



Cite this: *Phys. Chem. Chem. Phys.*, 2026, **28**, 6917

# Data-guided rational design of additives for halogenation of highly fluorinated naphthalenes: integrating fluorine chemistry and machine learning†

Naoya Ohtsuka,<sup>ab</sup> Muhammad Zhafran Mohd Aris,<sup>ac</sup> Toshiyasu Suzuki<sup>a</sup> and Norie Momiyama<sup>ab\*</sup>

Highly fluorinated aromatic compounds exhibit unique electronic structures, however their selective transformation remains a longstanding challenge. Halogenation of F<sub>7</sub> naphthalene previously required low temperatures (−40 to 0 °C) for high yields, while room-temperature reactions suffered from side reactions and decomposition. Here we present a data-guided framework for rational additive design enabling efficient halogenation under ambient conditions. Screening 25 functional additives revealed distinct groups, with effective ones affording halogenated products in yields above 50% and recovering in high rates. Machine learning models built from DFT-derived descriptors achieved strong predictive performance ( $R^2 = 0.90$ , RMSE = 10.9). Feature importance and SHAP analyses clarified the design criteria—moderately balanced functional-group charges and non-negative aromatic contributions—as critical for reactivity. Guided by these criteria, a chlorine-substituted additive, 1-chloro-4-(ethoxymethyl)benzene (**3a**), was designed, predicted to give >99% yield, and experimentally validated to deliver the iodinated product in 98% yield with 96% additive recovery at room temperature. Moreover, additive **3a** was effective in iodination, bromination, and chlorination, demonstrating the generality of the design principle. This study advances additive development from empirical trial-and-error to a predictive, machine learning-enabled strategy, offering guidelines for selective transformations of per-fluorinated arenes.

Received 15th September 2025,  
 Accepted 20th February 2026

DOI: 10.1039/d5cp03554f

rsc.li/pccp

## Introduction

Fluorine is the most electronegative element in the periodic table and exhibits extremely strong electron-withdrawing properties. When aromatic rings are highly fluorinated, they acquire electronic states fundamentally different from those of hydrocarbon-based aromatic compounds. These unique electronic features, often involving  $\sigma$ -holes and  $\pi$ -holes, significantly alter electron density distribution and molecular orbital characteristics (Fig. 1).<sup>1–6</sup>

Although highly fluorinated aromatic compounds display distinct physical properties and molecular recognition abilities,

their reactivity and molecular transformations are notoriously difficult to control. The efficient synthesis and transformation of perfluorohalogenated polycyclic aromatic compounds have long remained a challenging issue.<sup>7</sup>

Recently, we addressed this challenge by developing a synthetic method that combines metalation with an organo-magnesium base followed by halogenation, which enables the high-yield synthesis of perfluorohalogenated naphthalenes.<sup>8</sup> This approach makes it possible to access perfluoroiodo-

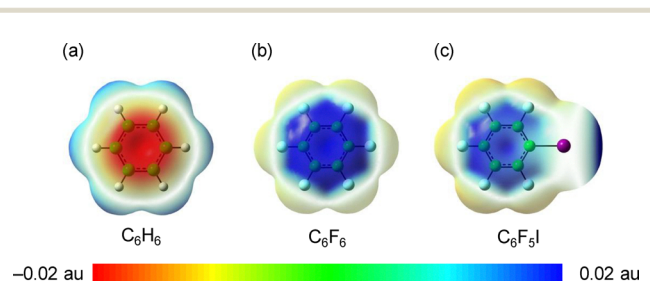


Fig. 1 Electrostatic potential maps of (a) benzene, (b) perfluorobenzene (C<sub>6</sub>F<sub>6</sub>), and (c) pentafluoroiodobenzene (C<sub>6</sub>F<sub>5</sub>I).

<sup>a</sup> Institute for Molecular Science, Okazaki, Aichi 444-8787, Japan.

E-mail: momiyama@ims.ac.jp

<sup>b</sup> SOKENDAI (Graduate University for Advanced Studies), Okazaki, Aichi 444-8787, Japan

<sup>c</sup> University of Malaya, Faculty of Science, Department of Biochemistry, Kuala Lumpur, 50603, Malaysia

† r Resnati, celebrating a career in fluorine and noncovalent chemistry on the occasion of his 70th birthday.



bromo-, and chloro-naphthalenes, which have scarcely been reported before, thereby opening new avenues for the molecular conversion of highly fluorinated aromatics. These reactions still require low-temperature conditions ( $-40$  to  $0$  °C) to achieve high yields, and their application under practical, room-temperature conditions remains difficult. Therefore, a new strategy is required to suppress side reactions and decomposition while enabling efficient transformations under mild conditions. In this context, the present study addresses these limitations by systematically evaluating the effects of diverse functional additives and applying machine learning (ML) to correlate additive structures with reactivity. By integrating experimental and computational analyses, we aim to establish a predictive framework that overcomes the constraints of low-temperature conditions and qualitative discussions in previous studies.

To overcome this limitation, we systematically investigated the effects of a broad range of additives using a functional group evaluation (FGE) kit. While the FGE kit has been widely applied to evaluate functional group compatibility and chemoselectivity in diverse organic reactions,<sup>9–16</sup> this study represents the first application of the kit to systematically investigate additive effects in halogenation of highly fluorinated aromatics. Consequently, we identified several additives that effectively suppressed undesired side reactions, enabling high-yield halogenation even at room temperature, which was not feasible previously. Further, we evaluated their roles more scientifically by comparing the product yields with the recovery of the additives. Although the additive effects have traditionally been discussed qualitatively or empirically,<sup>17–25</sup> our study provided a deeper and more quantitative perspective.

The novelty of this work lies in the quantitative and scientific analysis of additive effects using ML instead of qualitative or empirical considerations. We combine experimentally obtained yields with molecular descriptors derived from density functional theory (DFT) calculations and use them as input data, which enable us to identify molecular features that help promote the halogenation reaction. Previous studies have highlighted the utility of ML for predicting reaction performance,<sup>26,27</sup> modeling catalysts and reactivity,<sup>28–30</sup> and advancing organic synthesis through digital and AI-driven approaches.<sup>31–35</sup> Based on these developments, we present both the practical transformation of highly fluorinated aromatics and a methodology to clarify the additive effects through ML. Our study provides new design criteria for reaction development and offers insights into future strategies for molecular transformations by integrating fluorine chemistry with data science.

The remainder of this paper is organized as follows: Section 2 describes the experimental and computational methodology. Section 3 presents the results of additive screening, feature analysis using ML, and the design and validation of new additives. Section 4 provides the conclusions, including the implications and limitations of our findings as well as perspectives for future research.

## Methodology

### Chemicals

Anhydrous tetrahydrofuran (THF), methylene chloride ( $\text{CH}_2\text{Cl}_2$ ), and diethyl ether were obtained from Kanto Chemical Co., Inc. as a dehydrated solvent system.  $\text{F}_7$  naphthalene **1** was synthesized from octafluoronaphthalene in three steps, as listed in the SI. Bis(2,2,6,6-tetramethylpiperidyl)magnesium dilithium bromide ( $\text{TMP}_2\text{Mg}\cdot 2\text{LiBr}$ ) in THF was prepared from Mg powder following a reported procedure<sup>36</sup> (see SI). The FGE-kit<sup>9</sup> **A00–A24** was provided by Prof. Ohshima from Kyushu University. All other reagents were purchased from commercial suppliers and used as received without further purification.

### Materials synthesis

$\text{F}_7$  naphthalene **1** (51.0 mg, 0.200 mmol, and 1.0 equiv.) and a corresponding additive (1.0 equiv.) were dissolved in 1.5 mL of THF.  $\text{TMP}_2\text{Mg}\cdot 2\text{LiBr}$  (0.24 M in THF, 1.00 mL, 0.240 mmol, and 1.2 equiv.) was added at 25 °C, and the mixture was stirred for 30 min. Subsequently, molecular iodine (122 mg, 0.480 mmol, 2.4 equiv.) was added, and the reaction was stirred at 25 °C for 1 h. The mixture was then quenched with 2 M HCl aq. (5 mL) at 0 °C and extracted with  $\text{Et}_2\text{O}$  (10 mL  $\times$  3). The combined organic extracts were washed with saturated  $\text{Na}_2\text{SO}_3$  aqueous (5 mL) and brine (5 mL), dried over  $\text{Na}_2\text{SO}_4$ , filtered, and concentrated under reduced pressure. The crude residue was dissolved in  $\text{CDCl}_3$ , and dibromomethane (14  $\mu\text{L}$ , 0.200 mmol, and 1.0 equiv.) and hexafluorobenzene (23  $\mu\text{L}$ , 0.200 mmol, and 1.0 equiv.) were added as internal standards.

### Materials characterization

The products were characterized by  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{19}\text{F}$  NMR spectroscopy, infrared (IR), and high-resolution mass spectrometry (HRMS). The yield of product **2** and the recovery of the additive were determined *via*  $^1\text{H}$  and  $^{19}\text{F}$  nuclear magnetic resonance (NMR) spectroscopy. NMR spectra were recorded on JEOL spectrometers at ambient temperature:  $^1\text{H}$  (400 MHz),  $^{13}\text{C}$  (100 or 151 MHz), and  $^{19}\text{F}$  (376 MHz). Chemical shifts ( $\delta$ ) were reported in ppm relative to internal standards ( $\text{CDCl}_3$ :  $\delta = 7.26$  ppm for  $^1\text{H}$ ,  $\delta = 77.0$  ppm for  $^{13}\text{C}$ ;  $\text{DMSO}-d_6$ :  $\delta = 2.49$  ppm for  $^1\text{H}$ ) or an external standard ( $\alpha,\alpha,\alpha$ -trifluorotoluene:  $\delta = -63.72$  ppm for  $^{19}\text{F}$ ). IR spectra were recorded on a Jasco FT/IR-460 plus spectrometer using attenuated total reflectance. HRMS (FAB, 3-nitrobenzyl alcohol matrix) was performed using a JEOL JMS-700 instrument at the Instrument Center, Institute for Molecular Science.

### Computational details

The DFT calculations were performed using the M06-2X functional including Grimme's D3 dispersion correction.<sup>37</sup> Geometry optimizations were conducted using the SDD basis set for iodine and 6-311+G(d,p) for all other atoms, together with the SMD solvation model (THF). All optimized structures were confirmed as minima *via* a vibrational frequency analysis. Natural bond orbital (NBO) charges and electrostatic potential (ESP) maps were obtained from the optimized geometries.



All calculations were performed with Gaussian 16.<sup>38</sup> Full computational details are provided in the SI.

### Machine learning

The dataset was constructed from experimentally determined yields and molecular descriptors obtained from DFT calculations, including NBO charges and ESP values. A total of 26 regression models were evaluated using Datachemical Lab;<sup>39</sup> nonlinear support vector regression (NSVR) was identified as the best-performing model. Feature importance analyses such as cross-validation based permutation feature importance (CVPI)<sup>40</sup> and Shapley additive explanations (SHAP)<sup>41</sup> were conducted exclusively with the NSVR model. The dataset was randomly divided into 75 and 25% for the training and test sets, respectively, and five-fold cross-validation was performed. The predictive accuracy of each model was compared using the coefficient of determination ( $R^2$ ), root mean square error (RMSE), and mean absolute error (MAE) for the test set.

## Results and discussion

### Additive screening

In our previous study,<sup>8</sup> the iodination reaction of F<sub>7</sub> naphthalene **1** achieved high yield under low-temperature conditions (Scheme 1a). However, side reactions became predominant and the yield of the desired product **2a** dropped significantly to 4% (Scheme 1b) under ambient conditions, which remained a major limitation. In the present study, we selected this condition (magnesiumation at 25 °C for 30 min) for rigorously evaluating the effect of additives. Control experiments were performed to assess the reproducibility of the reaction system. The control additive **A00** was tested five times ( $n = 5$ ), and the average value was calculated. Subsequently, 25 additives (**A00–A24**) in the FGE kit<sup>9</sup> were evaluated. Each additive was tested in duplicate ( $n = 2$ ), and the results were compared with those of the control experiments ( $n = 5$ ) to examine the degree of variability. The mean of the duplicate values was used when the variance between them was negligible, that is, within the experimental error observed in the control reactions. Two additional



**Scheme 1** (a) Iodination of F<sub>7</sub> naphthalene **1** achieving high yield under low-temperature conditions in our previous study. (b) Same reaction under ambient conditions, where side reactions predominated and the yield of the desired product **2a** dropped to 4%.



**Fig. 2** Scatter plot showing the relationship between the recovery rate of the additives (x-axis) and yield of product **2a** (y-axis) in the FGE kit screening (**A00–A24**). The additives were classified into three distinct groups according to their behaviors.

experiments were conducted for cases where the variance was larger. Thus, a maximum of four runs was performed for each additive, and the mean values were used to obtain reliable data on the additive effects. The byproducts formed from the additives under the reaction conditions were isolated and purified, and their structures were confirmed by <sup>1</sup>H NMR analysis. In cases where the additive was not recovered, these byproducts were observed as the only identifiable products derived from the additive.

The evaluation revealed three distinct behaviors of the additives. As shown in the scatter plot in Fig. 2, the additives were clearly classified into three groups. **A04**, **A07**, **A13**, **A14**, **A15**, **A16**, **A20**, and **A22** clustered in the region with both no recovery and <1% yield, indicating the decomposition of the additives and no formation of product **2a** (Fig. 3a). **A01**, **A02**, **A03**, **A05**, **A06**, **A08**, **A12**, and **A17** showed high recovery but no detectable formation of product **2a** (Fig. 3b), forming a group with no promoting effect. In contrast, **A00**, **A09**, **A10**, **A11**, **A18**, **A19**, **A21**, **A23**, and **A24** formed product **2a** in yields above 5% and were distinguished as effective additives (Fig. 3c). These results indicate that certain additives contribute to suppressing side reactions, which enables the formation of product **2a** even under ambient conditions.

### Electronic and structural features of additives

We characterized and quantified the electronic and structural features common to effective additives based on the results of additive screening. In this ML analysis, additives that resulted in low or zero recovery were intentionally included, as they provide essential negative examples for quantitatively identifying molecular features unfavorable for productive iodination among the additives examined. To this end, a set of descriptors was defined based on the ESP and NBO analyses. The ESP





**Fig. 3** Additives classified into three groups. The chemical structures of the additives are shown together with the yield of product **2a** (%) and recovery rate of additive **A** (%). (a) Additives leading to decomposition with neither recovery nor product formation (**A04**, **A07**, **A13**, **A14**, **A15**, **A16**, **A20**, and **A22**). (b) Additives showing high recovery but no detectable product **2a** (**A01**, **A02**, **A03**, **A05**, **A06**, **A08**, **A12**, and **A17**). (c) Effective additives that afforded product **2a** in yields above 5% (**A00**, **A09**, **A10**, **A11**, **A18**, **A19**, **A21**, **A23**, and **A24**).

reflects the molecular surface charge distribution and is useful for predicting the strength and directionality of interactions with nucleophilic or electrophilic sites, providing insights for improving the yield of the desired product. In contrast, NBO charges quantitatively capture local electron density redistribution and are directly relevant to discussions on charge transfer and stability along the reaction coordinate. The descriptors defined here include ESP values around the Cl atom ( $\sigma$  and belt directions), ESP values at representative aromatic positions, ESP of the  $\pi$ -surface, and maximum and minimum ESP values

at functional groups, together with the corresponding NBO charges. In addition, as structural features, the type and position of substituents on the aromatic ring are also considered, which act in concert with electronic factors to define the characteristics of the reaction environment. These descriptors provide a quantitative basis to identify electronic and structural properties characteristic of effective additives and are employed as input variables for the ML analysis (see SI for details).

### Model selection and validation

The predictive performance of different regression models was evaluated by examining the yield of product **2a** predicted by the 26 ML models available in Datachemical Lab using ESP- and NBO-based descriptors as input variables. The performance of the top 10 models is summarized in Table 1. Among the examined models, NSVR provided the best performance ( $R^2 = 0.90$ , RMSE = 10.9, and MAE = 10.4). The NGPR3 ( $R^2 = 0.85$ , RMSE = 13.2, and MAE = 11.1) and NGPR5 ( $R^2 = 0.77$ , RMSE = 16.3, and MAE = 14.6) models showed relatively high performances; however, NSVR consistently provided the best results. This superior performance can be attributed to its ability to flexibly capture nonlinear relationships between the descriptors and reaction yield, which was essential for modeling the additive effects. Therefore, the NSVR model was selected for subsequent feature analysis and discussion.

### Feature importance analysis

The contribution of each descriptor to the prediction of the product **2a** yield was evaluated using the NSVR model (Fig. 4). Feature importance was calculated based on CVPFI. The most significant descriptors were associated with the electronic properties of the functional group, including NBO-FG max, ESP-FG mini, and ESP-FG max. These were followed by descriptors related to the electrostatic potential around the aromatic ring and chlorine atom, such as ESP-R1, ESP-R2, and ESP-Cl( $\sigma$ ). In contrast, descriptors corresponding to local NBO charges on the aromatic ring (NBO-R1, NBO-H1, and NBO-H2) exhibited only minor contributions. These results demonstrate the relative importance of each descriptor in determining the

**Table 1** Predictive performances of the top ten machine learning models examined for the additive dataset (**A00**–**A24**)<sup>a</sup>

Model	$R^2$ (test)	RMSE (test)	MAE (test)
NSVR	0.90	10.9	10.4
NGPR3	0.85	13.2	11.1
NGPR5	0.77	16.3	14.6
DT	0.77	16.5	11.2
NGPR9	0.75	17.1	15.1
NGPR7	0.74	17.6	15.5
NGPR1	0.72	18.1	15.7
NGPR2	0.72	18.1	15.7
RF	0.61	21.2	17.0
NGPR4	0.48	24.7	16.9

<sup>a</sup> The test set performance is evaluated using the coefficient of determination ( $R^2$ ), root mean square error (RMSE), and mean absolute error (MAE).





Fig. 4 (a) Feature importance of the descriptors calculated by the NSVR model using CVPFI. (b) Schematic showing correspondence between each descriptor and the structural parts of the additives. FG is an abbreviation for functional group.

predictive performance of the model and provide a basis for further interpretation in the subsequent SHAP analysis.

### SHAP analysis

The CVPFI analysis clarified the relative importance of each descriptor; however, it did not reveal their directional contributions to the yield. Therefore, SHAP analysis was introduced to visualize how each descriptor acts on the yield of product **2a**. The analysis was performed in Python using the NSVR model, which showed the highest predictive performance in the regression analysis with Datachemical Lab. The obtained predictive accuracy was comparable to that from the Datachemical Lab ( $R^2 = 0.86$ ), which confirmed sufficient reproducibility.

Fig. 5 shows the SHAP beeswarm plot. The descriptors are arranged in the descending order of importance on the vertical axis, and their SHAP values, representing both the magnitude and direction of their contributions to the yield, are plotted on



Fig. 5 SHAP beeswarm plot showing the contribution of each descriptor to the yield of product **2a**. Descriptors are ordered by their overall importance (vertical axis), and the SHAP values (horizontal axis) indicate both the magnitude and direction of their effects on the yield.

the horizontal axis. Similar to the CVPFI results, the electronic properties of the functional group (NBO-FG max, ESP-FG mini, and ESP-FG max) made the largest contributions. ESP-R1 was ranked among the top descriptors, which is consistent with its fourth position in the CVPFI results.

A closer inspection revealed that lower ESP-FG mini values were associated with negative contributions to the yield, whereas higher values contributed positively. In contrast, ESP-FG max contributed positively when its values were small and negatively when they were large. These results confirm that the yield decreases when the functional group bears excessively negative or positive charge, whereas moderate charges enhance the yield. In addition, an electrostatic environment without extreme imbalance around the aromatic ring and chlorine atom was also suggested to contribute to the higher yield of product **2a**. Local NBO descriptors on the aromatic ring (NBO-R1, NBO-H1, and NBO-H2) showed relatively minor contributions, suggesting a limited correlation with the yield. These results can be rationalized in terms of organic synthesis. When the charge at the functional group is extremely negative or positive, *in situ* generated Mg species are likely to react preferentially with the additive rather than engaging in the halogenation, which decreases the yield of product **2a**. In contrast, additives with moderate charges at the functional group do not induce such side reactions and instead enable the halogenation of the substrate to proceed smoothly.

A SHAP waterfall plot analysis was conducted for the additives that afforded product **2a** (A00, A09, A10, A11, A18, A19, A21, A23, and A24) to identify factors governing yield improvement. Among the six additives with recovery rates above 50% (A00, A09, A10, A19, A23, and A24), those with yields of product **2a** greater than 50% were defined as effective additives, those with lower yields were defined as ineffective additives, and the two groups were compared.

Fig. 6a presents representative effective additives: A00, A10, and A24. In all cases, descriptors related to the functional group (ESP-FG mini, ESP-FG max, and NBO-FG max) made strong positive contributions to yield. For A00, these descriptors dominated the prediction, indicating that functional-group properties were the primary factors governing yield improvement. Although ESP-R1 contributed negatively, its effect was outweighed by the strong positive contributions from the functional group. In A10, functional-group descriptors showed large positive effects, whereas additional positive contributions from ESP-R1 and ESP-R2 suggested a cooperative interaction between the functional group and aromatic ring. This synergistic feature distinguishes A10 from the other effective additives. In A24, ESP-R1 contributed negatively; however, this was offset by strong positive contributions from ESP-FG mini and NBO-FG max, which confirmed its classification as an effective additive dominated by functional-group effects.

In contrast, Fig. 6b shows representative ineffective additives: A09, A19, and A23. In these cases, strong negative contributions from descriptors associated with either the functional group or aromatic ring were observed. For A09, both NBO-C1 and ESP-R1 contributed negatively. For A19, the positive effects from





Fig. 6 SHAP waterfall plots illustrating contributions of individual descriptors to the yield of product **2a**. (a) Representative effective additives: **A00**, **A10**, and **A24**, where functional-group descriptors (ESP-FG mini, ESP-FG max, and NBO-FG max) make strong positive contributions. (b) Representative ineffective additives: **A09**, **A19**, and **A23**, where descriptors associated with the functional group or the aromatic ring exhibit strong negative contributions.

functional-group descriptors were counteracted by negative contributions from NBO-FG max and aromatic descriptors. For **A23**, ESP-FG mini, ESP-FG max, and ESP-R1 all contributed strongly in the negative direction, indicating that both the functional group and aromatic ring were electronically unfavorable. These results are consistent with the finding from the beeswarm plot that excessive negative or positive charges at the functional group can decrease yield. This supports the scenario in which *in situ* generated Mg species react with the additive rather than participating in the intended halogenation.

### Design guidelines for new additives

Several common factors for effective additives were identified from the SHAP analysis. Functional-group descriptors (ESP-FG mini, ESP-FG max, and NBO-FG max) consistently showed strong positive contributions, which indicates that functional-group charges must remain within a moderate range rather than being excessively biased. In addition, aromatic descriptors (ESP-R1 and ESP-R2) should not contribute strongly in the negative direction. These findings provide clear criteria for designing effective additives.

Another important aspect is that the electronic properties of the  $F_7$  naphthalene substrate **1** must be considered (Fig. 7). Substrate **1** is highly electron-deficient because of the multiple fluorine substituents, which results in a  $\pi$ -surface with



Fig. 7 Electronic properties of the  $F_7$  naphthalene substrate (**1**). (a) ESP map showing the electron-deficient  $\pi$ -surface induced by multiple fluorine substituents. (b) NBO charges highlighting the local charge distribution. DFT calculations were performed at the SMD(THF)/M06-2X-D3/6-311+G(d,p) level of theory.

predominantly positive ESP values. Such an electronic imbalance can promote side reactions or decomposition during the reaction.

For effective additives, waterfall plots demonstrated that descriptors related to the electronic state of the functional group contributed positively to the yield. This indicates that effective additives compensate for the excessively positive ESP of the substrate by providing complementary electronic features that stabilize the reaction pathway and promote selective formation of product **2a**. In contrast, ineffective additives

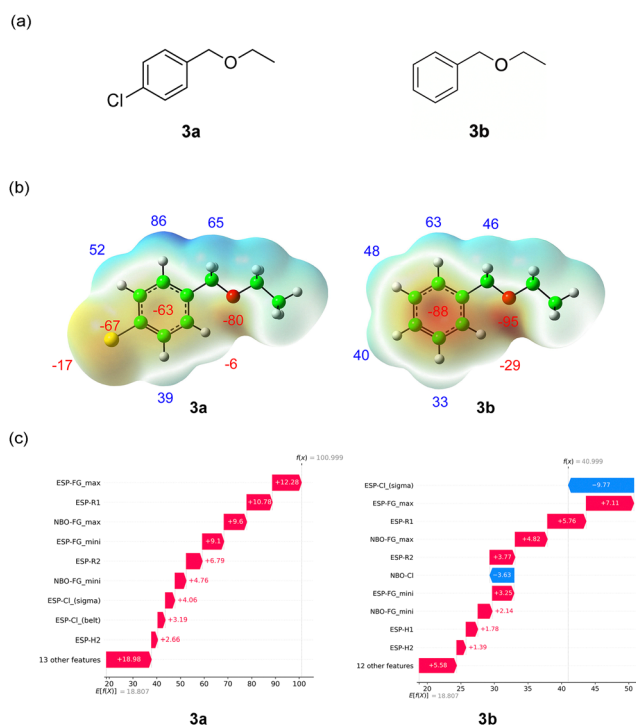


either exhibited excessively positive or negative functional-group charges or failed to achieve complementarity with the substrate ESP, preventing compensation for the substrate imbalance and allowing side reactions or additive decomposition to dominate, which resulted in lower yields.

Compared with previous empirical discussions of additive effects in organometallic reactions,<sup>17–25</sup> our ML-based framework provides a quantitative and predictive strategy. Unlike earlier studies that often relied on qualitative rationalization, the present analysis identifies explicit electronic and structural descriptors governing reactivity, thereby offering a generalizable design principle for reaction development. These results suggest that the following criteria are important for designing new additives: (i) functional-group descriptors should fall within a moderate range and contribute positively to the yield; (ii) aromatic contributions should not be strongly negative; and (iii) the additive should possess complementary electronic features that compensate for the intrinsic ESP imbalance of substrate **1**. Based on these criteria, new additive structures were designed, and their validity was examined through experimental and computational studies in the following section.

### Computational prediction and experimental validation

Two new additives were designed to validate the design guideline: additive **3a** and additive **3b** (Fig. 8a). The chlorine substituent on the aromatic ring of additive **3a** was expected to compensate for the electronic imbalance of the substrate **1**.



**Fig. 8** Computational and ML analyses of newly designed additives. (a) Structures of additives **3a** and **3b**. (b) ESP maps calculated by DFT (SMD(THF)/M06-2X-D3/6-311+G(d,p)); NBO charges are provided in the SI. (c) SHAP waterfall plots showing descriptor contributions to the predicted yields for **3a** and **3b**.

Additive **3b**, which lacks a chlorine substituent, was designed as a control to examine the contribution of chlorine and experimentally verify this hypothesis.

ESP maps and NBO charges were calculated by DFT for both additives (Fig. 8b for ESP maps, and SI for NBO charges), and the values were used as input for the existing ML model to predict the yields. For **3a**, the predicted yield was estimated to be >99%, which was consistent with the design guideline and suggested that **3a** would act as an effective additive. Although the predicted yield of **3b** was calculated as 41%, this value must only be considered as a reference. The training dataset did not include chlorine-free additives, and therefore, the model was forced to extrapolate into an unexplored chemical space, resulting in reduced accuracy.

SHAP waterfall plots were examined to further interpret these predictions (Fig. 8c). For **3a**, descriptors related to the functional group (ESP-FG max, ESP-R1, NBO-FG max, and ESP-FG mini) showed strong positive contributions, which is consistent with the criteria for effective additives. Although some descriptors of **3b** achieved minor positive effects, ESP-Cl( $\sigma$ ) and NBO-Cl exhibited strong negative contributions, which outweighed them. These results highlight the critical role of the chlorine substituent in determining additive performance.

Reactions were performed using additives **3a** and **3b** to experimentally validate the predictions from the SHAP analysis (Table 2). When **3a** was employed, the iodination of the substrate **1** afforded product **2a** in 98% yield, and the additive itself was recovered in 96% yield. These results experimentally demonstrated that **3a** was an effective additive consistent with the design guideline. In contrast, when **3b** was used, although the additive was recovered in 97% yield, product **2a** was not obtained at all. This finding clearly demonstrated that **3b**, which lacked the chlorine substituent, did not function as an effective additive in agreement with the waterfall plot analysis highlighting the importance of chlorine.

The structures depicted for the additive-derived species and the additive–Mg complexes represent hypothetical mechanistic models proposed to rationalize interaction trends between the additive and the Mg intermediate. To elucidate the origin of the

**Table 2** Experimental validation of additives **3a** and **3b** in the iodination of  $F_7$  naphthalene (**1**)<sup>a</sup>

Entry	Additive	Yield of <b>2a</b> <sup>b</sup> (%)	Recovery of <b>3</b> <sup>c</sup> (%)
1	<b>3a</b>	98	96
2	<b>3b</b>	0	97

<sup>a</sup> The reaction was carried out under the indicated conditions using 0.200 mmol of **1**, 1.0 equiv. of additive **3**, 1.2 equiv. of  $TMP_2Mg \cdot 2LiBr$ , and 2.4 equiv. of  $I_2$  in THF (1.5 mL). <sup>b</sup> Determined by  $^{19}F$  NMR using hexafluorobenzene as an internal standard. <sup>c</sup> Determined by  $^1H$  NMR using dibromomethane as an internal standard.





**Fig. 9** (a) ESP map of the magnesium intermediate ( $F_7$ -MgTMP) calculated by DFT (SMD(THF)/M06-2X-D3/6-311+G(d,p)). (b) Optimized geometry of the magnesium intermediate-**3a** complex obtained by DFT calculations (SMD(THF)/M06-2X-D3/6-311+G(d,p)). (c) IGMH isosurfaces visualizing noncovalent interactions in the complex (side views). Blue and green surfaces correspond to coordination (O–Mg) and weak interactions, respectively.

superior additive effect of **3a**, we conducted ESP analysis of the *in situ* generated magnesium intermediate. Subsequently, we optimized the geometry of magnesium intermediate-**3a** complex and then visualized it (Fig. 9) using the independent gradient model based on Hirshfeld partition of molecular density (IGMH)<sup>42</sup> method. The ESP map of the magnesium intermediate showed a highly positive potential around the magnesium center (Fig. 9a). This result indicated that the strongly negative region on the ether oxygen atom of **3a** (ESP\_FG\_mini, NBO\_FG\_mini) could coordinate to the magnesium center, thereby stabilizing the intermediate. Geometry optimization of the magnesium intermediate-**3a** complex confirmed this coordination and the formation of  $\pi$ - $\pi$  stacks between the aromatic rings (Fig. 9b). The calculated Gibbs free energy difference ( $\Delta G = -10 \text{ kcal mol}^{-1}$ ) indicated that the complex formation was thermodynamically favorable (see SI). Furthermore, the IGMH analysis revealed the occurrence of multiple noncovalent interactions, including the O–Mg coordination bond formation, CH...N interactions between the functional moiety and TMP unit,  $\pi$ - $\pi$  stacking, and Cl- $\pi$  interactions in the aromatic region (Fig. 9c). These interactions were consistent with the descriptors identified by the SHAP analysis (ESP\_FG\_max, ESP\_FG\_mini, NBO\_FG\_max, NBO\_FG\_mini), which highlighted that the electronic distribution around the functional group played a crucial role. The visualized  $\pi$ - $\pi$  and Cl- $\pi$  interactions also confirmed the contribution of the aromatic electronic features (ESP\_R1, ESP\_R2, ESP\_Cl(sigma), ESP\_Cl(belt)) to yield enhancement.

Overall, these results indicated that the cooperative electronic features of the functional group and aromatic ring stabilized the magnesium intermediate through multiple noncovalent interactions, thereby rationalizing the superior performance of **3a**.

**Table 3** Demonstration of the utility of additive **3a** in the halogenation of  $F_7$  naphthalene (**1**)<sup>a</sup>

Entry	X	2	Yield of <b>2</b> <sup>b</sup> (%)		Recovery of <b>3a</b> <sup>c</sup> (%)
			with <b>3a</b>	(without <b>3a</b> )	
1	I	<b>2a</b>	98	(4)	96
2	Br	<b>2b</b>	95	(2)	94
3	Cl	<b>2c</b>	96	(3)	96

<sup>a</sup> The reaction was carried out under the indicated conditions using 0.200 mmol of **1**, 1.0 equiv. of additive **3a**, 1.2 equiv. of  $\text{TMP}_2\text{Mg}\cdot 2\text{LiBr}$ , and the halogenation reagent in THF (1.5 mL). For details, see the SI. <sup>b</sup> Determined by  $^{19}\text{F}$  NMR using hexafluorobenzene as an internal standard. <sup>c</sup> Determined by  $^1\text{H}$  NMR using dibromomethane as an internal standard.

The utility of additive **3a** was further established not only in the iodination of  $F_7$  naphthalene (**1**) but also in bromination and chlorination reactions (Table 3). In all cases, the desired halogenated products **2** were obtained in high yields, and the additive **3a** was recovered in comparably high recovery yields. These results demonstrated that the design guideline proposed in this study is not limited to iodination but is generally applicable to other halogenation reactions. In other words, additives with complementary electronic properties that compensate for the intrinsic ESP imbalance of substrate **1** can universally function under different reaction conditions and with different halogen sources.

Collectively, SHAP analysis and additive screening clarified that effective additives require moderately balanced functional-group charges, positive contributions from ESP-FG and NBO-FG descriptors, and non-negative contributions from aromatic descriptors. These trends were rationalized by the electron-deficient nature of substrate **1**, which highlights the need for additives with complementary electronic features. We designed and validated additive **3a** by applying these data-derived criteria, which shows that machine learning shifts additive development from empirical trial-and-error to a predictive, generalizable strategy for halogenation.

## Conclusions

In this study, we established a data-driven framework integrating additive screening, electronic structure analysis, and machine learning for elucidating and predicting additive effects in the halogenation of  $F_7$  naphthalene. SHAP analysis revealed that moderately balanced functional-group charges and non-negative aromatic contributions are critical for yield improvement. Based on these criteria, additive **3a** was designed and experimentally validated, delivering high yields in iodination, bromination, and chlorination. These results underscored the utility of machine learning for rational additive design and



suggested a broadly applicable strategy for advancing organic synthesis beyond empirical approaches.

The findings also provide important implications: they demonstrate that explicit molecular descriptors can be directly connected to reactivity trends, offering a more predictive and interpretable framework than conventional qualitative rationalizations. Nevertheless, the present study is limited to F<sub>7</sub> naphthalene and a relatively narrow set of aromatic additives. Further investigations involving different fluorinated substrates, structurally diverse additives, and other reaction classes are required to examine the scope and generality of the methodology. Future work should also address the transferability of these design principles to broader contexts of reaction development.

Overall, our results highlight the potential of integrating fluorine chemistry with data science to establish predictive frameworks for reaction design, providing a foundation for more systematic and efficient molecular transformations in organic synthesis.

## Author contributions

Naoya Ohtsuka performed the preliminary investigation, curated the data, carried out the machine learning analysis and experimental validation, and prepared the original draft. Muhammad Zhafran Mohd Aris conducted the experimental screening. Toshiyasu Suzuki conducted the DFT calculations. Norie Momiyama supervised the project, contributed to conceptualization and funding acquisition, and revised the manuscript through review and editing.

## Conflicts of interest

There are no known competing financial conflicts to declare.

## Data availability

All data collected and analyzed are included in the manuscript and supplementary information (SI). Supplementary information includes experimental procedures and characterization data for synthesized compounds, DFT calculation details, machine learning, NMR spectra, and Cartesian coordinates of the optimized structures. See DOI: <https://doi.org/10.1039/d5cp03554f>.

## Acknowledgements

This work was supported by a Grant-in-Aid for Transformative Research Areas (A) "Digitalization-driven Transformative Organic Synthesis (Digi-TOS)" from the Ministry of Education, Culture, Sports and Technology (MEXT), Japan (Grant Number JP21H05218). Part of this study was conducted at the Institute for Molecular Science and supported by the Advanced Research Infrastructure for Materials and Nanotechnology in Japan (Organic Synthesis DX, Grant Numbers JPMXP1222MS0014,

JPMXP1222MS0042, and JPMXP1223MS0005) from MEXT, Japan, and by JST Moonshot R&D from the Japan Science and Technology Agency (JST) (Grant Number JPMJMS2236-7). We deeply thank Prof. Takashi Ohshima for his insightful discussions on this study.

## Notes and references

- 1 Y. Lu, Y. Liu, H. Li, X. Zhu, H. Liu and W. Zhu, Energetic effects between halogen bonds and anion- $\pi$  or lone pair- $\pi$  interactions: A theoretical study, *J. Phys. Chem. A*, 2012, **116**, 2591–2597.
- 2 N. Ma, Y. Zhang, B. Ji, A. Tian and W. Wang, Structural competition between halogen bonds and lone-pair- $\pi$  interactions in solution, *Chem. Phys. Chem.*, 2012, **13**, 1411–1414.
- 3 Y. Zhang, B. Ji, A. Tian and W. Wang, Competition between  $\pi$ - $\pi$  interaction and halogen bond in solution: A combined <sup>13</sup>C NMR and density functional theory study, *J. Chem. Phys.*, 2012, **136**, 141101.
- 4 X. R. Zhao, H. Wang and W. J. Jin, The competition of C-X...O = P halogen bond and  $\pi$ -hole...O = P bond between halopentafluorobenzenes C<sub>6</sub>F<sub>5</sub>X (X = F, Cl, Br, I) and triethylphosphine oxide, *J. Mol. Model.*, 2013, **19**, 5007–5014.
- 5 X. Q. Yan, X. R. Zhao, H. Wang and W. J. Jin, The competition of  $\sigma$ -hole...Cl<sup>-</sup> and  $\pi$ -hole...Cl<sup>-</sup> bonds between C<sub>6</sub>F<sub>5</sub>X (X = F, Cl, Br, I) and the chloride anion and its potential application in separation science, *J. Phys. Chem. B*, 2014, **118**, 1080–1087.
- 6 H. Wang, W. Wang and W. J. Jin,  $\sigma$ -Hole Bond vs  $\pi$ -Hole Bond: A Comparison Based on Halogen Bond, *Chem. Rev.*, 2016, **116**, 5072–5104.
- 7 A. V. Rozhkov, A. A. Eliseeva, S. V. Baykov, B. Galmés, A. Frontera and V. Y. Kukushkin, One-pot route to X-perfluoroarenes (X = Br, I) based on Fe<sup>III</sup>-assisted C-F functionalization and utilization of these Arenes as building blocks for crystal engineering involving halogen bonding, *Cryst. Growth Des.*, 2020, **20**, 5908–5921.
- 8 N. Ohtsuka, H. Ota, S. Sugiura, S. Kakinuma, H. Sugiyama, T. Suzuki and N. Momiyama, Perfluorohalogenated naphthalenes: Synthesis, crystal structure, and intermolecular interaction, *CrystEngComm*, 2024, **26**, 764–772.
- 9 N. Saito, A. Nawachi, Y. Kondo, J. Choi, H. Morimoto and T. Ohshima, Functional group evaluation kit for digitalization of information on the functional group compatibility and chemoselectivity of organic reactions, *Bull. Chem. Soc. Jpn.*, 2023, **96**, 465–474.
- 10 H. Kanda, A. Okabe, S. Harada and T. Nemoto, Systematic studies of functional group tolerance and chemoselectivity in carbene-mediated intramolecular cyclopropanation and intermolecular C-H functionalization, *Chem. Pharm. Bull.*, 2024, **72**, 313–318.
- 11 E. Sato, M. Fujii, K. Mitsudo and S. Suga, Alkynylation of aldehydes initiated by cathodic reduction, *ChemElectroChem*, 2024, **11**, e202300499.



- 12 S. Tamaki, T. Kusamoto and H. Tsurugi, Decarboxylative alkylation of carboxylic acids with easily oxidizable functional groups catalyzed by an imidazole-coordinated Fe<sub>3</sub> cluster under visible light irradiation, *Chemistry*, 2024, **30**, e202402705.
- 13 J. Choi, A. Nawachi, N. Saito, Y. Kondo, H. Morimoto and T. Ohshima, Evaluation of functional group compatibility and development of reaction-accelerating additives in ammonium salt-accelerated hydrazinolysis of amides, *Front. Chem.*, 2024, **12**, 1378746.
- 14 Y. Hisata, T. Washio, S. Takizawa, S. Ogoshi and Y. Hoshimoto, In-silico-assisted derivatization of triarylboranes for the catalytic reductive functionalization of aniline-derived amino acids and peptides with H<sub>2</sub>, *Nat. Commun.*, 2024, **15**, 3708.
- 15 A. Tagata, M. Sawamura and Y. Shimizu, Oxyamination of alkenes enabled by direct photoexcitation of fluorenone oxime derivatives, *Chem. Lett.*, 2024, **53**, upae216.
- 16 T. Matsuo, M. Sano, Y. Sumida and H. Ohmiya, Organic photoredox-catalyzed unimolecular PCET of benzylic alcohols, *Chem. Sci.*, 2025, **16**, 3150–3156.
- 17 T. Imamoto, N. Takiyama, K. Nakamura, T. Hatajima and Y. Kamiya, Reactions of Carbonyl Compounds with Grignard Reagents in the Presence of Cerium Chloride, *J. Am. Chem. Soc.*, 1989, **111**, 4392–4398.
- 18 E. M. Vogl, H. Geöger and M. Shibazaki, Towards Perfect Asymmetric Catalysis: Additives and Cocatalyst, *Angew. Chem., Int. Ed.*, 1999, **38**, 1570–1577.
- 19 F. F. Kneisel, M. Dochnahl and P. Knochel, Nucleophilic Catalysis of the Iodine–Zinc Exchange Reaction: Preparation of Highly Functionalized Diaryl Zinc Compounds, *Angew. Chem., Int. Ed.*, 2004, **43**, 1017–1021.
- 20 A. Krasovski and P. Knochel, A LiCl-Mediated Br/Mg Exchange Reaction for the Preparation of Functionalized Aryl- and Heteroaryl Magnesium Compounds from Organic Bromides, *Angew. Chem., Int. Ed.*, 2004, **43**, 3333–3336.
- 21 A. Bellomo, J. Zhang, N. Trongsirivat and P. J. Walsh, Additive effects on palladium-catalyzed deprotonative cross-coupling processes (DCCP) of sp<sup>3</sup> C–H bonds in diaryl-methanes, *Chem. Sci.*, 2013, **4**, 849–857.
- 22 L. Hong, W. Sun, D. Yang, G. Li and R. Wang, Additive Effects on Asymmetric Catalysis, *Chem. Rev.*, 2016, **116**, 4006–4123.
- 23 M. Balkenhohl and P. Knochel, Recent Advances of the Halogen–Zinc Exchange Reaction, *Chem. – Eur. J.*, 2019, **26**, 3688–3697.
- 24 Z. Lu, T. Li, S. R. Mudshinge, B. Xu and G. B. Hammond, Optimization of Catalysts and Conditions in Gold(I) Catalysis–Counterion and Additive Effects, *Chem. Rev.*, 2021, **121**, 8452–8477.
- 25 R. L. de Carvalho, E. B. Diogo, S. L. Homölle, S. Dana, E. N. de Silva Júnior and L. Ackermann, The crucial role of silver(I)-salts as additives in C–H activation reactions: overall analysis of their versatility and applicability, *Chem. Soc. Rev.*, 2023, **52**, 6359–6378.
- 26 D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher and A. G. Doyle, Predicting reaction performance in C–N cross-coupling using machine learning, *Science*, 2018, **360**, 186–190.
- 27 A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow and S. E. Denmark, Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning, *Science*, 2019, **363**, eaau5631.
- 28 F. Strieth-Kalthoff, F. Sandford, M. H. S. Segler and F. Glorius, Machine learning the ropes: principles, applications and directions in synthetic chemistry, *Chem. Soc. Rev.*, 2020, **49**, 6154–6168.
- 29 S. Matsubara, Digitization of Organic Synthesis –How Synthetic Organic Chemists Use AI Technology, *Chem. Lett.*, 2020, **50**, 475–481.
- 30 W. L. Williams, L. Zeng, T. Gensch, M. S. Sigman, A. G. Doyle and E. Anslyn, The Evolution of Data-Driven Modeling in Organic Chemistry, *ACS Cent. Sci.*, 2021, **7**, 1622–1637.
- 31 T. Hori, S. Kakinuma, N. Ohtsuka, T. Fujinami, T. Suzuki and N. Momiyama, Synthesis of Halogen-Bond-Donor-Site-Introduced Functional Monomers through Wittig Reaction of Perfluorohalogenated Benzaldehydes: Toward Digitalization as Reliable Strategy in Small-Molecule Synthesis, *Synlett*, 2023, 2455–2460.
- 32 S.-Q. Zhang, L.-C. Xu, S.-W. Li, J. C. Oliveira, X. Li, L. Ackermann and X. Hong, Bridging Chemical Knowledge and Machine Learning for Performance Prediction of Organic Synthesis, *Chem. – Eur. J.*, 2023, **29**, e202202834.
- 33 K. Takeda, N. Ohtsuka, T. Suzuki and N. Momiyama, Prediction Method for Reaction Yield of Deuteration of Polyfluoroperylene using Generative AI Techniques, *Comput. Aided Chem. Eng.*, 2024, **53**, 2689–2694.
- 34 S. A. A. Rizvi, J. Meng, M. E. I. Khan and X. Jiang, Machine learning advancements in organic synthesis: A focused exploration of artificial intelligence applications in chemistry, *Artif. Intell. Chem.*, 2024, **2**, 100049.
- 35 F. Hastedt, R. M. Bailey, K. Hellgardt, S. N. Yaliraki, E. A. del Rio Chanona and D. Zhang, Investigating the reliability and interpretability of machine learning frameworks for chemical retrosynthesis, *Digital Discovery*, 2024, **3**, 1194–1212.
- 36 N. Momiyama, H. Okamoto, M. Shimizu and M. Terada, Synthetic Method for 2,2'-Disubstituted Fluorinated Binaphthyl Derivatives and Application as Chiral Source in Design of Chiral Mono- Phosphoric Acid Catalyst, *Chirality*, 2015, **27**, 464–475.
- 37 A. V. Marenich, C. J. Cramer and D. G. Truhlar, Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.
- 38 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe,



- V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16, Revision A.03*, Gaussian, Inc., Wallingford CT, 2016.
- 39 Datachemical LAB (accessed August 27, 2024): <https://www.datachemicallab.com>.
- 40 H. Kaneko, Cross-validated permutation feature importance considering correlation between features, *Anal. Sci. Adv.*, 2022, **3**, 278–287.
- 41 S. M. Lundberg and S.-I. Lee, A unified approach to interpreting model predictions, *Adv. Neural Inf. Process. Syst.*, 2017, **30**, 4765–4774.
- 42 T. Lu and Q. Chen, Independent gradient model based on Hirshfeld partition: A new method for visual study of interactions in chemical systems, *J. Comput. Chem.*, 2022, **43**, 539–555.

