Check for updates

Advanced Materials

ADVANCED MATERIALS
www.advmat.de

RESEARCH ARTICLE OPEN ACCESS

Data-Driven Discovery of Quaternary Ammonium Interlayers for Efficient and Thermally Stable Perovskite Solar Cells

Jongbeom Kim[1] | Yang Jeong Park[2,3,4] | Chaehoon Jeon[1] | Nahye Shin[1] | Jaewang Park[1] | Seungun Lee[1] | Jino Im[5,6] | Sungroh Yoon[3,6] | Sang Il Seok[1,6]

[1]Department of Energy and Chemical Engineering, Ulsan National Institute of Science and Technology (UNIST), 50 UNIST-gil, Ulsan, Ulsan, Republic of Korea | [2]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA | [3]Department of Electrical and Computer Engineering, Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, Republic of Korea | [4]Department of Materials Science and Engineering, Ulsan National Institute of Science and Technology (UNIST), 50 UNIST-gil, Ulsan44919, Republic of Korea | [5]Korea Research Institute of Chemical Technology, Deajeon, Republic of Korea | [6]UNIST InnoCORE AI-Space Solar Initiative, Ulsan National Institute of Science and Technology (UNIST), 50 UNIST-gil, Ulsan, Republic of Korea

Correspondence: Jino Im (imjino@krict.re.kr) | Sungroh Yoon (sryoon@snu.ac.kr) | Sang Il Seok (seoksi@unist.ac.kr)

Received: 13 November 2025 | **Revised:** 26 February 2026 | **Accepted:** 9 April 2026

Keywords: active learning | alkylammonium halide | Bayesian optimization | interlayer | machine learning | perovskite solar cell

ABSTRACT

Interfacial engineering is essential for improving charge extraction and suppressing non-radiative recombination in perovskite solar cells (PSCs). Although numerous organic interfacial materials (IMs) have been explored, the vast molecular design space renders purely experimental screening inefficient. Here, we report on a machine learning-based framework that rapidly screens IMs using an in-house database. Six physicochemical descriptors capturing perovskite–molecule interactions were selected to train a Gaussian Process Regression model embedded in a Bayesian Optimization active learning loop. Post hoc interpretability revealed that thermally robust, higher-order alkylammonium cations are particularly beneficial for PSC interfaces. The model nominated 15 promising, previously untested IMs; one of them, tetra-n-hexyl-ammonium bromide, was experimentally incorporated into PSCs. Devices treated with this IM delivered a power-conversion efficiency of 25.31% under AM 1.5 G illumination and, remarkably, retained about 81.6% of the initial efficiency after 1508 h at 85°C, demonstrating enhanced thermal stability. These results demonstrate how an interpretable, data-driven strategy can accelerate the rational discovery of IMs, enabling the development of PSCs that combine record-level efficiency with outstanding long-term stability.

1 | Introduction

Perovskite solar cells (PSCs) have emerged as one of the most promising next-generation photovoltaic technologies [1], rapidly approaching power conversion efficiencies (PCEs) rivaling or surpassing those of established semiconductors [2–4]. However, their long-term stability under operational conditions remains a critical barrier to commercialization. Among the various degradation pathways, interfacial instability at the junction between the perovskite absorber and charge-transport layers is particularly detrimental. Poor energy-level alignment and defect-rich surfaces lead to non-radiative recombination, accelerated phase

Jongbeom Kim and Yang Jeong Park contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2026 The Author(s). *Advanced Materials* published by Wiley-VCH GmbH

degradation, and ultimately, rapid performance loss. Consequently, interfacial engineering to overcome such instabilities is central to achieving both high efficiency and durable PSCs [4–7].

Careful passivation of interfacial defects, such as vacancies, interstitials, and under-coordinated ions, combined with precise energy-level alignment, is known to suppress non-radiative recombination and promote efficient carrier extraction [4, 8, 9]. In the conventional n-i-p architecture, an ideal interlayer (IL) positioned between the perovskite absorber and the hole-transport layer (HTL) must satisfy three conditions: (i) its valence-band maximum (E_{VBM}) should provide a favorable positive offset (φ_{bi}) for hole extraction, (ii) its conduction-band minimum (E_{CBM}) must effectively block minority electrons, and (iii) its Fermi level (E_F) should induce beneficial band bending at the heterojunction. Bulky organic cations, in particular alkylammonium halides, fulfill these criteria by reacting with residual PbI_2 to form low-dimensional perovskite phases that combine tunable band gaps with enhanced defect passivation [4, 10, 11]. However, the design space for molecular ILs remains vast, encompassing countless possible chain lengths, functional substituents, and halide species [12–14]. Conventional experimental exploration, mainly driven by chemical intuition, has probed only a narrow fraction of this chemical space [15, 16], leaving many potentially superior candidates undiscovered. Recent evidence suggests that quaternary ammonium salts (R_4N^+), which are structurally robust and resistant to deprotonation, hold great potential to simultaneously improve efficiency and stability [17–20].

However, systematic identification of optimal candidates has been hindered by the lack of efficient, predictive design frameworks. Trial-and-error methods are too slow, while conventional machine learning approaches often demand large, heterogeneous datasets that are difficult to assemble for perovskite interfaces. To address this bottleneck, we adopt a data-driven active learning framework that integrates machine learning with targeted experiments [16, 21, 22]. By iteratively training a Gaussian process regression model on an in-house dataset collected under consistent device architectures, we enable reliable prediction of performance trends and rational exploration of chemical space. Importantly, this approach not only nominates promising new IL molecules but also provides post-hoc interpretability, revealing which molecular features govern stability and efficiency.

Here, using this strategy, we discover higher-order alkylammonium halides as a new guiding principle for interfacial stabilization. In particular, tetra-n-hexylammonium bromide (THABr) achieves a PCE of 25.31% and retains more than 81.6% of its initial efficiency after 1500 h at 85°C, underscoring the dual advantages of high efficiency and exceptional thermal robustness. These results demonstrate the power of a problem-driven, data-guided approach in accelerating the discovery of interfacial materials, thereby advancing PSCs toward practical deployment. While the present implementation focuses on commercially accessible ammonium salts to enable rapid experimental validation, the descriptor framework itself is not intrinsically limited to commercial compounds and can, in principle, be extended to a broader synthetically accessible chemical space.

2 | Result and Discussion

2.1 | Overview of the Search Process for Interfacial Materials

Our workflow for discovering interfacial materials (IMs) proceeds through three sequential stages.

2.1.1 | Stage 1 – Data Preprocessing

In the initial step, each molecule is encoded as a numerical vector that captures the molecular characteristics most relevant to IM performance. These characteristics—such as band alignment and defect passivation (Figure 1a)—are intrinsically determined by molecular structure and electronic properties [23]. Previous studies compared multiple feature sets and identified an optimal descriptor combination for predicting PSC efficiency [24]. Building on this foundation, the present study adopts six molecular descriptors: dipole moment (μ), molecular size (L_M), topological polar surface area (TPSA), and molecular shape indices derived from Kier's descriptors (κ_1, κ_2, κ_3) (Figure 1b) [25–27]. These descriptors were selected due to their well-established correlation with critical passivation mechanisms at perovskite interfaces, including facilitating band alignment, suppressing carrier recombination, and enhancing device stability [23]. By representing commercially available alkylammonium halides with these descriptors, 6D vectors were generated. To further optimize representation, principal component analysis (PCA) was applied, thereby reducing dimensionality and yielding a latent feature space. This reduction enhanced the generalizability of the surrogate model used in Bayesian Optimization (BO) [28], ultimately accelerating convergence toward promising IM candidates [29].

2.1.2 | Stage 2 – Material Recommendation and Experimental Validation

In the second stage (Figure 1c), material recommendation and validation were initiated by assembling an initial training dataset. This dataset was constructed by fabricating PSCs with 28 representative IMs selected through a descriptor-guided strategy rather than random sampling. Preliminary analysis of the full commercially available ammonium pool revealed that dipole moment (μ) and the first Kier shape index (κ_1) spanned the largest variance and exhibited strong relevance to interfacial energetics. Accordingly, we first selected 13 distinct cation backbones that collectively spanned a broad range of μ and κ_1 value and reflected structural motifs commonly reported in the literature. By further incorporating multiple halide species (I^-, Br^-, and Cl^-), we expanded this set to 28 materials, thereby ensuring chemically diverse yet descriptor-balanced coverage of the search space. The dataset size was determined by balancing chemical coverage with experimental precision. Rather than maximizing dataset size, we prioritized consistency and low-noise measurements, with all devices fabricated under identical architecture and strictly controlled conditions within a single laboratory. This strategy minimizes aleatoric variability and provides a reliable foundation

15214095, 0, Downloaded from https://advanced.onlinelibrary.wiley.com/doi/10.1002/adma.202522554 by Ulsan National Institute Of, Wiley Online Library on [23/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

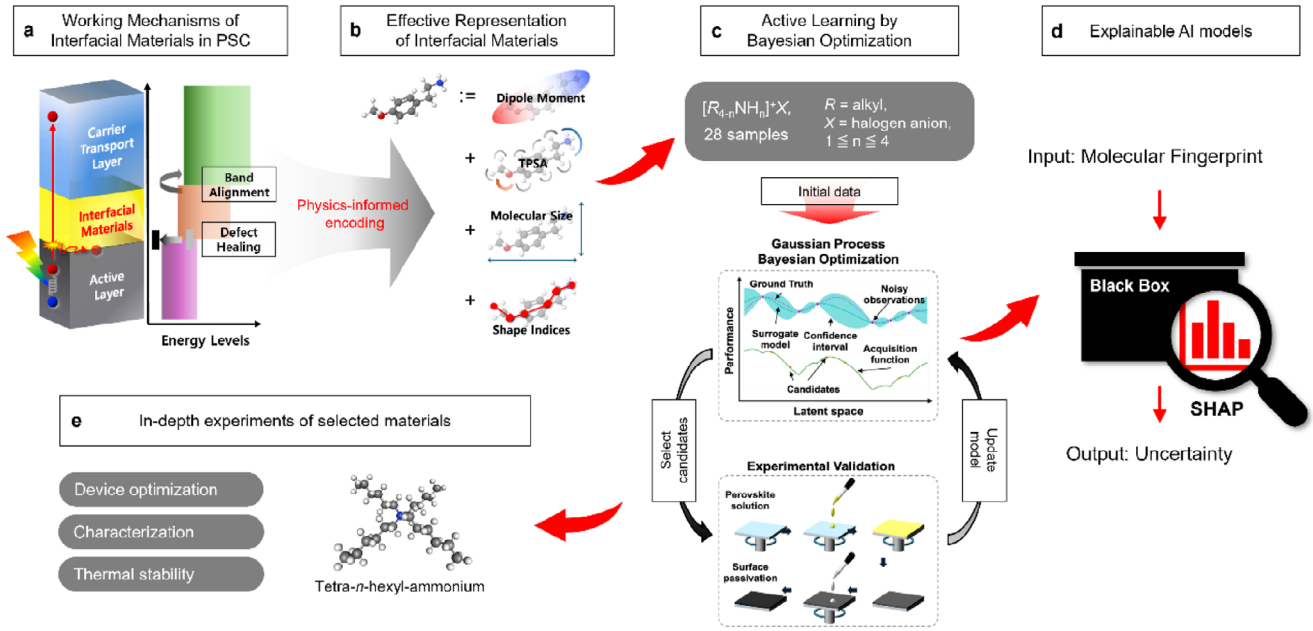

FIGURE 1 | Schematic diagram of the proposed active learning workflow. (a) A schematic picture of the working mechanisms of the interfacial materials of PSC. (b) An effective representation of interfacial materials using the dipole moment, TPSA, the molecular size, and Kier's shape indices. (c) A procedure of active learning, consisting of the construction of initial data following the representation, the Gaussian Process Bayesian Optimization (GPBO), and experimental validations. (d) Constructing an explainable AI model for the uncertainty evaluated from GPBO, by utilizing SHapley Additive exPlanations (SHAP). (e) Performing in-depth experiments for device optimization, characterization, and thermal stability of a selected material.

for surrogate modeling in a small-data active learning framework. A Gaussian Process Regression (GPR) model was trained as the surrogate model for the first BO round, using ammonium cation structural features to predict PCE without explicitly incorporating halide species. Accordingly, performance differences arising from distinct halogen anions were validated experimentally whenever possible. To account for unavoidable experimental variability inherent to perovskite solar cell fabrication and measurement, we incorporated a white-noise kernel term in the GPR surrogate model. This prevents minor device-to-device fluctuations from being overfitted as a meaningful signal and enables more realistic uncertainty calibration during BO. In each active learning (AL) cycle, new IM candidates were selected via BO using the upper confidence bound (UCB) acquisition function, after which experimental validation was performed under standardized PSC fabrication protocols to ensure comparability. The resulting validated data were integrated into the GPR model, progressively refining predictions with each cycle. After three AL iterations, several IMs delivering notably high PCE were identified, thereby confirming the effectiveness of this methodology. A more detailed, stepwise description of this process is provided in the following sections.

2.1.3 | Stage 3 – Post-Hoc Computational and Experimental Analysis

The final stage involved computational (Figure 1d) and experimental (Figure 1e) analysis to yield deeper insight into the AL process and its influence on device performance. Recognizing the inherent interpretability limitations of ML models, we trained independent models to predict "uncertainty" metrics derived from BO, using molecular fingerprints as input. This approach revealed the specific molecular features most responsible for IM selection at each iteration of the AL cycle. These insights guided subsequent device optimization employing the identified IMs. Experimental characterization demonstrated that PSCs fabricated with these optimized passivation molecules achieved efficiencies exceeding 25% while also exhibiting unprecedented thermal stability. Detailed explanations and stepwise analyses of these outcomes are provided in the following sections.

2.1.4 | Initial Round of Active Learning

In the initial round of AL, preliminary experiments were conducted to configure the inputs for training a GPR model, which serves as a surrogate model for BO. The Gaussian process Bayesian optimization (GPBO) identifies new candidates based on uncertainty. However, when public data from literature are used, aleatoric uncertainty—stemming from variations in experimental conditions across laboratories—can complicate the analysis. To mitigate this, we generated the entire dataset through experiments conducted in the same laboratory under consistent conditions, including an identical device structure. This approach ensured that variability was limited to the IMs themselves, minimizing the influence of external factors and enabling more reliable analysis.

For the initial AL inputs, we fabricated devices and measured the performance of PSCs incorporating 28 selected IMs (Figure 2). Among these, 13 alkylammoniums were chosen because of their broad structural and property variations [17, 30, 31]. Figure 2a illustrates the diversity of molecular structures among the selected alkylammoniums, which exhibit wide variations in dipole moments, molecular size, and molecular shape indices. The complete list of names

15214095, 0, Downloaded from https://advanced.onlinelibrary.wiley.com/doi/10.1002/adma.202522554 by Ulsan National Institute Of, Wiley Online Library on [23/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

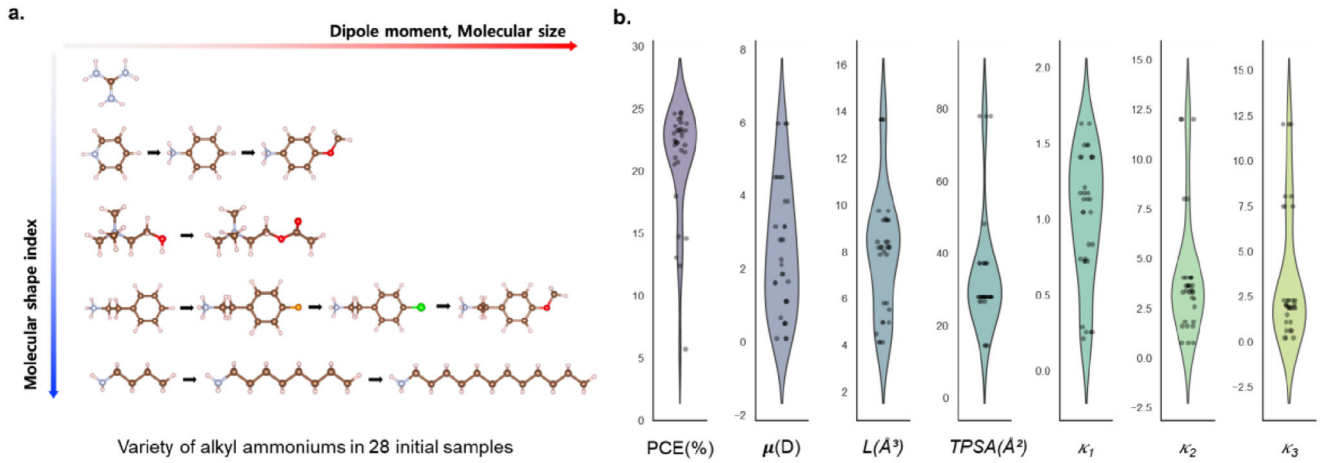

FIGURE 2 | Variety of alkylammoniums in initial samples. (a) Structural variations of alkylammoniums used as interfacial materials (IMs) in 28 initial data, focusing on dipole moment, molecular size, and molecular shape indices. (b) Statistical distribution of PCE and descriptors for the initial data. PCE was measured under AM 1.5G, including UV without AR film in 8 devices: conventional PSCs; FTO/SnO_2/Perovskite/IM/Spiro-OMeTAD/Au. (center line, median; small square, mean; box, 25%–75% standard deviation; whiskers, outliers).

and abbreviations is provided in Table S1. To establish the initial dataset, all materials other than IMs, as well as the fabrication procedures, were fixed within a conventional n-i-p device structure, specifically: *Glass/fluorine-doped tin oxide (FTO)/SnO_2/Perovskite/IM/2,2',7,7'-tetrakis[N,N-di(4-methoxyphenyl)amino]-9,9'-spirobifluorene (Spiro-OMeTAD)/Au*.

During the fabrication of PSCs with the 28 IMs, 4-methoxyphenethylammonium iodide (M-PEAI) was used as the control sample to evaluate the reproducibility of the preliminary experiments [32]. The control device consistently delivered stable efficiency across fabrication batches, thereby providing a reliable reference for comparing the performance of the 28 IMs in Figures S1 and S2. Notably, the initial PSC samples exhibited a broad PCE range of 5% to 24%, underscoring the profound impact of IM selection on PSC performance. In addition to variations in PCE, the values of all descriptors also showed substantial diversity across the samples, thereby facilitating effective learning in the predictive models (Figure 2b). We note that the present framework operates in a small-data regime, with only 28 experimentally measured IMs in the initial training set. To ensure robustness and avoid overfitting, we adopted physically grounded descriptors, applied PCA-based dimensionality reduction for numerical stability, and incorporated a white-noise kernel term in the GPR surrogate to account for unavoidable experimental variability. Each active learning iteration was followed by experimental validation under strictly standardized fabrication conditions.

2.1.5 | Experimental Validations of the Workflow

In subsequent rounds of AL, we iteratively selected new alkylammonium cation candidates via BO and experimentally validated their solar cell performance through the third round (Figure 3). Throughout the iterations, we identified the newly proposed materials and tracked the evolution of the BO search trend. Figure 3a shows a Uniform Manifold Approximation and Projection (UMAP) visualization of the search space across AL rounds for discovering new ammonium ions. Each panel represents the updated search space after each AL round, where newly recommended candidates (red) are experimentally evaluated. Gray nodes indicate unexplored candidate materials, while colored nodes represent experimentally tested materials, with color corresponding to the measured PCE. Figure 3b shows the experimental validations of solar cell performance for the suggested alkylammoniums in each round. In each active learning iteration, BO ranked the entire commercial ammonium pool using an identical UCB acquisition function, and the top 15 candidates were initially identified as the most informative molecules for experimental validation. In practice, however, experimental validation was constrained by reagent availability, shipping lead times, and fabrication throughput. Consequently, only the subset of top-ranked candidates that were immediately accessible was tested in each round, resulting in different validated batch sizes (6, 3, and 7 compounds across the three iterations). Importantly, the selection criterion and acquisition strategy remained unchanged throughout the process, and the surrogate model was updated only after incorporating experimentally validated results, thereby ensuring methodological consistency and preventing selection bias. Regarding PCE (Figure 3b), most of the newly suggested materials resulted in higher PCE values compared to the average PCE of the initial 28 alkylammoniums. This indicates that UCB can be a good choice for the acquisition function of BO. Similarly, for other metrics such as short-circuit current density (J_{SC}), open-circuit voltage (V_{OC}), and fill factor (FF), experimentally validated values of most of the selected alkylammoniums are higher than the average values of the initial set, as detailed in Figure S3. In addition, as the round progresses, the distribution tightens around higher PCE values, suggesting increased confidence in the GPR surrogate model's predictions.

The molecules recommended by BO differ significantly from the initial samples. As shown in Figure 3c and Table S2, BO predominantly suggested higher-order alkylammoniums (R_n-NH_{4-n}^+, $n \geq 2$) over primary alkylammoniums (R-NH_3^+) throughout the three iterations. Primary alkylammoniums have traditionally

15214095, 0, Downloaded from https://advanced.onlinelibrary.wiley.com/doi/10.1002/adma.202522554 by Ulsan National Institute Of, Wiley Online Library on [23/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

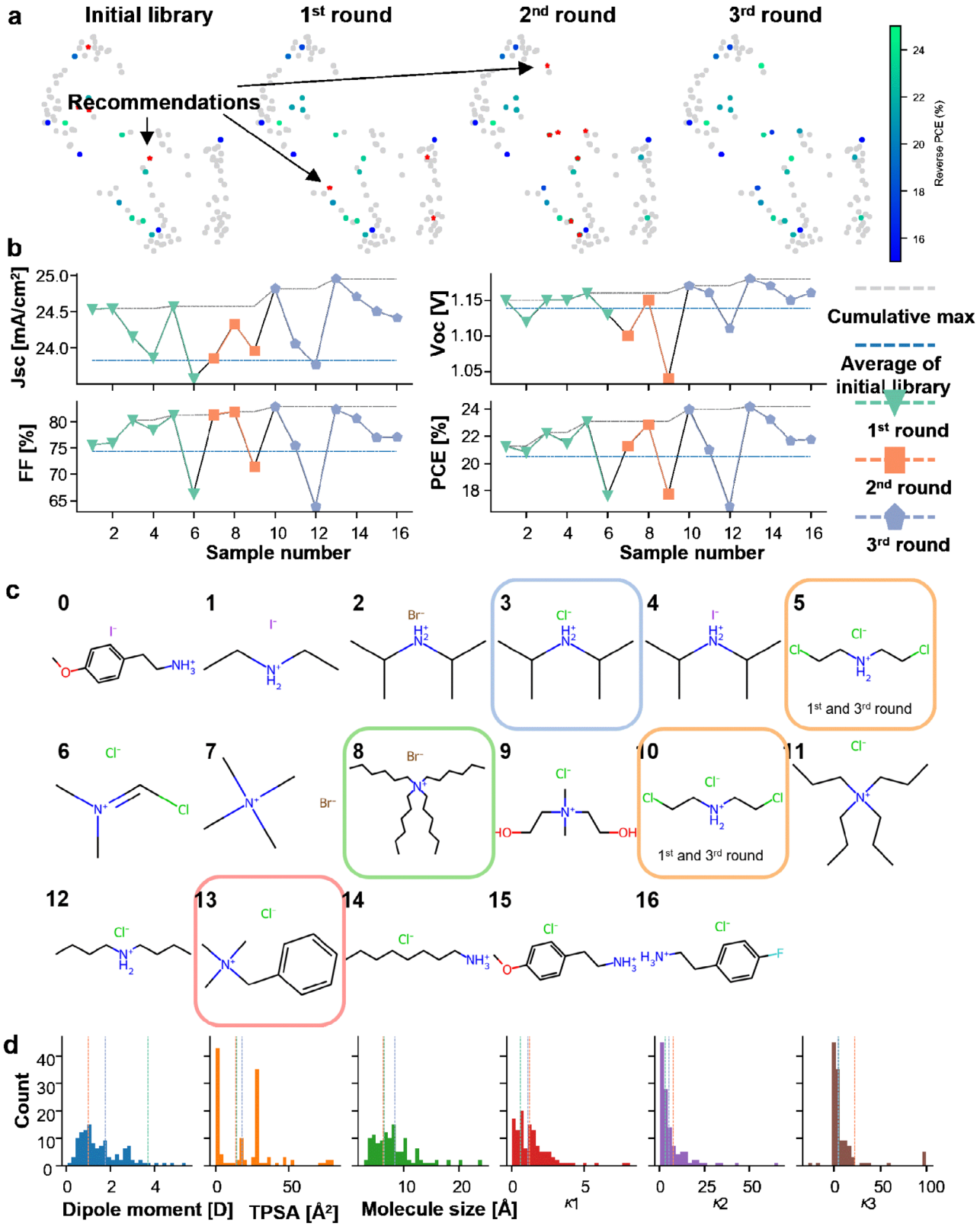

FIGURE 3 | Exploration of the chemical space using Gaussian process Bayesian optimization. (a) UMAP visualization of the explored chemical space across active learning rounds. Gray and red nodes indicate unexplored candidates and recommended candidates, respectively. Colored nodes represent experimentally tested materials, with color corresponding to the measured power conversion efficiency (PCE) according to the color bar. (b) Device performance of selected molecules during active learning rounds. (c) Molecular structures of recommended candidate materials in (b)-. (d) Molecular properties histogram of molecules to be explored by Bayesian optimization.

15214095, 0, Downloaded from https://advanced.onlinelibrary.wiley.com/doi/10.1002/adma.202522554 by Ulsan National Institute Of, Wiley Online Library on [23/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

been preferred for IMs due to their strong surface passivation capability, which enhances solar cell performance [33, 34]. Among the initial 28 samples, 18 samples were primary alkylammoniums. However, only three primary alkylammoniums were suggested during the AL iterations, with the remaining 12 candidates being higher-order alkylammoniums in Figure S4. Interestingly, three primary alkylammoniums were recommended in the third round, while only higher-order alkylammoniums were selected in the first and second rounds. This indicates that BO first explored materials with significantly different structures before returning to those similar to the initial samples in the final round. This pattern is more evident when examining the evolution of the descriptors for the selected alkylammoniums, as shown in Figure 3d. Among the six descriptors for IMs, the dipole moment and the κ_1 index vary significantly across iterations. The trends of both descriptors appear to be anti-correlated, and the κ_1 index, which shows greater variation than the initial samples, likely drives the BO selection process. The κ_1 index represents the compactness of the molecular shape, with lower values indicating linear, chain-like molecules and higher values indicating more spherical molecules [25–27]. As shown in Figure 3d, the κ_1 index increased steadily until the second round, then dropped in the third round, correlating well with changes in the selected molecules.

Through the iterations, we identified four candidate IMs along the Pareto frontier, as shown in Figure 3c. Two of these, bis(2-chloroethyl)ammonium chloride (BCACl) and benzyl trimethylammonium chloride (BTACl), have been previously reported as IMs [19, 20], while the other two candidates, di-iso-propylammonium chloride (Di-iso-PACl) and tetra-n-hexylammonium bromide (THABr), have not been previously reported as far as we are aware. All four candidates demonstrated strong solar cell performance, achieving over 22.5% PCE without anti-reflection film in experimental validations. We conducted further in-depth studies on these materials, which will be discussed in the final section. In the next section, we provide a detailed analysis of BO's recommendation process and its interpretation.

2.2 | Interpretation of the Suggestion Process

Interpreting the decisions of AI frameworks is crucial, as combining them with a researcher's intuition can lead to new discoveries in materials science. While our proposed method has successfully identified multiple materials, there are substantial challenges in model interpretability. Initially, we limited our GPR model training to data obtained from a single laboratory to reduce the substantial noise present in datasets collected from various sources in the literature. However, this small dataset lacks sufficient statistical power to ensure model interpretability. Additionally, due to the extremely limited number of data points, we opted to use a 2D PCA-compressed input derived from six descriptors selected in our preliminary work, rather than incorporating a large number of useful features. Furthermore, the linearly transformed features were not intuitively comprehensible. Therefore, there is a need for an alternative approach that can simplify features in an independent and meaningful manner, thereby enhancing the interpretability of model selection.

To achieve post-hoc interpretability, we propose training a separate 'interpretability model' to predict the uncertainty of the GPR model at each active learning round, as shown in Figure 4a. At each stage, we used the GPR model to compute uncertainty and UCB across the entire alkylammonium pool. We then introduced additional descriptors that are more intuitively understandable and potentially helpful for future IM design. These descriptors were manually adjusted to maximize independence based on a correlation matrix analysis (Figure S5). The interpretability model is trained using the optimized feature vector to explain the uncertainty and UCB across the ammonium ion pool. We also applied SHAP (SHapley Additive exPlanations), a cooperative game theory-based method, to assess feature importance in the interpretability model [35]. This analysis helps identify alkylammonium ions that are difficult for the model to predict (uncertainty) and those favored as next candidates (UCB). SHAP quantifies each descriptor's contribution to the model's predictions, effectively serving as a form of post-hoc feature disentanglement.

The developed interpretability model successfully predicted the uncertainty and UCB that are obtained from the GPR models of AL (Figure 4b; Figure S6). As rounds progress, more information is acquired, leading to a general reduction in uncertainty and a corresponding decrease in UCB. The predictive performance achieves an average R^2 of 0.8, consistently exceeding 0.6 across all rounds. This high predictive performance suggests that the interpretability model provides a stronger statistical justification for interpreting the GPR model's selection than does direct analysis of the GPR model itself. Then, we applied SHAP to each round's interpretability model and analyzed the contribution of each descriptor, as shown in Figure 4c (Figure S6 as well). Based on the analysis in the previous section, we explicitly examine the contribution of the alkyl substitution degree of N^+ (ASDN) and the number of rotatable bonds (RB) as descriptors. The former descriptor represents the number of alkyl groups attached to the four bonds of N^+, which is possibly highly correlated with the first Kier index, κ_1. We determined that the ASDN feature has the largest average absolute SHAP value, indicating that it makes the greatest contribution to UCB prediction. A positive SHAP value suggests an increase in the predicted UCB, implying that a higher number of alkyl groups correlates with a higher UCB. This is reasonable given that 18 out of the initial 28 ammonium ions contained only one alkyl group ($R\text{-}NH_3^+$). It is worth noting that this descriptor was not used in developing the GPR model in BO. The model exhibits curiosity about ammonium ions with multiple alkyl groups and explores them over three rounds. In the final round, the absolute SHAP value decreases, and its correlation with UCB diminishes, suggesting that this curiosity is largely resolved (Figure 4d; Figure S7). Next, the initial 28 species have a maximum of 10 RBs, whereas the full ammonium ion pool includes 22 ions with over 10 RBs. Consequently, ions with long chains tend to have high uncertainty, though they show relatively low correlation with UCB. Generally, a higher number of RBs increases molecular flexibility, which can facilitate film formation but may reduce the structural rigidity required for efficient charge transport. Given that UCB prediction is inherently a combination of PCE prediction and uncertainty prediction, this trade-off likely results in a weaker correlation between UCB and RB. The mixed SHAP distribution also reflects this trade-off. According to additional feature analyses in Figures S6–S9,

15214095, 0, Downloaded from https://advanced.onlinelibrary.wiley.com/doi/10.1002/adma.202522554 by Ulsan National Institute Of, Wiley Online Library on [23/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

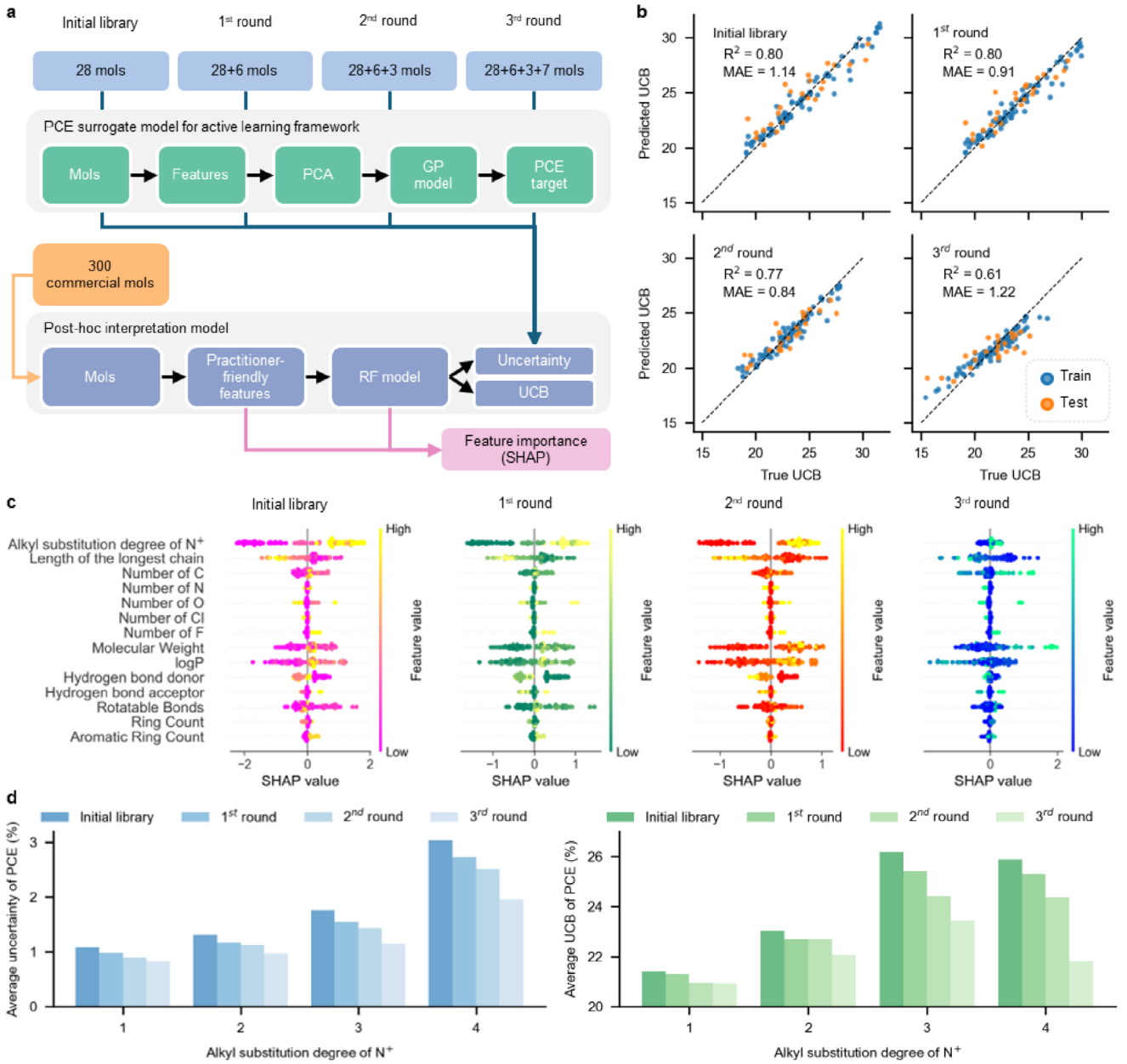

FIGURE 4 | Post-hoc interpretation of active learning framework. (a) A schematic diagram of the design process for post-hoc interpretation models. (b) Prediction performance of post-hoc interpretation models for each round. (c) SHAP importance according to active learning rounds. (d) Uncertainty and upper confidence bound according to the alkyl substitution degree of N^+.

even when uncertainty shows high variance, features are not affected if the variation in the UCB is not significant. Considering that such interpretations would have been difficult using only the GP surrogate model, we anticipate that the proposed method, leveraging the interpretability model, will substantially bridge the gap between GP surrogate model development and practitioners' understanding, as shown in Figures S7–S9. The consistent emergence of higher-order alkylammonium species across iterations, followed by experimentally verified improvements in efficiency and thermal stability, provides empirical evidence that the surrogate model captured physically meaningful structure–performance relationships rather than spurious correlations.

2.3 | In-Depth Analysis of the Suggested Materials

To explore how the ML-recommended materials contribute to performance enhancement, we categorized them by ASDN, which showed a strong correlation with the predicted performance of the IMs, and analyzed their corresponding device characteristics. Henceforth, primary ammonium (M-PEAI) will be denoted as Control, secondary ammonium (BCACl) as Reference, and quaternary ammonium (THABr) as Target for clarity. The J-V characteristics of the best-performing device, with anti-reflective coatings on the glass side, are shown in Figure 5a for reverse and forward bias sweeps. The J_{SC}, V_{OC}, and FF values calculated from the J-V curve at reverse scan mode were 25.94 mA

15214095, 0, Downloaded from https://advanced.onlinelibrary.wiley.com/doi/10.1002/adma.202522554 by Ulsan National Institute Of, Wiley Online Library on [23/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

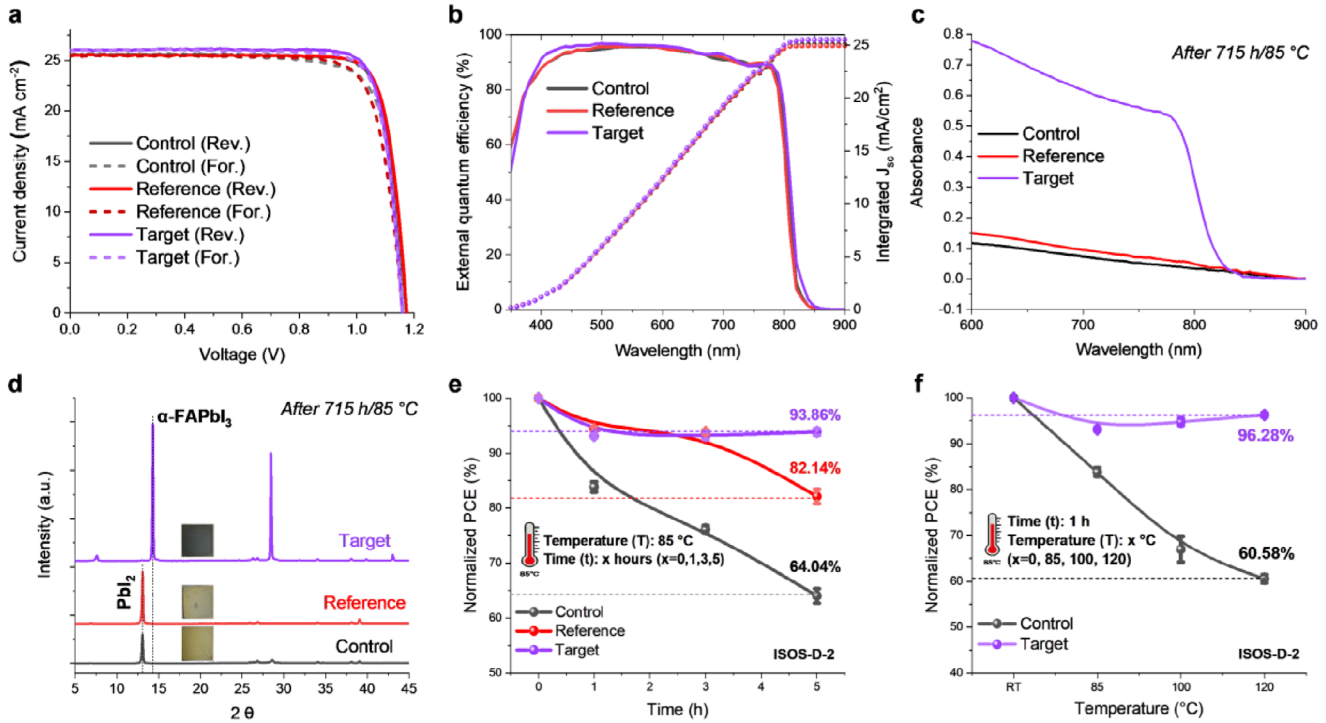

FIGURE 5 | Device characterization and performance. (a) J–V characteristics under AM 1.5G illumination. (b) External quantum efficiency (EQE) of the corresponding device. (c) UV–vis absorbance; (d) XRD patterns of thin films after thermal degradation at 85°C for 715 h. PCE distribution of devices after thermal degradation as a function of (e) time, and (f) temperature (error bars represent standard deviation across multiple independent devices; ISOS-D-2 conditions). Device structure: FTO/SnO_2/Perovskite/Interlayer/Spiro-OMeTAD/Au. The control, reference, and target exhibited initial average PCEs without anti-reflection (AR) film of 22.5%, 22.4%, and 22.5%, respectively. The data were obtained by taking 8 PSCs.

cm^{-2}, 1.16 V, and 84.12%, respectively. These factors correspond to a PCE of 25.31% under standard air mass (AM) 1.5G conditions, as detailed in Table S3. The IMs achieved PCEs exceeding 25%, with the target exhibiting a higher average PCE distribution than the control, validating its practical utility (Figure S10). Figure 5b shows the external quantum efficiency (EQE) for the same device as that of J-V. The current density of the target, obtained by integrating EQE over wavelength, is 25.51 mA cm^{-2}, in good agreement with that measured from the J-V curve using the solar simulator. To evaluate the impact of IM on the surface of $FAPbI_3$, photoluminescence (PL) and time-resolved PL (TRPL) measurements were performed on quartz substrates (Figure S11). Both PL intensity and carrier lifetime tended to increase in the order of control, reference, and target (Table S4), indicating suppressed non-radiative recombination [36]. To further clarify potential spectral shifts, we quantitatively extracted the PL peak positions from the measured spectra. The emission maxima of the control and reference samples were both located at 811 nm, whereas the target sample exhibited a peak at 813 nm (Figure S11). These values indicate that no blue shift occurred and that the emission peak positions remain essentially unchanged within the experimental resolution. The apparent visual shift in the unnormalized spectra likely originates from the significantly enhanced PL intensity and slight modification of the spectral profile in the target-treated sample. Therefore, the primary effect of the target treatment is attributed to defect suppression rather than bandgap modification.

To evaluate phase stability under thermal stress, we performed ultraviolet–visible spectroscopy (UV–vis) and X-ray diffraction (XRD) analysis during aging at 85°C. The thermal stability tests were conducted following ISOS-D-2 conditions, where unencapsulated devices were aged at 85°C in ambient air under dark conditions to ensure standardized comparability. After 715 h, the control and reference samples showed significant degradation, whereas the target sample maintained its structural and optical properties. As shown in Figure 5c and Figures S12 and S13, UV–vis spectra confirmed minimal absorbance loss in the target. Similarly, XRD patterns (Figure 5d; Figure S14) further revealed that only the target retained the α-$FAPbI_3$ black phase, while the other samples underwent phase transition to the non-perovskite yellow phase, accompanied by the formation of PbI_2. Prior to annealing (Figure S14), the (100) and (200) reflections of α-$FAPbI_3$ remain dominant under all conditions, indicating preserved bulk crystallinity. In the target sample, an additional diffraction peak appears at approximately $2\theta \approx 9.75°$, which is assigned to the higher-order (002) reflection of the 2D perovskite phase [37]. The coexistence of this peak with the (001) reflection at $2\theta \approx 5.06°$ suggests a preferentially oriented layered structure periodically stacked perpendicular to the substrate. The slight deviation from the reported theoretical (002) position ($2\theta \approx 10.12°$) can be attributed to variations in organic spacer tilt, residual solvent removal during annealing, or strain relaxation within the layered framework. After annealing, the 2D-related peak shifts from $2\theta \approx 5.06°$ to $2\theta \approx 4.93°$, corresponding to an increased interplanar spacing according to Bragg's law. This shift indicates an expansion of the interlayer gallery height, likely arising from thermally induced reorientation of the organic spacer cations and partial lattice relaxation within the quasi-2D structure [38]. During aging, the 2D crystalline phase observed in the control sample

15214095, 0, Downloaded from https://advanced.onlinelibrary.wiley.com/doi/10.1002/adma.202522554 by Ulsan National Institute Of, Wiley Online Library on [23/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

rapidly disappeared, whereas the target sample stably preserved the surface 2D phase (Figure S15). The superior durability of the target arises from the intrinsic chemical stability of quaternary ammonium cations, whose reduced propensity for deprotonation suppresses interfacial degradation and enhances thermal stability when employed as interfacial materials [39].

Moreover, we assessed the PCE of the full device (n-i-p) after thermal degradation as a function of time (Figure 5e; Figure S16) or temperature (Figure 5f; Figure S17), and correlated these results with scanning electron microscopy (SEM) images to observe the top surface (Figures S18 and S19). Notably, the target device exhibited the most stable performance, maintaining a smooth, uniform, and crack-free morphology over a wide range of aging times and temperatures. These results suggest that, rather than inducing the collapse of the surface 2D structures under thermal stress, the treatment promotes the stable retention of the 2D phases without deprotonation [40]. This interpretation is quantitatively supported by atomic force microscopy (AFM) analysis (Figure S20), which shows that the target sample exhibited the lowest surface roughness of 25.5 nm, compared to the pristine (36.0 nm), control (42.3 nm), and reference (36.6 nm) samples. The initial surface morphology varied depending on the halide species, with Cl^- producing a distinct surface texture, likely due to its smaller ionic radius and stronger binding interactions. After thermal aging, however, all films evolved into similar smooth morphologies, indicating that long-term stabilization is predominantly by the cation (Figure S21). Consistently, THAI- and THACl-treated devices achieved stable PCEs above 24%, comparable to that of the THABr-treated device (Figure S22). Finally, the long-term thermal stability of the unencapsulated devices was further evaluated by periodically measuring their performance under full solar illumination while aging on a hot plate. The target device retained over 81.6% of its initial efficiency after 1508 h, whereas the other devices rapidly degraded (Figure S23). Additionally, the long-term operational stability of the encapsulated device was evaluated by maximum power point tracking (MPPT) under continuous AM 1.5G illumination following the ISOS-L-1 protocol, without an ultraviolet cut-off filter, as depicted in Figure S24. The encapsulated target device retained approximately 80.2% of its initial efficiency after 730 h of operation and maintained 98.5% of its initial efficiency after 2000 h when stored under ISOS-D-1 conditions, with encapsulated devices stored in the dark at room temperature (20°C–25°C) and controlled humidity (~30% RH) (Figure S25).

Furthermore, DFT calculations support experimental observations, revealing strengthened van der Waals interactions between THA cations and the perovskite interface during 2D structure formation (Figure S26), which likely contributes to enhanced interfacial stability. Meanwhile, when THABr was applied to the perovskite surface in the p–i–n architecture, it also effectively passivated defects and enhanced thermal stability (Figure S27). However, unlike in the n–i–p configuration, THABr in the p–i–n structure is located at the interface between the perovskite and the electron transport layer (C_{60}). Therefore, careful consideration of the distinct interfacial configurations in the two device architectures is required to maximize charge transfer and minimize interfacial losses.

3 | Conclusion

In summary, we introduce an ML–guided framework that accelerates the discovery of IMs for PSCs. Systematic screening of alkylammonium halides identified key molecular descriptors—namely, a low dipole moment and a high κ_1 shape index—associated with superior performance. The model favors symmetric, higher-order cations that enhance both efficiency and stability. The identified THABr achieved a PCE of 25.31% and preserved over 81.6% of its efficiency after 1508 h at 85°C, underscoring the simultaneous optimization of efficiency and durability. Beyond this single material, our strategy expands the design space of complex salts, demonstrating how machine learning can surpass trial-and-error discovery and inform the rational design of next-generation optoelectronic materials.

4 | Materials and Methods

4.1 | Materials

Tin(IV) chloride pentahydrate ($SnCl_4 \bullet 5H_2O$; 98%), citric acid ($C_6H_8O_7$; 99.5%), potassium hydroxide (KOH; 99.99%), and titanium diisopropoxide bis(acetylacetonate) (75 wt.% in isopropanol) were purchased from Sigma–Aldrich. Formamidinium chloride (FACl; 99.99%) was purchased from Solar Materials. Aqueous hydrogen peroxide solution (H_2O_2; 35 wt.%) was purchased from Alfa Aesar. Ammonium hydroxide solution (NH_4OH; 25 wt%) was purchased from DAEJUNG. All reagents were used without purification. The following materials were used in the experiments: formamidinium iodide (FAI; 99.99%, Greatcell solar), lead(II) iodide (PbI2; 99.99%, TCI), methylammonium chloride (MACl; Merck), propylamine hydrochloride (PACl; Sigma–Aldrich), 4-methoxyphenethylammonium iodide (MeO-PEAI; 99%, Greatcell Solar), Phenethylammonium iodide (PEAI, Greatcell Solar), 4-Fluoro-Phenethylammonium iodide (FPEAI, Greatcell Solar), Phenethylammonium Bromide (PEABr, Greatcell Solar), n-Butylammonium iodide (BAI, Greatcell Solar), n-Octylammonium iodide (OAI, Greatcell Solar), n-Dodecylammonium iodide (DAI, Greatcell Solar), Guanidinium iodide (GuaI, Greatcell Solar), Pyridinium iodide (PyI, Greatcell Solar), Phenylammonium iodide (PAI, Greatcell Solar), 4-Chlorophenethylammonium iodide (ClPEAI, Greatcell Solar), n-Dodecylammonium bromide (DABr, Greatcell Solar), 4-Fluoro-Phenethylammonium bromide (F-PEABr, Greatcell Solar), Guanidinium bromide (GuaBr, Greatcell Solar), 4-Chlorophenethylammonium bromide (ClPEABr, Greatcell Solar), Acetylcholine chloride (ACCl, TCI), Acetylcholine bromide (ACBr, TCI), Acetylcholine iodide (ACI, TCI), n-Octylammonium chloride (OACl, Greatcell Solar), 4-Fluoro-Phenethylammonium chloride (F-PEACl, Greatcell Solar), 4-Methoxy-Phenethylammonium Bromide (M-PEABr, Greatcell Solar), Phenethylammonium chloride (PEACl, Greatcell Solar), Pyridinium Bromide (PyBr, Greatcell Solar), 4-Methoxy-Phenylammonium Bromide (M-PABr, Greatcell Solar), Phenylammonium Chloride (PnACl, Greatcell Solar), Guanidinium Chloride (GuaCl, Greatcell Solar), Pyridinium Chloride (PyCl, Greatcell Solar), Choline chloride (CHCl, Greatcell Solar), Diethylammonium iodide (Sigma–Aldrich),

15214095, 0, Downloaded from https://advanced.onlinelibrary.wiley.com/doi/10.1002/adma.202522554 by Ulsan National Institute Of, Wiley Online Library on [23/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Di-iso-Propylammonium bromide (Greatcell Solar), Di-isopropylammonium chloride (TCI), Di-isopropylammonium iodide (Greatcell Solar), Citric acid diammonium salt (99+%, Alfa aesar), Bis(2-chloroethyl)ammonium chloride (Sigma–Aldrich), (Chloromethylene)dimethylammonium chloride (96%, Sigma–Aldrich), Tetramethylammonium bromide (98%, Alfa aesar), Tetra-n-hexylammonium bromide (98%, Alfa aesar), Bis(2-hydroxyethyl)dimethylammonium chloride (99%, Fisher scientific), Tetrapropylammonium chloride (94%, Sigma–Aldrich), Di-Butylammonium chloride (Greatcell Solar), Benzyltrimethylammonium chloride (97%, Sigma–Aldrich), 4-Methoxy-Phenethylammonium chloride(Greatcell Solar), N,N-dimethylformamide (DMF; 99.8%, Sigma–Aldrich), dimethyl sulfoxide (DMSO; 99.9%, Sigma–Aldrich), 2-propanol (99.5%, Sigma–Aldrich), 2-methoxyethanol (2-ME; 99.8%, Sigma–Aldrich), chlorobenzene (99.8%, Sigma–Aldrich), acetonitrile (ACN; 99.8%, Sigma–Aldrich), Spiro-OMeTAD (Lumtec.), 4-tert-butylpyridine (t-BP; Sigma–Aldrich), bis(trifluoromethane)sulfonimide lithium salt (Li-TFSI; 99.95%, Sigma–Aldrich), FK209 Co(III) TFSI salt (Lumtec.), and ethyl ether (99.0%, SAMCHUN). Poly(triarylamine) (PTAA) was purchased from EM Index. 4-isopropyl-4′-methyldiphenyliodoniumtetrakis (pentafluorophenyl)borate (DPI-TPFB,98%, TCI).

4.1.1 | Preparation of SnO_2 Colloidal Solution

$SnCl_4$•$5H_2O$ (20 mmol) and citric acid (5 mmol) were dissolved in 100 mL of hydrogen peroxide aqueous solution. NH_4OH solution (110 mL; 25 wt%) was added to the solution with constant stirring. The reaction mixture was aged at 70°C for 12 h. Care should be taken to avoid sudden temperature changes, which could cause an explosion. The resulting white precipitate was separated by centrifugation and washed four times with deionized water. The washed SnO_2 was dispersed in deionized water to a concentration of 5.6 wt.% and the pH was adjusted to approximately 10 by adding a few drops of KOH solution. The mixture was peptized by stirring to obtain a transparent colloidal solution.

4.1.2 | Preparation of SnO_2-FACl Solution

The colloidal SnO_2 solution was first diluted to a concentration of 2.8 wt.% with deionized water containing 10 vol% of hydrogen peroxide. The diluted solution was ultrasonicated for around 1 h to activate the surface of SnO_2 colloids, and then FACl was added. Prior to spin-coating, the solution was filtered using a 0.45 μm PVDF filter.

4.2 | Device Fabrication

FTO glass (Asahi VU glass, 12–13 $\Omega{\cdot}cm^{-2}$) substrates were sequentially sonicated in acetone, detergent, and ethanol for 20 min each. A several-nanometer-thick layer of TiO_2 was deposited by spray pyrolysis at 450°C using a 0.2 M titanium diisopropoxide bis(acetylacetonate) solution. Next, the SnO_2-FACl solution was spin-coated onto the substrate at 3000 rpm for 40 s, followed by annealing at 150°C for 1 h in air. The perovskite precursor solutions were prepared by dissolving 1.42 M $FAPbI_3$ powder with ACCl (0.75 mol%), PACl (4 mol%) and MACl (35 mol%) in a mixed solvent of DMF: DMSO = 8:1. The perovskite solutions were then spin-coated onto the substrate at 1000 rpm for 10 s and 5000 rpm for 15 s; 1 mL of ethyl ether was dripped onto the substrate during spinning. All perovskite layers were heat-treated at 120 °C for 1 h. After cooling, 14 mM of the interfacial material (IM) was prepared by dissolving it in 2-propanol (IPA). The solutions were spin-coated on the perovskite-coated substrate at 5000 rpm for 30 s and annealed at 100°C for 5 min. The Spiro-OMeTAD solution, prepared by mixing 90 mg mL^{-1} of Spiro-OMeTAD in chlorobenzene with 37 μL of t-BP, 23μL of Li-TFSI salt (516 mg mL^{-1} in ACN), and 9 μl of Co-TFSI salt (375 mg mL^{-1} in ACN), was spin-coated at 3500 rpm for 30 s. A Poly(triarylamine) (PTAA, EM Index, Mn = 17 500 $g{\cdot}mol^{-1}$)/Chlorobenzene (12 mg/1 mL) solution with additives of DPI-TPFB 3 mg was spin-coated on the perovskite layer at 3000 rpm for 30 s. Finally, a 70-nm-thick gold counter electrode was deposited by thermal evaporation.

4.3 | Characterization

The optical properties of the films were measured using a UV-2600 (Shimadzu) UV-vis spectrometer. The crystal structure was analyzed using X-ray diffraction (XRD, D/MAX2500V/PC, Rigaku). Steady-state photoluminescence (PL) and time-resolved photoluminescence (TRPL) spectra were measured using a commercial time-correlated single photon counting (TCSPC) setup (FluoTime 300, PicoQuant GmbH) equipped with a PMA-C-192-M detector and high-resolution excitation monochromators. The J-V characteristics of the devices were measured using a Keithley 2420 source meter under illumination by a solar simulator (Newport, Oriel Sol3A class AAA) with an AM 1.5G filter and 100 mW cm^{-2} irradiation intensity. The measurements were conducted with a starting voltage of −0.2 V and an ending voltage of 1.5 V. The parameters for the measurement were set as follows: the number of points was 170, the voltage settling time was 10 ms, and the NPLC (Number of Power Line Cycles) was 0.01 (Fast mode). The active area was determined by placing a metal mask in front of the solar cell to avoid overestimating the photocurrent density. A spectral mismatch factor of 1.05 was used for all J-V measurements. An antireflective (AR) film (NANOECOWAY CO. Ltd.) was applied to the surface to measure high-efficiency devices. The AR film has a thickness of 0.2–0.3 μm and a transmittance exceeding 95%. Its reflectance, haze, and surface resistance are 0.8% ± 0.2%, 0.7% ± 0.1%, and 10^8–10^{10} Ω, respectively. External quantum efficiency (EQE) was measured using an internal quantum efficiency system (Oriel, IQE 200B) under irradiation by a 100 W Xenon lamp. The film morphology and thickness were measured using scanning electron microscopy (FE-SEM, SU-7000, SU8220) operated at 10 kV.

4.4 | Photoluminescence (PL) Spectra and Time-Resolved Photoluminescence (TRPL) Analysis

The film was illuminated from the front side using 520 nm excitation. The carrier lifetime values were determined using the bi-exponential equation: $Y = A_1\exp(-t/\tau_1) + A_2\exp(-t/\tau_2)$, where τ_1 and τ_2 represent the fast and slow decay times associated with the trap-assisted non-radiative and radiative recombination

15214095, 0, Downloaded from https://advanced.onlinelibrary.wiley.com/doi/10.1002/adma.202522554 by Ulsan National Institute Of, Wiley Online Library on [23/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

processes, respectively. The discrepancies in the PL and carrier lifetime in the measurement can be attributed to variations in charge transfer to the interlayer rather than the recombination of charges generated in the perovskite.

4.5 | Thermal Stability Device Test

For short-term thermal stability tests (5 h), a sample with an FTO/SnO_2/perovskite/IM stack was placed on a hot plate at 85°C for several hours. After this thermal aging step, the device was completed by sequentially depositing Spiro-OMeTAD as the hole transport layer (HTL) and a gold electrode. For long-term stability evaluation, a full device using thermally stable PTAA as the HTL was aged directly on the hot plate under the same conditions. All stability measurements were conducted in accordance with relevant International Summit on Organic Photovoltaic Stability (ISOS) protocols to ensure standardization and comparability. Unencapsulated thermal aging tests followed ISOS-D-2 conditions (85°C in ambient air, dark), operational stability under illumination followed ISOS-L-1 conditions (continuous AM 1.5G illumination at maximum power point), and shelf-life stability tests followed ISOS-D-1 conditions (dark storage at room temperature and controlled humidity).

4.6 | Device Stability Test

The stability test of continuous maximum power point tracking operation under 1 sun illumination at AM 1.5G (Xe lamp, 450 W) for the encapsulated devices was performed under an inert atmosphere (N_2) by fixing the voltage at V_{MPP} and then tracking the current output.

4.7 | Interaction Mechanism of THABr

We performed DFT calculations to examine the interactions between THA^+ species and the $FAPbI_3$ surface. For computational simplicity, we focused on the THA^+ cation. Because THA^+ is expected to be positively charged in solution, we modeled THA^+ binding at an FA-vacancy site on the $FAPbI_3$ surface to maintain charge neutrality. As shown in Figure S26a,b, the most stable configuration features THA^+ adopting a nearly perpendicular orientation relative to the surface. The calculated adsorption energy for this geometry is −0.1351 eV, suggesting that the adsorption is energetically feasible and may occur readily under relevant conditions. The stabilization is primarily attributed to van der Waals interaction via multiple hydrogen atoms of THA^+ and surface iodide anions of $FAPbI_3$. Motivated by this strong interaction, we further explored how THA^+ may facilitate the formation of a 2D perovskite phase on the 3D $FAPbI_3$ layer. As shown in Figure S26c, we evaluated several candidate geometries and found that the most stable structure corresponds to two THA^+ cations bound nearly perpendicular to the 3D/2D interface. In this geometry, van der Waals interaction can be strengthened by increasing the number of interacting hydrogen and iodine atoms at the interface, and this preference is consistent with the single-ion adsorption result above. Moreover, the calculated binding energy of the THA^+-based 2D perovskite on the 3D perovskite surface is substantially lower than that of the corresponding MPEA-based 2D perovskite (Figure S26b), which supports the experimentally observed enhanced thermal stability of the target device. All calculations were performed using the VASP code with a 350 eV energy cutoff for the plane-wave basis and the PBE exchange-correlation functional. For k-point sampling, a single k-point of (0.25,0.25,0.0) was used. During the structure optimization, the unit cell parameters were fixed, and only internal coordinates were relaxed, with a force criterion of 0.03 eV/A. The Tkatchenko-Scheffler method was employed to model van der Waals interactions.

4.8 | Details of the Active Learning Framework

4.8.1 | Acquiring the Available Molecule List

The materials pool for this study is restricted to commercially available ammonium halides, including several well-known surface treatment molecules (STMs). Although the chemical search space could theoretically be expanded using generative models or other computational design tools, such approaches would substantially increase costs due to the inclusion of chemically invalid structures and the need for new material synthesis. To ensure practical feasibility and experimental workability, we limited our investigation to commercially accessible compounds. This strategy enables efficient screening within a well-defined chemical space while simplifying subsequent experimental validation. To construct the candidate library, we searched for all compounds containing the keyword 'ammonium ion' in the databases of Sigma-Aldrich, Alfa Aesar, Merck, and Greatcell Solar Materials, thereby restricting the search to readily available materials. The initial search yielded 194, 735, and 202 compounds from these suppliers, respectively, resulting in a total of 1,131 entries. The corresponding CAS numbers were converted to SMILES [41, 42]. Entries that could not be converted were excluded, leaving 434 compounds. The SMILES strings were then parsed to verify whether each molecule contained a positively charged ammonium ion. Compounds containing metal ions or salt forms were excluded. For example, cases in which an ammonium moiety appeared as part of the anion were removed from the dataset. After eliminating duplicates, a final set of 115 unique ammonium ions was obtained for subsequent analysis.

4.8.2 | Features for Surface Treatment Materials

For the subsequent machine learning study, ammonium halides were encoded as numerical vectors using property-based molecular descriptors. Ideally, such descriptors should be universally applicable to unexplored materials while retaining clear physicochemical interpretability. In a previous study by Lee et al. [24], property-based descriptors were shown to enable the development of highly accurate predictive models while facilitating mechanistic insight into structure–property relationships. Following this strategy, we employed six descriptors to represent ammonium cations as candidate surface passivation materials: molecular size (MS), dipole moment (DM), topological polar surface area (TPSA), and Kier's shape indices (κ_1, κ_2, and κ_3). TPSA describes the electrical and geometric properties of various types of compounds using functional group contributions from large databases. The κ_i indices, a concept used in molecular

15214095, 0, Downloaded from https://advanced.onlinelibrary.wiley.com/doi/10.1002/adma.202522554 by Ulsan National Institute Of, Wiley Online Library on [23/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

modeling and cheminformatics, represent the 3D structural information of a molecule, including its size, shape, and the relative positions of atoms within the molecule. These indices quantify structural information such as linearity, ring structure, and branching. Molecular structures were represented using SMILES notation and processed using the RDKit cheminformatics toolkit. TPSA and Kier shape indices were computed using established topological algorithms implemented in RDKit. Molecular size was quantified based on molecular weight and structural topology descriptors. Dipole moments were calculated at the density functional theory (DFT) level using the B3LYP functional with the 6–31G(d) basis set as implemented in Gaussian software. The anion type of ammonium halides is not considered a direct molecular feature. Instead, separate predictive models were developed for each anion type to account for its influence.

4.8.3 | Dimensionality Reduction by PCA

Bayesian optimization (BO) is a model-based approach highly effective for global optimization of black-box functions. Standard BO consists of two key steps: (1) estimating the black-box function from data using a probabilistic surrogate model and (2) maximizing an acquisition function that balances exploration and exploitation. However, as the amount of available data decreases and the dimensionality of the input space increases, the acquisition function tends to become mostly flat, with sharp peaks in certain regions [43], making the data collection process more challenging. High-dimensional data can often be mapped to a lower-dimensional manifold with reduced intrinsic dimensionality. In our study, we calculated the properties of commercially available molecules and found significant correlations among their features. Dimensionality reduction techniques can broadly be classified into linear and nonlinear transformations. While nonlinear embeddings can achieve higher compression ratios, addressing the relationship between features and performance in perovskite solar cells—an area still not fully understood—and considering the constraints of limited experimental resources, determining the most appropriate dimensionality reduction technique is beyond the scope of this study. To simplify the problem, we employed PCA, a well-established dimensionality reduction technique. Given a dataset $X \in \mathbb{R}^{n\times d}$, where n is the number of samples and d is the number of features, the dataset is mean-centered by subtracting the mean of each feature:

$$\mathbf{X}_{centered} = \mathbf{X} - \mu, \mu = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$$

where $\mathbf{x}_i$ is the i-th sample. The covariance matrix $\mathbf{C}$ is computed as:

$$\mathrm{C} = \frac{1}{n}\mathbf{X}^{\top}_{centered}\mathbf{X}_{centered}$$

The eigenvalues λ_k and eigenvectors v_k of $\mathbf{C}$ are calculated: $\mathbf{C}_{v_k} = \lambda_k v_k$. The eigenvectors represent the directions of the principal components, and the eigenvalues indicate the amount of variance captured by each component. The top p eigenvectors corresponding to the largest p eigenvalues are selected to form the projection matrix $W \in \mathbb{R}^{d\times p}$:

$$\mathbf{W} = \left[v_1, v_2, \ldots, v_p\right]$$

The original dataset is then projected into the lower-dimensional space as:

$$\mathbf{Z} = \mathbf{X}_{centered}\mathbf{W}W, Z \in \mathbb{R}^{n\times p}$$

In this study, PCA was applied to reduce the 6D feature space of molecular descriptors into a lower-dimensional latent space. The first two principal components explained 69.14% of the total variance in the commercial ammonium pool, while the first three components explained 83.30%. To determine the appropriate dimensionality, we compared the cross-validated predictive performance of the Gaussian process surrogate using 2D, 3D, and full 6D representations. The predictive accuracy and uncertainty calibration were comparable across these settings, with no measurable performance gain observed when increasing dimensionality beyond two components. In the absence of performance improvement, we adopted a 2D representation to enhance numerical stability and reduce the risk of overfitting in the small-data Bayesian optimization regime. Lower-dimensional embeddings mitigate the risk of acquisition-function flattening and sharp local peaks that commonly arise in high-dimensional surrogate-assisted optimization under limited data. To further assess representativeness, we evaluated the applicability domain using the squared Mahalanobis distance (D_M^2) relative to the training distribution (Figure S28). The 95% confidence threshold was defined based on the χ^2 distribution with six degrees of freedom. This analysis confirmed that the experimentally measured samples occupy a representative region of the descriptor space while maintaining sufficient coverage of structurally diverse candidates. Furthermore, although the present study restricts the candidate pool to commercially available ammonium halides to ensure experimental feasibility, the selected physicochemical descriptors represent general interfacial interaction characteristics and are, in principle, extendable to synthetically accessible ammonium derivatives. However, expanding to broader chemical spaces may require additional, or computationally scalable, descriptors to maintain predictive reliability under limited-data conditions.

4.9 | Gaussian Process Regression

Gaussian Process Regression (GPR) is a non-parametric Bayesian framework for regression that provides a flexible and probabilistic approach to modeling complex relationships between input variables and target responses. Given a set of N independent observations $\mathcal{D} = \{(\boldsymbol{x_i}, \boldsymbol{y_i})\}_{i=1}^{N}$, GPR assumes that the observations can be represented as samples from a multivariate Gaussian distribution, which is governed by a covariance function $k(x, x')$, also known as a kernel function. This kernel function encodes assumptions about the function we wish to learn and defines the properties of the Gaussian process. GPR is implemented in scikit-learn package [44]. GPR models the relationship between input features x and output y as a distribution over functions:

$$f(x) \sim \mathcal{GP}\left(m(x), k_f(x, x')\right)$$

15214095, 0, Downloaded from https://advanced.onlinelibrary.wiley.com/doi/10.1002/adma.202522554 by Ulsan National Institute Of, Wiley Online Library on [23/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where $m(x)$is the mean function (assumed to be zero without loss of generality) and $k_f(x,x')$is the covariance function (kernel) that encodes similarity between input points.

Because PSC performance measurements inherently contain small but unavoidable experimental noise—even under highly standardized fabrication conditions—explicit modeling of observation noise is essential, particularly in small-data regimes. Without an explicit noise term, the surrogate model may misinterpret minor experimental fluctuations as a structured signal, leading to mis-calibrated predictive uncertainty during active learning.

To account for experimental observation noise, we adopt an additive Gaussian noise model:

$$y_i = f(x_i) + \epsilon_i, \epsilon_i \sim \mathcal{N}\left(0, \sigma_n^2\right)$$

where σ_n^2 represents the variance of independent and identically distributed observation noise. This formulation is equivalent to augmenting the signal kernel with a white-noise term:

$$k_y(x, x') = k_f(x, x') + \sigma_n^2 \delta_{xx'}$$

where $\delta_{xx'}$is the Kronecker delta function. Accordingly, the covariance matrix of the training outputs becomes:

$$K_y(X, X) = K_f(X, X) + \sigma_n^2 I$$

where $I$is the identity matrix.

Given the training inputs $X$and outputs y, the predictive distribution at a new input $x_*$is Gaussian:

$$f_* \mid X, y, x_* \sim \mathcal{N}\left(\mu_*, \sigma_*^2\right)$$

with predictive mean and variance given by:

$$\mu_* = k_*^T \left(K_f(X, X) + \sigma_n^2 I\right)^{-1} y$$

$$\sigma_*^2 = k_f(x_*, x_*) - k_*^T \left(K_f(X, X) + \sigma_n^2 I\right)^{-1} k_*$$

where $k_* = k_f(X,x_*)$ denotes the covariance vector between the training inputs and the test input.

In this work, the hyperparameters of the kernel function, including the noise variance σ_n^2, were optimized by maximizing the log marginal likelihood of the observed data. All GPR implementations were carried out using the scikit-learn package. The inclusion of an explicit white-noise term is particularly important in perovskite solar cell studies, where residual fabrication variability may otherwise be misinterpreted as a structured signal, leading to overconfident uncertainty estimates in small-data regimes.

4.10 | Bayesian Optimization

BO is a model-based approach for the global optimization of a black-box function $f(x)$. In this study, BO was employed as part of an active learning framework to identify promising ammonium halide candidates for surface treatment of PSCs. GPR serves as a probabilistic surrogate model to approximate the black-box objective function, defined here as the PCE of PSCs. Before applying BO, the molecular descriptor data are reduced in dimensionality using PCA, as described in the previous section, to simplify the input space for the Gaussian process model. The upper confidence bound (UCB) acquisition function was used to balance the exploration of unknown regions and the exploitation of areas predicted to yield high performance.

$$\alpha_{UCB}(x) = \mu(x) + \kappa \cdot \sigma(x)$$

where κ controls the trade-off between exploration and exploitation. Then we can select the next evaluation point:

$$x_{next} = arg\max_x \alpha(x)$$

Evaluation of $f(x_{next})$ will update the model. The process continues until a stopping criterion is met.

4.11 | Statistical Analysis

All data were processed and analyzed using Origin 2020 software (OriginLab). Data are presented as mean ± standard deviation (SD). The sample size for each experiment and the specific statistical methods used to determine significance are described in the respective figure legends. Before analysis, data were evaluated for outliers and normality to ensure the validity of the statistical tests. A P-value less than 0.05 was considered statistically significant.

Author Contributions

Jongbeom Kim, Yang Jeong Park, Jino Im, and Sang Il Seok conceived this work and designed the experiment. Jongbeom Kim fabricated the PSCs with various interfacial materials and characterized the perovskite films. Yang Jeong Park implemented the active learning loop and proposed the post-hoc interpretation method. Jino Im simulated the theoretical calculation. Nahye Shin measured XRD analysis. Chaehoon Jeon conducted UV-vis absorbance spectra. Jaewang Park and Seungun Lee discussion. Jongbeom Kim, Yang Jeong Park, Jino Im, and Sang Il Seok wrote the draft of the manuscript, and all authors contributed feedback and comments to revise the manuscript. Sang Il Seok directed and supervised the study.

Acknowledgements

This work was supported by financial support from the Basic Science Research Program (NRF-2018R1A3B1052820) and the Nano & Material Technology Development Program (RS-2024-00449449 and RS-2024-00356670) through the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT & Future Planning (MSIP).

15214095, 0, Downloaded from https://advanced.onlinelibrary.wiley.com/doi/10.1002/adma.202522554 by Ulsan National Institute Of, Wiley Online Library on [23/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Funding

This work was supported by financial support from the Basic Science Research Program (RS-2018-NR030954, RS-2024-00356670, and RS-2026-25497753) and the Nano & Material Technology Development Program (RS-2024-00449449) through the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT & Future Planning (MSIP). J.I., S.Y., and S.I.S. acknowledge the support from the InnoCORE program of the Ministry of Science and ICT(1.260007.01).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. J. H. Heo, S. H. Im, J. H. Noh, et al., "Efficient Inorganic–Organic Hybrid Heterojunction Solar Cells Containing Perovskite Compound and Polymeric Hole Conductors," *Nature Photonics* 7 (2013): 486–491.

2. G. Seo, J. J. Yoo, S. Nam, et al., "Efficient and Luminescent Perovskite Solar Cells Using Defect-Suppressed SnO_2 via Excess Ligand Strategy," *Nature Energy* 10 (2025): 774–784.

3. Y. Wang, C. Lu, M. Liu, et al., "Solvent-Assisted Reaction for Spontaneous Defect Passivation in Perovskite Solar Cells," *Nature Photonics* 19 (2025): 985–991.

4. J. Kim, J. Park, J. Lim, et al., "Susceptible Organic Cations Enable Stable and Efficient Perovskite Solar Cells," *Joule* 9 (2025): 101879.

5. Z. Wang, T. Xu, N. Li, et al., "Interfacial Engineering With Trivalent Cations for Efficient and Stable Inverted Inorganic Perovskite Solar Cells," *Energy & Environmental Science* 17 (2024): 7271–7280.

6. D. Y. Heo, W. J. Jang, and S. Y. Kim, "Recent Review of Interfacial Engineering for Perovskite Solar Cells: Effect of Functional Groups on the Stability and Efficiency," *Materials Today Chemistry* 26 (2022): 101224.

7. S. S. Shin, B.-W. Park, J. H. Noh, and S. I. Seok, "Interlayer Engineering In Metal Halide Perovskite Photovoltaics," *Nature Photonics* 20 (2026): 11–23.

8. Z. Wang, C. Gong, C. Zhang, et al., "Recent Advances in Interfacial Engineering for High-Efficiency Perovskite Photovoltaics," *DeCarbon* 8 (2025): 100107.

9. S. You, F. T. Eickemeyer, J. Gao, et al., "Bifunctional Hole-Shuttle molecule for Improved Interfacial Energy Level Alignment and Defect Passivation in Perovskite Solar Cells," *Nature Energy* 8 (2023): 515–525.

10. D. He, P. Chen, J. A. Steele, et al., "Homogeneous 2D/3D Heterostructured Tin Halide Perovskite Photovoltaics," *Nature Nanotechnology* 20 (2025): 779–786.

11. H. Rao, S. Ye, T. Salim, et al., "Selective Templating Growth of Chemically Inert Low-Dimensional Interfaces for Perovskite Solar Cells," *Nature Energy* 10 (2025): 991–1000.

12. Y. Boeije, W. T. M. Van Gompel, Y. Zhang, et al., "Tailoring Interlayer Charge Transfer Dynamics in 2D Perovskites With Electroactive Spacer Molecules," *Journal of the American Chemical Society* 145 (2023): 21330–21343.

13. S. Teale, M. Degani, B. Chen, E. H. Sargent, and G. Grancini, "Molecular Cation and Low-Dimensional Perovskite Surface Passivation in Perovskite Solar Cells," *Nature Energy* 9 (2024): 779–792.

14. S. Ye and T. C. Sum, "Inert Low-Dimensional Interfaces for Perovskite Solar Cells," *Nature Energy* 10 (2025): 932–933.

15. H. Chen, C. Liu, J. Xu, et al., "Improved Charge Extraction in Inverted Perovskite Solar Cells With Dual-Site-Binding Ligands," *Science* 384 (2024): 189–193.

16. J. Y. Zhang, J. C. Wu, V. M. Le Corre, J. A. Hauch, Y. C. Zhao, and C. J. Brabec, "Advancing Perovskite Photovoltaic Technology Through Machine Learning-Driven Automation," *Infomat* 7 (2025): 70005.

17. J. W. Yoo, E. Noh, J. Jang, et al., "R_4N^+ and Cl^- Stabilized α-Formamidinium Lead Triiodide and Efficient Bar-Coated Mini-Modules," *Joule* 7 (2023): 797–809.

18. X. P. Zheng, B. Chen, J. Dai, et al., "Defect Passivation in Hybrid Perovskite Solar Cells Using Quaternary Ammonium Halide Anions and Cations," *Nature Energy* 2 (2017): 17102.

19. W. Shao, H. Wang, S. Fu, et al., "Tailoring Perovskite Surface Potential and Chelation Advances Efficient Solar Cells," *Advanced Materials* 36 (2024): 2310080.

20. N. Guan, Y. Zhang, W. Chen, et al., "Deciphering the Morphology Change and Performance Enhancement for Perovskite Solar Cells Induced by Surface Modification," *Advanced Science* 10 (2023): 2205342.

21. R. Q. Zhang, B. Motes, S. Tan, et al., "Machine Learning Prediction of Organic–Inorganic Halide Perovskite Solar Cell Performance From Optical Properties," *ACS Energy Letters* 10 (2025): 1714–1724.

22. J. Wu, L. Torresi, M. Hu, et al., "Inverse Design Workflow Discovers Hole-Transport Materials Tailored for Perovskite Solar Cells," *Science* 386 (2024): 1256–1264.

23. Y. Pu, Z. Dai, Y. Zhou, et al., "Data-Driven Molecular Encoding for Efficient Screening of Organic Additives in Perovskite Solar Cells," *Advanced Functional Materials* 36 (2026): 06672.

24. S. Lee, Y. J. Park, J. Kim, J. Im, S. Yoon, and S. I. Seok, "Data-Driven Analysis on Perovskite Solar Cell Devices," *Current Applied Physics* 68 (2024): 98–107.

25. L. B. Kier, "Shape Indexes of Orders One and Three From Molecular Graphs," *Quantitative Structure-Activity Relationships* 5, (Wiley, 1986), 1–7.

26. L. B. Kier, "Inclusion of Symmetry as a Shape Attribute in Kappa Index Analysis," *Quantitative Structure-Activity Relationships*, 6, (Wiley, 1987), 8–12.

27. L. B. Kier, "A Shape Index From Molecular Graphs," *Quantitative Structure-Activity Relationships* 4 (1985): 109–116.

28. S. Ullah, D. A. Nguyen, H. Wang, S. Menzel, B. Sendhoff, and T. Bäck, presented at 2020 IEEE Symposium Series on Computational Intelligence (SSCI), (IEEE, 2020), 2965–2974.

29. L. Pretsch, I. Arsenyev, E. Raponi, and F. Duddeck, Twofold Adaptive Design Space Reduction for Constrained Bayesian Optimization of Transonic Compressor Blades, Proc. ASME Turbo Expo 2024, GT2024–121848.

30. J. Park, J. Kim, H.-S. Yun, et al., "Controlled Growth of Perovskite Layers With Volatile Alkylammonium Chlorides," *Nature* 616 (2023): 724–730.

31. H. Kim, S.-U. Lee, D. Y. Lee, et al., "Optimal Interfacial Engineering With Different Length of Alkylammonium Halide for Efficient and Stable Perovskite Solar Cells," *Advanced Energy Materials* 9 (2019): 1902740.

32. J. Zhuang, P. Mao, Y. Luan, et al., "Interfacial Passivation for Perovskite Solar Cells: The Effects of the Functional Group in Phenethylammonium Iodide," *ACS Energy Letters* 4 (2019): 2913–2921.

33. W. Gao, J. Ding, Q. Ma, et al., "Synergistic Modulation of Orientation and Steric Hindrance Induced by Alkyl Chain Length in Ammonium Salt Passivator Toward High-Performance Inverted Perovskite Solar Cells and Modules," *Advanced Materials* 37 (2025): 2413304.

34. L. Zheng, L. Shen, Z. Fang, et al., "Reducing the Surface Reactivity of Alkyl Ammonium Passivation Molecules Enables Highly Efficient Perovskite Solar Cells," *Advanced Energy Materials* 13 (2023): 2301066.

35. S. M. Lundberg and S.-I. Lee, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, (Curran Associates Inc., 2017).

36. L. Krückemeier, B. Krogmeier, Z. Liu, U. Rau, and T. Kirchartz, "Understanding Transient Photoluminescence in Halide Perovskite Layer Stacks and Solar Cells," *Advanced Energy Materials* 11 (2021): 2003489.

15214095, 0, Downloaded from https://advanced.onlinelibrary.wiley.com/doi/10.1002/adma.202522554 by Ulsan National Institute Of, Wiley Online Library on [23/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

37. C. C. Stoumpos, D. H. Cao, D. J. Clark, et al., "Ruddlesden–Popper Hybrid Lead Iodide Perovskite 2D Homologous Semiconductors," *Chemistry of Materials* 28 (2016): 2852–2867.

38. J. Zhang, J. Wu, Y. Zhao, et al., "Revealing the Crystallization and Thermal-Induced Phase Evolution in Aromatic-Based Quasi-2D Perovskites Using a Robot-Based Platform," *ACS Energy Letters* 8 (2023): 3595–3603.

39. S. Song, S. J. Yang, J. Choi, et al., "Surface Stabilization of a Formamidinium Perovskite Solar Cell Using Quaternary Ammonium Salt," *ACS Applied Materials & Interfaces* 13 (2021): 37052–37062.

40. X. Shen, X. Lin, Y. Peng, et al., "Two-Dimensional Materials for Highly Efficient and Stable Perovskite Solar Cells," *Nano-Micro Letters* 16 (2024): 201.

41. D. Weininger, "SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules," *Journal of Chemical Information and Computer Sciences* 28 (1988): 31–36.

42. D. Weininger, A. Weininger, and J. L. Weininger, "SMILES. 2. Algorithm for Generation of Unique SMILES Notation," *Journal of Chemical Information and Computer Sciences* 29 (1989): 97–101.

43. S. Rana, C. Li, S. Gupta, V. Nguyen, and S. Venkatesh, in *Proceedings of the 34th International Conference on Machine Learning*, 70, eds, P. Doina, T. Y. Whye. (PMLR, 2017).

44. F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research* 12 (2011): 2825–2830.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File: adma73087-sup-0001-SuppMat.docx.

15214095, 0, Downloaded from https://advanced.onlinelibrary.wiley.com/doi/10.1002/adma.202522554 by Ulsan National Institute Of, Wiley Online Library on [23/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License