Data-scarce synthesis-by-design of ferroelectric Dion–Jacobson 2D hybrid organic–inorganic perovskites

Lulu Wang[*1] Weixin He[*1] Shurogn Yuan[2,3] Kian Ping Loh[1,3], Pengfei Ou[1,4]

[*]Equal contribution [1][Department of Chemistry, National University of Singapore, 3 Science Drive 3, Singapore 117543] [2][Institute of High Energy Physics, Chinese Academy of Sciences, 19 B Yuquan Rd, Beijing 100049, China] [3][Department of Applied Physics, The Hong Kong Polytechnic University, 11 Yuk Choi Road, Hung Hom, Kowloon, Hong Kong SAR, China] [4][NUS Artificial Intelligence Institute, National University of Singapore, 11 Research Link, Singapore 119391]

Correspondence to: [Kian Ping Loh] chmlohkp@nus.edu.sg; [Pengfei Ou] pengfei.ou@nus.edu.sg

1. Introduction

Ferroelectric Dion–Jacobson (DJ) two-dimensional hybrid organic–inorganic perovskites (2D HOIPs) are promising for low-voltage ferroelectric electronics and polar spin–orbit–coupled functionalities, because their polarity is chemically programmable through organic spacers and strong organic–inorganic coupling. Yet synthesis is fundamentally bottlenecked by two compounding scarcities: firstly, only 11 DJ-phase 2D HOIPs have been experimentally confirmed as ferroelectric, limiting reliable structure–property knowledge; secondly, across the vast divalent-spacer space, most candidates fail to crystallize into the DJ phase, so "designing ferroelectricity" without accounting for phase formability is rarely actionable. We address this by coupling an interpretable ferroelectric structure map (learned under extremely small-data conditions) with an explicit DJ-formability screen, enabling spacer selection that is both ferroelectric-consistent and synthetically realizable.

2. Workflow overview

Our synthesis-by-design workflow integrates three decisions: (i) Define a ferroelectric-favored regime from scarce exemplars using symbolic regression over physically motivated descriptors; (ii) Expand and filter divalent spacers in a chemistry-aware manner to obtain a tractable, synthesizable candidate library; (iii) Screen DJ phase formability explicitly through MLIP-accelerated structural exploration and inorganic-slab connectivity checks, then prioritize targets that are both DJ-robust and ferroelectric-like, followed by design-guided synthesis and validation.

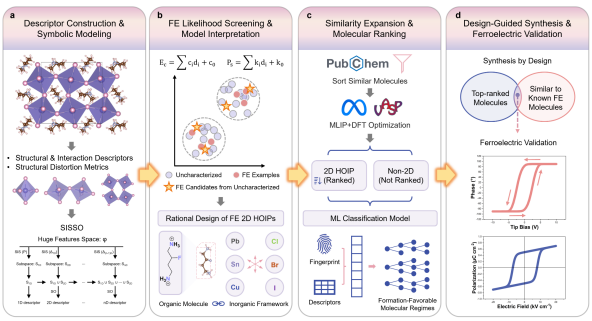

Fig. 1: Synthesis-by-design workflow. SISSO ferroelectric mapping → PubChem spacer expansion and filtering → MLIP DJ-formability screening (connectivity) → ferroelectric-like selection → DFT refinement → synthesis and ferroelectric validation of (s-FBDA)PbI_4.

3. Results and discussion

3.1 Related work

This work builds on and differs from two active directions. First, recent random-structure-search and crystal-structure-prediction efforts for 2D HOIPs generate candidate geometries but do not, by themselves, convert sparse ferroelectric knowledge into spacer-level synthetic decisions. Second, universal MLIPs provide scalable relaxation and screening. Still, without an interpretable ferroelectric map and explicit DJ-connectivity checks, they can prioritize low-energy structures that are not experimentally reachable DJ phases or not ferroelectric-consistent. Our key contribution is to couple interpretable, rule-bounded ferroelectric regime learning under extreme data scarcity with phase-formability screening that directly targets DJ slab connectivity, closing the loop with experimental validation in a newly synthesized ferroelectric DJ HOIP.

3.2 Interpretable ferroelectric mapping under extreme data scarcity

Starting from the 11 confirmed ferroelectric DJ-phase 2D HOIPs, we quantify each structure with 37 descriptors that encode organic geometry, inorganic framework parameters, organic–inorganic interfacial interactions, and octahedral/lattice distortion metrics. We then use SISSO-based symbolic regression to derive closed-form relations for two experimentally relevant ferroelectric figures of merit: coercive field (E_c) and spontaneous polarization (P_s). The symbolic models show strong interpolation accuracy on the known set (RMSE 0.02 for E_c and 0.03 for P_s), and we use them as rule-bounded, semi-quantitative criteria rather than absolute predictors, enforcing polar symmetry, descriptor-range consistency with the 11 references, and SISSO-output consistency to define a ferroelectric-favored regime for screening new DJ structures.

3.3 Similarity-guided spacer expansion and DJ-formability ranking

To move beyond the narrow spacer chemistry represented in the ferroelectric DJ literature, we perform a PubChem similarity expansion (≥90% Tanimoto similarity to curated DJ spacer motifs), yielding 83,924 molecules. A stepwise filtering protocol—redundancy removal, elemental screening, geometry optimization, and descriptor-based size/shape gating guided by SISSO trends—produces 1,012 plausible diammonium spacers for downstream assessment. Because the SISSO trends consistently favor Pb–I hosts, we fix the inorganic slab to Pb–I for all structural formability screening to provide a consistent, permissive lattice environment.

We then explicitly assess DJ-phase formability with MLIP-accelerated structural exploration. For each of the 1,012 spacers, we generate 200 randomized DJ-like initial configurations and relax them using a universal MLIP (FairChem). The 10 lowest-energy relaxed structures per spacer are evaluated by an inorganic-slab connectivity criterion (corner-sharing Pb–I connectivity and DJ-bridging linkage). Spacers are ranked by the number of their top-10 structures that preserve DJ connectivity. In total, 338/1,012 (~33%) spacers yield at least one connectivity-preserving structure and are classified as DJ-formable; within this set, 12 spacers preserve DJ connectivity in 10/10 low-energy structures (high DJ-formability). One of these high-confidence spacers corresponds to a Pb–I 2D HOIP that has already been reported experimentally, providing an internal validation of the screening strategy. (Overall compute scale: 202,400 MLIP relaxations.)

3.4 Learning what controls DJ-formability

To understand why some spacers are DJ-formable while others are not, we train a Random Forest classifier using each molecule's PubChem-881 fingerprint together with the physically motivated descriptors. To minimize wasted synthesis attempts, we select the decision threshold by prioritizing precision; the resulting model achieves strong performance (accuracy 0.94, PR-AUC 0.80, ROC-AUC 0.91). SHAP analysis identifies three synthesis-relevant trends: elongated yet slender spacers (higher length, lower width/height) increase formability; excessive bulk (high van der Waals volume/molecular weight) reduces formability; and higher polarity (TPSA) stabilizes the organic–inorganic interface and improves formability. These trends translate into simple, actionable spacer-selection heuristics for DJ-phase synthesis campaigns.

3.5 Selecting ferroelectric DJ synthesis targets and experimental validation

Among the DJ-formable set, we prioritize spacers that lie in the same chemical–structural neighborhood as known ferroelectric spacers. Embedding DJ-formable spacers into a joint t-SNE space (PubChem-881 + descriptors), followed by descriptor-distribution checks and hierarchical clustering, yields two top synthesis targets at the intersection of "phase-robust" and "ferroelectric-like": CID 147066 and CID 151574811. After DFT re-optimization of their top-10 MLIP-ranked DJ-like structures and comparison of average formation energies, we select CID 147066 (s-2-fluorobutane-1,4-diamine; s-FBDA) for experimental validation.

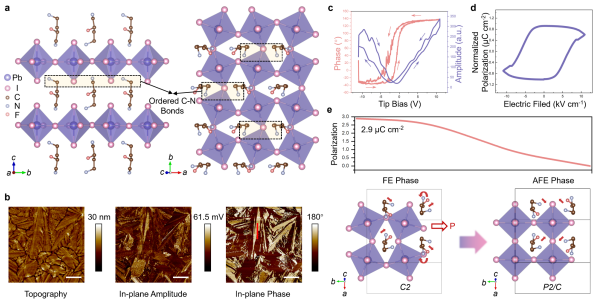

Fig. 2: Experimental/theoretical validation of (s-FBDA)PbI$_4$. Polar crystal structure; SHG; PFM domain switching; macroscopic *P*–*E* loop (E_c, P_s); Berry-phase polarization and Rashba splitting.

Acknowledgments

P.O. acknowledges support from the National University of Singapore through the Presidential Young Professorship (PYP) Start-Up Grant (A-0010024-00-00) and the White Space Fund (A-0010024-01-00). K.P.L. acknowledges funding support from Hong Kong's Research Grants Council GRF project ID P0046892 (funding reference 1530223) and HK PolyU grant no. P0043087, as well as Singapore's Ministry of Education Tier 1 Grant "A new class of 2D metal-free dielectric" (A80026710000).

References

[1] Karimitari, N. et al. Accurate Crystal Structure Prediction of New 2D Hybrid Organic–Inorganic Perovskites. J. Am. Chem. Soc. 146, 27392–27404 (2024).

[2] Wood, B. M. et al. UMA: A Family of Universal Models for Atoms (2025).

[3] National Center for Biotechnology Information. PubChem Subgraph Fingerprint (2009).

[4] Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001).

[5] Lundberg, S. M.; Lee, S.-I. A unified approach to interpreting model predictions. NeurIPS 30 (2017).

[6] King-Smith, R. D.; Vanderbilt, D. Theory of polarization of crystalline solids. Phys. Rev. B 47, 1651–1654 (1993).

[7] Manchon, A. et al. New perspectives for Rashba spin–orbit coupling. Nat. Mater. 14, 871–882 (2015).

[8] Ai, Y. et al. H/F substitution for advanced molecular ferroelectrics. Trends Chem. 3, 1088–1099 (2021).