Beyond Rapid Nucleation: Unveiling the Role of Solvent-Precursor Interactions in Antisolvent-Free Perovskite Fabrication

Yunfan Wang[1], Weiren Zhao[2], Juncheng Wang[3], Zhuoqiong Zhang[4,*], Ziyao Yue[1], Yuxuan Zhang[1], Yiting Jiang[5], Zixin Zeng[1], Jie Zeng[1], Hainam Do[6], Hong Lu[6,7], Yuanhang Cheng[8,*], Johnny C. Ho[1], Tom Wu[4], Hin Lap Yip[1,5], Dewei Zhao[3,*], Sai-Wing Tsang[1,9,*]

[1] *Department of Materials Science and Engineering, City University of Hong Kong, Hong Kong SAR 999077, P.R. China.*

[2] *Bioinformatics Center, Institute for Chemical Research, Kyoto University, Kyoto, Japan*

[3] *College of Materials Science and Engineering & Institute of New Energy and Low-Carbon Technology & Engineering Research Center of Alternative Energy Materials and Devices, Ministry of Education, Sichuan University, Chengdu, China.*

[4] *Department of Applied Physics, The Hong Kong Polytechnic University, Hong Kong SAR 999077, P.R. China.*

[5] *School of Energy and Environment, City University of Hong Kong, Hong Kong SAR 999077, P.R. China.*

[6] *Department of Chemical and Environmental Engineering, University of Nottingham Ningbo China, Ningbo 315100, China.*

[7] *Perovskite Research Institute, Yangzhou Precision Systems Inc., Yangzhou 225000, China.*

[8] *School of New Energy, Nanjing University of Science and Technology, Jiangyin, Jiangsu 21443, China.*

[9] *Center of Super-Diamond and Advanced Films (COSDAF), and Hong Kong Institute for Clean Energy, City University of Hong Kong, Hong Kong SAR 999077, P.R. China.*

[*]*Corresponding author: Sai-Wing Tsang (saitsang@cityu.edu.hk); Dewei Zhao (*dewei_zhao@hotmail.com*); Zhuoqiong Zhang (18481965@life.hkbu.edu.hk); Yuanhang Cheng (yhcheng@njust.edu.cn)*

These authors contributed equally: Yunfan Wang, Weiren Zhao, Juncheng Wang, Zhuoqiong Zhang

Experimental methods

Machine learning. Codes are available at *https://github.com/ZWR0/Solvent-Selection*.

Materials. All the organic solvents, including DMF, DMSO, NMP, DMI, DMPU, TMU, and NEP were purchased from Sigma-Aldrich. FAI, methylammonium chloride (MACl), and phenethylammonium iodide (PEAI) were purchased from GreatCell Solar. PCBM, BCP, and CsI were purchased from Xi'an Polymer Light Technology Corp. PbI_2, $PbCl_2$, and Me-2PACz were purchased from TCI Development Co., Ltd. The ITO glass was purchased from Advanced Electron Technology Co. Ltd.

AFP film fabrication. The precursor solution was prepared by dissolving 2 mmol $FA_{0.92}Cs_{0.0}8PbI_3$ (316.43 mg FAI, 41.57mg CsI, 922.02 mg PbI_2) and 0.1 mmol additive (6.75 mg MACl and 27.81 mg $PbCl_2$) in 1 mL DMF solvent. For the NMP-treated sample, 2 mmol (192 μL) NMP was added as an additive and labeled "NMP" or "NMP-treated". For the optimized precursor, 2 mmol mixed solvent (128 μL NMP and 80 μL DMPU, 2:1 mol ratio) was added, and the sample was labeled "DMPU" or "DMPU-treated". The prepared precursor was spin-coated on the glass or ITO substrate at 5000 rpm for 50 s. The substrate was then immediately annealed at 180 °C for 10 s, followed by 150 °C for 10 min. During the film fabrication, the T_A in the glove box was controlled by the water-cooling system and a hotplate, with a temperature variation of ± 0.5 °C at each stabilized temperature.

AFPSCs fabrication. The ITO substrate (with pattern) was cleaned with detergent, deionized water, and ethyl alcohol, respectively. The Me-2PACz layer was spin-coated on the ITO substrate (3000 rpm for 30 s), followed by annealing at 100 °C for 10 min. Then, the perovskite precursors were spin-coated on the Me-2PACz layer at 5000 rpm for 50 s. After annealing at 180 °C for 10 s, the film was transferred to another hot plate (150 °C, 10 min). The perovskite films were then spin-coated with PEAI at 4000 rpm for 30 s and annealed at 100 °C for 10 min for passivation. Afterward, the PCBM precursor was deposited on the perovskite layer at 1800 rpm for 40 s. Then, the BCP precursor was spin-coated at 5000 rpm for 40 s. Finally, 80 nm thick silver was deposited by thermal evaporation.

Large module fabrication. The fabrication process began with laser scribing to ablate the P1 pattern on a 30 cm × 30 cm FTO substrate. NiOx was then sputtered as the hole transport layer via magnetron sputtering, followed by slot-die coating of the perovskite precursor solution onto the FTO/NiOx substrate. The sample was rapidly transferred to a vacuum system (<10 Pa) for 30 seconds and subsequently annealed at 150 °C for 30 minutes. A 20 nm C60 layer was deposited via thermal evaporation, followed by ALD deposition of SnO_2. The P2 pattern was scribed near the P1 line, and a 100 nm Ag electrode was evaporated. Finally, the P3 pattern was scribed after Ag deposition.

Characterization. The SEM was performed by Field Emission Scanning Electron Microscope (Tescan MAIA3). The FTIR results were obtained from the PerkinElmer FTIR Spectrometer. The DLS was conducted by a dynamic light scattering particle size Analyzer (Malven Zetasizer Nano

ZS). The XRD tests were conducted by a D2 Phaser instrument using Cu Kα (λ = 0.154 nm) radiation. The PL imaging was measured by the WITec alpha300 R Raman System. The PL and TRPL were measured by Edinburgh Instruments FLS1000 System. The in-situ optical techniques were home-built as reported in the previous work, and the spectrum of the halogen lamp for in-situ absorption is presented in **Fig. S31**.[1-3]

The *J-V* curves were recorded using a Keithley SMU source meter under AM 1.5G illumination (IVS-KA6000 sunlight simulator from Enlitech) with a standard silicon solar cell (NREL-certified) as the reference, at a temperature of 25 °C in air. The cell illuminated area during measurement was 0.0468 cm_2, and the module illuminated area during measurement was 693 cm^2, which are determined by using a non-reflective metal mask. The *J-V* scan of cells was performed over a voltage range of -0.02 V to 1.30 V with a step size of 0.02 V and a delay time of 0.01 s. The *J-V* scan of modules was performed over a voltage range of -1.5 V to 42 V with a step size of 0.05 V and a delay time of 0.02 s. The EQE spectrum was obtained using a solar cell spectral response measurement system (QER3-011, Enli Technology Co., Ltd., Taiwan), with light intensity at each wavelength calibrated by a standard monocrystalline silicon cell. MPP tracking was carried out under a 1-sun white LED array in air.

DFT calculation. The electronic structure calculations were performed using first-principles DFT implemented in the Vienna Ab initio Simulation Package (VASP).[4] We employed the projector augmented-wave (PAW) method to describe electron-ion interactions. For exchange-correlation effects, the generalized gradient approximation (GGA) with the Perdew-Burke-Ernzerhof (PBE) functional was utilized throughout the study.[5,6] In the computational setup, the electronic wavefunctions were expanded using projector-augmented-wave formalism with an energy cutoff of 450 eV. For Brillouin zone integration, only the gamma-point was considered in all simulations. To minimize periodic perturbations from adjacent images, a large supercell (20 × 20 × 20 Å) was employed, with the cell dimensions kept fixed throughout subsequent calculations. To identify the global energy minimum that accurately simulated the interaction between solvent molecules and the target system, all initial models were fully relaxed until the maximum force on any atom fell below 0.05 $eV \cdot Å^{-1}$. During a complete structural optimization cycle, the electronic structure is first determined via a self-consistent field calculation, using a stringent energy convergence threshold of 10^{-4} eV. The resulting interatomic forces guide the generation of the next structural configuration. This iterative process between electronic structure determination and geometry updates continues until the system satisfies the convergence criteria, yielding the final optimized binding configuration. To properly account for dispersion forces, we incorporated van der Waals corrections through Grimme's DFT-D3 method with zero-damping.

Machine learning details

The workflow of the ML-guided solvent engineering strategy

As illustrated in **Fig. S1**, this study presents an integrated workflow that combines ML with experimental validation to accelerate the identification of optimal solvents within a defined chemical space, while also revealing the underlying mechanisms by which these solvents influence the crystallization of AFP. As shown in **Fig. S1a**, this workflow starts with systematic compilation and preprocessing of a comprehensive solvent dataset, emphasizing the critical role of both dataset quality and size. A high-quality dataset ensures reliable and accurate predictions, while a larger dataset broadens the chemical space explored, thereby increasing the likelihood of identifying optimal solutions.[7] As for rigorous preprocessing, it can ensure data reliability for subsequent analyses. Next, unsupervised ML algorithms are employed to analyze the intrinsic physicochemical descriptors, clustering solvents into groups with shared properties.[8] This clustering enables the targeted experimental validation of representative candidates, streamlining the optimization process (**Fig. S1b**). Feedback from the experimental results is then used to refine the dataset by integrating empirical observations, ensuring alignment between computational predictions and practical performance (**Fig. S1c**). Building on this validated dataset, supervised classification models are trained to predict solvent efficacy across the chemical space. High-throughput virtual screening prioritizes top candidates for detailed experimental characterization, enabling faster and more accurate identification of target solvents (**Fig. S1d**). This approach combines computational analysis with experiments, establishing a unified framework that speeds up solvent discovery (detailed code is given in the Methods).

Solvent selection

We initially compiled data for 348 solvents from published literature and retrieved their corresponding simplified molecular input line entry specification (SMILES) using PubChemPy.[9-14] The structural validity of each molecule was verified to ensure data accuracy. Physicochemical properties were calculated using RDKit, providing a comprehensive set of molecular descriptors.[15] Following preprocessing steps—including duplicate removal, missing value imputation, and standardization—a refined dataset was established. The final dataset comprises 348 solvents described by 168 physicochemical descriptors (**Fig. S2**), serving as the foundation for subsequent analyses. In general, physicochemical descriptors are subjected to dimensionality reduction using Principal Component Analysis (PCA) to simplify subsequent data processing.[16] However, we observed that the first two principal components (PC1: 15.59%, PC2: 8.89%) captured only 24.48% of the total variance after PCA processing, as shown in **Fig. S3**. This result indicates a substantial information loss (~75.52%) during the dimensionality reduction process. Furthermore, the correlation between descriptors was investigated, as shown in **Fig. S4**. The correlation heatmap reveals negligible correlations among most variables, highlighting the inherent multidimensionality of the dataset. Although PCA simplifies the data structure, it risks compromising critical information retention and exacerbating challenges associated with high-dimensional data. Therefore, unlike the conventional approach of performing PCA for dimensionality reduction prior to clustering, we opted to preserve maximal information by bypassing PCA and directly utilizing the standardized high-

dimensional dataset for subsequent analyses.[17] Based on the dataset, unsupervised clustering was conducted to elucidate relationships among solvents. The optimal number of clusters was determined using a consensus approach that combined the elbow method, silhouette score, and Calinski-Harabasz index, as shown in **Fig. S5**. This multi-criteria strategy ensured a reliable and robust determination of the cluster count for subsequent analyses. To further enhance accuracy, scores from these metrics were normalized and averaged across cluster numbers (2-50), ultimately identifying an optimal cluster count of 23 by the comprehensive approach. (**Fig. S6**).

Experimental validation

The DMSO is not suitable for AFP fabrication. Most of the grouped solvents in cluster 17th are suitable for AFP fabrication (**Fig. S8-11**).

Prediction

Clustering based solely on descriptors proved insufficient for accurately identifying target solvents. To improve prediction accuracy, experimentally validated solvent data were used as labels to train supervised learning classification models. The solvent descriptors were improved by incorporating both RDKit-calculated properties and experimentally derived features from PubChem, providing a more comprehensive representation of each solvent, thereby enhancing the robustness of the models and the reliability of their predictions. To construct and validate our machine learning models, we employed a suite of classification algorithms and an ensemble voting strategy to refine prediction outcomes. Solvents were experimentally labeled as 'good' or 'bad' based on their ability to facilitate the formation of high-quality, pinhole-free perovskite films and enhance the power conversion efficiency (PCE) of the resulting devices under antisolvent-free conditions. The labeled training set comprised representative 'good' solvents (e.g., NMP, DMI, TMU) and 'bad' solvents (e.g., NEP, DMSO), ensuring balanced representation. Given the limited size of the labeled dataset, we utilized Leave-One-Out Cross-Validation (LOOCV) and five-fold cross-validation (Fig. S12–13) to rigorously evaluate all models from the LazyPredict library. To ensure predictive robustness, an ensemble was constructed using only the **56 models** that demonstrated both exceptional performance and generalizability. Specifically, a model was included in the ensemble only if it achieved 100% accuracy on the test set during the validation phase (with a 5:1 training-to-test split). These selected high-performance models were then deployed to screen the entire database of 348 solvents. The final probability of a solvent being categorized as 'good' was quantified as the ratio of models predicting a positive label to the total number of ensemble models.

General Design Rules for Solvent Screening

“Based on the comprehensive ML framework and systematic experimental feedback, we have established a set of multi-dimensional, descriptor-based design rules for identifying the "Gold Solvent" in antisolvent-free (ASF) perovskite processing. These rules bridge data-driven predictions with a chemically intuitive screening protocol. The enrichment of nitrogen and oxygen lone pairs—quantified by descriptors such as fr_amide and fr_C_O_noCOO—ensures strong coordination and high binding energy with the lead-halide framework, as confirmed by our DFT calculations. Additionally, the PEOE-derived electrostatic surface area (PEOE_VSA12) provides a physical basis for stabilizing the precursor-solvent intermediate phase over a wide ambient temperature window. To facilitate the discovery of next-generation efficient and green solvent systems, we propose a

three-step high-throughput screening (HTS) protocol:

1. Functional Group Filtration (Coordination Capability) Candidate solvents must possess a high density of Lewis basic sites to ensure effective precursor interaction. The primary filter requires a frequency of amide or non-carboxylic carbonyl functionalities of ≥ 1 (i.e., fr_amide ≥ 1 or fr_C_O_noCOO ≥ 1)
2. Electronic Environment Refinement (Local Polarization) Sufficient local polarization is critical for the robust stabilization of PbI_2-solvent adducts. Solvents should feature a specific electrostatic surface area distribution characterized by PEOE_VSA12 > 2.5.
3. Solubility & Polarity Optimization (Phase Stability) To ensure high precursor solubility and optimal solvent-solute compatibility, the solvent must exhibit high hydrophilicity/polarity. A lipophilicity threshold of MolLogP < 1.0 is required to maintain a stable solvation shell during the drying process.

These quantitative guidelines not only explain the exceptional performance of the solvents identified in Cluster 17 (e.g., DMI, DMPU) but also offer a transferable and predictive framework for the future design of antisolvent-free perovskite fabrication systems."

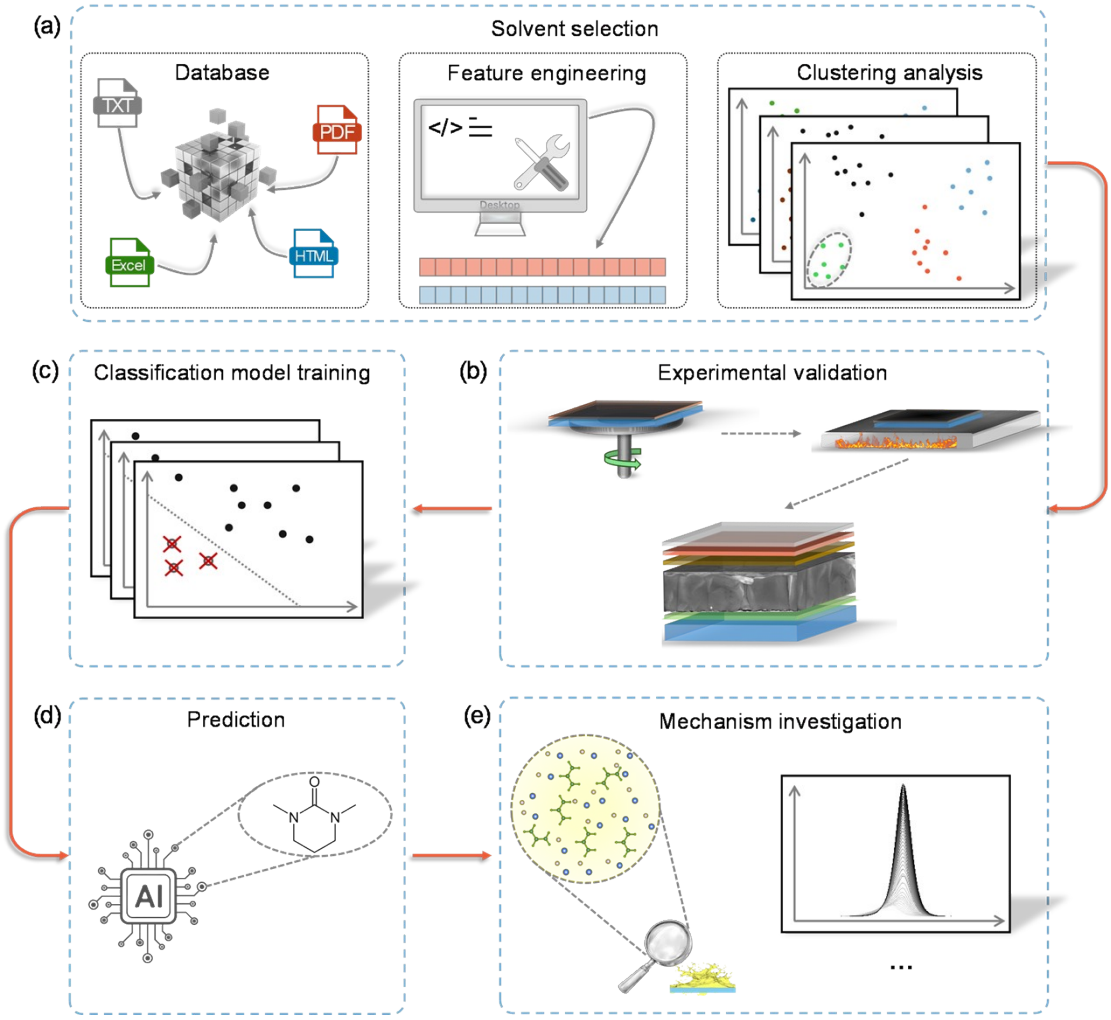
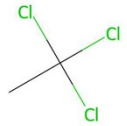
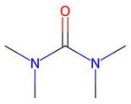
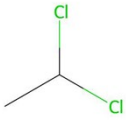
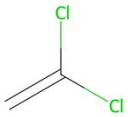


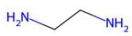
Fig. S1 The solvent dataset used in this work.

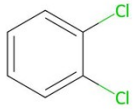

Solvent 1


Solvent 2

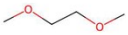

Solvent 3

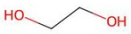

Solvent 4

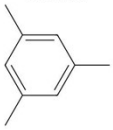

Solvent 5


Solvent 6


Solvent 7

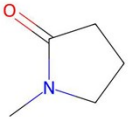

Solvent 8

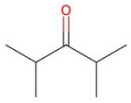

Solvent 9

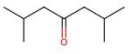

Solvent 10


Solvent 11


Solvent 12

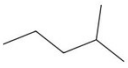

Solvent 13

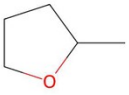

Solvent 14


Solvent 15

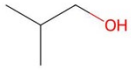

Solvent 16

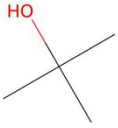

Solvent 17

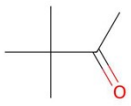

Solvent 18

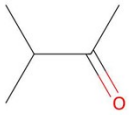

Solvent 19

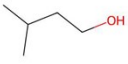

Solvent 20

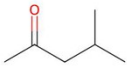

Solvent 21

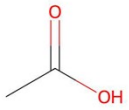

Solvent 22

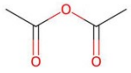

Solvent 23

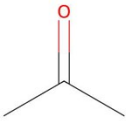

Solvent 24

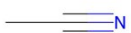

Solvent 25

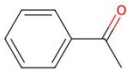

Solvent 26

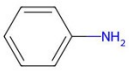

Solvent 27

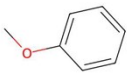

Solvent 28

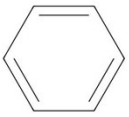

Solvent 29

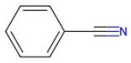

Solvent 30

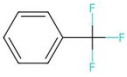

Solvent 31

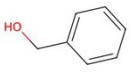

Solvent 32

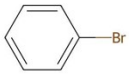

Solvent 33


Solvent 34

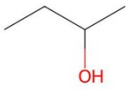

Solvent 35

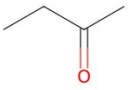

Solvent 36


Solvent 37


Solvent 38

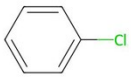

Solvent 39

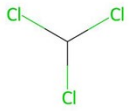

Solvent 40

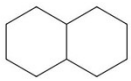

Solvent 41

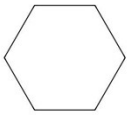

Solvent 42

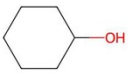
Solvent 43

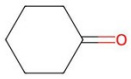

Solvent 44

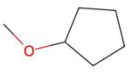

Solvent 45


Solvent 46

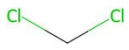

Solvent 47

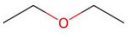

Solvent 48


Solvent 49

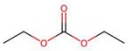

Solvent 50

Solvent 51


Solvent 52


Solvent 53


Solvent 54

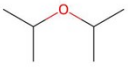

Solvent 55

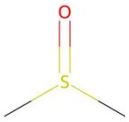
Solvent 56

Solvent 57

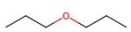
Solvent 58

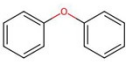

Solvent 59

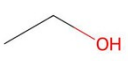

Solvent 60

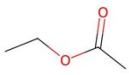

Solvent 61

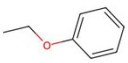
Solvent 62

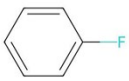

Solvent 63

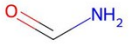

Solvent_64

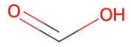

Solvent_65

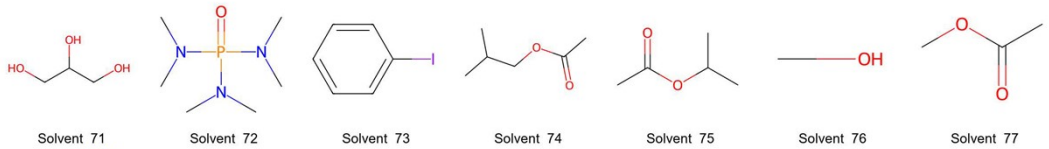

Solvent_66


Solvent_67

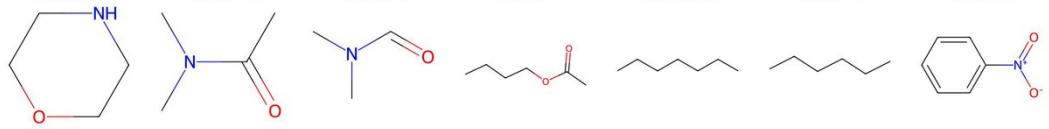

Solvent_68


Solvent_69

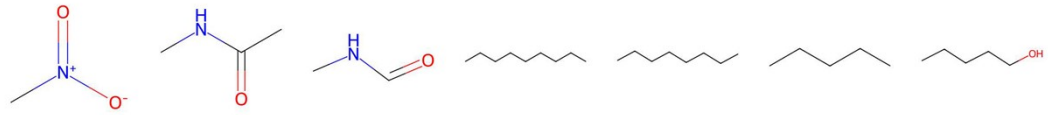

Solvent_70



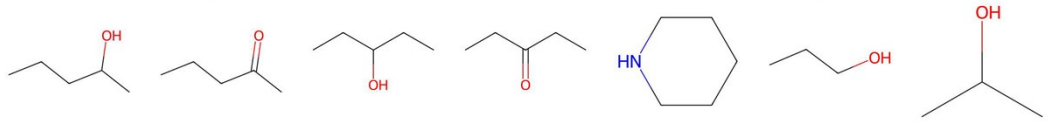
Solvent 71 Solvent 72 Solvent 73 Solvent 74 Solvent 75 Solvent 76 Solvent 77



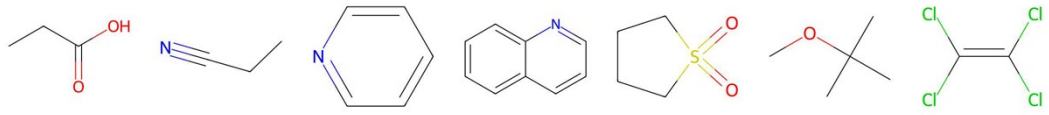
Solvent 78 Solvent 79 Solvent 80 Solvent 81 Solvent 82 Solvent 83 Solvent 84



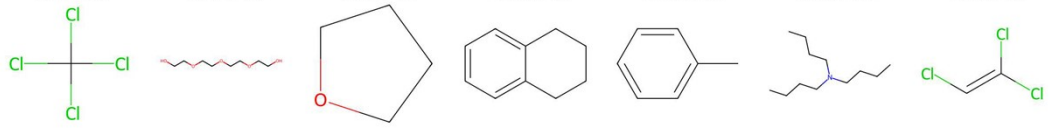
Solvent 85 Solvent 86 Solvent 87 Solvent 88 Solvent 89 Solvent 90 Solvent 91



Solvent 92 Solvent 93 Solvent 94 Solvent 95 Solvent 96 Solvent 97 Solvent 98

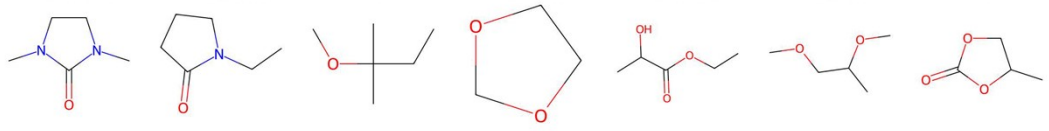


Solvent 99 Solvent 100 Solvent 101 Solvent 102 Solvent 103 Solvent 104 Solvent 105



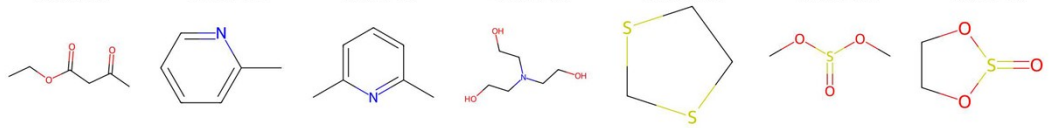
Solvent 106 Solvent 107 Solvent 108 Solvent 109 Solvent 110 Solvent 111 Solvent 112

Solvent 113 Solvent 114 Solvent 115 Solvent 116 Solvent 117 Solvent 118 Solvent 119

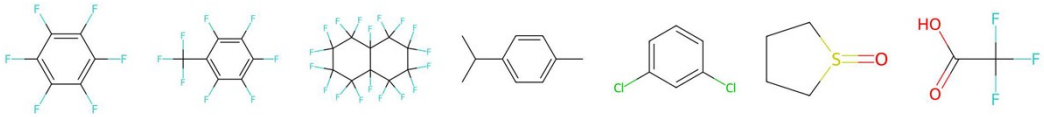


Solvent 120 Solvent 121 Solvent 122 Solvent 123 Solvent 124 Solvent 125 Solvent 126

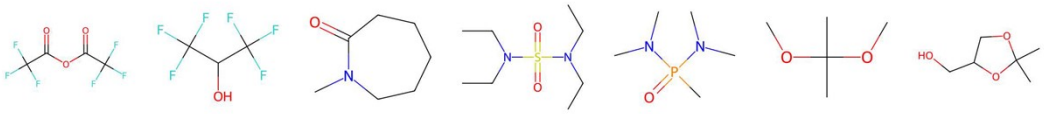
Solvent 127 Solvent 128 Solvent 129 Solvent 130 Solvent 131 Solvent 132 Solvent 133



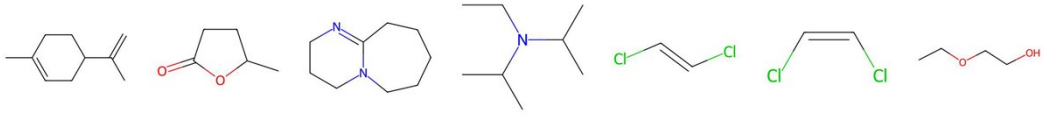
Solvent_134 Solvent_135 Solvent_136 Solvent_137 Solvent_138 Solvent_139 Solvent_140



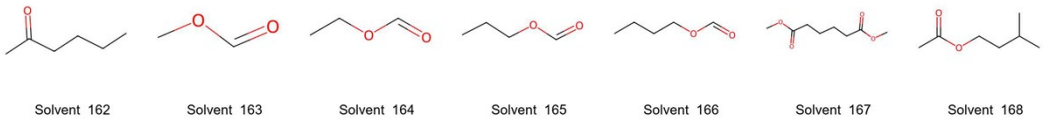
Solvent 141 Solvent 142 Solvent 143 Solvent 144 Solvent 145 Solvent 146 Solvent 147



Solvent 148 Solvent 149 Solvent 150 Solvent 151 Solvent 152 Solvent 153 Solvent 154

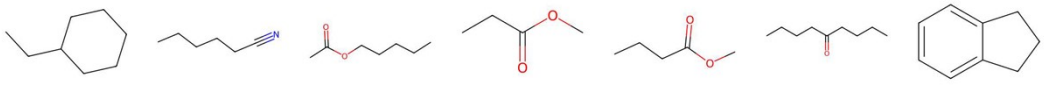


Solvent 155 Solvent 156 Solvent 157 Solvent 158 Solvent 159 Solvent 160 Solvent 161

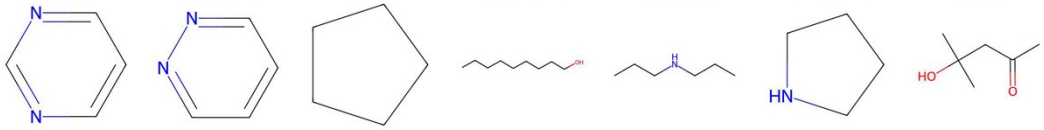


Solvent 162 Solvent 163 Solvent 164 Solvent 165 Solvent 166 Solvent 167 Solvent 168

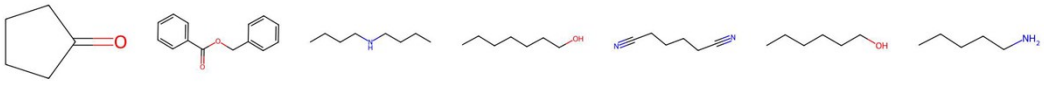
Solvent 169 Solvent 170 Solvent 171 Solvent 172 Solvent 173 Solvent 174 Solvent 175



Solvent 176 Solvent 177 Solvent 178 Solvent 179 Solvent 180 Solvent 181 Solvent 182

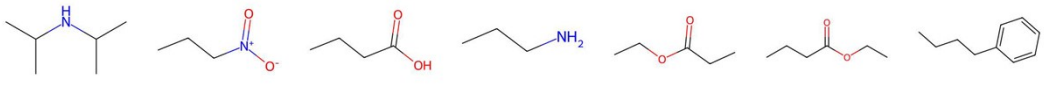


Solvent 183 Solvent 184 Solvent 185 Solvent 186 Solvent 187 Solvent 188 Solvent 189

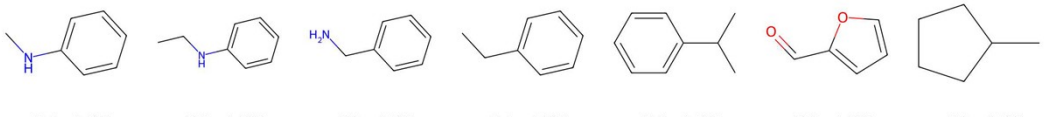


Solvent 190 Solvent 191 Solvent 192 Solvent 193 Solvent 194 Solvent 195 Solvent 196

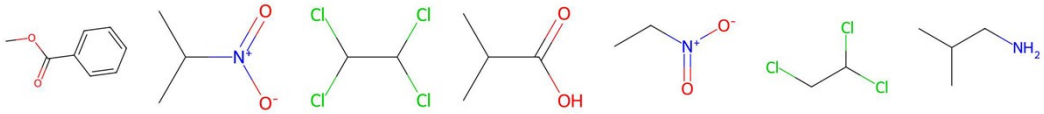
Solvent 197 Solvent 198 Solvent 199 Solvent 200 Solvent 201 Solvent 202 Solvent 203



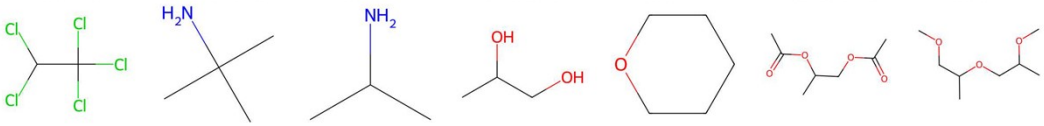
Solvent_204 Solvent_205 Solvent_206 Solvent_207 Solvent_208 Solvent_209 Solvent_210



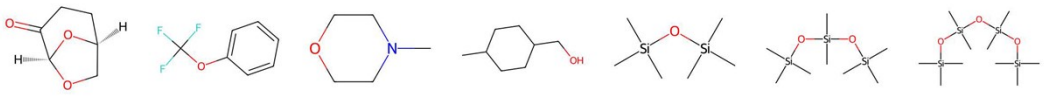
Solvent 211 Solvent 212 Solvent 213 Solvent 214 Solvent 215 Solvent 216 Solvent 217



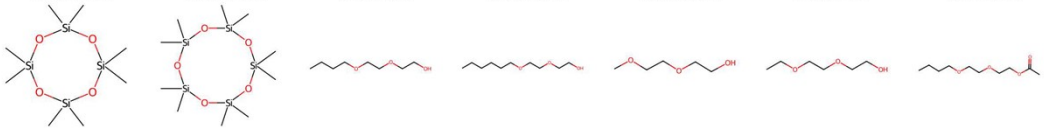
Solvent 218 Solvent 219 Solvent 220 Solvent 221 Solvent 222 Solvent 223 Solvent 224



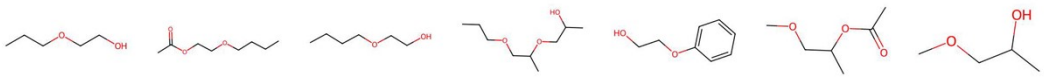
Solvent 225 Solvent 226 Solvent 227 Solvent 228 Solvent 229 Solvent 230 Solvent 231



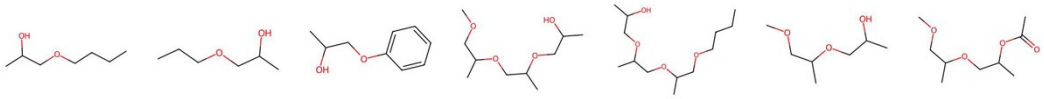
Solvent 232 Solvent 233 Solvent 234 Solvent 235 Solvent 236 Solvent 237 Solvent 238



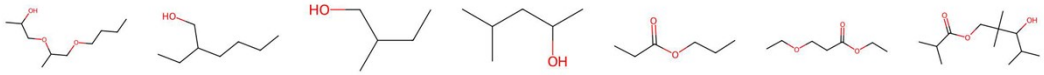
Solvent 239 Solvent 240 Solvent 241 Solvent 242 Solvent 243 Solvent 244 Solvent 245



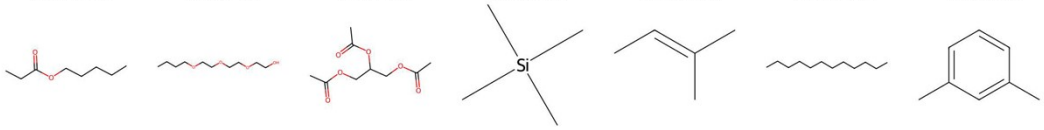
Solvent 246 Solvent 247 Solvent 248 Solvent 249 Solvent 250 Solvent 251 Solvent 252



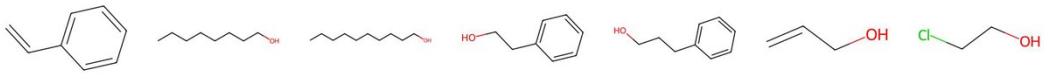
Solvent 253 Solvent 254 Solvent 255 Solvent 256 Solvent 257 Solvent 258 Solvent 259



Solvent 260 Solvent 261 Solvent 262 Solvent 263 Solvent 264 Solvent 265 Solvent 266



Solvent 267 Solvent 268 Solvent 269 Solvent 270 Solvent 271 Solvent 272 Solvent 273



Solvent_274 Solvent_275 Solvent_276 Solvent_277 Solvent_278 Solvent_279 Solvent_280

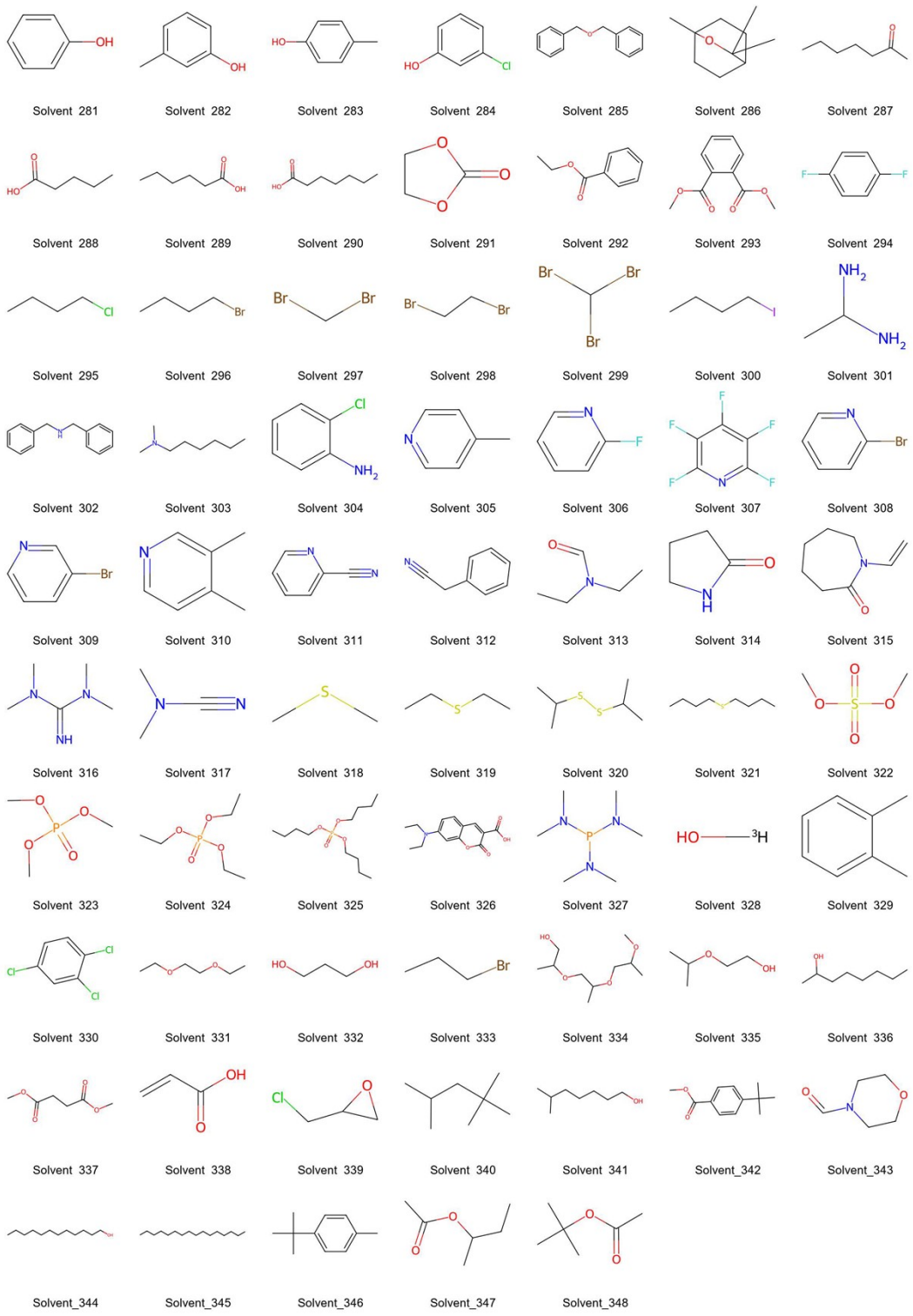


Fig. S2 The solvent dataset used in this work

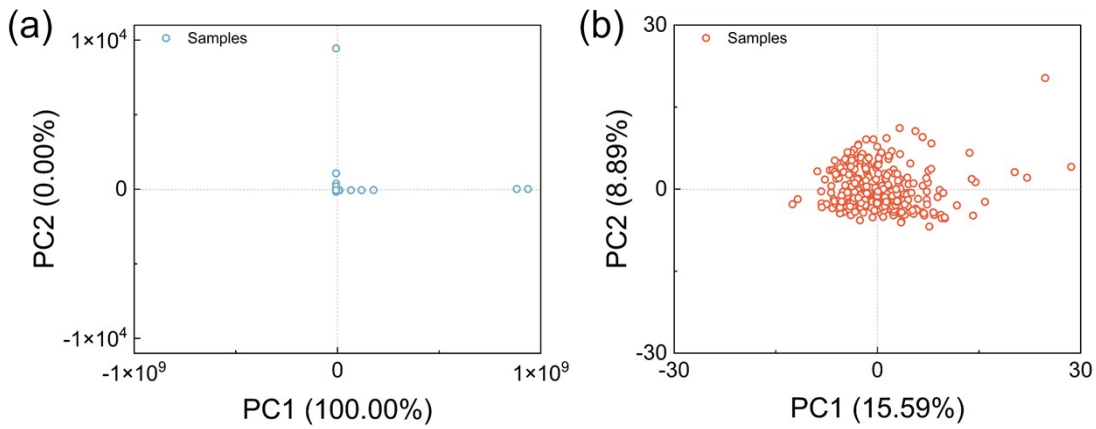


Fig. S3 Visualization of solvent data via CA dimensionality reduction.

It is essential to perform data standardization before PCA dimensionality reduction. Without standardization, PCA results can be misleading due to the disproportionate influence of variables with larger scales on the covariance matrix. For instance, in our unstandardized dataset, it was observed that the first principal component (PC1) accounted for 100% of the variance while the second principal component (PC2) contributed 0%, as shown in **Fig. S3**. This indicated significant scale differences across dimensions, where variables with larger magnitudes dominated the analysis, overshadowing other important yet smaller features within the dataset and reducing the overall interpretability and effectiveness of the PCA. Upon performing data standardization, PC1 and PC2 accounted for 15.59% and 8.89% of the total variance respectively. This adjustment highlights a more balanced representation of the major features within the dataset without any single variable dominating the PCA outcome.

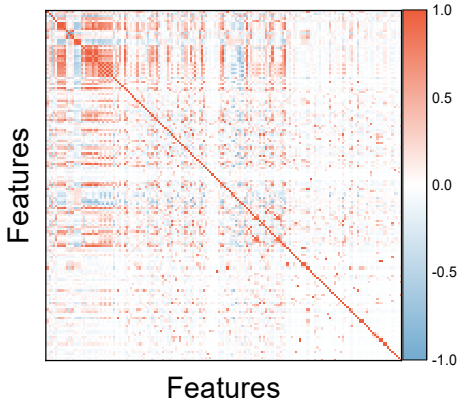


Fig. S4 Correlation matrix for the 168 selected descriptors of the 348 solvents.

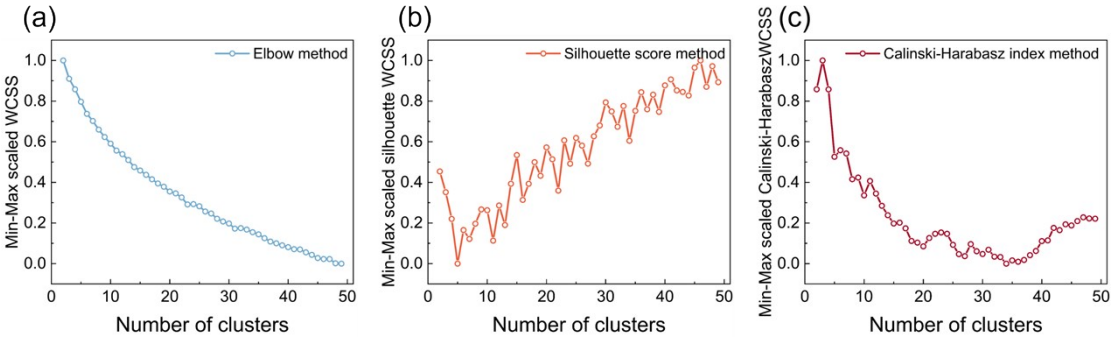


Fig. S5 Optimal number of clusters was determined using a consensus approach that combined the **(a)** elbow method, **(b)** silhouette score, and **(c)** Calinski-Harabasz index.

The KMeans algorithm was implemented using scikit-learn with the following parameters: init='k-means++', n_init=10, max_iter=300, and tol=1e-4. All features were standardized to zero mean and unit variance prior to clustering. To determine the optimal number of clusters (k), we evaluated cluster solutions across a range of k = 2 to 50. For each k, three widely used internal validation metrics were computed: Elbow Method, Silhouette Score and Calinski-Harabasz (CH) Index. Individual analyses suggested different optimal values: the Elbow method indicated $k \approx 14$, the Silhouette score peaked at k = 46, and the CH index reached its maximum at k = 3. To integrate these complementary criteria, we normalized each metric (min-max scaling) and computed a composite score as the sum of the normalized values. The value of k = 23 was selected as it maximized this aggregated score, representing a balanced trade-off among all three evaluation criteria.

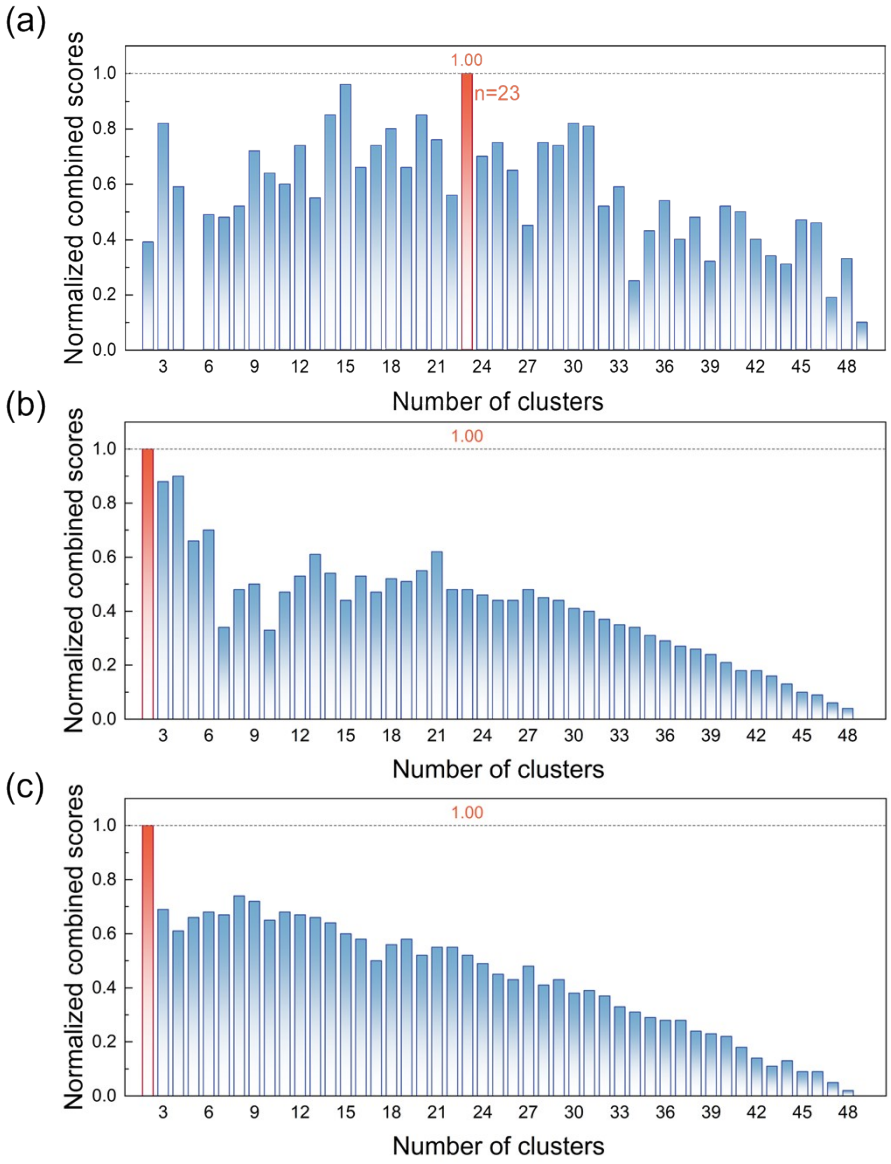


Fig. S6 Clustering results obtained by **(a)** RDKit, **(b)** Mordred descriptors, and **(c)** molecular fingerprints.

In addition to utilizing descriptors calculated by RDKit for clustering analysis, this study also explored the use of Mordred-derived physicochemical property descriptors and RDKit-generated molecular fingerprints representing molecular structural information for separate clustering analyses (**Fig. S7**). The determination of the optimal number of clusters was based on a weighted combination of the elbow method, silhouette score, and Calinski-Harabasz index, as previously described. The results indicated that clustering solely based on Mordred descriptors, which capture the physicochemical properties of molecules, or based on molecular fingerprints, which reflect molecular structural information, yielded suboptimal results. Specifically, both approaches could only cluster the 348 solvent molecules into two groups, providing limited insights for subsequent analyses. In contrast, clustering using the 168 descriptors calculated by RDKit, which encompass both physicochemical and structural information, resulted in more meaningful clusters, offering valuable insights that facilitate further analysis.

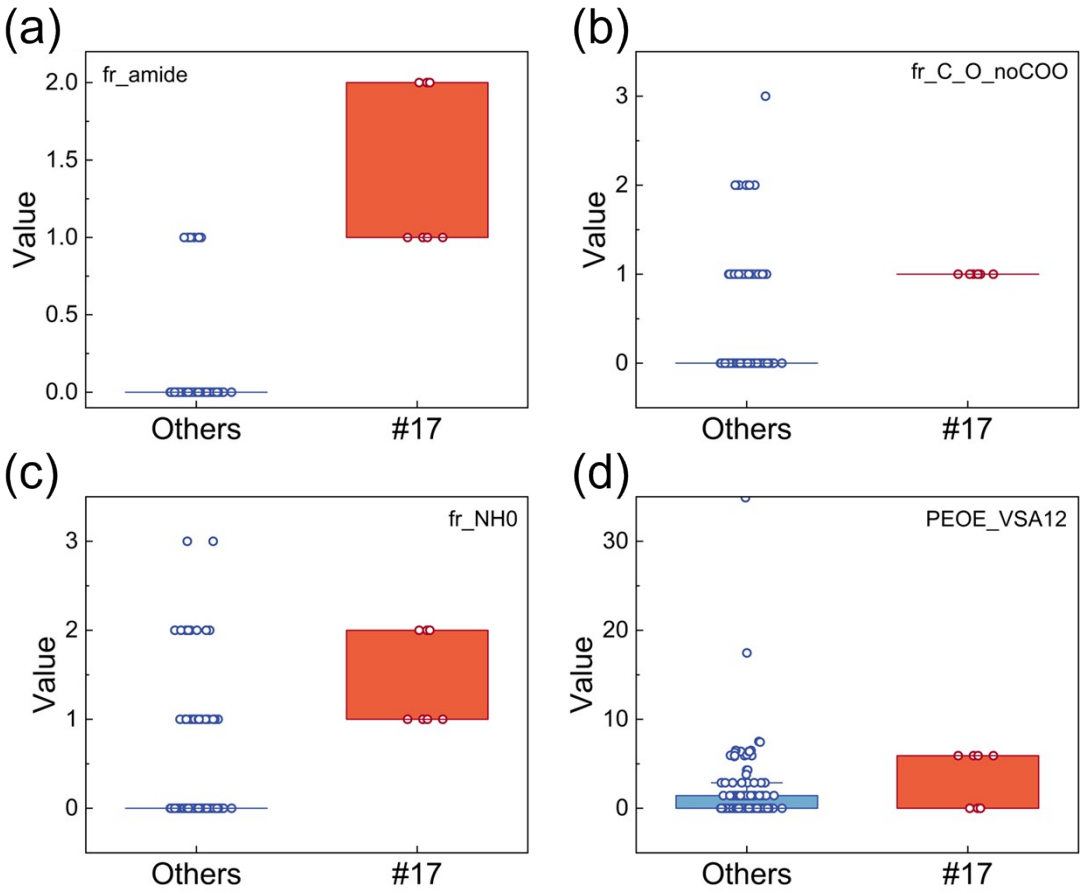


Fig. S7 Statistical distribution of key molecular descriptors for Cluster 17 and the global dataset

The boxplots illustrate the significant enrichment of coordination-related functional groups and electronic features in Cluster 17. **(a)** fr_amide and **(b)** fr_C_O_noCOO reflect the high density of Lewis basic sites (carbonyl and amide groups) that facilitate strong solvent-precursor coordination. **(c)** fr_NH0 indicates the prevalence of tertiary structures, which are essential for forming stable intermediate phases. **(d)** PEOE_VSA12 represents the specific surface area with high partial charge, correlating with the localized polarization required for effective solvation of Pb^{2+} ions.

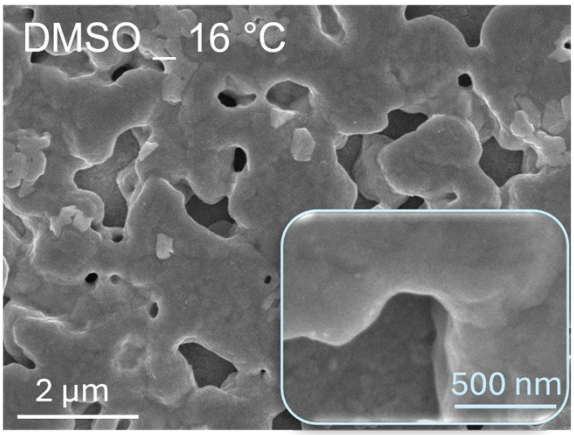


Fig. S8 SEM images of DMSO-treated AFP films fabricated at a T_A of 16 °C.

Although the DMSO has been widely used in other perovskite solution-processed methods, it is unsuitable for the antisolvent-free system, consistent with the solvent clustering results by ML.

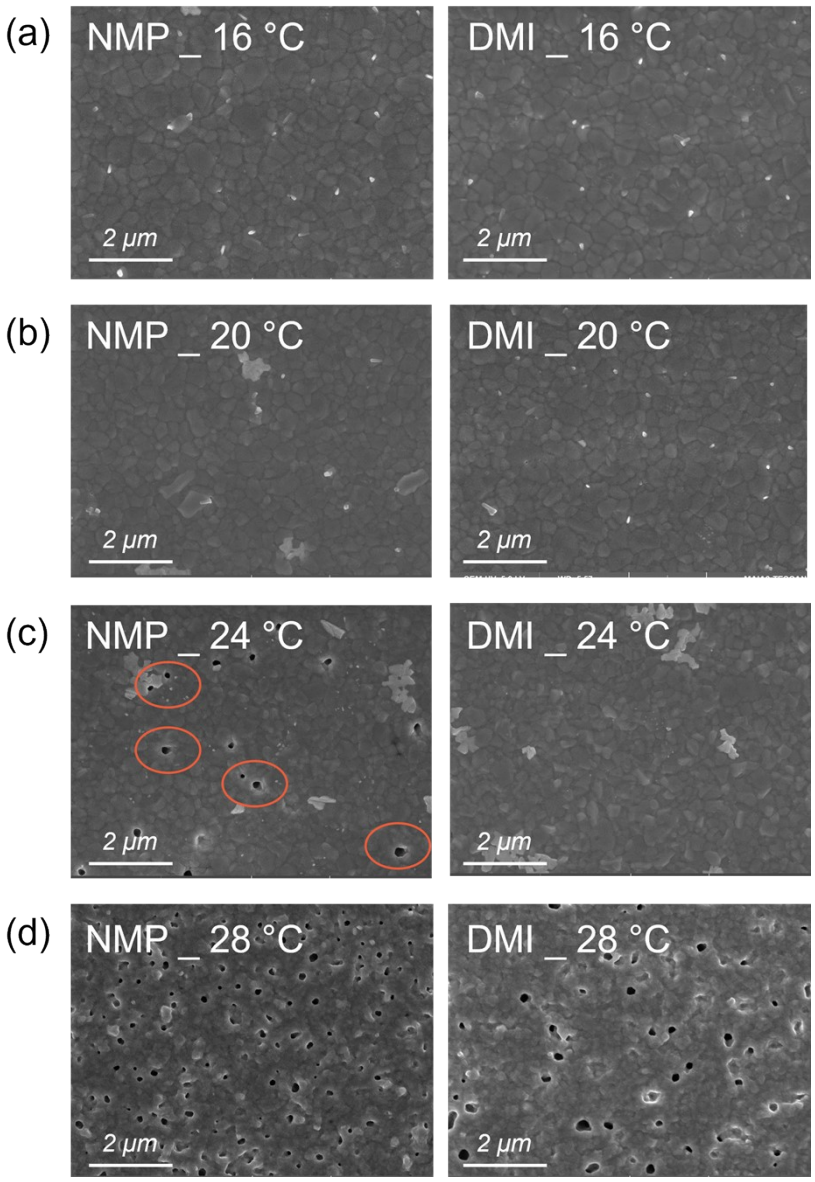


Fig. S9 SEM images of different solvent-treated AFP films fabricated at various T_A. The scale bar is 2 μm.

Fig. S9 shows the top-view scanning electron microscopy (SEM) images of AFP films fabricated with precursor solutions using different solvents at a T_A of 16 °C. The films fabricated with NMP and DMI exhibit smooth, dense surfaces with grain sizes of approximately 500 nm. On the other hand, the pin-hole has even been formed on the NMP-treated films at a T_A of 24 °C, while DMI can effectively suppress pinhole formation at this T_A.

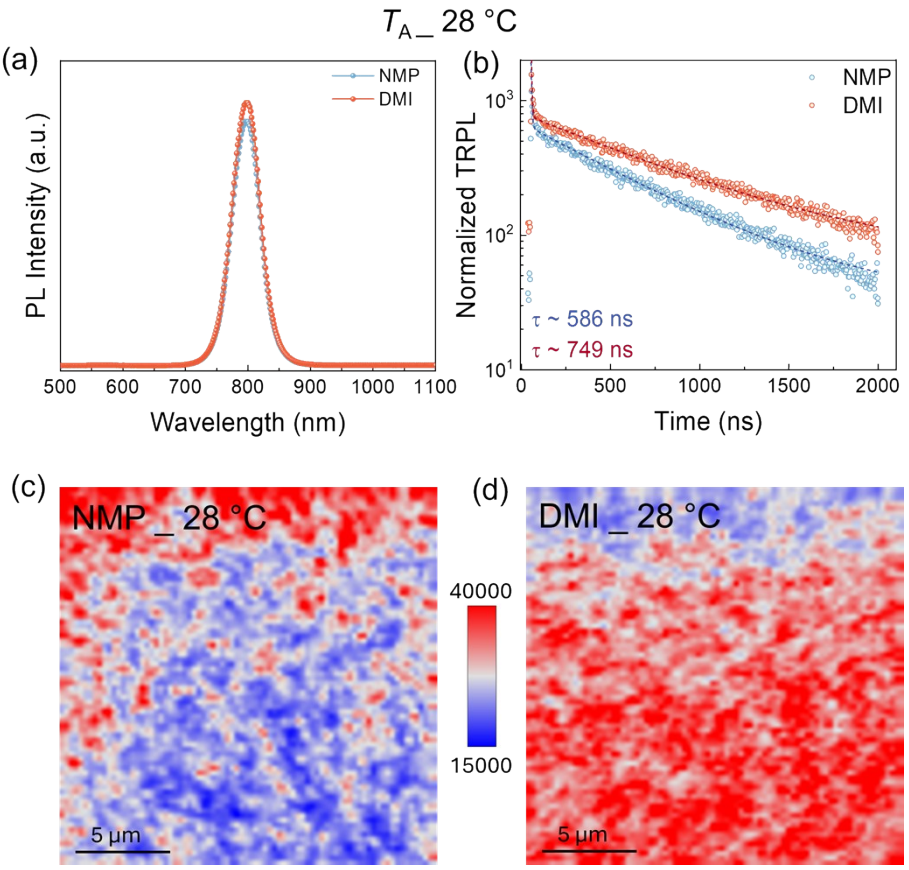


Fig. S10 a, Steady PL results, **(b)** TRPL results, and **(c-d)** PL intensity mapping images of different solvents-treated AFP films fabricated at a T_A of 28 °C. The scale bar is 5 μm.

Pinholes in the perovskite film generally increase defect states, resulting in significant non-radiative recombination. As shown in **Fig. S10a**, the DMI-treated film demonstrates higher steady-state photoluminescence (PL) intensity compared to that in the NMP-treated film, indicating a reduction in non-radiative recombination. Meanwhile, time-resolved PL measurements further confirm an extended PL lifetime of 749 ns for the DMI-treated film, which is longer than that of the NMP-treated film (586 ns), as shown in **Fig. S10b**. These enhancements in PL intensity and carrier lifetime are consistent with the higher quality and pinhole-free morphology of the DMI-treated films. In addition, PL intensity mapping images (**Fig. S10c-d**) also demonstrate a more uniform and higher fluorescence intensity in the DMI-treated film, reflecting homogeneous grain distribution and reduced non-radiative recombination, consistent with the steady-state PL results.

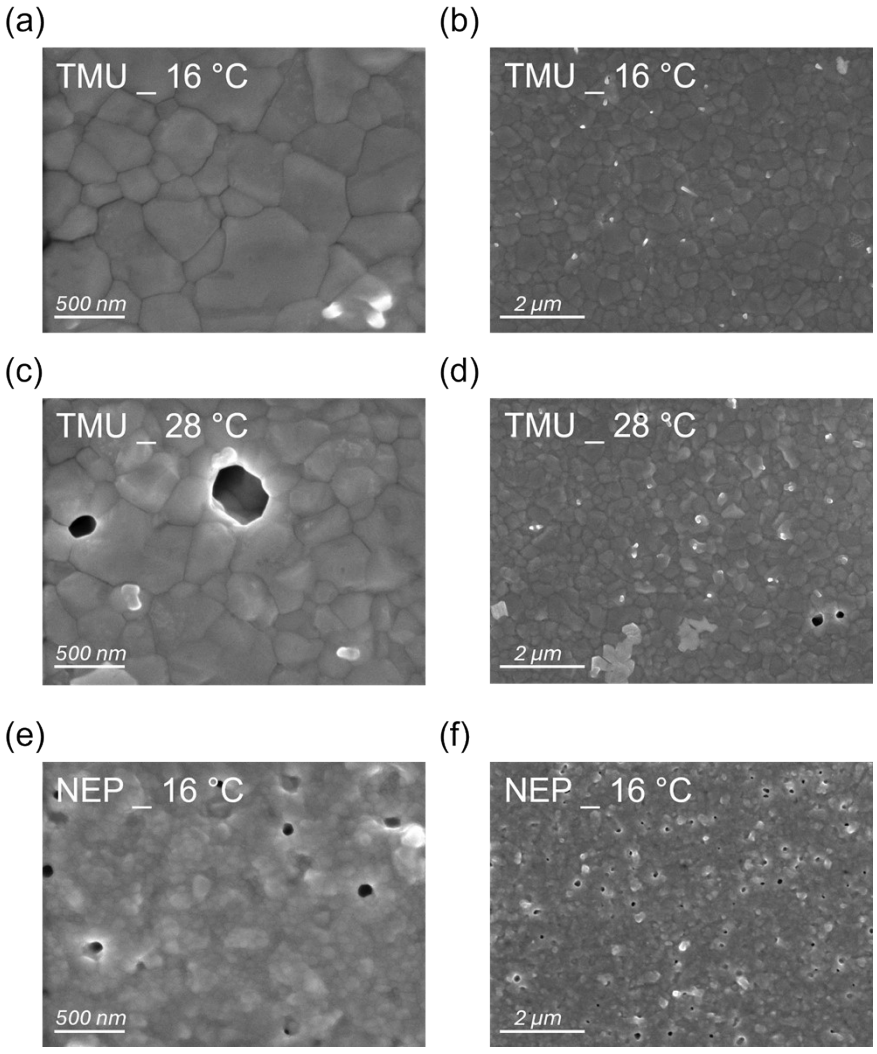


Fig. S11 SEM images of TMU-treated and NEP-treated AFP films fabricated at different T_A conditions.

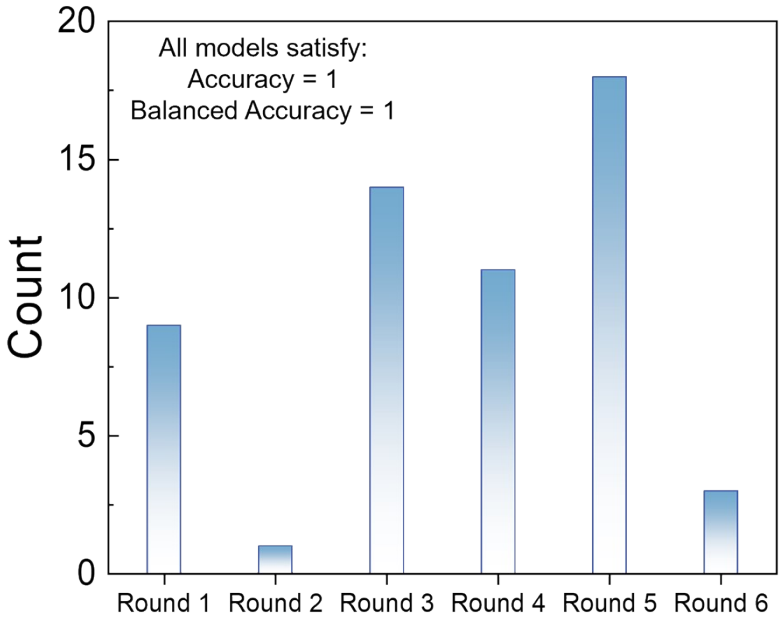


Fig. S12 Number of models with perfect accuracy in each training round.

In the process of training and selecting classification models, we employed a 5-fold cross-validation approach to randomly partition the solvent training dataset. The random seed was set to 42 to ensure the reproducibility of experimental results. To rapidly evaluate the performance of various models, we utilized the `lazypredict` Python package to train the partitioned data. From each fold, the model achieving the highest test accuracy (100%) was selected and recorded for subsequent multi-model ensemble prediction. Given the small size of the dataset (only six data points), leave-one-out cross-validation (LOOCV) was further applied to retrain and evaluate the previously recorded models over six iterations. This process yielded 56 models across different training folds that achieved perfect classification accuracy (test accuracy = 1). These high-performing models were retained for downstream prediction tasks. For predicting unknown solvents, we adopted a model voting-based strategy: the ratio of the number of models predicting a given solvent as a "good solvent" to the total number of models was used as the probability estimate for that solvent being a "good solvent." Subsequently, we ranked all solvents based on the calculated probabilities and screened out the most promising candidates. This ensemble method not only fully leveraged the predictive power of multiple models but also effectively enhanced the stability and reliability of the prediction results.

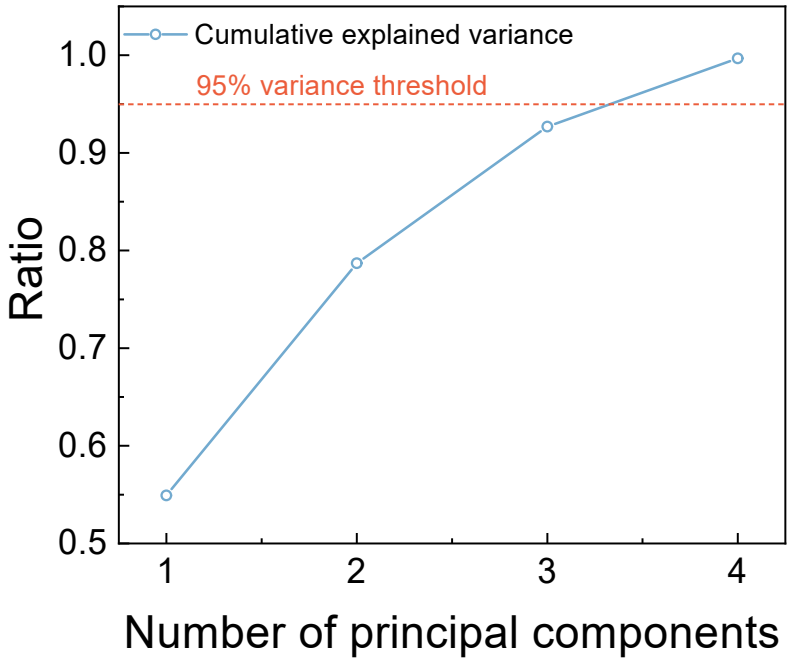


Fig. S13 Cumulative explained variance of principal components.

To enable a more comprehensive comparison, we performed PCA on the six samples and used the resulting low-dimensional features to train classification models under leave-one-out cross-validation (LOOCV). This yielded 24 models with perfect accuracy (1.0), which is fewer than the number obtained using the original (non-PCA) features.

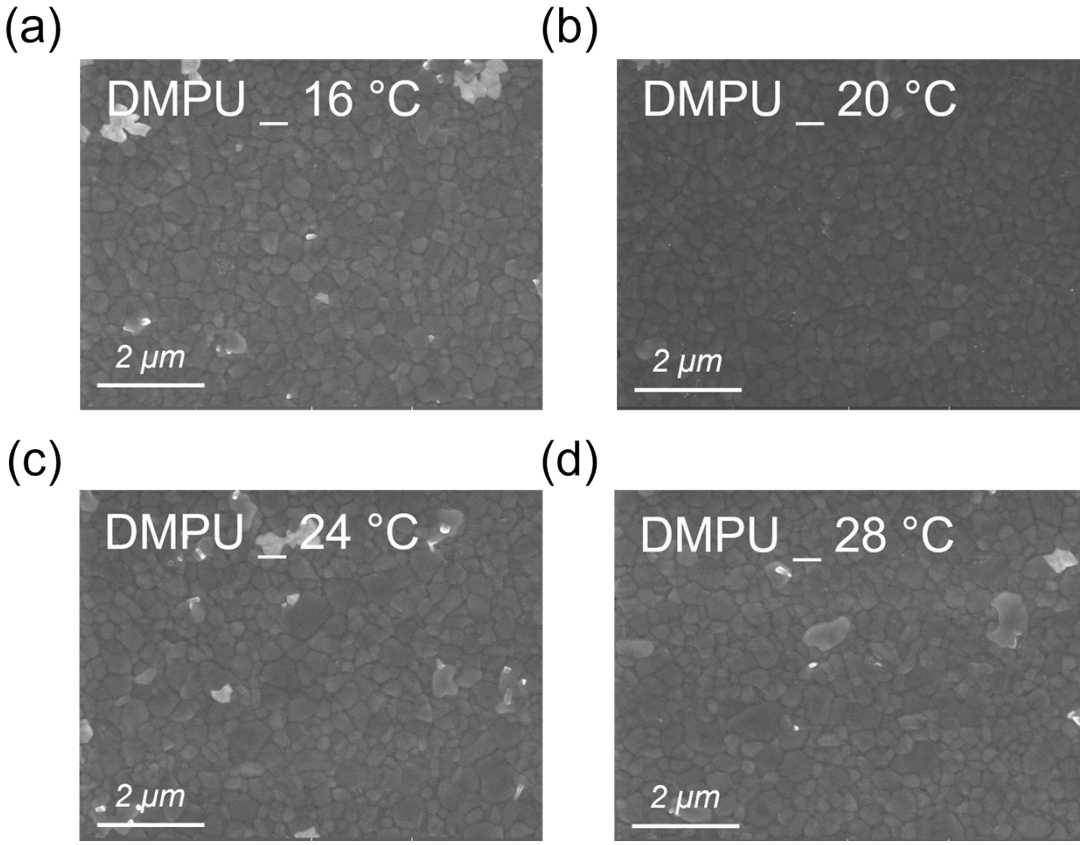


Fig. S14 SEM images of DMPU-treated AFP films fabricated at different T_A. The scale bar is 2 μm.

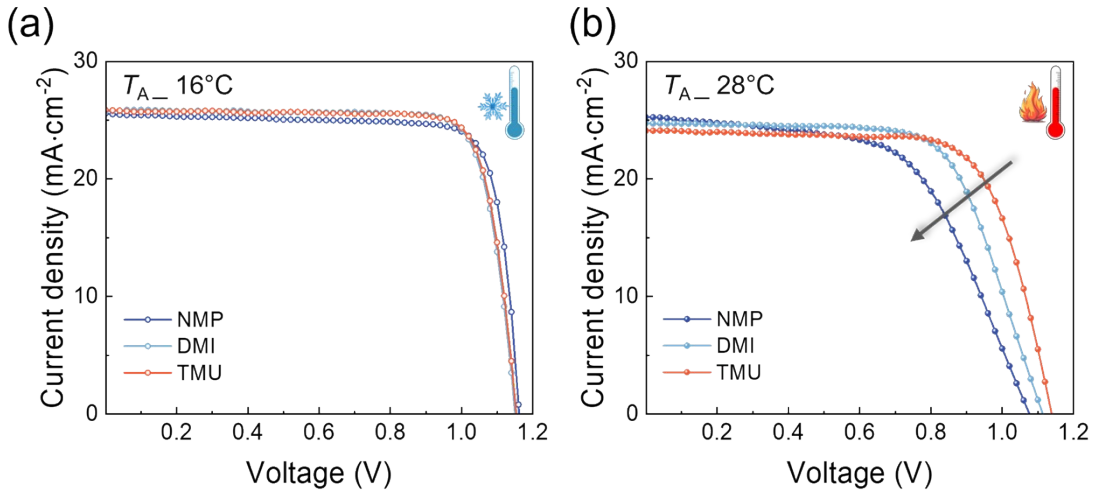


Fig. S15 *J-V* curves of different solvent-treated AFPSCs fabricated at a T_A of **(a)** 16 °C and **(b)** 28 °C.

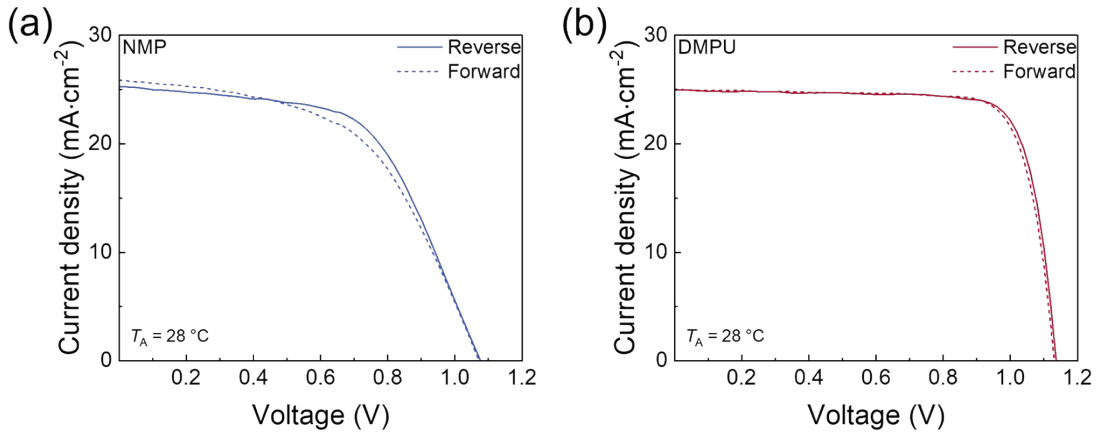


Fig. S16 The Forward and reverse scanning *J-V* curves of **(a)** NMP- and **(b)** DMPU-treated AFPSCs fabricated at a T_A of 28 °C.

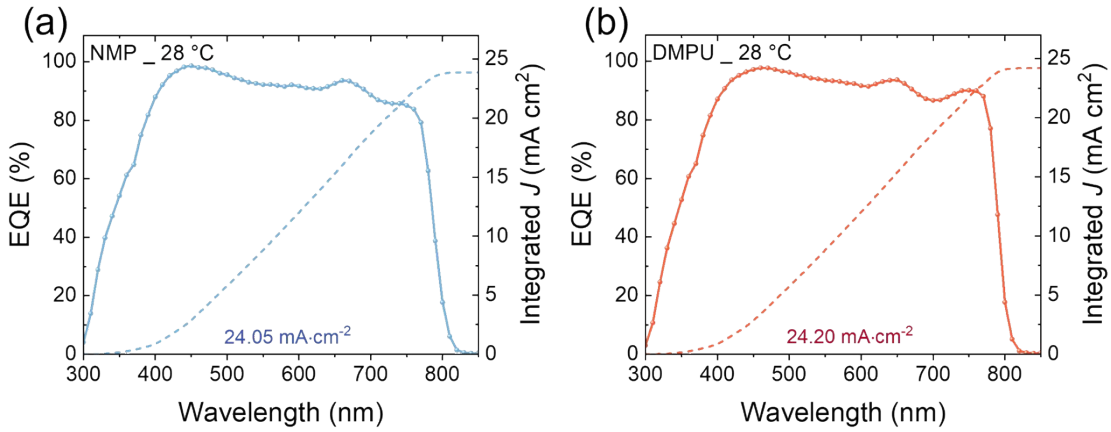


Fig. S17 The external quantum efficiency spectra of **(a)** NMP- and **(b)** DMPU-treated AFPSCs fabricated at a T_A of 28 °C.

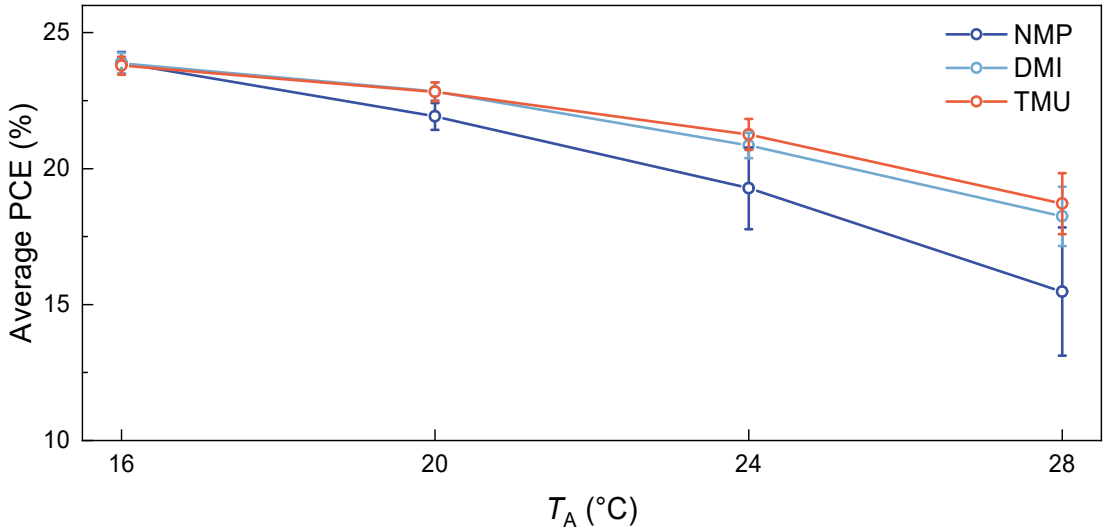


Fig. S18 PCE statistics of different solvent-treated AFPSCs fabricated at different T_A conditions.

According to **Fig. S15** and **S18**, in addition to DMPU, the other ML-selected solvents (DMI and TMU) also demonstrated the ability to widen the T_A processing window of AFPSCs.

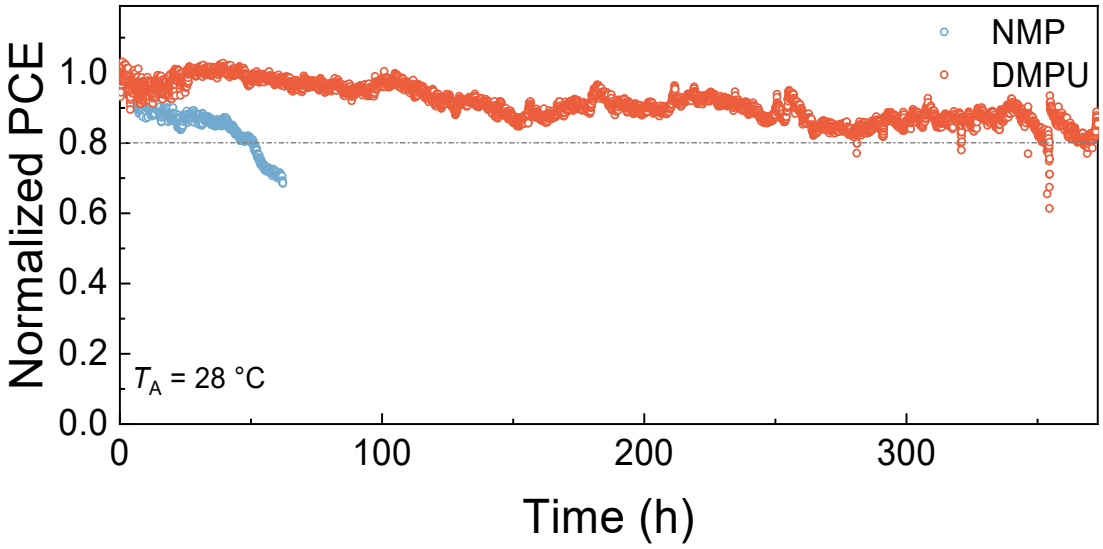


Fig. S19 The maximum power point tracking of encapsulated devices fabricated at a T_A of 28 °C.

We performed maximum power point (MPP) tracking of encapsulated AFPSCs. The DMPU-treated AFPSC maintains over 80% of its initial PCE after more than 350 h of continuous operation in the air (25 °C, 50 - 60% RH), indicating better stability.

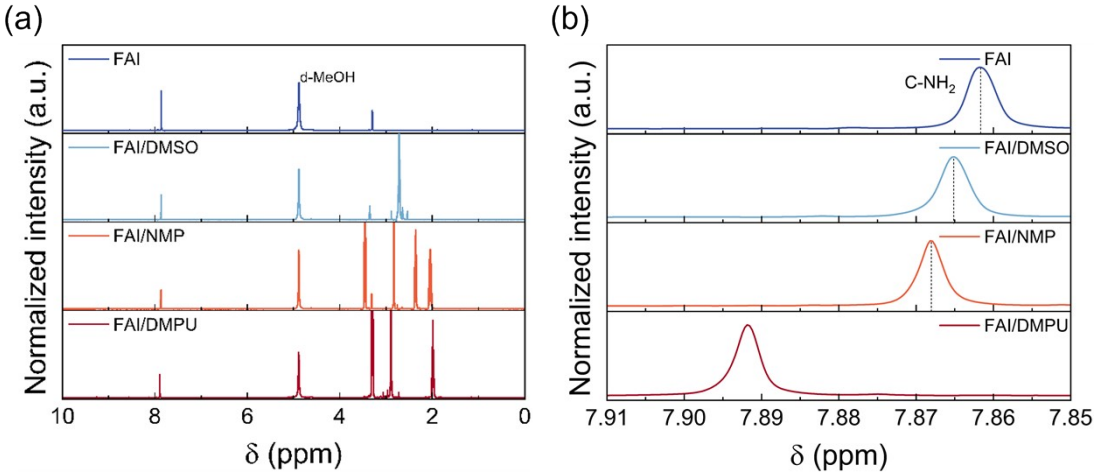


Fig. S20 (a) Complete ^{1}H NMR spectra of FAI, FAI/DMSO, FAI/NMP, and FAI/DMPU. **(b)** The C-NH_2 group in ^{1}H NMR spectra. During the ^{1}H NMR test, deuterated methanol (d-MeOH) was used as the solvent. Its characteristic peak is located at approximately 4.87 ppm

To further investigate the interaction between FAI and the solvent, we have provided the nuclear magnetic resonance (1H NMR) spectra results. As shown in Fig. R1, the C=NH_2 peak of FAI at δ = 7.86 ppm shifted to a downfield (higher ppm) position after mixing with NMP or DMPU, compared to DMSO. This indicates that more electron cloud of the hydrogen atoms in FA is pulled toward the solvent molecules, supporting the stronger hydrogen bonding interactions between NMP/DMPU and FA+ than between DMSO and FA^{+}, which is in line with the FTIR and DFT calculation results.

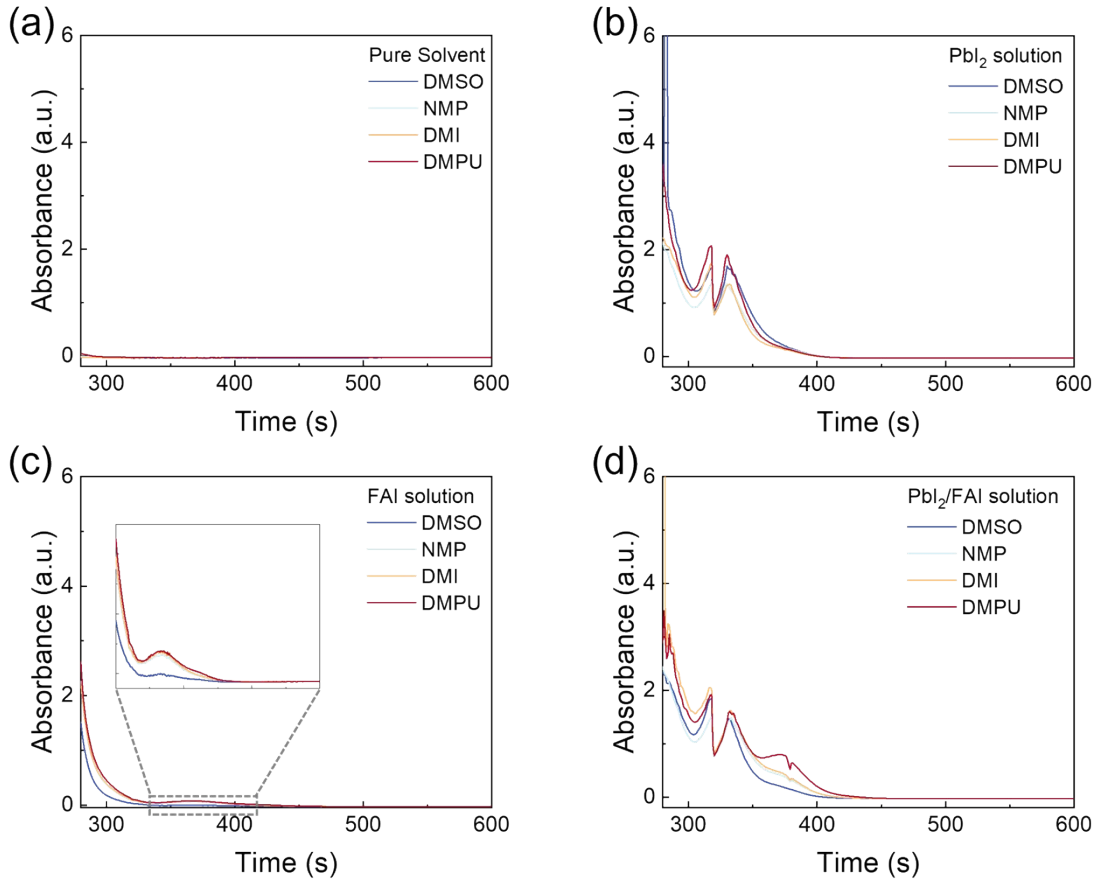


Fig. S21 UV-vis absorption spectra of the **(a)** pure solvents, **(b)** PbI_2 solutions, **(c)** FAI solutions, **(d)** FAI/PbI_2 solution.

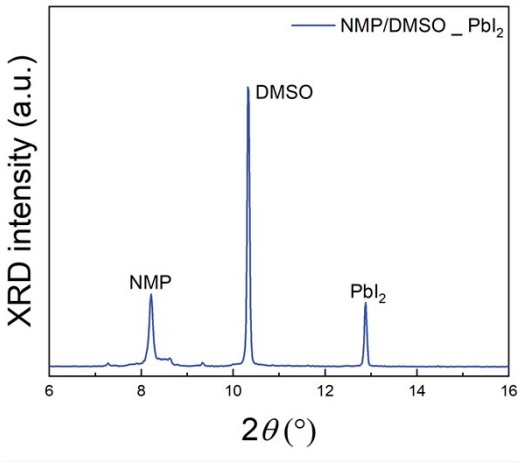


Fig. S22 XRD spectra of the solvent-adduct fabricated from the precursor with NMP/DMSO and PbI_2.

As shown in **Fig. S22**, the diffraction intensity of the DMSO-based adduct is higher than that of the NMP-based adducts, indicating that PbI_2 is inclined to form adducts with DMSO in the mixture solution.

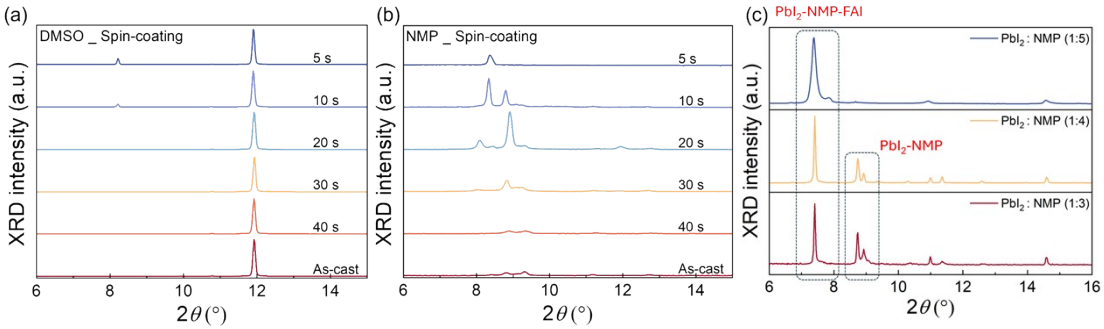


Fig. S23 Time-resolved XRD spectra of the **(a)** DMSO-treated and **(b)** NMP-treated AFP films during the spin-coating process. The T_A is 16 °C. **(c)** The XRD spectra of the film prepared from the precursor with PbI_2, FAI, and various NMP contents.

Such a stable adduct as an intermediate phase can significantly affect the phase evolution of FA-based AFP during the spin-coating process, as shown in **Fig. S23**. Time-resolved XRD measurements reveal distinct phase evolution of DMSO- and NMP-treated AFP films. For the DMSO-treated film, peaks corresponding to the intermediate phase are observed only within the first 10 seconds, accompanied by a prominent δ-phase peak. In contrast, the intermediate phase in the NMP-treated film remains stable throughout the entire spin-coating process, indicating that the NMP-based adduct is more stable in an FA-rich environment, consistent with the above static XRD results. However, such time-resolved XRD measurements in air are limited by low temporal resolution and are easily affected by moisture, interfering with the analysis of AFP phase evolution. To overcome these challenges and gain accurate insights into how these adducts influence AFP crystallization, in situ optical techniques were employed in an N_2 glovebox. As shown in **Figure S23c**, as the NMP content decreased, shifting the molar ratio of PbI_2:NMP from 1:5 to 1:3, the relative concentration of PbI_2 increased. This higher PbI_2 concentration promotes the formation of the NMP-PbI_2 adduct phase, which explains the XRD differences between **Figure 3b** and **Figure S23b**.

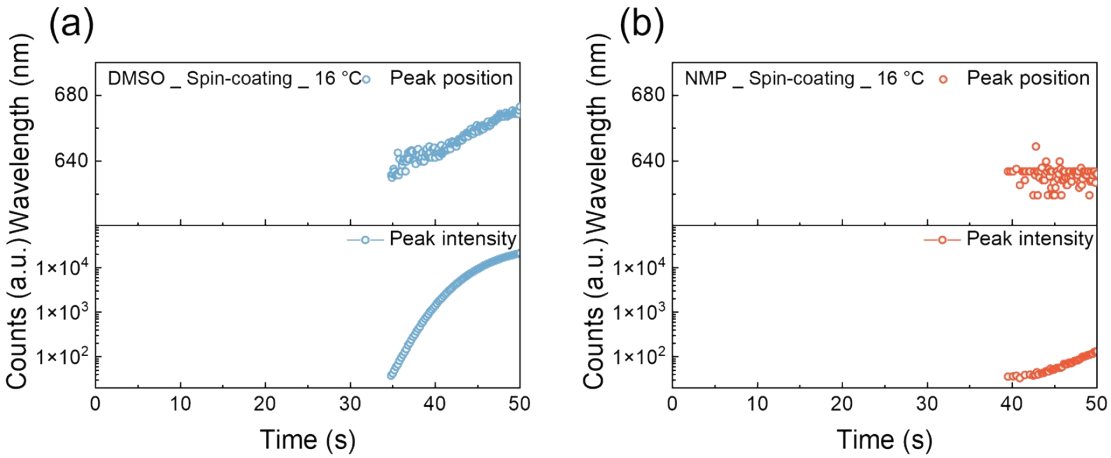


Fig. S24 PL intensity and peak position evolution of **(a)** DMSO-treated and **(b)** NMP-treated AFP films during the spin-coating process. The T_A is 16°C.

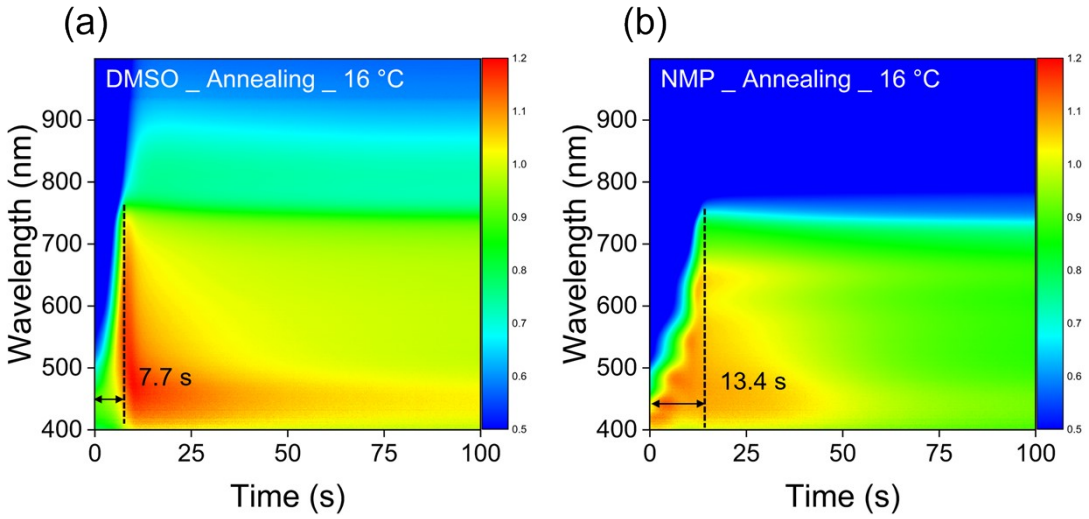


Fig. S25 In-situ absorption results of **(a)** DMSO-treated and **(b)** NMP-treated AFP films during the annealing process. The T_A is 16°C.

The in-situ absorption results during the subsequent thermal annealing further support the different nucleation dynamics between the NMP and DMSO-treated samples. The full transition to the perovskite bulk phase in the NMP-treated sample is completed at 13.4 s, as opposed to 7.7 s for the DMSO-treated sample, indicative of the retention of NMP-based adducts and retarded intermolecular exchange process in the NMP-treated sample.

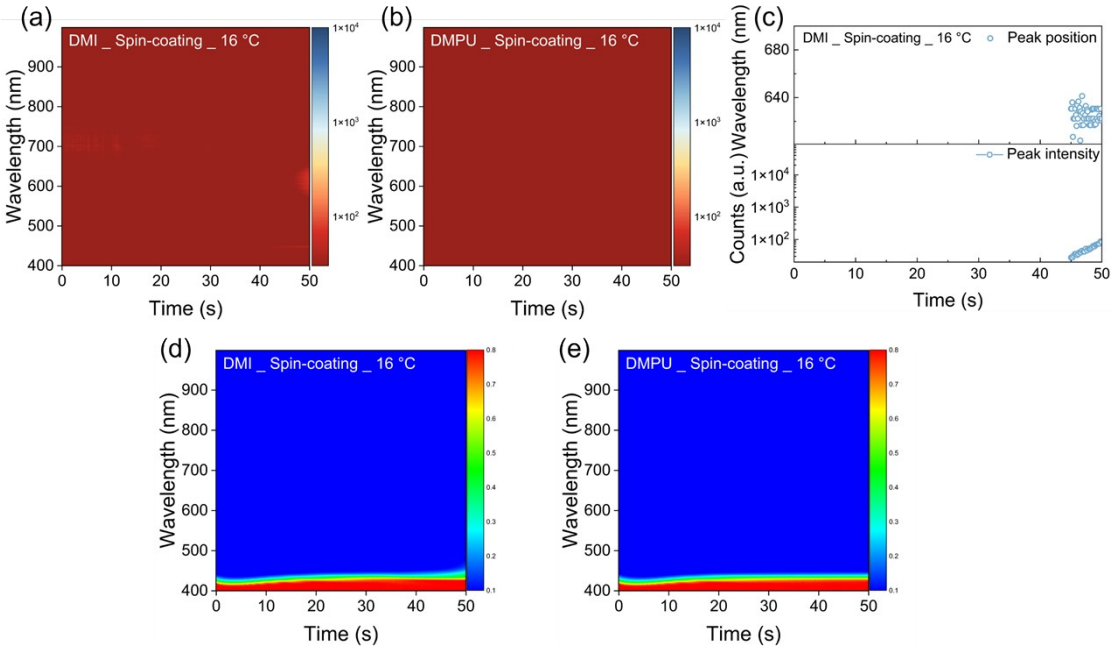


Fig. S26 In-situ **(a-b)** PL and **(d-e)** absorption results of different solvent-treated AFP films during the spin-coating process. **c**, PL intensity and peak position evolution of DMI-treated AFP films during the spin-coating process. The T_A is 16°C.

No PL signal was detected in DMPU-treated film during spin-coating.

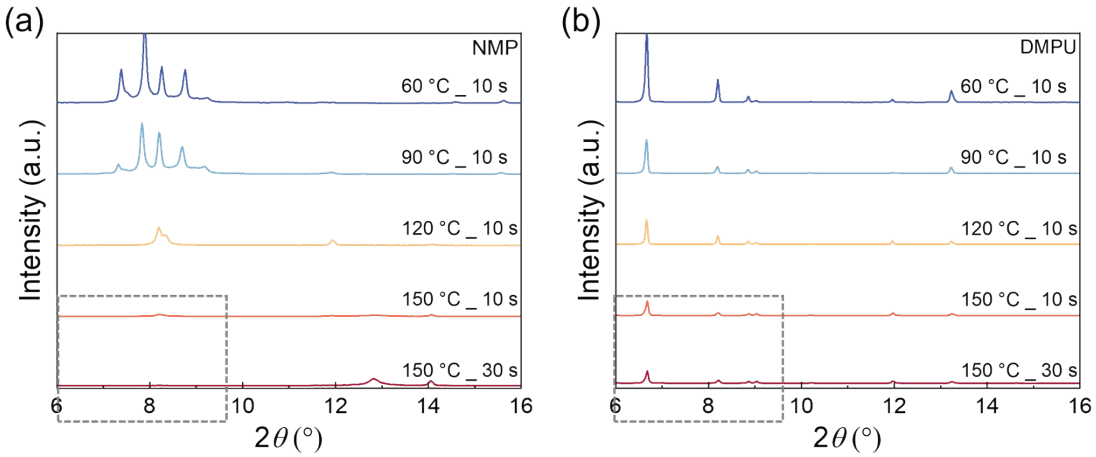


Fig. S27 XRD patterns of different solvent-based adducts under annealing.

To further investigate the temperature effect on the performance differences of AFPSCs prepared by different ML-selected solvents, we evaluated the stability of solvent-based adducts by XRD measurements. As shown in **Supplementary Fig. 27**, the diffraction peaks of the NMP-based adduct remain strong and sharp after annealing at 90 °C for 10 s, indicating that its structure is relatively stable at this stage. However, as the temperature increases, the peak intensity decreases significantly, and at 150 °C, the signal of the adduct phase disappears entirely, with the diffraction peak of PbI_2 (12.9°) becoming detectable. This suggests that the NMP-based adduct is destroyed and transitions to the PbI_2 phase at higher temperatures. In contrast, DMPU-based adducts exhibit greater thermal stability, particularly the DMPU-based adduct, which retains its structure even after annealing at 150 °C for 30 seconds.

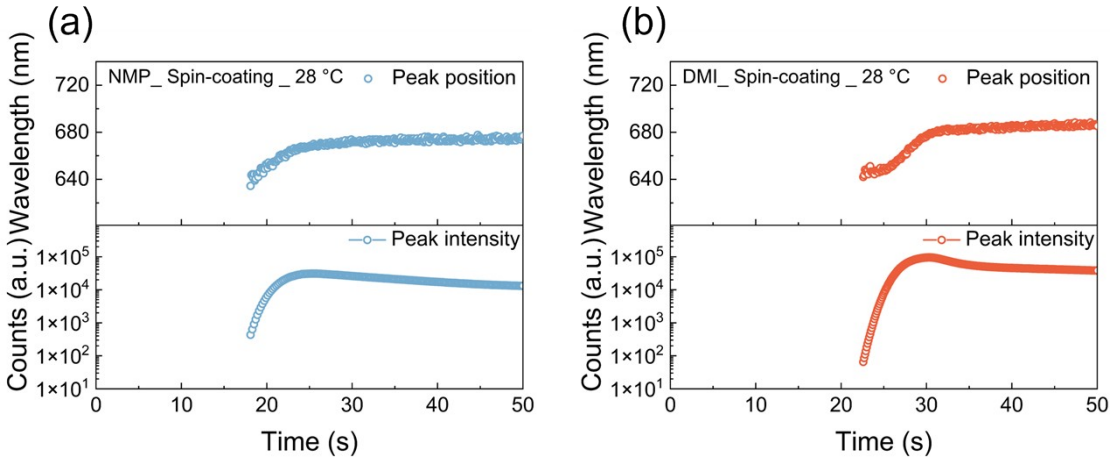


Fig. S28 PL intensity and peak position evolution of **(a)** NMP-treated and **(b)** DMI-treated AFP films during the spin-coating process. The T_A is 28°C.

No PL signal was detected in DMPU-treated film during spin-coating.

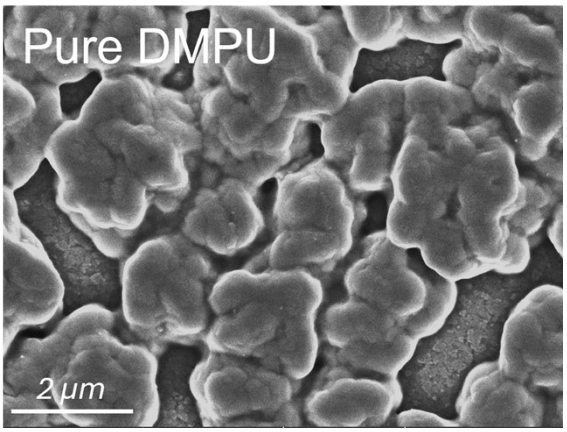


Fig. S29 SEM images of AFP films prepared from the precursor with the only additive of DMPU. The T_A is 16 °C.

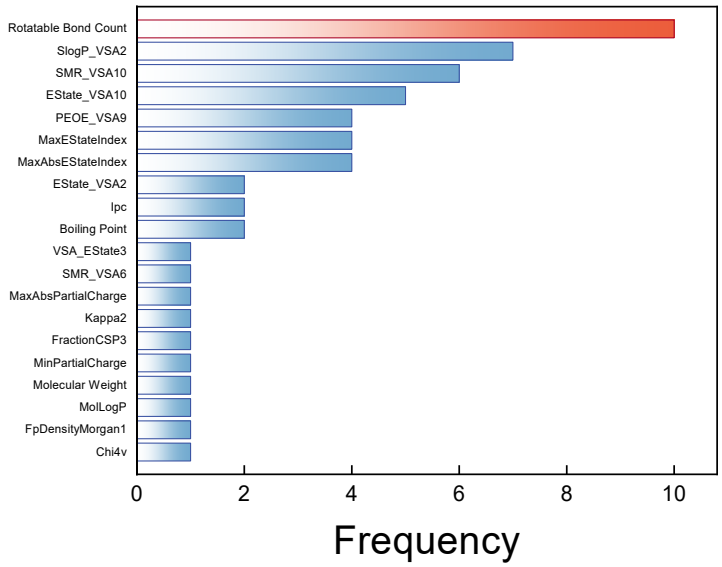


Fig. S30 Most important feature based on SHapley additive exPlanations (SHAP).

To assess feature contributions, SHAP analysis was performed on the 56 classification models that achieved perfect accuracy (1.0) under leave-one-out cross-validation. For each model, the most influential feature was identified, yielding a set of 56 top-ranked features. The frequency of each feature's occurrence was then computed and visualized in a bar plot. Due to the limited sample size, the stability of the SHAP-based importance ranking is constrained. Therefore, while these results provide a preliminary indication of feature relevance, they are not interpreted as definitive. Mechanistic insights are instead primarily derived from DFT-based analyses.

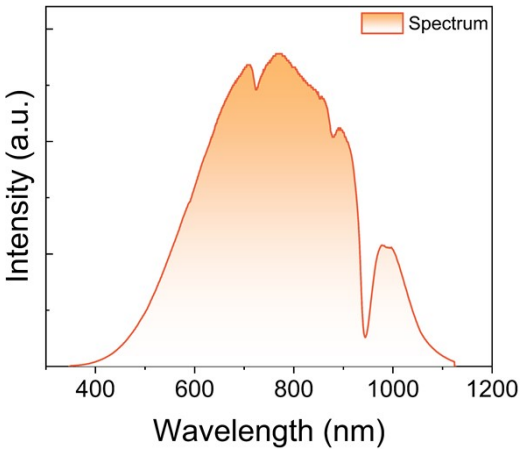


Fig. S31 The light source spectrum used in in-situ absorption test.

Tab. S1 Photovoltaic parameters of different solvent-treated AFPSCs fabricated at different T_A conditions.

	J_{SC} (mA·cm^{-2})	V_{OC} (V)	FF (%)	PCE (%)
NMP _ 16 °C	25.53	1.16	81.20	24.05
NMP _ 28 °C	25.27	1.07	57.72	15.60
DMPU _ 16 °C	25.58	1.16	82.99	24.62
DMPU _ 28 °C	24.94	1.14	79.18	22.51

Tab. S2 Photovoltaic parameters of different solvent-treated PSCs fabricated at 28 °C.

	J_{SC} (mA·cm^{-2})	V_{OC} (V)	FF (%)	PCE (%)
NMP _ forward	25.85	1.07	53.03	14.67
NMP _ reverse	25.27	1.07	57.72	15.60
DMPU _ forward	24.99	1.13	78.51	22.17
DMPU _ reverse	24.94	1.14	79.18	22.51

Tab. S3 Quantitative comparison of key molecular descriptors between Cluster 17 and the global solvent dataset.

	Global Mean	*Cluster 17 Mean*	Deviation (%)
fr_amide	0.0517	1.4286	2661.9
fr_C_O_noCOO	0.2730	1.0000	266.3
fr_NH0	0.1925	1.4286	642.0
PEOE_VSA12	0.8503	3.3755	297.0
MolLogP	1.3486	0.5722	-57.6

The quantitative divergence of Cluster 17 from the global dataset is summarized in Table S1. Most notably, the high-performing Cluster 17 exhibits an enrichment of amide groups (fr_amide) by over 2600% and non-carboxylic carbonyl groups (fr_C_O_noCOO) by 266% compared to the average organic solvent. These descriptors are directly linked to the solvent's Donor Number and its ability to form strong Lewis acid-base adducts with PbI_2. The drastic statistical separation in PEOE_VSA12 (nearly 300% deviation) further confirms that these solvents provide a unique electrostatic environment that stabilizes the precursor solution, preventing premature and disordered crystallization in the antisolvent-free process. In addition to electronic fingerprints, MolLogP was utilized to evaluate the lipophilicity and solvation capability of the solvents. As shown in Tab. S3, Cluster 17 exhibits a significantly lower MolLogP average (0.57) compared to the global dataset (1.35), representing a 57.6% decrease. This enhanced hydrophilic character (polar character) ensures optimal solubility of ionic perovskite precursors and facilitates the formation of a homogeneous solution, which is essential for maintaining a stable solvation shell during the antisolvent-free evaporation process.

REFERENCES

1 Wang, Y. *et al.* Intermediate State Formation Extends the Ambient Temperature Processing Window of Solution-Processed Perovskite Solar Cells. *ACS Energy Lett.* **10**, 647-657 (2025).

2 He, A. *et al.* Unraveling the Impact of Bromine Incorporation on the Hybrid Perovskite Film Phase Evolution Mechanism by In Situ Characterization. *Adv. Energy Mater.* **13**, 2300957 (2023).

3 Jiang, Y. *et al.* Crystallization Kinetics Regulation for Strain and Morphology Management Enables Efficient Tin Perovskite Solar Cells. *Advanced Functional Materials*, 2504541 (2025).

4 Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15-50 (1996).

5 Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953 (1994).

6 Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).

7 Toniato, A., Schwaller, P., Cardinale, A., Geluykens, J. & Laino, T. Unassisted noise reduction of chemical reaction datasets. *Nat. Mach. Intell.* **3**, 485-494 (2021).

8 Esterhuizen, J. A., Goldsmith, B. R. & Linic, S. Uncovering electronic and geometric descriptors of chemical activity for metal alloys and oxides using unsupervised machine learning. *Chem. Catalysis* **1**, 923-940 (2021).

9 Marcus, Y. The properties of organic liquids that are relevant to their use as solvating solvents. *Chem. Soc. Rev.* **22**, 409-416 (1993).

10 Krasnov, L., Mikhaylov, S., Fedorov, M. & Sosnin, S. Bigsoldb: Solubility dataset of compounds in organic solvents and water in a wide range of temperatures. (2023).

11 Diorazio, L. J., Hose, D. R. & Adlington, N. K. Toward a more holistic framework for solvent selection. *Org. Process Res. Dev.* **20**, 760-773 (2016).

12 WEININGER, D.-S. a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem Inf. Model.*, 1549-9596

13 Geoffrey, B. *et al.* A program to automate the discovery of drugs for west nile and dengue virus—programmatic screening of over a billion compounds on pubchem, generation of drug leads and automated in silico modelling. *J. Biomol. Struct. Dyn.* **40**, 4293-4300 (2022).

14 Kim, S. *et al.* PubChem 2023 update. *Nucleic Acids Res.* **51**, D1373-D1380 (2023).

15 Bento, A. P. *et al.* An open source chemical structure curation pipeline using RDKit. *J. Cheminformatics* **12**, 1-16 (2020).

16 Abdi, H. & Williams, L. J. Principal component analysis. *Wiley Interdiscip. Rev.-Comput. Stat.* **2**, 433-459 (2010).

17 Hueffel, J. A. *et al.* Accelerated dinuclear palladium catalyst identification through unsupervised machine learning. *Science* **374**, 1134-1140 (2021).