111. DOI
82. Noh JH, Im SH, Heo JH, Mandal TN, Seok SI. Chemical management for colorful, efficient, and stable inorganic-organic hybrid nanostructured solar cells. *Nano Lett* 2013;13:1764-9. DOI PubMed
83. Tao S, Schmidt I, Brocks G, et al. Absolute energy level positions in tin- and lead-based halide perovskites. *Nat Commun* 2019;10:2560. DOI PubMed PMC
84. Butler KT, Frost JM, Walsh A. Band alignment of the hybrid halide perovskites $CH_3NH_3PbCl_3$, $CH_3NH_3PbBr_3$ and $CH_3NH_3PbI_3$. *Mater Horiz* 2015;2:228-31. DOI

85. Lu X, Zhao Z, Li K, et al. First-principles insight into the photoelectronic properties of Ge-based perovskites. *RSC Adv* 2016;6:86976-81. DOI
86. Braly IL, Stoddard RJ, Rajagopal A, et al. Current-induced phase segregation in mixed halide hybrid perovskites and its impact on two-terminal tandem solar cell design. *ACS Energy Lett* 2017;2:1841-7. DOI
87. Dong X, Chao L, Niu T, et al. Phase-pure engineering for efficient and stable formamidinium-based perovskite solar cells. *Solar RRL* 2022;6:2200060. DOI
88. Otero-Martínez C, Imran M, Schrenker NJ, et al. Fast A-site cation cross-exchange at room temperature: single-to double- and triple-cation halide perovskite nanocrystals. *Angew Chem Int Ed Engl* 2022;61:e202205617. DOI PubMed PMC
89. Boldyreva AG, Zhidkov IS, Tsarev S, et al. Unraveling the impact of hole transport materials on photostability of perovskite films and p-i-n solar cells. *ACS Appl Mater Interfaces* 2020;12:19161-73. DOI PubMed
90. Xie LQ, Chen L, Nan ZA, et al. Understanding the cubic phase stabilization and crystallization kinetics in mixed cations and halides perovskite single crystals. *J Am Chem Soc* 2017;139:3320-3. DOI
91. Zhong Y, Yang J, Wang X, et al. Inhibition of ion migration for highly efficient and stable perovskite solar cells. *Adv Mater* 2023;35:e2302552. DOI PubMed
92. Mathew P, Cho J, Kamat PV. Ramifications of ion migration in 2D lead halide perovskites. *ACS Energy Lett* 2024;9:1103-14. DOI
93. Zhang L, Wang S, Jiang Y, Yuan M. Stable and efficient mixed-halide perovskite LEDs. *ChemSusChem* 2024;17:e202301205. DOI PubMed
94. Sun H, Liu S, Liu X, et al. Suppressed phase segregation with small A-site and large X-site incorporation for photostable wide-bandgap perovskite solar cells. *Small Methods* 2024;17:e2400067. DOI PubMed
95. Wang C, Qu D, Zhou B, et al. Self-healing behavior of the metal halide perovskites and photovoltaics. *Small* 2024;20:e2307645. DOI PubMed
96. Wang K, Yao Q, Zhang J, et al. Short-range migration of A-site cations inhibit photoinduced phase segregation in $FA_xMA_yCs_{1-x-y}PbI_{3-z}Br_z$ single crystals. *J Phys Chem C* 2021;125:23050-7. DOI
97. Cho J, Kamat PV. Photoinduced phase segregation in mixed halide perovskites: thermodynamic and kinetic aspects of Cl–Br segregation. *Adv Opt Mater* 2021;9:2001440. DOI
98. Knight AJ, Herz LM. Preventing phase segregation in mixed-halide perovskites: a perspective. *Energy Environ Sci* 2020;13:2024-46. DOI
99. Chen S, Hou Y, Chen H, et al. Exploring the stability of novel wide bandgap perovskites by a robot based high throughput approach. *Adva Energy Mater* 2018;8:1701543. DOI
100. Cao J, Fang Z, Liu S. Tailoring the Cs/Br ratio for efficient and stable wide-bandgap perovskite solar cells. *Solar RRL* 2023;7:2200955. DOI