CHEMISTRY OF MATERIALS

pubs.acs.org/cm Article

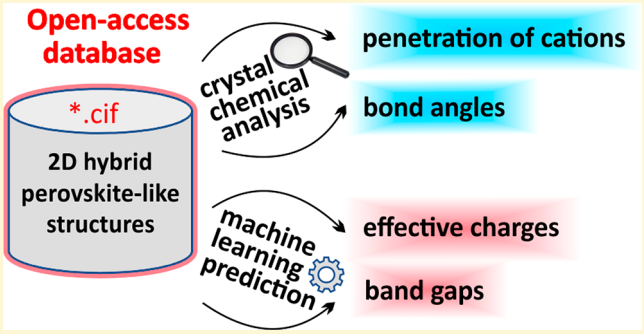
Database of Two-Dimensional Hybrid Perovskite Materials: Open-Access Collection of Crystal Structures, Band Gaps, and Atomic Partial Charges Predicted by Machine Learning

Ekaterina I. Marchenko,[#] Sergey A. Fateev,[#] Andrey A. Petrov, Vadim V. Korolev, Artem Mitrofanov, Andrey V. Petrov, Eugene A. Goodilin, and Alexey B. Tarasov*

Cite This: *Chem. Mater.* 2020, 32, 7383−7388 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We describe a first open-access database of experimentally investigated hybrid organic−inorganic materials with a two-dimensional (2D) perovskite-like crystal structure. The database includes 515 compounds, containing 180 different organic cations, 10 metals (Pb, Sn, Bi, Cd, Cu, Fe, Ge, Mn, Pd, and Sb) and 3 halogens (I, Br, and Cl) known so far and will be regularly updated. The database contains a geometrical and crystal chemical analysis of the structures, which are useful for revealing quantitative structure−property relationships for this class of compounds. We show that the penetration depth of the spacer organic cation into the inorganic layer and M−X−M bond angles increase in the number of inorganic layers (n). The machine learning model is developed and trained on the database for the prediction of a band gap with accuracy within 0.1 eV. Another machine learning model is trained for the prediction of atomic partial charges with accuracy within 0.01 e. We show that the predicted values of band gaps decrease with an increase of n and with an increase of M−X−M angles for single-layered perovskites. In general, the proposed database and machine learning models are shown to be useful tools for the rational design of new 2D hybrid perovskite materials.



INTRODUCTION

The family of organic−inorganic layered structures, often referred as "2D hybrid perovskites" (2D, two-dimensional), is derived from the perovskite structural type and exhibits unprecedented structural flexibility, which opens prospects for the design of various innovative materials for photovoltaics and optoelectronics. This class of materials shows a set of unique functional properties, such as record-breaking yield of photo- and electroluminescence, broad white-light emission, tunable narrow emission, and excellent photoconductivity.[1,2]

2D hybrid halide perovskites can be regarded as a product of cutting the three-dimensional (3D) parent compound along a specific crystallographic plane[3] as reflected by the general formula $(A')_{2/q}A_{n-1}B_nX_{3n+1}$, where $[A']^{q+}$ represents a singly (q = 1) or doubly (q = 2) charged organic spacer cation, A^+ is a small singly charged cation (such as Cs^+, $CH_3NH_3^+$, or $[HC(NH_2)_2]^+$), B^{2+} = Pb^{2+}, Ge^{2+}, Sn^{2+}, etc., X^- = Cl^-, Br^-, and I^-, and n is the number of layers of corner-shared octahedra within a perovskite slab.

Prediction of material properties based on crystal structures has come to be a useful approach for the directed rational design of the materials. Factual data sets, especially crystallographic databases, are an important tool for structural design since they deliver primary information needed for further analysis. The widespread use of high-throughput density functional theory (HT-DFT) calculations in materials science has contributed to the emergence of extensive databases[4−8] containing optimized crystal structures and the values of the most important physicochemical properties. Such projects stimulate the discovery of fundamentally new materials and the use of already known compounds in new areas. Recently, a data-driven approach has been fruitfully used to extract quantitative structure−property relationships (QSPR) hidden in publicly available materials repositories.[9,10] The screening of available data sets for specific applications was significantly accelerated by rapidly developing machine learning (ML) models.[11−19] Despite the growing interest in 2D hybrid perovskites, there is still no specific database for these structures.

In this work, we present for the first time an open-access regularly updated database of 2D hybrid perovskites containing experimental values of the band gap and structural parameters

Received: June 1, 2020
Revised: August 11, 2020
Published: August 11, 2020

https://dx.doi.org/10.1021/acs.chemmater.0c02290
Chem. Mater. 2020, 32, 7383−7388

as well as a number of structural descriptors with charges and band gap values calculated by a machine learning approach that can be used to derive the correlations "chemical composition−structure−property". Currently, the database includes 515 compounds composed of a halometallate framework consisting of cations M^{n+} of different metals (including Pb, Sn, Bi, Cd, Cu, Fe, Ge, Mn, Pd, and Sb) and halide anions X^- (Cl^-, Br^-, and I^-) as a central atom and ligands of a $[MX_6]$ octahedron, respectively, and 180 different organic spacer cations (protonated nitrogen- or phosphorus-containing organic bases) occupying cuboctahedral voids or interlayer space. In sum, the database contains 414 (100) layered perovskites, 27 (110) layered perovskites, 7 (111) layered perovskites, and 67 structures that are not ascribed to a specific structure derived from the perovskite structural type and containing metal−halide octahedral frameworks with different topologies. There are (100) structures with a different number of octahedra layers (n) including 310 items with n = 1 (single-layered) and multilayered structures: 50 items with n = 2, 34 items with n = 3, 11 items with n = 4, 3 items with n = 5, 1 item with n = 6, and 2 items with n = 7. Among the (100) structures, iodide compounds dominate (236 records) over bromide (103) and chloride (73) compounds. A progressive extension of the data set is expected as compounds with new organic cations continue to emerge.

METHODS

The present database of 2D hybrid perovskites contains both the information about the crystal structure of the materials (like the first-level databases[20]) and different parameters of these materials calculated from our crystal chemical analysis and predicted from machine learning models.

The structural information such as chemical formulas of compounds, space groups of symmetry, the number of octahedra layers (n), and the types of organic cations was extracted from *.cif files of experimentally investigated compounds supplementing the cited articles or from crystallographic data presented in another form. In the latter case, the *.cif files were manually created for convenience and for further data analysis and processing. Visualization of crystal structures was carried out using the VESTA program.[21]

Topological and Crystal Chemical Analysis. Having performed a topological and crystal-chemical analysis of the collected structures (*.cif), we derived the following parameters: the M−X bond length, X−M−X and M−X−M bond angles, and penetration depth of spacer organic cations into cuboctahedral voids of the layer of corner-shared octahedra ("perovskite layers").

For an individual representative series of related compounds (such as the series of (100) layered perovskites with the same spacer cation and different n), such parameters as volumes distortion of BX_6 octahedra and Voronoi−Dirichlet polyhedra of organic cations were calculated.

These parameters and volumes of Voronoi−Dirichlet polyhedra of organic cations that characterize the relative atomic size in crystal structures are obtained from a geometrical analysis of the crystal structures from the data set performed using the ToposPro program package[22] (see Supporting Information).

To determine the degree of distortion of BX_6 octahedra, we used the equation introduced by Alonso et al.[23] and commonly used for evaluation of the distortion degree of layered perovskites

$$\Delta d = \frac{1}{6}\sum\left[\frac{d_n - d}{d}\right]^2 \quad (1)$$

where d_n is the individual B−X distances and d is the arithmetic mean values of the individual B−X distances.

The values of the penetration depth of spacer organic cations into the inorganic layer of corner-shared octahedra (hereinafter, "penetration") were calculated for each structure with integer occupancies of N and X atoms according to the methodology described in ref 24. The penetration depth was defined as the average distance between the protonated nitrogen atoms of the spacer organic cation and the plane of the axial halogen atoms of the perovskite slab.

Development of Machine Learning Models. First machine learning model (MLM1) for predicting the band gap values was developed and trained on the reliable set of 2D hybrid perovskite structures from the database and corresponding band gap values calculated by DFT. The electronic structure was calculated by the DMol3 module of the Materials Studio software package[25,26] using the atomic basis DNP+ (Double Numerical plus polarization with the addition of diffuse functions) with spin−orbital coupling. The calculated values of the band gaps are in good agreement with the experimental data (Figure 1) and therefore was used as a training set for MLM1.

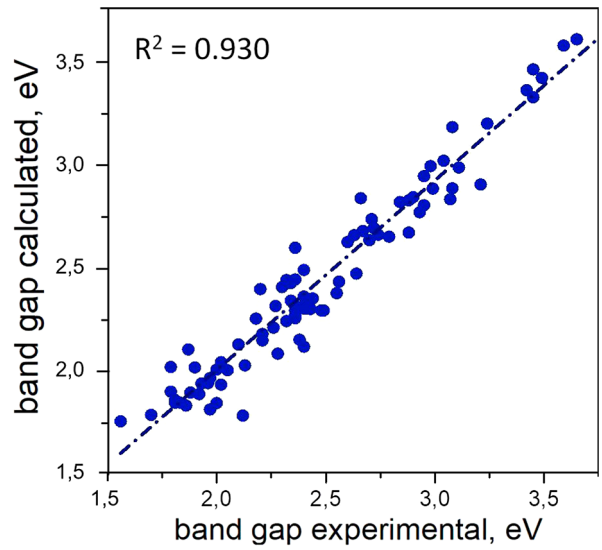


Figure 1. Band gaps (theoretical(DFT) vs experimental) of structures from the train set. Only the structures with reported experimental values of the band gap are shown here. The band gap values are listed in the Supporting Information file.

The second machine learning model for predicting partial atomic charges (MLM2) was trained on the computation-ready, experimental metal−organic frameworks (CoRE MOFs) database.[27−29] Atomic point charges for 2932 MOFs[30] obtained with the DDEC charge partitioning method were used for the presented training/validation model.

Data representation for partial charge prediction does not require averaging since this quantity is atomwise; that is, it corresponds to a distinct atom. We used elemental properties and structural descriptors (Voronoi tessellation-based features) to characterize an atomic environment.[31] The suitable structure representation is a critical point for high-performance QSPR modeling.[32,33] To represent hybrid organic−inorganic perovskites for band gap prediction, we used the smooth overlap of atomic positions (SOAP) kernel[34] (more detailed information on SOAP is presented in the Supporting Information). The local environment of each atom is determined as a fixed length vector. We introduced the fast average kernel over all atoms as descriptors for the crystal structures.

A detailed description of the ML pipeline for partial charge prediction is provided in ref 35. The gradient boosting decision tree method implemented with the XGBoost library[36] was used for both end points. We tested the performance on an external test set (20% and 10% of structures from the initial set) with 5- and 10-fold cross-validation for band gap and partial charge prediction, respectively.

https://dx.doi.org/10.1021/acs.chemmater.0c02290
Chem. Mater. 2020, 32, 7383−7388

RESULTS AND DISCUSSION

The information presented in the database for each compound is stored with open access at www.pdb.nmse-lab.ru.

Each entry of the database contains the following information about the given compound:

(1) CIF file;
(2) General formula;
(3) Number of octahedra layers;
(4) Space group symmetry;
(5) Type of layered perovskite structure (such as Ruddlesden–Popper (RP) and Dion–Jacobson (DJ));
(6) M–X bond lengths;
(7) M–X–M and X–Pb–X angles;
(8) Penetration depth of spacer organic cations into inorganic layers;
(9) Reference to the original work;
(10) DOI of the reference;
(11) Calculated band gaps;
(12) Experimental optical band gaps;
(13) Partial atomic charges.

We analyzed the structural parameters of compounds, which primarily determine the electronic structure of these compounds: distortions of the inorganic framework and penetration depth of spacer organic cations into inorganic layers. The distortions of the inorganic framework can be divided into two parts. The first is the distortion of the MX_6 octahedra (Δd) themselves (Figure 2), represented by X–Pb–X angles. The second parameter is octahedral tilting which can be determined from the M–X–M angle(s).[24] We analyzed the

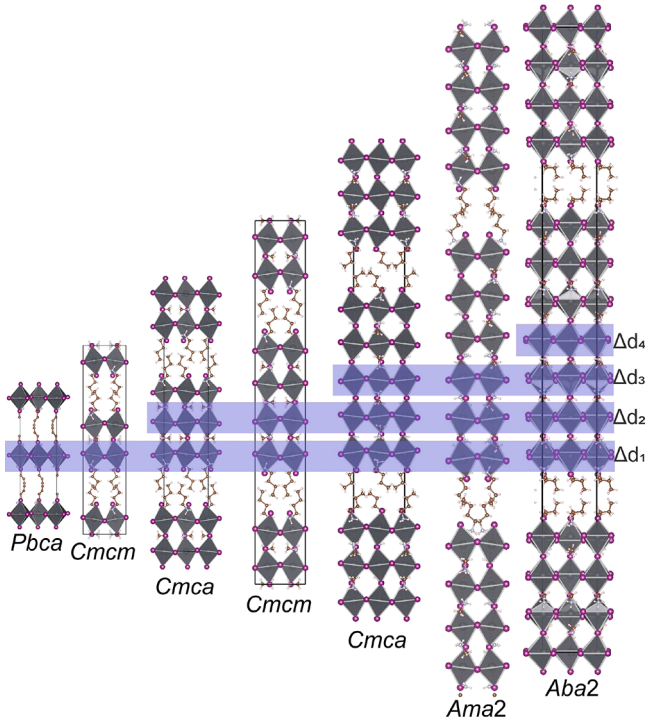


Figure 2. Series of 2D lead iodide perovskite structures (n = 1–7) with butylammonium (BA^+) as a spacer organic cation $BA_2(CH_3NH_3)_{1-n}Pb_nI_{3n+1}$. The layers of octahedra with corresponding distortions Δd_{1-4} are highlighted. Carbon, hydrogen, nitrogen, and iodine atoms are shown in dark brown, light pink, gray, and purple, respectively. PbI_6 octahedra are shown in gray.

Δd of different octahedral layers (Δd_1, Δd_2, Δd_3, Δd_4) as shown for some series of (100) 2D hybrid perovskites with different cations (Figure S3 in the Supporting Information), such as n-butylammonium (BA^+), octane-1,8-diammonium (ODA^{2+}), guanidinium (GUA^+), and 4-(aminomethyl)-piperidinium ($4AMP^+$) (see Table S2 in the Supporting Information). It was found that the octahedra of the layers neighboring with the spacer organic cations are the most distorted ones in the multilayer structures (n > 1). To investigate further this dependency, we estimated the volumes of the Voronoi–Dirichlet polyhedra (V_{VDP})[37] for spacer organic cations and revealed that it decreases with an increase of the Δd_1 (Figure S4 in the Supporting Information, Table S3).

The increase of another distortion parameter, the average M–X–M bond angle, toward an ideal 180° indicates a lower perovskite layer tilting. This should lead to the decrease of the band gap with the decreasing of the tilting for the series of homological 100 perovskites with different spacer organic cations. Indeed, we show that the predicted band gap values decrease with increasing M–X–M angles (Figure 3).

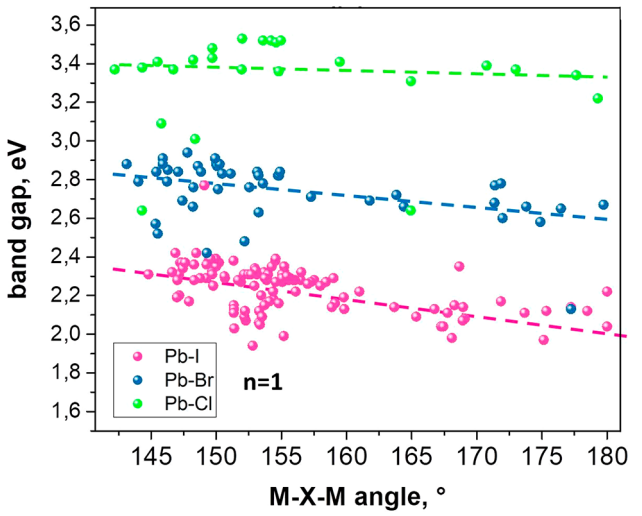


Figure 3. Dependence of the calculated band gap by ML on the Pb–X–Pb angles for single-layered (100) compounds $(A')_{2/q}A_{n-1}B_nX_{3n+1}$ with different spacer organic cations $[A']^{q+}$ from the database. The dotted lines are given for eye guidance only.

To couple the parameters of an inorganic framework with the features of organic cations, we calculate the values of the penetration depths of spacer organic cations into inorganic layers. It was shown for 10 series of (100) perovskites with n = 1–4 and Pb–I inorganic framework with different organic cations that with an increase in the number of layers, the penetration and the deviation angle of Pb–X–Pb bonds increase (Figures 4a,b). An increase in the penetration with an increase of n can be explained from an electrostatic point of view. Indeed, with an increase of n, the dielectric constant of the perovskite slab increases.[2] Consequently, the attraction between such a slab and an "electric double layer" formed by the charged planes of organic cations (+) and terminal halides (−) should also enhance, resulting in greater mutual penetration of the organic cation and the layer. The large scatter in the values of penetration for structures with the same n is simply explained by the simultaneous presence in the

https://dx.doi.org/10.1021/acs.chemmater.0c02290
Chem. Mater. 2020, 32, 7383–7388

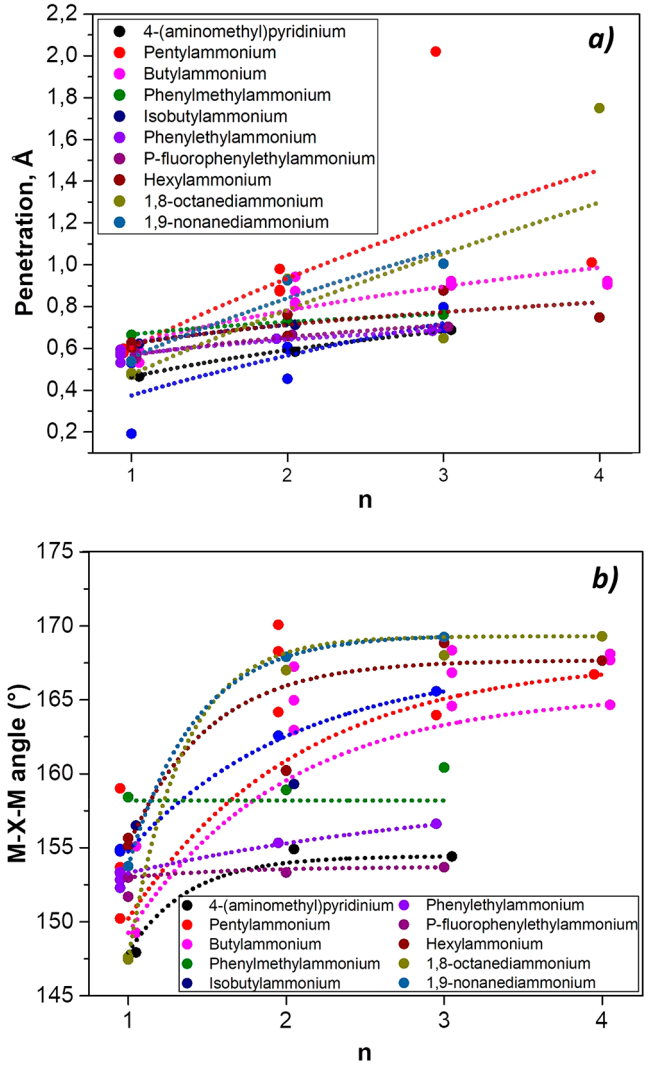


Figure 4. Penetration depths of spacer organic cation (a) and M−X−M bond angles (b) for a series of 2D lead halide perovskites $(A')_{2/q}A_{n-1}B_nX_{3n+1}$ with different spacer organic cations $[A']^{q+}$ listed in the legend for n = 1−4. The points are offset along the x axis from the n values for clarity.

database of different structures for one compound including those refined at different temperatures.

Moreover, we found a correlation between the composition of the organic cation and its penetration depth into the structure for a number of compounds with the same inorganic framework. So, for (100) single-layered compounds structures with Pb−I(Br) and Sn−I inorganic frameworks the lowest values of penetration depth are observed for spacer cations with relatively short carbon chain containing heteroatoms (see Figures S5, S6, S9, and S11).

To show the potential of the data set information for prediction of the parameters relevant for the rational design of hybrid perovskite materials, we implemented two machine learning models (MLM): MLM1 for band gap (essential for photovoltaic applications) calculation and MLM2[35] for atomic partial charges (necessary for the development of force fields models for high-throughput simulations).

Using our MLM1 algorithm, we predicted the band gap for all compounds of our database without partial occupations of M and X crystallographic positions with good accuracy: the root-mean-square deviation (RMSD) and mean absolute deviation (MAD) were 0.136 and 0.103 eV, respectively (Figure 5). Giving a smaller mismatch, our machine learning model outperforms another similar model predicting band gap values for organic crystal structures available to date.[38]

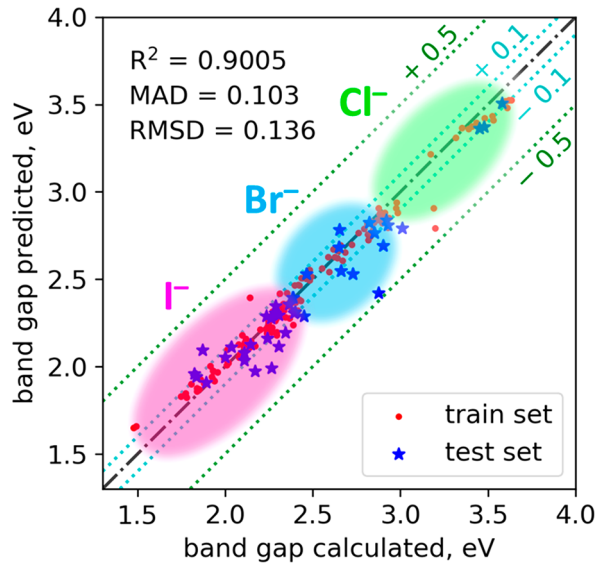


Figure 5. Band gaps of 2D hybrid halide perovskites calculated by DFT and predicted by the MLM1 model trained on the reliable set of structures from the database.

Importantly, our machine learning model predicts a clear trend of decreasing band gaps with an increase of n for all known series of 2D lead halide perovskites compounds $(A')_{2/q}A_{n-1}B_nX_{3n+1}$ with different spacer organic cations (Figure 6). The predicted trend fairly matches with a fundamental dependency of band gaps with increasing of the number of layers.[2] This implies that the band gaps for multilayer perovskites having the highest promise for a photovoltaic application can be predicted correctly. In

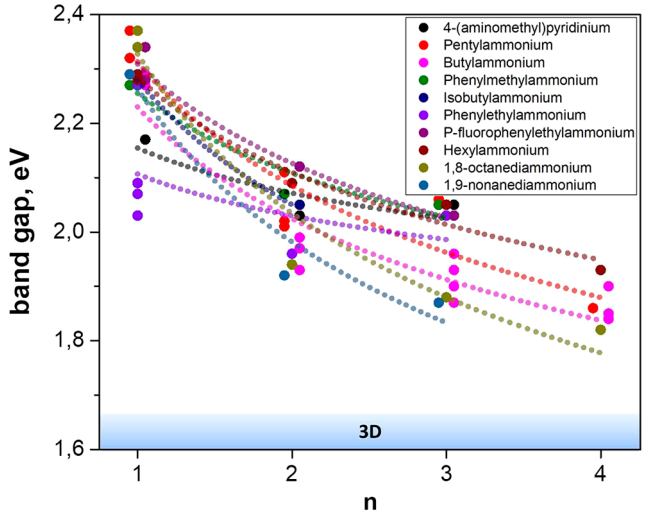


Figure 6. Band gaps calculated by the MLM1 model for a series of 2D hybrid halide perovskites $(A')_{2/q}A_{n-1}B_nX_{3n+1}$ with different spacer organic cations $[A']^{q+}$ listed in the legend for n = 1−4. The points are offset along the x axis from the n values for clarity.

https://dx.doi.org/10.1021/acs.chemmater.0c02290
Chem. Mater. 2020, 32, 7383−7388

addition, it can be seen that the slope of the trends of band gaps from n significantly depends on the tilting of the octahedra and type of spacer cation. We assume that this dependency originates from the differences in M−X−M bond angles for various series of compounds in Figure 6. Such correlations have key importance for fine-tuning of the optoelectronic properties of the layered perovskite materials.

We believe that the proposed high-throughput machine learning approach has many potential applications and can be especially useful, for instance, for band gap adjustment of light-absorbing materials used in tandem solar cells.

CONCLUSIONS

We have constructed the database of 515 layered hybrid perovskites compounds. On the basis of the crystal chemical analysis, we found several novel structural correlations. Particularly, we found that the V_{VDP} of the spacer organic cation in the structure decreases with an increasing distortion of neighboring octahedra. We show for 10 series of (100) perovskites with n = 1−4 with different spacer organic cations that with an increase in the number of layers, the penetration and the deviation angle of the Pb−X−Pb bonds increase.

We developed a machine learning model for band gap calculation and successfully trained it on the data set of our database. The proposed model predicts correctly the fundamental trend of decreasing band gap values with an increase of the number of inorganic layers (n), which is essential for tuning the band gap from 3D structures to multilayered or monolayered structures. A successful application of the developed ML model on the experimental structural data of 2D hybrid halide perovskites collected in our database opens the prospect for high-throughput searching of essential quantitative structure−property relationships (QSPR) for layered hybrid halide perovskites. We believe that the newly revealed QSPRs will become a useful tool for the development of effective photovoltaic and optoelectronic devices based on hybrid perovskites.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemmater.0c02290.

Details of the Voronoi−Dirichlet approach, smooth overlap of atomic positions (SOAP) kernel, organic cation penetration in (100) single-layer perovskites from the database, and the band gap values for (100) single-layer perovskites from the database calculated by the ML model (PDF)

AUTHOR INFORMATION

Corresponding Author

Alexey B. Tarasov − *Laboratory of New Materials for Solar Energetics, Faculty of Materials Science and Department of Chemistry, Lomonosov Moscow State University, Moscow 119991, Russia;* orcid.org/0000-0003-4277-1711; Email: alexey.bor.tarasov@yandex.ru

Authors

Ekaterina I. Marchenko − *Laboratory of New Materials for Solar Energetics, Faculty of Materials Science and Department of Geology, Lomonosov Moscow State University, Moscow 119991, Russia*

Sergey A. Fateev − *Laboratory of New Materials for Solar Energetics, Faculty of Materials Science, Lomonosov Moscow State University, Moscow 119991, Russia*

Andrey A. Petrov − *Laboratory of New Materials for Solar Energetics, Faculty of Materials Science, Lomonosov Moscow State University, Moscow 119991, Russia;* orcid.org/0000-0003-0368-8800

Vadim V. Korolev − *Department of Chemistry, Lomonosov Moscow State University, Moscow 119991, Russia; Science Data Software, LLC, Rockville, Maryland 20850, United States;* orcid.org/0000-0001-6117-5662

Artem Mitrofanov − *Department of Chemistry, Lomonosov Moscow State University, Moscow 119991, Russia; Science Data Software, LLC, Rockville, Maryland 20850, United States;* orcid.org/0000-0001-8891-6862

Andrey V. Petrov − *Institute of Chemistry, Saint-Petersburg State University, Saint-Petersburg 198504, Russia;* orcid.org/0000-0002-4650-4891

Eugene A. Goodilin − *Laboratory of New Materials for Solar Energetics, Faculty of Materials Science and Department of Chemistry, Lomonosov Moscow State University, Moscow 119991, Russia*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemmater.0c02290

Author Contributions

#E.I.M. and S.A.F. contributed equally to this work. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

The database is available free of charge at http://pdb.nmse-lab.ru. The interface for band gap predictions is available at https://eg.scidatasoft.com/. The algorithm for calculation of effective charges is available at http://mof.scidatasoft.com/api/ui/.

ACKNOWLEDGMENTS

This work was financially supported by a grant from the Russian Science Foundation, project number 19-73-30022. The research is carried out using the equipment of the shared research facilities of HPC Computing Resources at Lomonosov Moscow State University. The DFT calculations were carried out using computational resources provided by the Resource Center "Computer Center of SPbU". The authors thank Dr. Pavel Ivlev and Artem Ordinartsev for their assistance in data acquisition.

REFERENCES

(1) Mao, L.; Stoumpos, C. C.; Kanatzidis, M. G. Two-Dimensional Hybrid Halide Perovskites: Principles and Promises. *J. Am. Chem. Soc.* **2019**, *141* (3), 1171−1190.

(2) Katan, C.; Mercier, N.; Even, J. Quantum and Dielectric Confinement Effects in Lower-Dimensional Hybrid Perovskite Semiconductors. *Chem. Rev.* **2019**, *119* (5), 3140−3192.

(3) Wang, J.; Dong, J.; Lu, F.; Sun, C.; Zhang, Q.; Wang, N. Two-Dimensional Lead-Free Halide Perovskite Materials and Devices. *J. Mater. Chem. A* **2019**, *7* (41), 23563−23576.

(4) Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R. H.; Nelson, L. J.; Hart, G. L. W.; Sanvito, S.; Buongiorno-Nardelli, M.; et al. AFLOWLIB.ORG: A Distributed Materials Properties

Repository from High-Throughput Ab Initio Calculations. *Comput. Mater. Sci.* **2012**, *58*, 227−235.
(5) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1* (1), 011002.
(6) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65* (11), 1501−1509.
(7) Borysov, S. S.; Geilhufe, R. M.; Balatsky, A. V. Organic Materials Database: An Open-Access Online Database for Data Mining. *PLoS One* **2017**, *12* (2), e0171501.
(8) Haastrup, S.; Strange, M.; Pandey, M.; Deilmann, T.; Schmidt, P. S.; Hinsche, N. F.; Gjerding, M. N.; Torelli, D.; Larsen, P. M.; Riis-Jensen, A. C.; et al. The Computational 2D Materials Database: High-Throughput Modeling and Discovery of Atomically Thin Crystals. *2D Mater.* **2018**, *5* (4), 042002.
(9) Schmidt, J.; Marques, M. R. G.; Botti, S.; Marques, M. A. L. Recent Advances and Applications of Machine Learning in Solid-State Materials Science. *npj Comput. Mater.* **2019**, *5* (1), 83.
(10) Himanen, L.; Geurts, A.; Foster, A. S.; Rinke, P. Data-Driven Materials Science: Status, Challenges, and Perspectives. *Adv. Sci.* **2019**, *6* (21), 1900808.
(11) Carrete, J.; Li, W.; Mingo, N.; Wang, S.; Curtarolo, S. Finding Unprecedentedly Low-Thermal-Conductivity Half-Heusler Semiconductors via High-Throughput Materials Modeling. *Phys. Rev. X* **2014**, *4* (1), 1−9.
(12) Sparks, T. D.; Gaultois, M. W.; Oliynyk, A.; Brgoch, J.; Meredig, B. Data Mining Our Way to the next Generation of Thermoelectrics. *Scr. Mater.* **2016**, *111* (May), 10−15.
(13) Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D. G.; Wu, T.; et al. Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach. *Nat. Mater.* **2016**, *15* (10), 1120−1127.
(14) Li, Z.; Wang, S.; Chin, W. S.; Achenie, L. E.; Xin, H. High-Throughput Screening of Bimetallic Catalysts Enabled by Machine Learning. *J. Mater. Chem. A* **2017**, *5* (46), 24131−24138.
(15) Mansouri Tehrani, A.; Oliynyk, A. O.; Parry, M.; Rizvi, Z.; Couper, S.; Lin, F.; Miyagi, L.; Sparks, T. D.; Brgoch, J. Machine Learning Directed Search for Ultraincompressible, Superhard Materials. *J. Am. Chem. Soc.* **2018**, *140* (31), 9844−9853.
(16) Choudhary, K.; Bercx, M.; Jiang, J.; Pachter, R.; Lamoen, D.; Tavazza, F. Accelerated Discovery of Efficient Solar Cell Materials Using Quantum and Machine-Learning Methods. *Chem. Mater.* **2019**, *31* (15), 5900−5908.
(17) Schmidt, J.; Shi, J.; Borlido, P.; Chen, L.; Botti, S.; Marques, M. A. L. Predicting the Thermodynamic Stability of Solids Combining Density Functional Theory and Machine Learning. *Chem. Mater.* **2017**, *29* (12), 5090−5103.
(18) Cai, Y.; Xie, W.; Teng, Y. T.; Harikesh, P. C.; Ghosh, B.; Huck, P.; Persson, K. A.; Mathews, N.; Mhaisalkar, S. G.; Sherburne, M.; et al. High-Throughput Computational Study of Halide Double Perovskite Inorganic Compounds. *Chem. Mater.* **2019**, *31* (15), 5392−5401.
(19) Rajan, A. C.; Mishra, A.; Satsangi, S.; Vaish, R.; Mizuseki, H.; Lee, K. R.; Singh, A. K. Machine-Learning-Assisted Accurate Band Gap Predictions of Functionalized Mxene. *Chem. Mater.* **2018**, *30* (12), 4031−4038.
(20) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72* (2), 171−179.
(21) Momma, K.; Izumi, F. VESTA 3 for Three-Dimensional Visualization of Crystal, Volumetric and Morphology Data. *J. Appl. Crystallogr.* **2011**, *44* (6), 1272−1276.
(22) Blatov, V. A.; Shevchenko, A. P.; Proserpio, D. M. Applied Topological Analysis of Crystal Structures with the Program Package Topospro. *Cryst. Growth Des.* **2014**, *14* (7), 3576−3586.
(23) Alonso, J. A.; Martínez-Lope, M. J.; Casais, M. T.; Fernández-Díaz, M. T. Evolution of the Jahn-Teller Distortion of MnO6 Octahedra in RMnO3 Perovskites (R = Pr, Nd, Dy, Tb, Ho, Er, Y): A Neutron Diffraction Study. *Inorg. Chem.* **2000**, *39* (5), 917−923.
(24) Du, K. Z.; Tu, Q.; Zhang, X.; Han, Q.; Liu, J.; Zauscher, S.; Mitzi, D. B. Two-Dimensional Lead(II) Halide-Based Hybrid Perovskites Templated by Acene Alkylamines: Crystal Structures, Optical Properties, and Piezoelectricity. *Inorg. Chem.* **2017**, *56* (15), 9291−9302.
(25) Delley, B. From Molecules to Solids with the DMol3 Approach. *J. Chem. Phys.* **2000**, *113* (18), 7756−7764.
(26) Delley, B. An All-electron Numerical Method for Solving the Local Density Functional for Polyatomic Molecules. *J. Chem. Phys.* **1990**, *92* (1), 508−517.
(27) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal-Organic Frameworks: A Tool to Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26* (21), 6185−6192.
(28) Nazarian, D.; Camp, J. S.; Chung, Y. G.; Snurr, R. Q.; Sholl, D. S. Large-Scale Refinement of Metal-Organic Framework Structures Using Density Functional Theory. *Chem. Mater.* **2017**, *29* (6), 2521−2528.
(29) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; et al. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal−Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64* (12), 5985−5998.
(30) Nazarian, D.; Camp, J. S.; Sholl, D. S. A Comprehensive Set of High-Quality Point Charges for Simulations of Metal-Organic Frameworks. *Chem. Mater.* **2016**, *28* (3), 785−793.
(31) Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N. E. R.; Bajaj, S.; Wang, Q.; Montoya, J.; Chen, J.; Bystrom, K.; Dylla, M.; et al. Matminer: An Open Source Toolkit for Materials Data Mining. *Comput. Mater. Sci.* **2018**, *152*, 60−69.
(32) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114* (10), 1−5.
(33) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine Learning in Materials Informatics: Recent Applications and Prospects. *npj Comput. Mater.* **2017**, *3* (1), 54.
(34) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87* (18), 1−16.
(35) Korolev, V.; Mitrofanov, A.; Marchenko, E.; Eremin, N.; Tkachenko, V.; Kalmykov, S. Transferable and Extensible Machine Learning Derived Atomic Charges for Modeling Metal-Organic Frameworks. Preprint. https://arxiv.org/abs/1905.12098
(36) Chen, T.; Guestrin, C. *XGBoost*. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*; ACM Press: New York, 2016; pp 785−794. DOI: 10.1145/2939672.2939785.
(37) Blatov, V. A. Voronoi−Dirichlet Polyhedra in Crystal Chemistry: Theory and Applications. *Crystallogr. Rev.* **2004**, *10* (4), 249−318.
(38) Olsthoorn, B.; Geilhufe, R. M.; Borysov, S. S.; Balatsky, A. V. Band Gap Prediction for Large Organic Crystal Structures with Machine Learning. *Adv. Quantum Technol.* **2019**, *2* (7−8), 1900023.

https://dx.doi.org/10.1021/acs.chemmater.0c02290
Chem. Mater. 2020, 32, 7383−7388