Supplementary Material

Machine learning-enabled chemical space exploration of all-inorganic perovskites for photovoltaics

Jin-Soo Kim[1], Juhwan Noh[1*], Jino Im[1*]

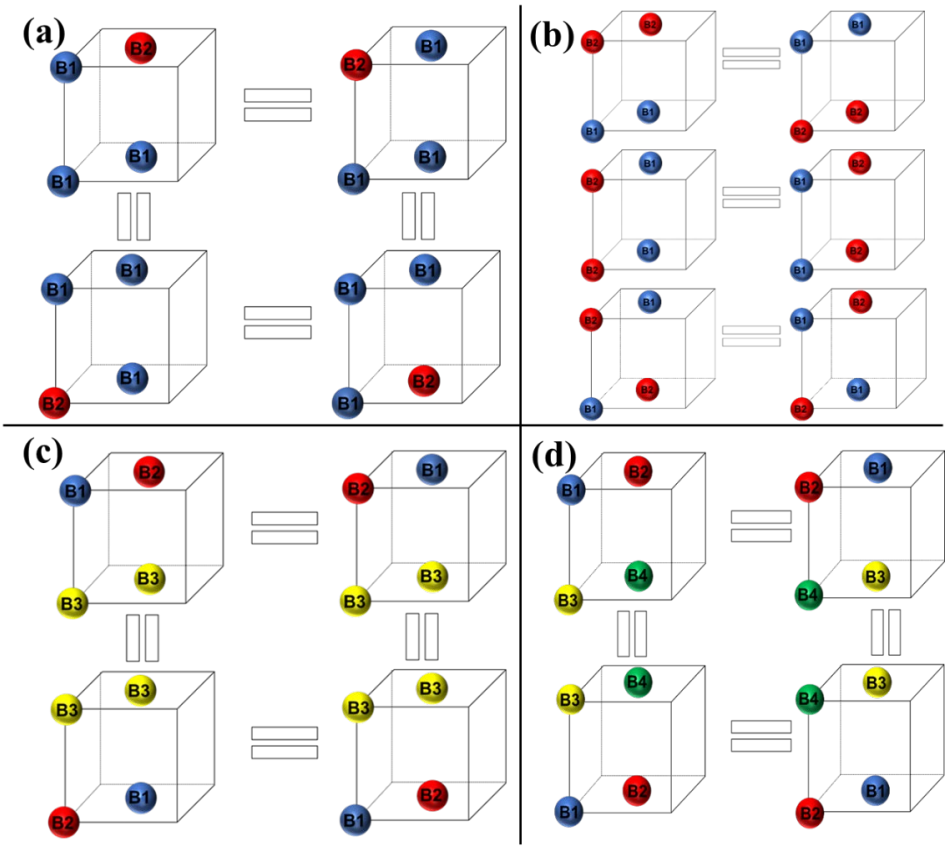
[1]Chemical Data-driven Research Center, Korea Research Institute of Chemical Technology, Daejeon 34114, Republic of Korea. E-mail: js.kim@krict.re.kr, jhwan@krict.re.kr, imjino@krict.re.kr

*Corresponding author

Supplementary Note 1. Details of obtaining 3,159 non-equivalent B-mixed ABX3 structures for the training data

We constructed 3159 unique B-site-mixed ABX_3 structures using our in-house code (https://github.com/KRICT-DATA/Perov_CGCNN/blob/main/DFT_training_data/run.sh). Since we use an identical initial structure, which has *Pnma* space group, construction of non-equivalent configurations is straightforward. For a simple description, we will explain four different cases, where B-site is unary, binary, ternary, and quaternary. When B-site is unary, a composition of B-site element is fixed as one, and thus, only one configuration is possible. In total, the number of non-equivalent configurations in the unary case equals to the possible number of elemental combinations, which is calculated to be is ${}_3C_1 \times {}_6C_1 \times {}_3C_1 = 54$.

When B-site is binary, the number of non-equivalent configurations differs depending on a composition. In the case of binary, the possible combinations of elements are ${}_3C_1 \times {}_6C_2 \times {}_3C_1 = 135$. Within the four B-sites in the unit cell, ratios of 1:3, 2:2, and 3:1 for B_1:B_2 exist. At 1:3 and 3:1, the number of non-equivalent configurations is 1, because the four configurations shown in the **Supplementary Figure 1a** are equivalent to each other. Note that the atomic positions of B-sites of the crystal structure as CIF format in the **Supplementary Note 6**. At 2:2, the number of non-equivalent configurations is 3, because the three among the six ($\because {}_4C_2=6$) cases are equivalent to each other (see **Supplementary Figure 1b**). In total, when B-site is binary, the number of non-equivalent configurations is $135 \times (1+3+1) = 675$.



Supplementary Figure 1. Equivalent atomic configurations when (a) B_1:B_2=1:3, (b) B_1:B_2=2:2, (c) B_1:B_2:B_3=1:1:2, and (d) B_1:B_2:B_3:B_4=1:1:1:1. Blue, red, yellow and green balls represent B_1, B_2, B_3, and B_4, respectively.

When B-site is ternary, the possible number of elemental combinations is ${}_3C_1 \times {}_6C_3 \times {}_3C_1 = 180$. For each elemental combination, the possible compositions are in ratios of B_1:B_2:B_3 = 1:1:2, 1:2:1, and 2:1:1. Within each composition, the total count of configurations sums up to 12 when including equivalent configurations (∵ ${}_4C_2 \times {}_2C_1 = 12$). The number of non-equivalent configurations is 3, because the four configurations shown in the **Supplementary Figure 1c** are equivalent to each other (3=12÷4). Hence, when B-site is ternary, the count of non-equivalent configurations would be 180×(3+3+3) = 1620.

When B-site is quaternary, the possible number of elemental combinations is ${}_3C_1 \times {}_6C_4 \times {}_3C_1 = 135$. For each elemental combination, the possible composition is in ratios of B_1:B_2:B_3:B_4= 1:1:1:1. Within this composition, the total count of configurations sums up to 24 when including equivalent configurations (∵ 4!=24). The number of non-equivalent configurations is 6, because the four configurations shown in the **Supplementary Figure 1d** are equivalent to each other (6=24÷4). Hence, when B-site is quaternary, the count of non-equivalent configurations would be 135×6 = 810.

Therefore, the total number of non-equivalent configurations is 54+675+1620+810=3159.

Supplementary Note 2. Details of classification accuracy comparison of Bartel[1] and Filip's[2] tolerance factor

The Filip's tolerance factor pertains to double perovskites, assuming 1:1 ratio of B and B`. Since our study assumes compositions other than 1:1, the formula to calculate average octahedral factor ($\bar{\mu}$) and the octahedral mismatch ($\Delta\mu$) in the paper by Filip et al[2] had to be modified as follows.

$$\bar{\mu} = (\sum_{i=1}^{N} x_i r_{B_i})/r_X$$

$$\Delta\mu = \sum_{i=1}^{N_B-1} \sum_{j=i+1}^{N_B} \frac{|x_i r_{B_i} - x_j r_{B_j}|}{r_X}$$

where x_i and r_{B_i} are composition and ionic radius of ion *i*. N_B= 2, 3, and 4 for binary, ternary, and quaternary systems for B-site.

We applied the geometric limits for the formability of perovskites summarized in the paper by Filip et al[2], for several materials experimentally confirmed to exist as perovskites. Such materials are $CsGe_{0.25}Sn_{0.75}Br_3$, $CsGe_{0.5}Sn_{0.5}Br_3$, $CsGe_{0.75}Sn_{0.25}Br_3$, $CsGe_{0.5}Sn_{0.5}I_3$, $CsPb_{0.3}Sn_{0.7}Br_3$, $CsPb_{0.5}Sn_{0.5}Br_3$, $CsPb_{0.7}Sn_{0.3}Br_3$, and $CsGe_{0.1}Pb_{0.9}I_3$. The conditions predicted 7 out of 8 as non-perovskites, whereas Bartel's tolerance factor predicted 6 out of 8 as perovskites. From these results, we can observe that Bartel's tolerance factor predicts formability of perovskites closer to the experimental results. We attached the Python code for obtaining these results (https://github.com/KRICT-DATA/Perov_CGCNN/blob/main/Codes_for_figues/tolerance.py).

Supplementary Note 3. Details of computing 5.6×10^{11} atomic configurations

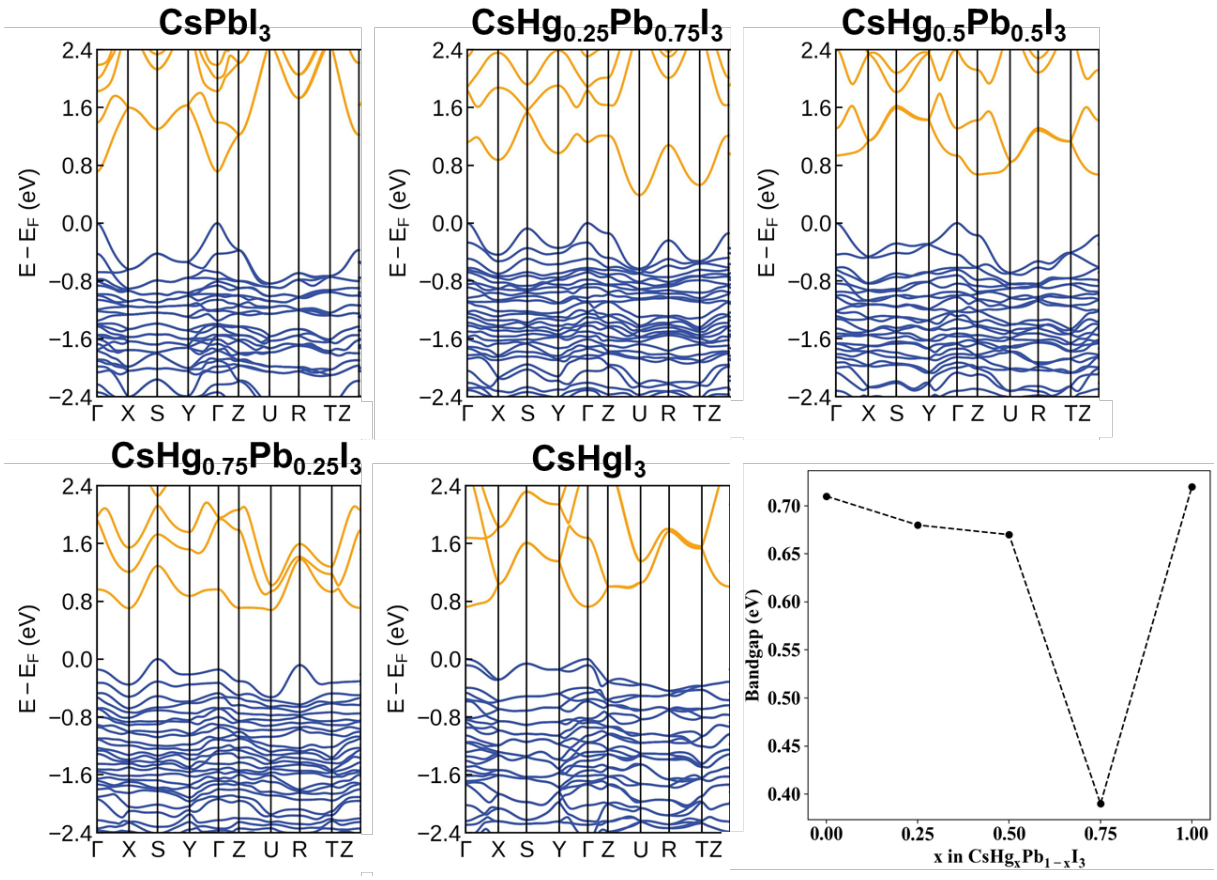
The count of 5.6×10^{11} was calculated using the following Python code. This code calculates the number of configurations when B-site is binary, ternary, and quaternary. Taking the example of binary for easier understanding, the total number of atomic configurations when the B-site is binary alloys is the product of (the number of all possible arrangements of two types of elements occupying 16 B-sites) and (the number of elemental combinations of binary alloy for the B-site). The same principle applies to the calculation for ternary and quaternary cases as well. We attached the Python code for obtaining this result (https://github.com/KRICT-DATA/Perov_CGCNN/blob/main/Codes_for_figues/count.py).

Supplementary Note 4. Analysis of the performance degradation in bandgap predictions

We analyzed performance degradation in **Supplementary Figure 8b** compared to **Figure 4b**, and we found that data with incorrect band type predictions show relatively high prediction error as shown in **Supplementary Table 1**. Specifically, while the trained CGCNN model predicted all 110 selected compounds as non-indirect bandgap materials, 75 compounds were incorrectly predicted showing relatively inaccurate prediction performance (i.e. high MAE) for bandgap. In particular, we note that among the latter 75 data with misclassified band type, 68 data included Hg atom in B-site (5 of 35 data with correctly classified band type contain Hg atom in B-site). Here, we note that compounds including Hg were frequently selected through the screening process due to their thermodynamic stability. From this, we speculate that the introduction of Hg into the B-site caused changes in the electronic structure that made it challenging for the ML model to predict. **Supplementary Figure 2** below illustrates how difficult it becomes to predict band type and bandgap when Hg is included. Both $CsPbI_3$ and $CsHgI_3$ have direct bandgaps around 0.7 eV, but for intermediate compositions, all exhibit indirect bandgaps. Especially, the bandgap of $CsHg_{0.75}Pb_{0.25}I_3$ sharply decreases. Due to these characteristics of Hg, it seems that our ML model did not adequately learn the band type and bandgap when Hg was included.

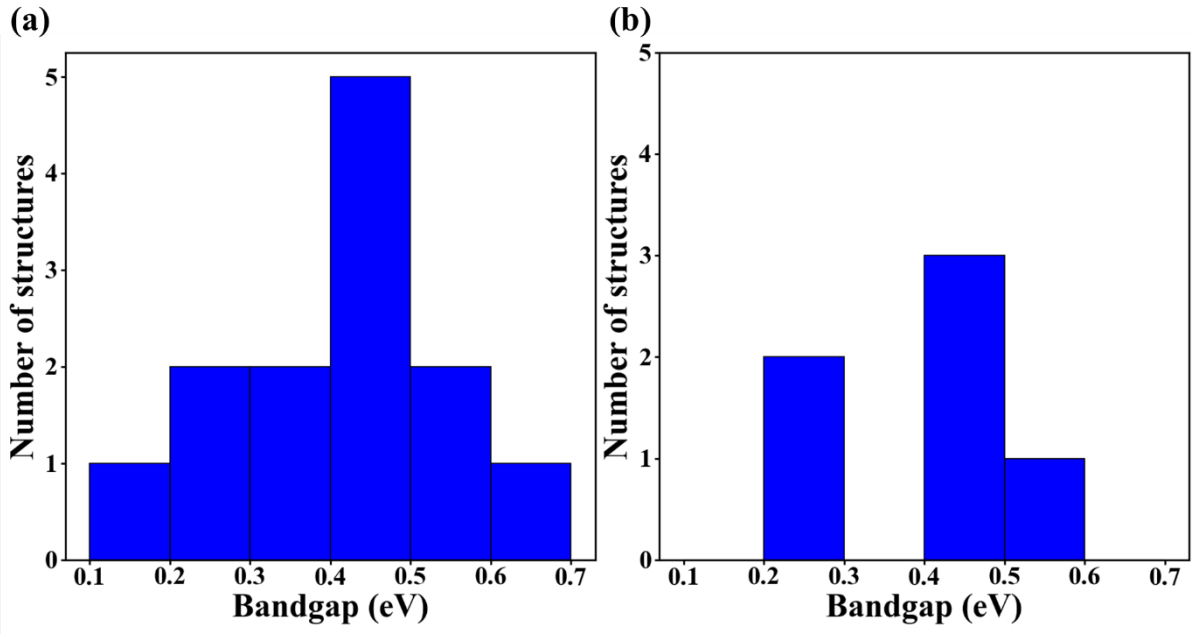
Supplementary Table 1. Bandgap prediction mean absolute error (MAE) and R^2 comparison between data with correct band type prediction and data with incorrect band type prediction among the 110 selected data. We also listed MAE and R^2 of the entire 110 data as references.

	Whole data	Correct band type prediction	Incorrect band type prediction
MAE (eV)	0.142	0.076	0.178
R^2	0.023	0.606	-2.720



Supplementary Figure 2. Band structure plot of $CsHg_xPb_{1-x}I_3$ (x=0, 0.25, 0.5, 0.75, 1.0) and bandgap as a function of Hg ratio.

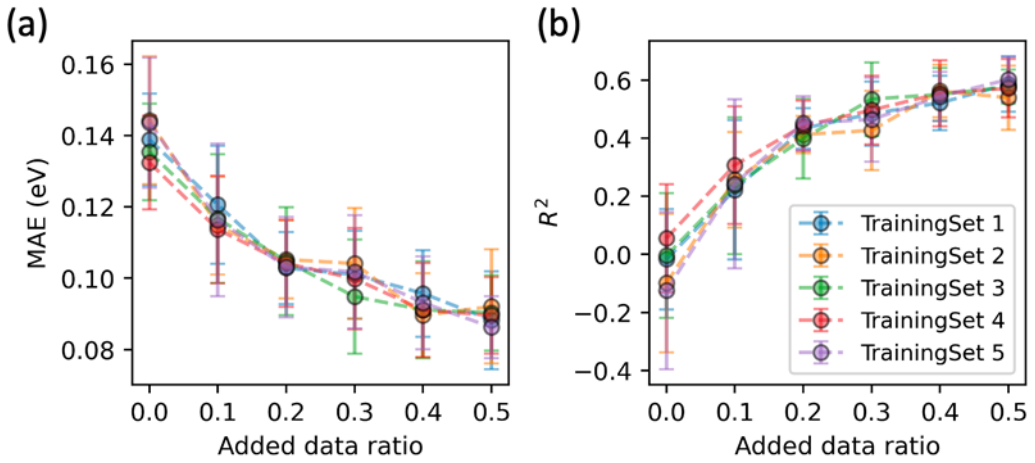
In addition, even within the same composition, electronic band property varies as a function of atomic configurations, as shown in **Supplementary Figure 3**. That is, using the atomic configurations within the training domain (i.e., 4 atoms for B-site) possibly causes limited prediction capacity when exploring the expanded atomic configurations with 16 atoms for B-site. **Supplementary Figure 3a** and **3b** depict the distribution of bandgaps for different atomic arrangements with 16 and 4 B-site atoms, respectively. As described, the distribution in **Supplementary Figure 3a** is wider than in **3b**. Thus, predicting bandgaps for a structure with 16 atoms for B-site using a model trained by structures with 4 atoms for B-site would likely result in significant errors.



Supplementary Figure 3. Distribution of the PBESol-calculated bandgap values for (a) 12 randomly generated disordered structures with a fixed composition of $Cs_{16}Cd_2Ge_9Hg_3Sn_2Cl_{48}$ (b) 6 non-equivalent structures with a fixed composition of $Cs_4Cd_1Ge_1Hg_1Sn_1Cl_{16}$.

Supplementary Note 5. Strategy to enhance the bandgap prediction reliability

One practically applicable strategy to enhance the prediction reliability could be adding training data randomly selected from the target domain after the DFT calculations as similarly done in active learning strategy. To validate the efficacy of the latter strategy, we analyzed the performance enhancement by adding new data obtained from the target domain after conducting a small amount of DFT calculations as shown in **Supplementary Figure 4**. Specifically, we added X % of 110 selected candidates (X = 0, 10, 20, 30, 40, and 50) to the training domain remaining 50 % of the 110 selected candidates as the test data. We then computed the mean absolute error (MAE, **Supplementary Figure 4a**) and R^2 score (**Supplementary Figure 4b**) for the test data after training the model for 5000 epochs. In **Supplementary Figure 4**, we plotted the averaged results after conducting five independent trials of the latter procedures (i.e., splitting training set within 3,159 original data as well as adding randomly selected data to the fixed training set) to reduce the effect of randomness, and the error bar in each figure means standard deviation. As described, we could observe a clear improvement in both MAE and R^2 score by adding more DFT-calculated data to the training data. In particular, we could observe a significant improvement in both MAE and R^2 score by only using 20 % of 110 selected candidates indicating the efficacy of the proposed strategy. While we currently validate the proposed strategy for the 110 selected candidates, we expect a further improvement of prediction reliability for the bandgap prediction by utilizing the widely used active learning scheme.



Supplementary Figure 4. Prediction performance analysis (i.e., mean absolute error; MAE, and R^2 score) of hold-out test set randomly selected 50 % of 110 candidate materials by adding X % of DFT-calculated data (X = 0, 0.1, 0.2, 0.3, 0.4 and 0.5) within the remaining 50 % of 110 candidate materials.

Supplementary Note 6. Crystal structure used for CGCNN training (.CIF format)

```
data_image0
_chemical_formula_structural       Cs4Pb4I12
_chemical_formula_sum              "Cs4 Pb4 I12"
_cell_length_a          8
_cell_length_b          8.5
_cell_length_c          11.6
_cell_angle_alpha       90
_cell_angle_beta        90
_cell_angle_gamma       90

_space_group_name_H-M_alt       "P 1"
_space_group_IT_number          1

loop_
   _space_group_symop_operation_xyz
   'x, y, z'

loop_
   _atom_site_type_symbol
   _atom_site_label
   _atom_site_symmetry_multiplicity
   _atom_site_fract_x
   _atom_site_fract_y
   _atom_site_fract_z
   _atom_site_occupancy
   Cs   Cs1          1.0   0.54005   0.09701   0.75000   1.0000
   Cs   Cs2          1.0   0.45995   0.90299   0.25000   1.0000
   Cs   Cs3          1.0   0.04220   0.40063   0.24999   1.0000
   Cs   Cs4          1.0   0.95780   0.59937   0.75001   1.0000
   Pb   Pb1          1.0   0.00000   0.00000   0.50000   1.0000
   Pb   Pb2          1.0   0.50000   0.50000   0.50000   1.0000
   Pb   Pb3          1.0   0.00000   0.00000   0.00000   1.0000
   Pb   Pb4          1.0   0.50000   0.50000   0.00000   1.0000
   I    I1           1.0   0.31844   0.18727   0.43899   1.0000
   I    I2           1.0   0.68156   0.81273   0.56101   1.0000
   I    I3           1.0   0.17678   0.69199   0.43853   1.0000
   I    I4           1.0   0.82322   0.30801   0.56147   1.0000
   I    I5           1.0   0.68155   0.81274   0.93899   1.0000
   I    I6           1.0   0.31845   0.18726   0.06101   1.0000
   I    I7           1.0   0.82322   0.30803   0.93854   1.0000
   I    I8           1.0   0.17678   0.69197   0.06146   1.0000
   I    I9           1.0   0.87492   0.03392   0.24999   1.0000
   I    I10          1.0   0.12508   0.96608   0.75001   1.0000
   I    I11          1.0   0.36350   0.45884   0.75000   1.0000
   I    I12          1.0   0.63650   0.54116   0.25000   1.0000
```

Supplementary Table 2. List of Materials Project (MP ID) or Open Quantum Materials Database (OQMD ID) IDs and the DFT (PBESol)-calculated total energy of the AX and BX_2 phases.

Formula	MP-ID or OQMD ID	Total energy (eV)
CsBr	mp-571222	-25.46
CsCl	mp-573697	-27.51
CsI	mp-614603	-23.09
KBr	mp-23251	-25.84
KCl	mp-23193	-28.09
KI	mp-22898	-23.26
RbBr	mp-22867	-25.42
RbCl	mp-23295	-27.58
RbI	mp-22903	-22.95
$CdBr_2$	mp-27934	-14.90
$CdCl_2$	mp-22881	-25.27
CdI_2	mp-571579	-63.77
$GeBr_2$	mp-1184708	-20.40
$GeCl_2$	oqmd-1343439	-89.99
GeI_2	mp-1540935	-28.30
$HgBr_2$	mp-571558	-16.63
$HgCl_2$	mp-22855	-25.20
HgI_2	mp-583213	-157.12
$PbBr_2$	mp-2456335	-41.22
$PbCl_2$	mp-862871	-22.57
PbI_2	mp-567542	-37.35
$SnBr_2$	mp-1272139	-10.15
$SnCl_2$	mp-569152	-44.86
SnI_2	mp-27194	-27.58
$ZnBr_2$	mp-647579	-23.13
$ZnCl_2$	mp-22909	-35.72
ZnI_2	mp-27161	-216.05

Supplementary Table 3. Hyperparameters of crystal graph convolution neural network (CGCNN) used in this study.

Hyperparameter	Value
Number of properties used in atom feature vector	9
Number of convolutional layers	3
Length of learned atom feature vector	92
Number of hidden features in convolutional layers	64
Number of hidden features after pooling	128
Number of hidden layers after pooling	1
Optimizer	Stochastic gradient descent
Batch size	256
Learning rate	0.01

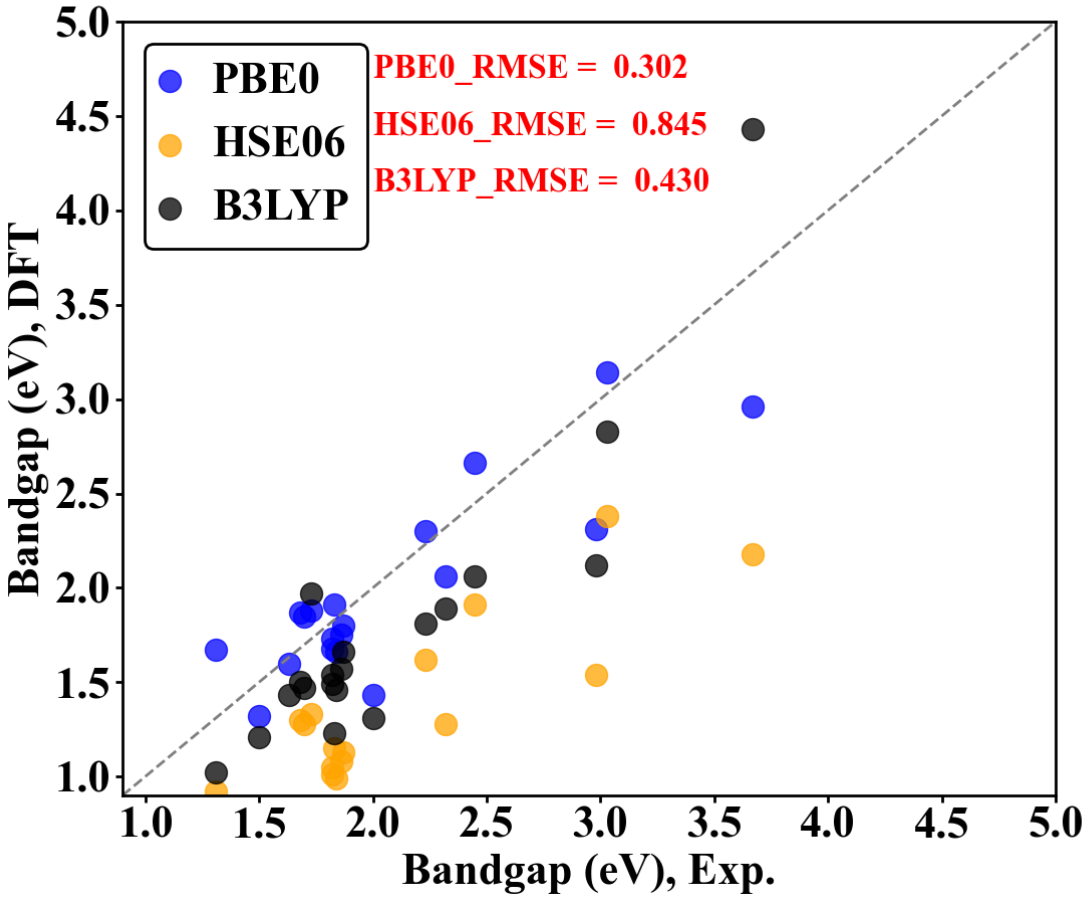
Supplementary Table 4. Comparison between experimental (E_{gap}^{exp}), PBE0-calculated (E_{gap}^{PBE0}), HSE06-calculated (E_{gap}^{HSE06}), and B3LYP-calculated (E_{gap}^{B3LYP}) bandgaps.

Formula	E_{gap}^{exp} (eV)	E_{gap}^{PBE0} (eV)	E_{gap}^{HSE06} (eV)	E_{gap}^{B3LYP} (eV)
$CsPbI_3$	1.73[3]	1.88	1.33	1.97
$CsPbBr_3$	2.23~2.446[4–7]	2.66	1.91	2.06
$CsPbCl_3$	3.03[8]	3.14	2.38	2.83
$CsGeI_3$	1.63[9]	1.60	0.83	1.43
$CsGeBr_3$	2.47[10], 2.32[11]	2.06	1.28	1.89
$CsGeCl_3$	3.67[11]	2.96	2.18	4.43
$CsSnI_3$	1.31[9]	1.67	0.92	1.02
$CsSnBr_3$	1.83[10], 1.75[12]	1.91	1.15	1.23
$CsSnCl_3$	2.98[13]	2.31	1.54	2.12
$CsGe_{0.5}Sn_{0.5}I_3$	1.50[9]	1.32	0.67	1.21
$CsGe_{0.5}Sn_{0.5}Br_3$	2.0[10]	1.43	0.75	1.31
$CsPb_{0.5}Sn_{0.5}Br_3$	1.87[14]	1.80	1.13	1.66
$CsPb_{0.9}Sn_{0.1}Br_3$	2.23[14]	2.30	1.62	1.81
$CsPb_{0.4}Sn_{0.6}Br_3$	1.86[14]	1.75	1.08	1.57
$CsPb_{0.3}Sn_{0.7}Br_3$	1.82[14]	1.73	1.05	1.54
$CsPb_{0.2}Sn_{0.8}Br_3$	1.82[14]	1.68	1.01	1.49
$CsPb_{0.1}Sn_{0.9}Br_3$	1.84[14]	1.66	0.99	1.46
$CsGe_{0.1}Pb_{0.9}I_3$	1.698	1.85	1.28	1.47
$CsGe_{0.05}Pb_{0.95}I_3$	1.679	1.87	1.30	1.50

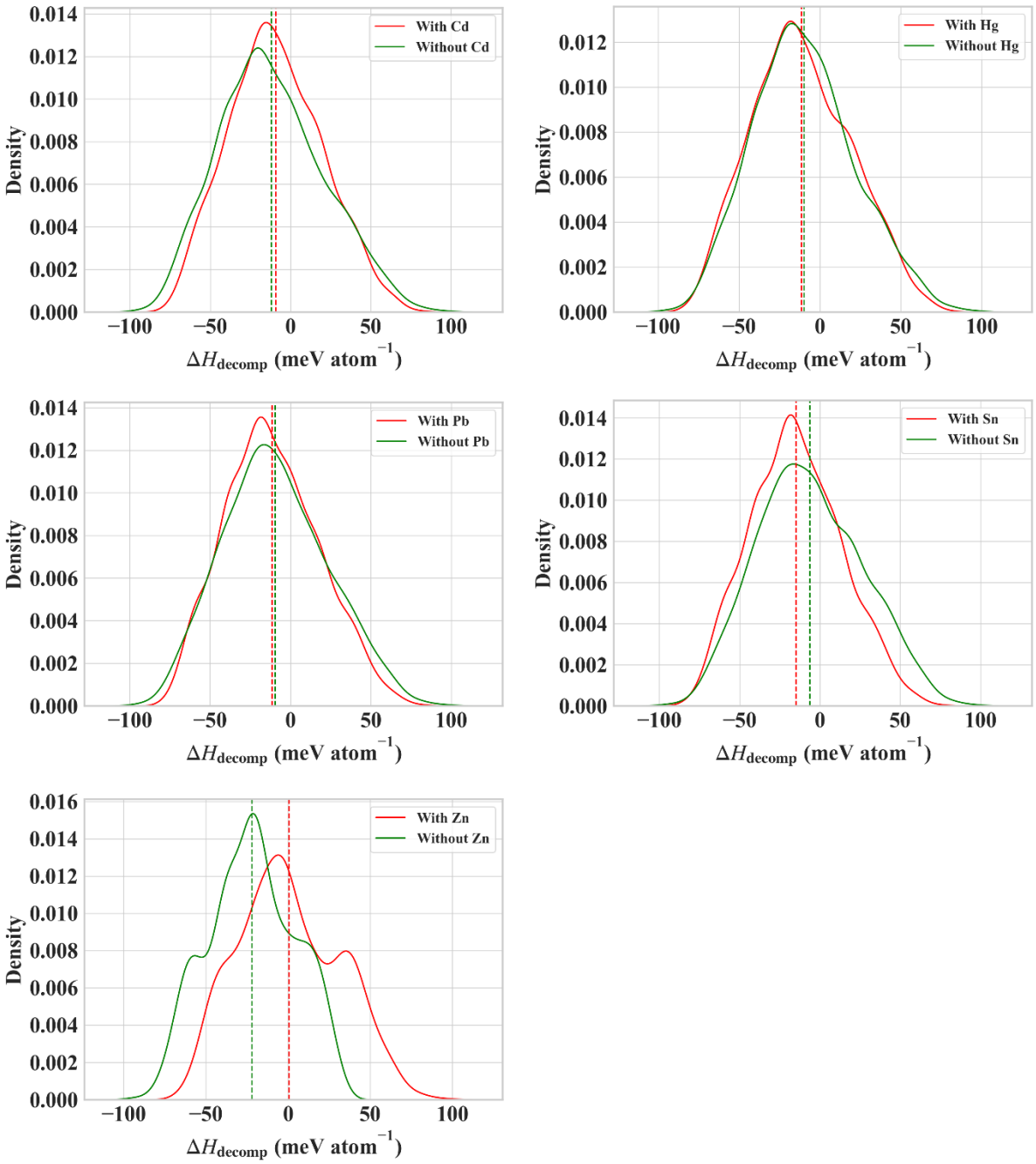
Supplementary Table 5. Comparison between CGCNN- and density functional theory-calculated thermodynamic and electronic properties of 31 compounds selected during screening. The band edges of all selected compounds lie on the gamma point. m_e^* and m_h^* represent the effective masses of the electrons and holes, respectively. We present the minimum and maximum effective masses obtained from calculations in different directions in reciprocal space. The highlighted rows correspond to the compositions belong to **Table 1** in the main manuscript.

Formula	$\Delta H_{\text{decomp}}^{\text{CGCNN}} - T\Delta S_{\text{mix}}$ (meV atom^{-1})	$\Delta H_{\text{decomp}}^{\text{PBEsol}} - T\Delta S_{\text{mix}}$ (meV atom^{-1})	τ	Band type (PBEsol)	$E_{\text{gap}}^{\text{CGCNN}}$ (eV)	$E_{\text{gap}}^{\text{PBEsol}}$ (eV)	$E_{\text{gap}}^{\text{PBE0}}$ (eV)	$m_{\text{e,min}}^*, m_{\text{e,max}}^*$ (m_e)	$m_{\text{h,min}}^*, m_{\text{h,max}}^*$ (m_e)
$CsGe_{0.8125}Sn_{0.1875}Br_3$	-99.79	-101.1	4.18	Direct	0.13	0.27	1.57	0.076, 0.076	0.043, 0.073
$CsGe_{0.75}Sn_{0.25}Br_3$	-99.96	-101.27	4.12	Direct	0.11	0.23	1.52	0.072, 0.072	0.040, 0.070
$CsGe_{0.5625}Sn_{0.4375}Br_3$	-98.18	-98.58	4.00	Direct	0.07	0.11	1.39	0.063, 0.11	0.037, 0.052
$CsGe_{0.5}Sn_{0.5}Br_3$	-97.12	-97.66	3.97	Direct	0.03	0.14	1.43	0.067, 0.10	0.040, 0.051
$CsGe_{0.4375}Sn_{0.5625}Br_3$	-95.14	-95.18	3.95	Direct	0.07	0.13	1.4	0.067, 0.10	0.041, 0.054
$CsGe_{0.375}Sn_{0.625}Br_3$	-92.92	-92.72	3.94	Direct	0.07	0.18	1.46	0.067, 0.16	0.044, 0.061
$CsGe_{0.3125}Sn_{0.6875}I_3$	-46.58	-47.46	4.17	Direct	0.17	0.2	1.34	0.070, 0.10	0.051, 0.061
$CsGe_{0.25}Sn_{0.75}I_3$	-46.34	-47.27	4.17	Direct	0.18	0.23	1.37	0.072, 0.12	0.054, 0.069
$CsGe_{0.1875}Sn_{0.8125}I_3$	-45.56	-46.32	4.17	Direct	0.18	0.22	1.36	0.073, 0.11	0.055, 0.069
$CsPb_{0.5}Sn_{0.5}Br_3$	-76.79	-76.72	4.06	Direct	0.45	0.45	1.80	0.13, 0.19	0.12, 0.13
$CsPb_{0.375}Sn_{0.625}Br_3$	-78.08	-78.2	4.05	Direct	0.43	0.43	1.75	0.13, 0.17	0.11, 0.12
$CsPb_{0.3125}Sn_{0.6875}Br_3$	-78.19	-78.36	4.05	Direct	0.41	0.41	1.73	0.13, 0.16	0.096, 0.11
$CsPb_{0.25}Sn_{0.75}Br_3$	-77.93	-78.07	4.05	Direct	0.41	0.39	1.7	0.13, 0.15	0.092, 0.10
$CsGe_{0.375}Pb_{0.625}I_3$	-29.73	-30.94	4.17	Direct	0.49	0.53	1.77	0.11, 0.14	0.11, 0.14
$CsGe_{0.0625}Pb_{0.3125}Sn_{0.625}Br_3$	-84.17	-83.95	4.01	Direct	0.38	0.39	1.71	0.12, 0.16	0.094, 0.10

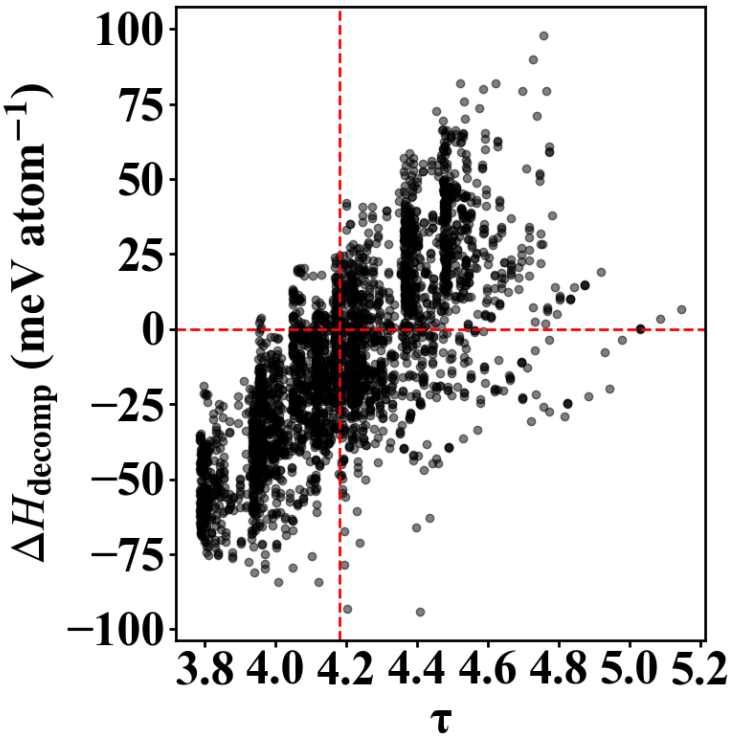
$CsGe_{0.25}Pb_{0.5}Sn_{0.25}Br_3$	-88.08	-88.07	3.95	Direct	0.48	0.5	1.86	0.10, 0.24	0.088, 0.12
$CsGe_{0.125}Pb_{0.5}Sn_{0.375}Br_3$	-84.92	-84.55	3.99	Direct	0.44	0.46	1.8	0.12, 0.17	0.10, 0.11
$CsGe_{0.0625}Pb_{0.25}Sn_{0.6875}Br_3$	-84.06	-83.87	4.01	Direct	0.37	0.37	1.69	0.12, 0.15	0.089, 0.098
$CsGe_{0.1875}Pb_{0.5}Sn_{0.3125}Br_3$	-86.85	-86.63	3.96	Direct	0.46	0.5	1.86	0.11, 0.23	0.10, 0.13
$CsGe_{0.0625}Pb_{0.5}Sn_{0.4375}Br_3$	-81.96	-81.66	4.02	Direct	0.44	0.46	1.80	0.13, 0.18	0.11, 0.12
$CsGe_{0.4375}Pb_{0.0625}Sn_{0.5}Br_3$	-98.20	-98.23	3.95	Direct	0.1	0.18	1.48	0.076, 0.10	0.049, 0.060
$CsGe_{0.0625}Pb_{0.375}Sn_{0.5625}Br_3$	-83.81	-83.63	4.01	Direct	0.4	0.41	1.74	0.13, 0.16	0.099, 0.11
$CsGe_{0.25}Pb_{0.5}Sn_{0.25}I_3$	-45.91	-46.99	4.17	Direct	0.41	0.49	1.69	0.12, 0.16	0.11, 0.13
$CsGe_{0.3125}Pb_{0.5}Sn_{0.1875}I_3$	-43.86	-44.85	4.17	Direct	0.41	0.44	1.65	0.11, 0.15	0.10, 0.13
$CsGe_{0.25}Pb_{0.5625}Sn_{0.1875}I_3$	-42.97	-44.17	4.17	Direct	0.46	0.5	1.71	0.12, 0.15	0.11, 0.14
$CsCd_{0.0625}Ge_{0.25}Pb_{0.1875}Sn_{0.5}I_3$	-54.49	-55.16	4.17	Direct	0.28	0.44	1.61	0.11, 0.23	0.11, 0.13
$CsCd_{0.0625}Ge_{0.1875}Pb_{0.5}Sn_{0.25}I_3$	-48.16	-49.39	4.18	Direct	0.49	0.64	1.84	0.15, 0.15	0.20, 0.20
$CsGe_{0.5625}Hg_{0.3125}Sn_{0.125}Cl_3$	-103.41	-108.21	3.88	Direct	0.12	0.33	1.75	0.27, 1.86	0.26, 0.59
$CsCd_{0.0625}Ge_{0.1875}Pb_{0.1875}Sn_{0.5625}I_3$	-54.35	-55.15	4.17	Direct	0.28	0.46	1.63	0.16, 0.16	0.098, 0.15
$CsCd_{0.0625}Ge_{0.25}Pb_{0.5}Sn_{0.1875}I_3$	-46.57	-47.52	4.17	Direct	0.48	0.55	1.75	0.12, 0.20	0.12, 0.18
$CsCd_{0.125}Ge_{0.5625}Hg_{0.1875}Sn_{0.125}Cl_3$	-106.60	-110.31	3.89	Direct	0.31	0.44	1.86	0.36, 2.06	0.30, 0.37



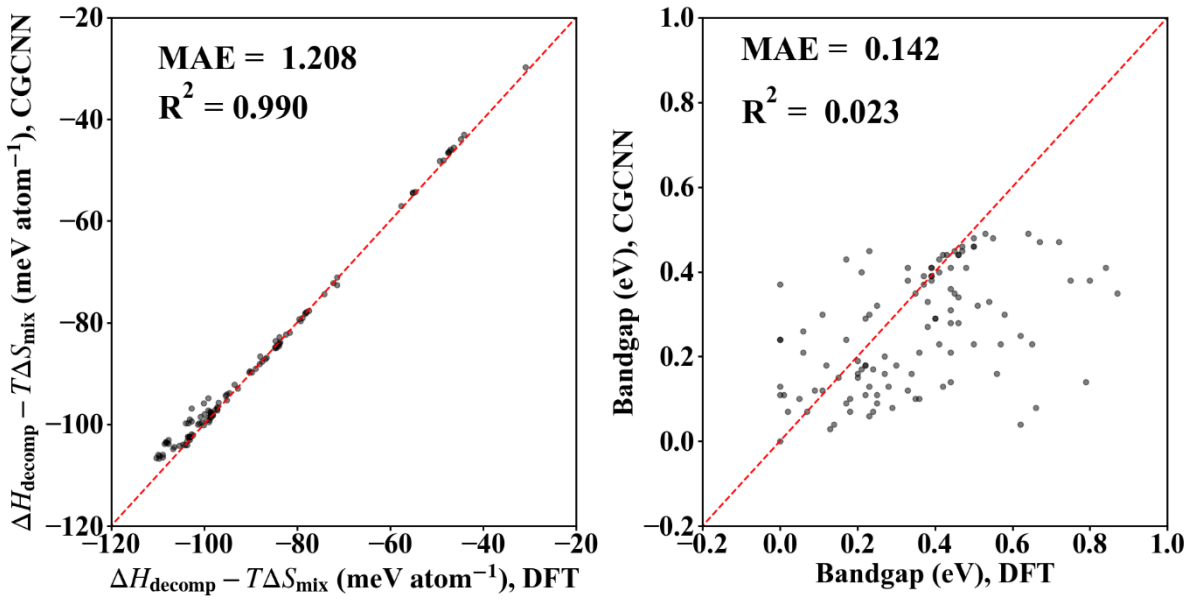
Supplementary Figure 5. PBE0-level and HSE06-level estimates of bandgaps compared with collected experimental measurements for 19 compounds shown in **Supplementary Table 4**. Root mean squared error (RMSE) of each functional compared to the experimental value are presented.



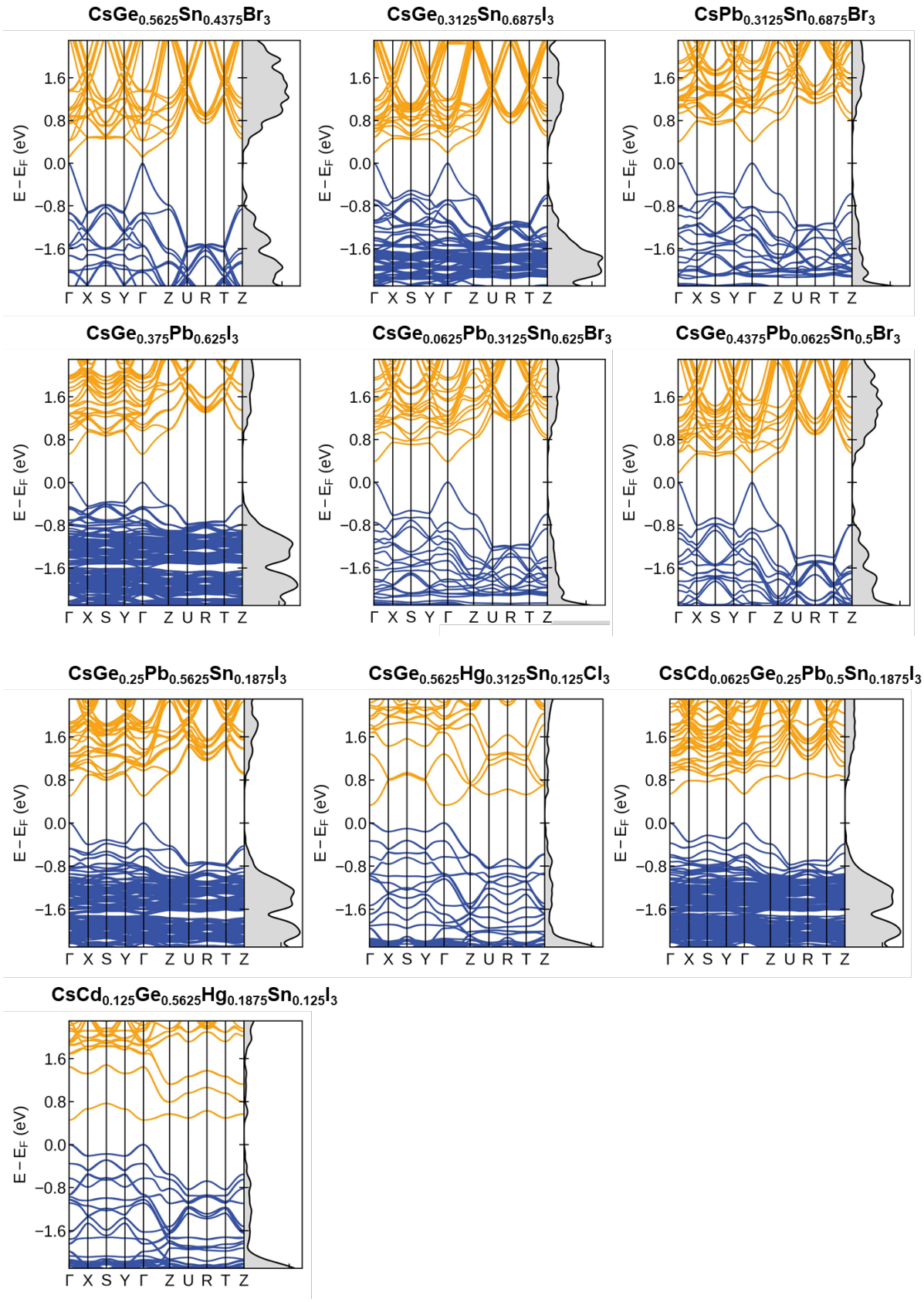
Supplementary Figure 6. Difference in distribution of the decomposition enthalpy ($\Delta H_{\mathrm{decomp}}$) with the existence of Cd, Hg, Pb, Sn, and Zn in the training dataset. Dashed lines represent the mean value of each decomposition energy distribution.



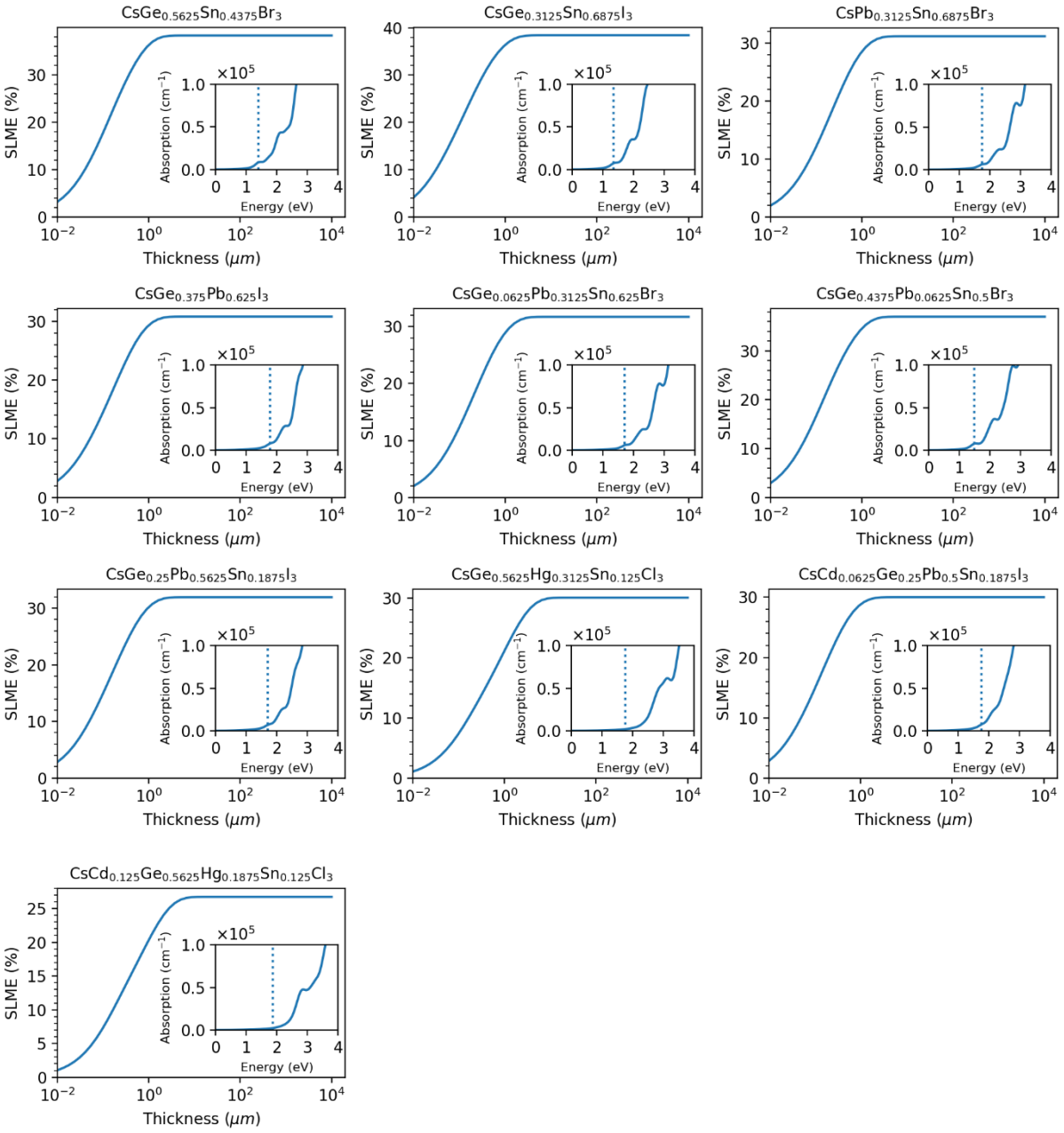
Supplementary Figure 7. Decomposition enthalpy ($\Delta H_{\mathrm{decomp}}$) versus Bartel's tolerance factor (τ) for the PBEsol training dataset.



Supplementary Figure 8. Parity plot between CGCNN-predicted and DFT (PBEsol)-calculated (a) $\Delta H_{\text{decomp}} - T\Delta S_{\text{mix}}$ and (b) bandgap for the selected 110 compounds throughout the screening procedure.



Supplementary Figure 9. PBESol-computed electronic band structure along with density of states of the 10 compounds presented in Table 1.



Supplementary Figure 10. PBE0-spin-orbit coupling-computed spectroscopic limited maximum efficiency as a function of the sample thickness and optical absorption spectra (inset) for the 10 promising compounds presented in Table 1. The vertical dotted lines represent the bandgaps.

Supplementary references

1. Bartel, C. J. *et al.* New tolerance factor to predict the stability of perovskite oxides and halides. *Sci. Adv.* **5**, eaav0693 (2019).
2. Filip, M. R. & Giustino, F. The geometric blueprint of perovskites. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 5397–5402 (2018).
3. Zhang, J., Hodes, G., Jin, Z. & Liu, S. (Frank). All-Inorganic $CsPbX_3$ Perovskite Solar Cells: Progress and Prospects. *Angew Chem Int Ed* **58**, 15596–15618 (2019).
4. Mannino, G. *et al.* Temperature-Dependent Optical Band Gap in $CsPbBr_3$, $MAPbBr_3$, and $FAPbBr_3$ Single Crystals. *J. Phys. Chem. Lett.* **11**, 2490–2496 (2020).
5. Maji, P., Sadhukhan, P. & Das, S. Optoelectronic properties of facile synthesized orthorhombic cesium lead bromide (CsPbBr3). *J Mater Sci: Mater Electron* **31**, 17100–17109 (2020).
6. Brescia, R. *et al.* Bandgap determination from individual orthorhombic thin cesium lead bromide nanosheets by electron energy-loss spectroscopy. *Nanoscale Horiz.* **5**, 1610–1617 (2020).
7. Paul, T. *et al.* Tunable cathodoluminescence over the entire visible window from all-inorganic perovskite $CsPbX_3$ 1D architecture. *J. Mater. Chem. C* **6**, 3322–3333 (2018).
8. He, Y. *et al.* Demonstration of Energy-Resolved γ-Ray Detection at Room Temperature by the $CsPbCl_3$ Perovskite Semiconductor. *J. Am. Chem. Soc.* **143**, 2068–2077 (2021).
9. Chen, M. *et al.* Highly stable and efficient all-inorganic lead-free perovskite solar cells with native-oxide passivation. *Nat Commun* **10**, 16 (2019).
10. Kama, A., Tirosh, S., Itzhak, A., Ejgenberg, M. & Cahen, D. New Pb-Free Stable Sn–Ge Solid Solution Halide Perovskites Fabricated by Spray Deposition. *ACS Appl. Energy Mater.* **5**, 3638–3646 (2022).
11. Schwarz, U., Wagner, F., Syassen, K. & Hillebrecht, H. Effect of pressure on the optical-absorption edges of CsGeBr 3 and CsGeCl 3. *Phys. Rev. B* **53**, 12545–12548 (1996).
12. Gupta, S., Bendikov, T., Hodes, G. & Cahen, D. $CsSnBr_3$, A Lead-Free Halide Perovskite for Long-Term Solar Cell Application: Insights on SnF_2 Addition. *ACS Energy Lett.* **1**, 1028–1033 (2016).
13. Ali, Md. S., Das, S., Abed, Y. F. & Basith, M. A. Lead-free $CsSnCl_3$ perovskite nanocrystals: rapid synthesis, experimental characterization and DFT simulations. *Phys. Chem. Chem. Phys.* **23**, 22184–22198 (2021).
14. Schwartz, H. A. *et al.* Band-Gap Tuning in All-Inorganic $CsPb_xSn_{1-x}Br_3$ Perovskites. *ACS Appl. Mater. Interfaces* **13**, 4203–4210 (2021).