ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Prediction of Photochromism of Salicylideneaniline Crystals Using a Data Mining Approach

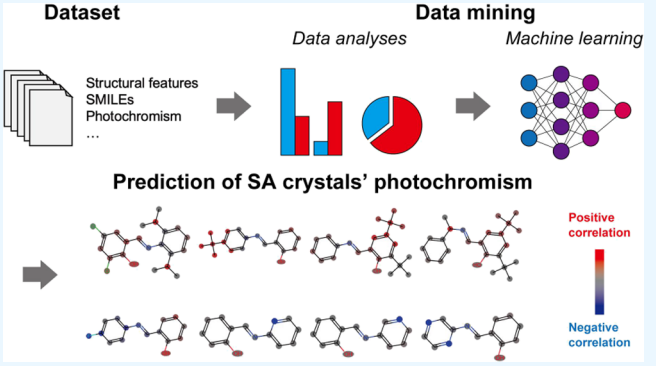
Shodai Hasebe, Kan Hatakeyama-Sato, Kenichi Oyaizu, Toru Asahi, and Hideko Koshima*

Cite This: *ACS Omega* 2024, 9, 1463−1471

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Salicylideneanilines (SAs) are photochromic compounds that undergo enol−keto photoisomerization in the solid state. Research over the past 60 years has revealed empirically that SAs with steric and planar conformations tend to be photochromic and nonphotochromic, respectively. However, increasing counterexamples in the recent literature raise questions about the nature of the relationship between structure and photochromism in SA crystals and whether the photochromism of SA crystals is predictable. This study is the first to construct a data set on SA crystals and conduct a comprehensive analysis to investigate the relationship between molecular and crystal structures and photochromism. A data mining approach revealed that the dihedral angle is the most dominant structural parameter for photochromism, followed by the Hirshfeld surface volume. SAs with neutral bulky hydrocarbon groups, such as the *tert*-butyl group, tend to be photochromic because such SAs have steric conformation and a loosely packed structure. In contrast, SAs with fluorine, pyridine, and pyrazine are less likely to be photochromic due to their planar conformation and densely packed structures. The photochromism of the SA crystals in our data set was predicted with high accuracy (>85%) using machine learning. The results of this study provide a useful reference for designing SA crystals with desired photochromic properties.



INTRODUCTION

Photochromic compounds have gained significant interest due to their capacity to undergo reversible transformations between two isomers possessing distinct absorption spectra upon light irradiation. This unique property enables them to undergo reversible color changes upon exposure to light, rendering them promising candidates for applications in sensors, switches, and actuators.[1−3] Salicylideneaniline (SA) and its derivatives are T-type (thermally and photochemically reversible) photochromic compounds that undergo enol to *trans*-keto photoisomerization.[4−7] Under ultraviolet (UV) light irradiation, the enol-form molecule is vertically excited and undergoes excited-state intramolecular proton transfer to produce the excited *cis*-keto form, followed by pedal motion, an out-of-plane motion around the imine bond, which guides the excited *cis*-keto form to the ground state *trans*-keto form (Figure 1).[8−11]

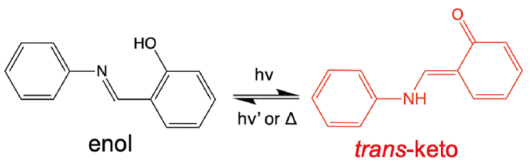
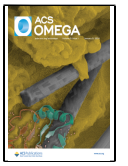


Figure 1. Enol−keto photoisomerization of salicylideneaniline (SA).

SAs exhibit either photochromic or nonphotochromic properties in the crystalline state.[4−7] Previous studies have revealed that the photochromic properties of SA crystals are related to the molecular conformation rather than the chemical properties of the compounds. Cohen et al.[4] first reported that steric (nonplanar) SAs are more likely to be photochromic, whereas planar SAs are not. Subsequent investigations confirmed the validity of this observation, revealing that enol-keto photoisomerization involves pedal motion, necessitating ample spatial freedom around the molecule; i.e., a steric conformation is preferable for the photochromic characteristics.[6,7,12] Johmoto et al.[13] proposed the "SA planarity rule", a quantitative criterion for photochromism that SAs with a dihedral angle (between two benzene rings) of >30° are photochromic, whereas those with a dihedral angle of <20° are nonphotochromic. Carletta et al.[14] discovered that the C/C contact ratio in the form of the C−C% value, i.e., the percentage of $\pi-\pi$ stacking interactions relative to all

Received: October 9, 2023
Revised: November 28, 2023
Accepted: December 13, 2023
Published: December 22, 2023



intermolecular interactions, is also a reliable indicator. Sugiyama et al.[15] revealed that in addition to the dihedral angle, it is important to consider the average free available space and packing index, which are parameters used for determining an open or close-packed structure; however, they considered the dihedral angle to be the dominant factor for photochromism. Mercier et al.[16] reported that when the length of the C−O bond at the salicyl plane is <1.34 Å, the crystals are nonphotochromic; they also proposed that SAs having many strong, short contacts around the photochromic center are less likely to be photochromic.

Although these criteria are useful for predicting and understanding the photochromic properties of SA crystals based on their crystal structures, it remains uncertain which structural parameters are the most important. It is also unclear whether these criteria can be applied to all SA crystals and their derivatives. Furthermore, few studies have attempted a comprehensive analysis of the relationship between the molecular and crystal structures of SA crystals and their photochromic properties. Understanding these hidden relationships through the detailed analysis of known SA crystals would be beneficial for predicting crystal structures and photochromism from molecular structures prior to the fabrication of single crystals, thereby enabling more efficient development of new SA crystals with desired properties.

In this study, we constructed the first data set containing molecular and crystal structural information and the photochromism properties of over 331 SA molecules, based on 46 previous studies[7,10−54] published within the past 60 years. All 331 SAs are organic molecular crystals, of which 260 SAs are single components and the remaining 71 SAs are multicomponent. The data set was analyzed using a data mining approach to uncover hidden patterns and relationships between molecular and crystal structures and photochromism. We also constructed a machine learning model to predict the photochromism of the SA crystals.

RESULTS AND DISCUSSION

Re-Evaluation of Previous Photochromism Rules. Before investigating the relationship between the crystal and chemical structures of SA crystals and their photochromic properties, we first re-evaluated the three SA photochromism rules reported to date. The following SAs were omitted for re-evaluation: (1) multicomponent, (2) the X-ray diffraction measurement temperature is below 273 K, (3) photochromism is not reported, and (4) structural features (the dihedral angle, C/C contact ratio, or C−O length) are not reported or cannot be calculated from the cif files. Figure 2a illustrates the dihedral angle rule,[13] showing eight photochromic crystals with $\varphi < 20°$ and six nonphotochromic crystals with $\varphi > 30°$, for a total of 14 exceptions among 96 crystals (15%, Table 1). Figure 2b illustrates the C/C contact ratio rule,[14] in which six crystals are nonphotochromic with a C−C% value of ≤2.6% and three samples are photochromic with a C−C% value of ≥5.2%, for a total of nine exceptions among 77 crystals (12%, Table 1). Figure 2c illustrates the C−OH length rule[16] in which 11 crystals among 17 crystals with a short C−O length of <1.34 Å are photochromic. This high exception ratio could be explained by variation in the C−O length, depending on whether the researcher refined the crystal structure as having an enol or *cis*-keto form, which results in longer or shorter C−O length, respectively. In addition to the high exception ratio, this C−OH length rule cannot explain the photochromism of

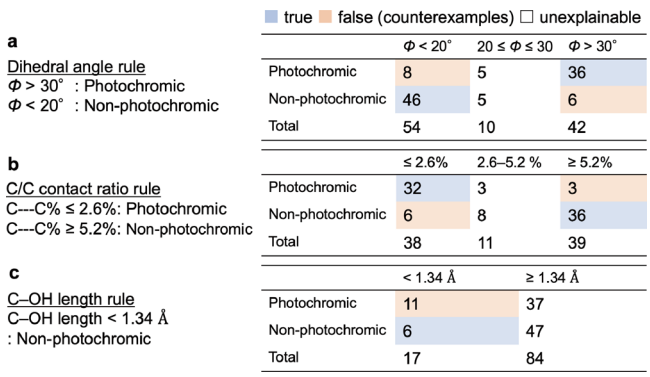


Figure 2. Re-evaluation of previously reported SA photochromic rules. (a) Dihedral angle rule.[13] (b) C/C contact ratio rule.[14] (c) C−OH length rule.[16]

SA crystals with a long C−O length of ≥1.34 Å (83%, Table 1), which makes it not attractive for predicting photochromism in SA crystals.

Relationship between Crystal Structures and Photochromism. The dihedral angle rule (Figure 2a) and the C/C contact ratio rule (Figure 2b) can explain the photochromism of 85−88% SA crystals; however, 12−15% crystals do not follow the rules and 9−13% are unexplainable (Table 1). To unveil the hidden relationships between the crystal structure and photochromism, we prepared a data set comprising the molecular structure, crystal structure, and photochromism of SA crystals and analyzed the set using Orange.[55] The following SAs were excluded from the analysis: (1) multicomponent, (2) the X-ray diffraction measurement temperature is below 273 K, (3) photochromism is not reported, and (4) at least one of 12 structural features (see below) are not reported or cannot be calculated from the cif files. As a result, 81 SAs were used for analysis to unveil the relationship between the crystal structure and photochromism. First, we calculated Gini and the χ^2 values of 12 structural features related to photochromism and ranked the features according to their effect on SA photochromism; the structural features are defined in Table S1. Gini values represent inequality among values of a frequency distribution, taking values from 0 (no bias) to 1 (extreme bias). The χ^2 values indicate the dependence between the feature and class as measured by χ^2 statistics. Higher Gini and χ^2 values indicate a strong correlation with the feature and photochromism.

The dihedral angle showed the highest Gini and χ^2 values of 0.27 and 34.41, followed by the Hirshfeld surface volume, packing index, void volume per molecule, density, and C/C contact ratio (Figure 3a). Photochromic SAs exhibited a higher dihedral angle, a higher Hirshfeld surface volume, and a higher void volume per molecule as well as a lower packing index, a lower density, and a smaller C/C contact ratio compared to nonphotochromic SAs (Figure 3b−g, Table S3). Previous studies have reported that nonplanar and loosely packed SAs tend to be photochromic,[4−7] which is consistent with our analysis results, i.e., a high dihedral angle (Figure 3b) and void volume per molecule (Figure 3e) and a low packing index (Figure 3d), low density (Figure 3f), and low C/C contact ratio (Figure 3g). This study is the first to reveal the Hirshfeld surface volume as a new indicator of the photochromism of SAs (Figure 3c). The distributions of the six remaining features with low Gini and χ^2 values are shown in Figure S4. Nonphotochromic SAs showed stronger intermolecular interactions around the photochromic center than photo-

https://doi.org/10.1021/acsomega.3c07859
ACS Omega 2024, 9, 1463−1471

Table 1. Number and Percentage of SA Crystals That Follow (True), Do Not Follow (False), and Are Unexplainable by the Photochromism Rules

features	figure	true	false	explainable	unexplainable
		(true/true + false)	(false/true + false)	(explainable/total)	(unexplainable/total)
dihedral angle	2a	82 (85%)	14 (15%)	96 (91%)	10 (9.4%)
C/C contact ratio	2b	68 (88%)	9 (12%)	77 (87%)	11 (13%)
C−OH length	2c	6 (35%)	11 (65%)	17 (17%)	84 (83%)
dihedral angle and void volume per molecule	4a, b	69 (99%)	1 (1.4%)	70 (86%)	11 (14%)
dihedral angle and Hirshfeld surface volume	4c, d	68 (99%)	1 (1.4%)	69 (85%)	12 (15%)

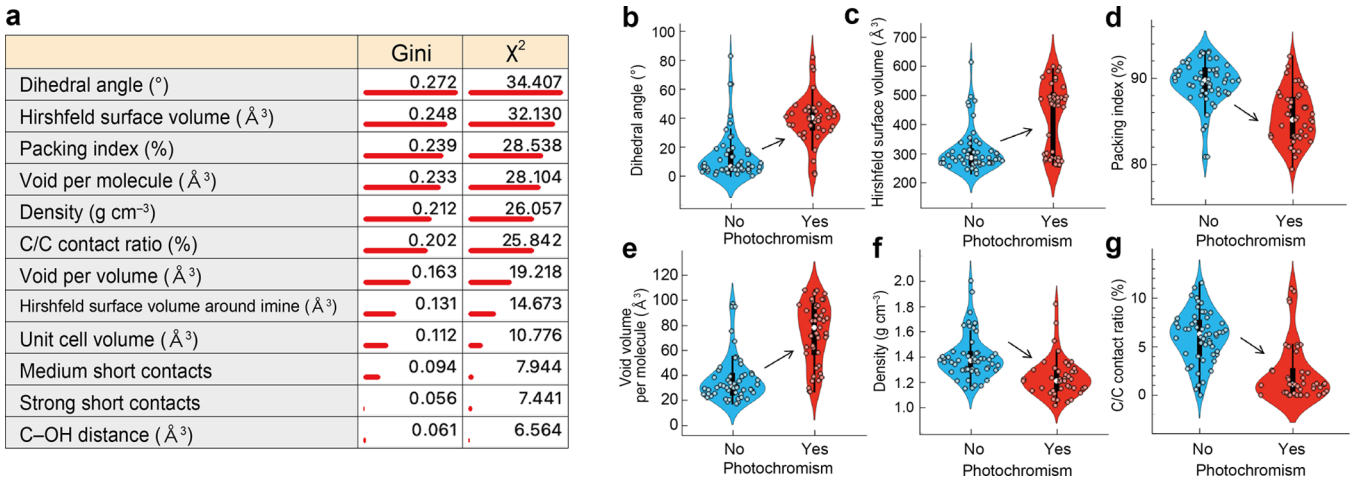


Figure 3. (a) Gini and χ^2 values of 12 structural features of photochromism. (b−g) Distributions of the (b) dihedral angle, (c) Hirshfeld surface volume, (d) packing index, (e) void volume per molecule, (f) density, and (g) C/C contact ratio values of 81 SA crystals. blue: nonphotochromic, red: photochromic.

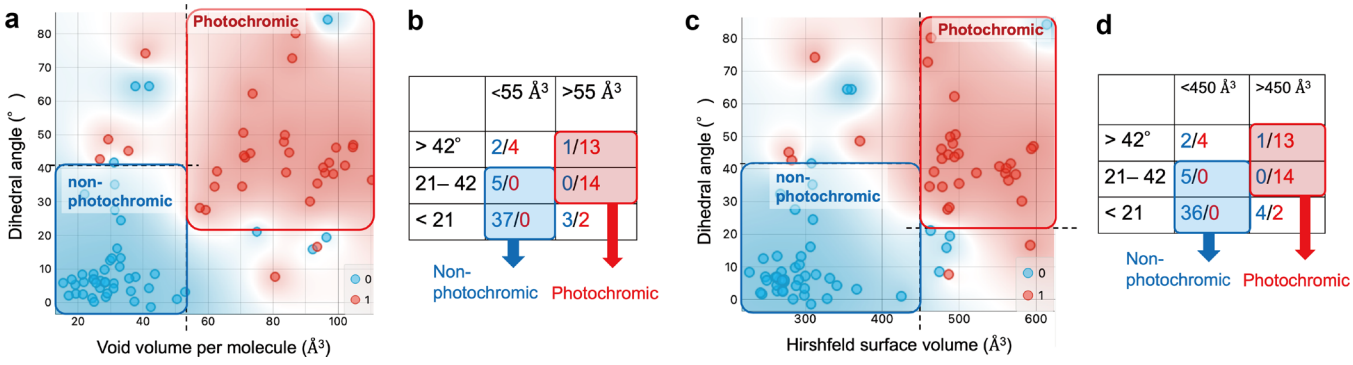


Figure 4. (a, c) Combination plot of (a) dihedral angle and void volume per molecule and (c) dihedral angle and Hirshfeld surface volume (n = 81). Blue: nonphotochromic; red: photochromic. (b, d) Distribution of the number of SA crystals based on (b) dihedral angle (row) and void volume per molecule (column) and (d) dihedral angle (row) and Hirshfeld surface volume (column). (b, d) Blue and red numbers in the cells indicate the number of nonphotochromic and photochromic SAs, respectively.

chromic SAs (Figure S4e), as these short contacts may inhibit photoisomerization.[16]

Although the six features shown in Figure 3b−g are good indicators of the photochromism of SA crystals, using only one feature is insufficient to explain SA photochromism with high accuracy (Figure 2a,b and Table 1). Therefore, intelligent data visualization was performed using Orange[55] to determine the best feature pairs for the accurate classification of photochromic and nonphotochromic SAs. Figure 4a shows a combination plot of the dihedral angle and void volume per molecule for 81 SA crystals; photochromic SAs mainly have a high dihedral angle and void volume per molecule values, whereas these values are low in nonphotochromic SAs. Figure 4b displays the distribution of the number of SA crystals based on these parameters.

In this study, we propose new SA photochromism rules by combining two structural features. Crystals with a high dihedral angle (>21°) and void volume per molecule (>55 Å^3) are photochromic (27 of 28) and those with a low dihedral angle (<42°) and void volume per molecule (<55 Å^3) are nonphotochromic (42 of 42). This proposed guideline was remarkably accurate (true: 99%, false: 1.4%, Figure 4a, b, Table 1) compared to the conventional photochromism rules (true: ≤ 88%, false: ≥ 12%, Figure 2, Table 1). A strong positive correlation [Pearson coefficient (R^2) = 0.95; Figures S5 and S6] confirmed that photochromism could be explained by a combination of the dihedral angle and Hirshfeld surface

https://doi.org/10.1021/acsomega.3c07859
ACS Omega 2024, 9, 1463−1471

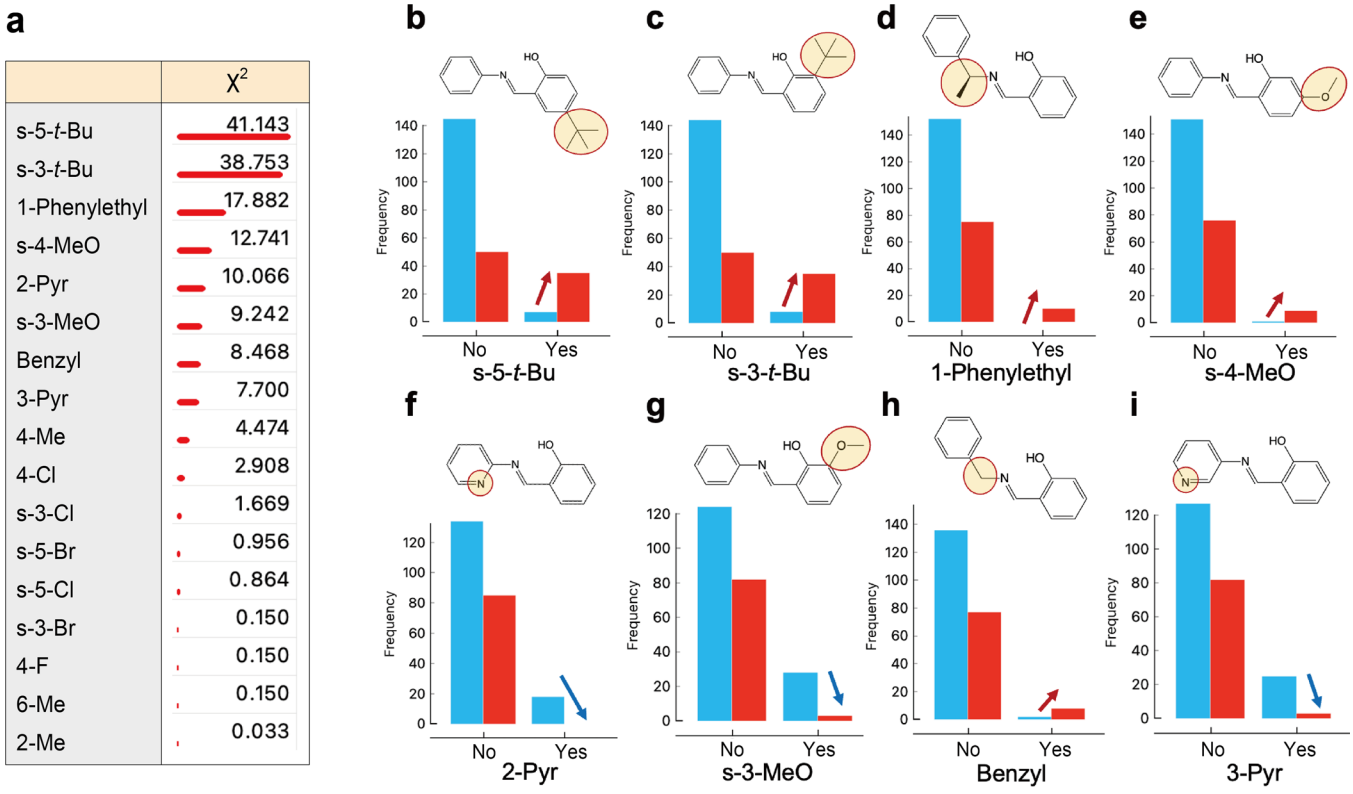


Figure 5. (a) χ^2 values of 17 features of molecular structures (substituents) with potential influence on photochromism. (b–i) Eight substituents with the highest χ^2 values. Frequency distribution of photochromic (red) and nonphotochromic (blue) SA crystals with (1) and without (0) (b) s-5-*t*-Bu, (c) s-3-*t*-Bu, (d) 1-phenylethyl, (e) s-4-MeO, (f) 2-Pyr, (g) s-3-MeO, (h) benzyl, and (i) 3-Pyr groups (n = 237).

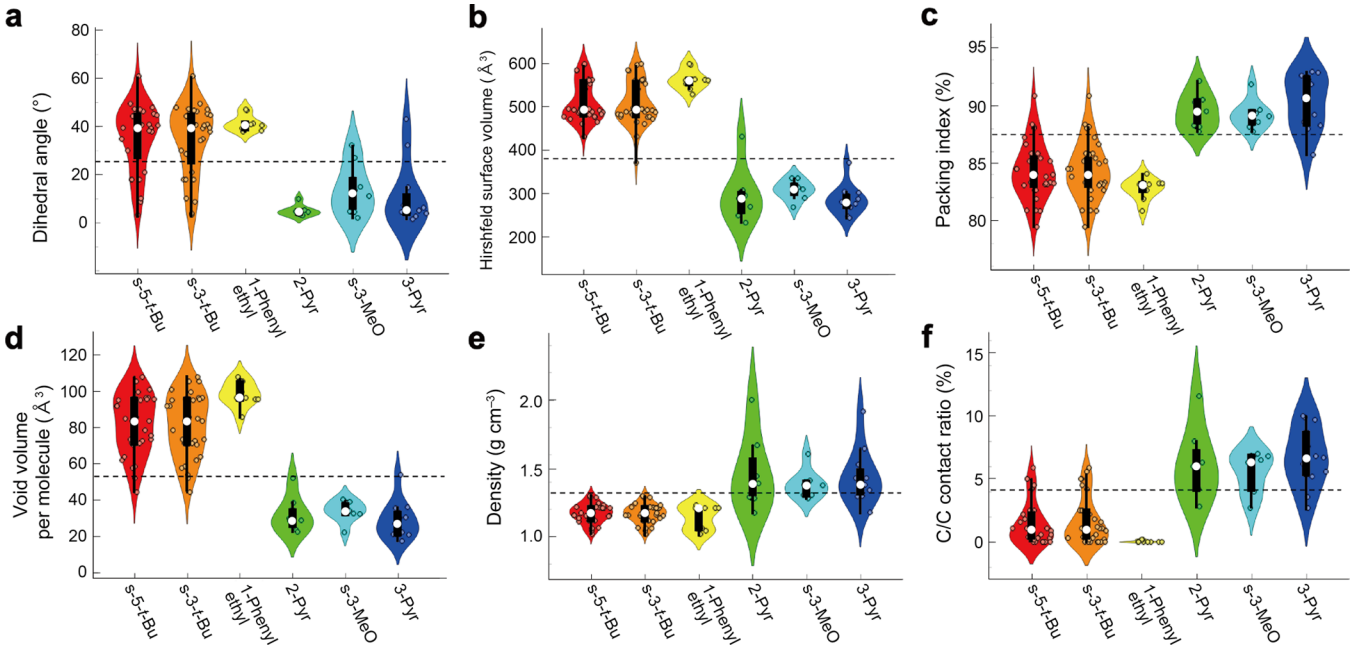


Figure 6. Distribution of (a) dihedral angle, (b) Hirshfeld surface volume, (c) packing index, (d) void volume per molecule, (e) density, and (f) C/C contact ratio values of 81 SA crystals for each substituent. Red: s-5-*t*-Bu; orange: s-3-*t*-Bu; yellow: 1-phenylethyl; green: 2-Pyr; blue: s-3-MeO; dark blue: 3-Pyr. Dashed lines indicate average values of 81 SA crystals.

volume (true: 99%, false: 1.4%, Figure 4c,d, Table 1). The void volume per molecule and Hirshfeld surface volume values were the main parameters to determine photochromism of SA crystals with intermediate dihedral angle values (21–42°) possibly because these two parameters directly represent spatial freedom of these crystals, which is necessary for photoisomerization, rather than the dihedral angle. The only counterexample was the crystal of *N*-(3,5-dichlorosalicylidene)-2,4,6-tri-*tert*-butylaniline,[25] which is not photochromic despite its large dihedral angle (82.9°), Hirshfeld surface

https://doi.org/10.1021/acsomega.3c07859
ACS Omega 2024, 9, 1463−1471

volume (614.9 Å^3), and void volume per molecule (97.05 Å^3). In this case, the bulky *tert*-butyl groups at the 2,6-positions are oriented toward the imine moiety, which prevents its pedal motion, resulting in nonphotochromic properties.[25]

Relationship between Molecular Structures and Photochromism. In the previous section, we revealed the relationship between the crystal structure and photochromism. Next, we focused on the relationship between the molecular structure and photochromism. To identify the substituents influencing photochromism, the χ^2 value for 17 such substituents (Figure S2) was calculated. The salicyl-5-*tert*-butyl (s-5-*t*-Bu) group had the highest χ^2 value of 41.1, followed by salicyl-3-*tert*-butyl (s-3-*t*-Bu, 38.8), 1-phenylethyl (17.9), salicyl-4-methoxy (s-4-MeO, 12.7), 2-pyridine (2-Pyr, 10.1), salicyl-3-methoxy (s-3-MeO, 9.24), benzyl (8.47), and 3-pyridine (3-Pyr, 7.70) (Figure 5a). Given that the critical values for 1 degree of freedom are 10.8 and 6.63 for significance levels of 0.1% and 1%, respectively, the possibility that there is no correlation between photochromism and the existence of an influential substituent was <0.1% for s-5-*t*-Bu, s-3-*t*-Bu, 1-phenylethyl, and s-4-MeO and <1% for 2-Pyr, s-3-MeO, benzyl, and 3-Pyr, respectively. Figure 5b−i shows the frequency distribution of photochromic and nonphotochromic SA crystals with and without representative substituents with high χ^2 values. We found that s-5-*t*-Bu, s-3-*t*-Bu, 1-phenylethyl, s-4-MeO, and benzyl increased the possibility of photochromism, whereas 2-Pyr, s-3-MeO, and 3-Pyr decreased the possibility. s-5-*t*-Bu and s-3-*t*-Bu exhibited almost the same trend (Figure 5b,c) because 31 of 32 SAs with s-3-*t*-Bu have s-5-*t*-Bu as well. In particular, all 10 SAs with 1-phenylethyl were photochromic (Figure 5d), whereas all 18 SAs with 2-Pyr were nonphotochromic (Figure 5f).

To rationalize the influence of the representative substituents shown in Figure 5a on photochromism, we examined how each substituent affects crystal structure features (Figure 6 and Tables S6−S15). The s-3-*t*-Bu and s-5-*t*-Bu groups showed an increased void volume per molecule and a lower packing index (Figure 6c,d, Tables S6 and S7). This result aligns with the previous report by Kawato et al.[56] that *t*-Bu groups increase the free space of the crystal. In addition, *t*-Bu groups increased the dihedral angle and Hirshfeld surface volume and reduced the density and C/C contact ratio, indicating that *t*-Bu groups make the SA crystal structure more steric (nonplanar) and less dense (Figure 6a,b,e,f, Tables S6 and S7). These trends also applied to 1-phenylethyl (Figure 6, yellow bars, Table S8), in which a greater shift in the featured values was observed compared with that of the *t*-Bu group. Thus, the insertion of 1-phenylethyl could be the more reliable approach to form steric and photochromic SA crystals. Indeed, all nine of the SAs with 1-phenylethyl were photochromic (Figure 5d).

In contrast, 2-Pyr showed a significantly reduced mean dihedral angle, from 25.8° (all SAs) to 5.03° (Figure 6a, Table S9), in good agreement with a previous report by Moustakali-Mavridis et al.[19] that the N atom in the pyridine ring undergoes an intramolecular interaction with the H of the imine, locking the salicyl and pyridine ring in the plane. Additionally, 2-Pyr decreased the Hirshfeld surface volume and void volume per molecule and increased the packing index, density, and C/C contact ratio, resulting in denser, more planar structures (Figure 6b−f, Table S9). Similarly, 3-Pyr shifted the crystal structure features as well as 2-Pyr, resulting in denser planar structures (Figure 6 and Table S11). Hadjoudis et al.[21] reported that 3-Pyr creates denser structures that are not as planar as 2-Pyr because the N atom in the pyridine ring is not involved in the intramolecular interaction with the H of the imine, which is consistent with our dihedral angle results (2-Pyr ≤9.96° and 3-Pyr ≤43.07°, Figure 6a). In fact, no SAs with 2-Pyr were photochromic (Figure 5f) according to our findings, whereas 3 of 28 SAs with 3-Pyr were found to be photochromic (Figure 5i). Our results also revealed that the s-3-MeO group contributes to denser, more planar structures (Figure 6, blue bars, Table S10). Although we were unable to investigate the influence of s-4-MeO and benzyl groups on photochromism due to insufficient structural information for the SAs with s-4-MeO and benzyl groups, we have illustrated the causal relationships between the molecular and crystal structures and photochromism in Figure 7.

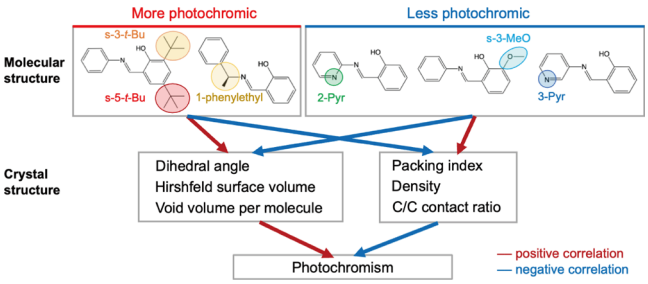


Figure 7. Causal relationships among the molecular structure, crystal structure, and photochromism. Red and blue arrows represent positive and negative correlations, respectively.

Machine Learning-Based Validation of Molecular Structural Influences on Crystal Structures. As part of our comprehensive analysis of the influence of molecular structures on crystal structures, we visualized these structures using a machine learning approach. First, we constructed a machine learning model to predict the crystal structural features, including dihedral angle, Hirshfeld surface volume, packing index, void volume per molecule, density, and C/C contact ratio, based on molecular structures. A gradient-boosted decision tree machine learning model implemented in XGBoost[57] was employed for the model; XGBoost is used in various material science fields as it solves nonlinear regression problems, easily interprets the predicted output, and outperforms other algorithms in terms of prediction accuracy.[58−60] The molecular structures were represented as Morgan fingerprints for machine learning.

Table 2 shows the R^2 and mean absolute error values for predictions of the dihedral angle, Hirshfeld surface volume, packing index, void volume per molecule, density, and C/C contact ratio. The high R^2 scores of the training (>0.99) and testing (>0.80) data for the Hirshfeld surface volume and void volume per molecule indicated that the model was capable of accurately predicting these values from the molecular structure.

For the Hirshfeld surface volume and void volume per molecule, we visualized influential substructures using SHapley Additive exPlanations (SHAP),[61] a game theoretical platform used to explain the output of a machine learning model (Figure S12). Figure 8 visualizes the influence of each atom on the Hirshfeld surface volume and void volume per molecule based on SHAP values. The *t*-Bu groups and 1-phenylethyl increased the Hirshfeld surface volume and void volume per molecule (Figure 8, red atoms), in good agreement with the data analysis results shown in Figures 6 and 7. Similarly, isopropyl

https://doi.org/10.1021/acsomega.3c07859
ACS Omega 2024, 9, 1463−1471

Table 2. Prediction of Pearson's Correlation Coefficient (R^2) and Mean Absolute Error Values for Predictions of Structural Features from Molecular Structures Using XGBoost, a Machine Learning Model[a]

	dihedral angle	Hirshfeld surface volume	packing index	void volume per molecule	density	C/C contact ratio
R^2 train	0.92 ± 0.05	1.00	0.80 ± 0.02	0.96 ± 0.02	0.97 ± 0.03	0.79 ± 0.01
R^2 test	0.48 ± 0.34	0.94 ± 0.09	0.62 ± 0.10	0.86 ± 0.06	0.80 ± 0.11	0.54 ± 0.08
MAE train	4.58 ± 1.84	4.24 ± 2.61	1.21 ± 0.05	4.42 ± 0.90	0.02 ± 0.01	1.23 ± 0.04
MAE test	10.14 ± 1.42	13.50 ± 5.87	1.69 ± 0.20	7.60 ± 1.95	0.05 ± 0.02	1.66 ± 0.18

[a]Mean ± standard deviation, n = 5.

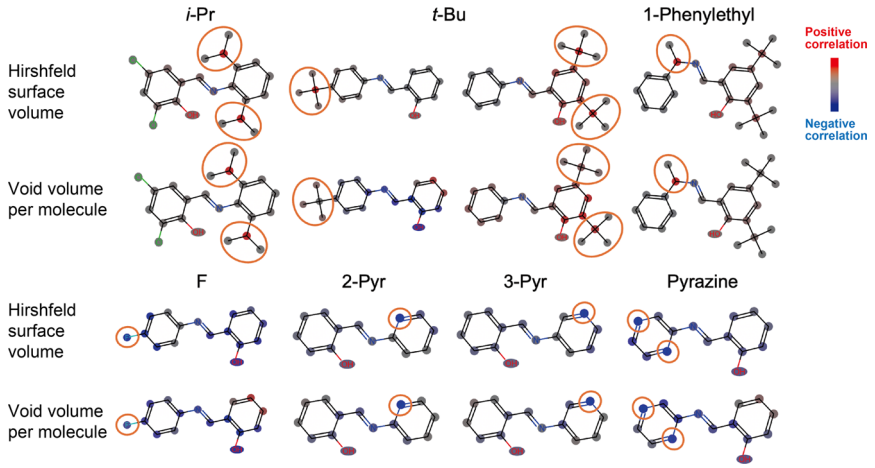


Figure 8. Visualization of structures influencing the Hirshfeld surface volume and void volume per molecule using SHAP. Red, gray, and blue atoms indicate positive, zero, and negative correlations with the structural features, respectively.

(*i*-Pr) groups are composed of electrically neutral and bulky hydrocarbon groups, which enable a more steric and loose conformation of molecules (Figure 8, red atoms). In contrast, 2-Pyr and 3-Pyr were coincident with the data analysis results shown in Figures 6 and 7 (Figure 8, blue atoms). The visualization also revealed that pyrazine has the same effect as 2/3-Pyr (Figure 8) because *ortho*-N in the pyrazine ring forms an intramolecular interaction with the imine H, locking the salicyl and pyrazine rings in the same planes.[44] Additionally, the fluoro (F) group decreases the Hirshfeld surface volume (Figure 8), unlike other halogens (Cl and Br) (Figure S13a,b,g), which is contrary to our expectation that halogens increase the Hirshfeld surface volume because they have larger van der Waals radii than hydrogen.

To rationalize the negative correlation between fluorine and the Hirshfeld surface volume, the crystal structures of nonsubstituted and halogen (X)-substituted SAs (X = F, Cl, and Br) were compared. We found that SAs with halogens have more intermolecular interactions than nonsubstituted SAs (Figure S13e,f), which offset the increase in the Hirshfeld surface volume by halogens (Figure S13c,d). As a result, SAs with fluorine, the halogen with the shortest van der Waals radius, have the smallest Hirshfeld surface volume compared with Cl and Br and even nonsubstituted SAs (Figure S13a,b).

Machine Learning-Based Prediction of Photochromism Based on Molecular/Crystal Structures. SA crystal photochromism was predicted from molecular or structural features using the XGBoost binary classification algorithm.[57]

Prediction of Photochromism from Crystal Structures. Figure 9a (upper panel) shows the prediction performance from 12 crystal structural features in Figure 3a. Both the accuracy and F_1 score were >0.85, indicating that photochromism was predicted with high accuracy from the crystal structures. Figure 9b shows the contributions of the 12 structural features quantified by SHAP values. Like indications based on the Gini and χ^2 values in Figure 3a, the dihedral angle was the most influential factor for photochromism, followed by the Hirshfeld surface volume. Notably, the Hirshfeld surface volume around the imine bond was the third most influential one, suggesting that SA crystals with higher values are more likely to be photochromic, because larger void space around the imine bond is associated with the facilitation of enol−keto pedal motion.[32]

Prediction of Photochromism from Molecular Structures. Figure 9a (lower panel) shows the prediction performance from the molecular structure or molecular fingerprint. Although slightly inferior to the prediction from crystal structures, both accuracy and F_1 score were >0.80 and photochromism was predicted adequately. Figure 9c shows the influence of each atom on photochromism in terms of color, based on the SHAP values, with red, gray, and blue indicating positive, zero, and negative correlations, respectively. *i*-Pr, *t*-Bu, and 1-phenylethyl moieties are indicated in red, such that SAs comprising these groups are more likely to be photochromic, in accordance with the results shown in Figures 5 and 8. In the figures, F, 2-Pyr, 3-Pyr, and pyrazine are indicated in blue, suggesting that SAs with these constituents are more likely to be nonphotochromic (in accordance with the results shown in Figures 5 and 8).

In summary, our data analysis clarified that SAs with *t*-Bu and 1-phenylethyl tend to have more steric and less dense structures, making the crystals more photochromic. In contrast, SAs with 2-Pyr, 3-MeO, and 3-Pyr have less steric and more dense structures, making the crystals less photochromic (Figures 5−7). Machine learning results revealed that the *i*-Pr group also makes the SA crystal structures more steric and less dense, whereas pyrazine and the fluoro (*F*) group have the opposite effect (Figures 8 and 9). Thus, SAs with electrically

https://doi.org/10.1021/acsomega.3c07859
ACS Omega 2024, 9, 1463−1471

a

	Accuracy	F_1 score
Prediction from crystal structures	0.88(0.04)	0.87(0.04)
Prediction from molecular structures	0.85(0.03)	0.84(0.05)

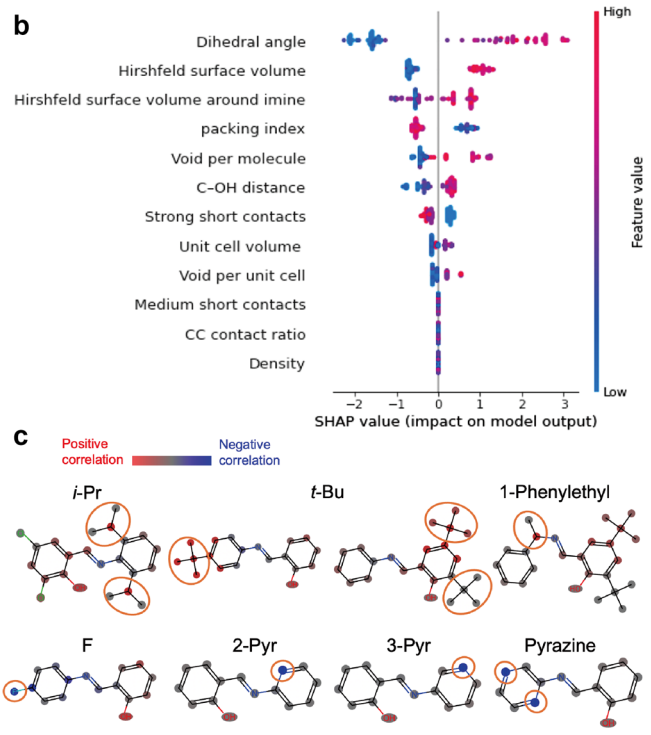


Figure 9. Prediction of photochromism using XGBoost. (a) Accuracy and F_1 score for the testing data (mean ± standard deviation, n = 5). (b) SHAP summary dot plot displaying SHAP values for the prediction of photochromism from crystal structures. (c) Visualization of influential substructures on photochromism using SHAP, where red, gray, and blue atoms indicate positive, zero, and negative correlations with photochromism, respectively.

neutral and bulky hydrocarbon groups tend to show photochromic attributes, whereas those with substituents forming intermolecular interactions are more likely nonphotochromic.

CONCLUSIONS

Information on the molecular and crystal structures of over 300 SA crystals and their photochromic properties were collected from 46 published studies to construct an SA crystal data set. The data set was analyzed using Orange, an open-source data mining toolkit, to reveal hidden relationships between the molecular and crystal structures of SAs and their photochromic characteristics. Among 12 structural parameters, the dihedral angle was shown to have the highest influence on photochromism, accompanied by the Hirshfeld surface volume, packing index, void volume per molecule, density, and C/C contact ratio. Through a combination of the dihedral angle with the void volume per molecule or Hirshfeld surface volume, the photochromism of SA crystals was predicted accurately compared with the conventional dihedral angle rule and the C/C contact ratio rule. Additionally, we statistically revealed that the *tert*-butyl group and 1-phenylethyl increase the dihedral angle, Hirshfeld surface volume, and void volume per molecule, while decreasing the packing index, density, and C/C contact ratio, leading to more photochromic tendencies. In contrast, the 2-pyridine, 3-methoxy, and 3-pyridine moieties drive the six structural parameters in the opposite direction, toward being less photochromic. Machine learning results revealed that the *tert*-butyl and isopropyl groups, as well as 1-phenylethyl, which are electrically neutral and bulky hydrocarbon groups, increase the possibility of photochromism due to the increased Hirshfeld surface volume and void volume per molecule, whereas pyridine, pyrazine, and the fluoro group create the opposite effect due to the formation of intermolecular interactions around nitrogen or fluorine. Furthermore, the photochromism of SA crystals was predicted with a high accuracy of >85% from molecular or crystal structures with our machine learning model. These findings are anticipated to facilitate a deeper understanding of the solid-state reactions of photochromic compounds and the design of novel photochromic solid materials with the desired properties. Extending the targeting materials to cocrystals and other photochromic compounds such as azobenzene and diarylethene and investigating the relationships between structural parameters and other photochromic characteristics such as the isomerization rate constants and the intensity of photochromism are future research themes to be addressed.

EXPERIMENTAL SECTION

Preparation of the Data Set. The data set of SA crystals was prepared by manually collecting information from 46 articles found using Google Scholar and the Cambridge Crystallographic Data Centre (CCDC, last access: 1-23-2023). Values for temperature, space group, unit cell volume, and density were referenced from cif files deposited with the CCDC. Values for the dihedral angle, C−OH distance, and number of strong/short contacts were obtained from articles or calculated using Mercury v3.10.3. Values for the C/C contact ratio, Hirshfeld surface, void volume per unit cell/molecule, and packing index were obtained from articles or calculated using CrystalExplorer v17.5.[62] Chemical structures were recorded using a simplified molecular-input line-entry system (SMILES). Detailed information on the data set is provided in Figures S1 and S2 and Table S1.

Data Analyses. Data analyses and visualization were conducted using Orange v3.30.1,[55] an open-source visual programming software package for data visualization, machine learning, data mining, and data analysis, unless otherwise described. The detailed analysis procedure is described in Figures S3, S7, and S8 and Tables S2, S4, and S5.

Machine Learning. Molecular structures represented as SMILES were converted from the Morgan fingerprint (2048 bits) using the *rdkit* package in Python 3.9.13.[63] A gradient-boosted, decision-tree machine learning model implemented in XGBoost 1.7.2 was employed for the prediction. To optimize the hyperparameters and evaluate the prediction accuracy of the model, double cross-validation (5-fold) was performed. Detailed information is provided in Figures S9−S11 and Table S16. For features that were accurately predicted in all five testing data sets, all data were used for training to rebuild the model. The SHAP model[61] was used to compute values used for visualizing influential substructures on the features. Detailed information is provided in Figure S12.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c07859.

https://doi.org/10.1021/acsomega.3c07859
ACS Omega 2024, 9, 1463−1471

Additional experimental details on the data set, data analyses, and machine learning (PDF)

Data set and codes for machine learning (ZIP)

AUTHOR INFORMATION

Corresponding Author

Hideko Koshima − *Research Organization for Nano & Life Innovation, Waseda University, Tokyo 162-0041, Japan;* orcid.org/0000-0002-3489-6357; Email: h.koshima@kurenai.waseda.jp

Authors

Shodai Hasebe − *Department of Advanced Science and Engineering, Graduate School of Advanced Science and Engineering, Waseda University, Tokyo 169-8555, Japan;* orcid.org/0000-0002-2537-5453

Kan Hatakeyama-Sato − *School of Materials and Chemical Technology, Tokyo Institute of Technology, Tokyo 152-8552, Japan;* orcid.org/0000-0003-1959-5430

Kenichi Oyaizu − *Department of Advanced Science and Engineering, Graduate School of Advanced Science and Engineering, Waseda University, Tokyo 169-8555, Japan;* orcid.org/0000-0002-8425-1063

Toru Asahi − *Department of Advanced Science and Engineering, Graduate School of Advanced Science and Engineering, Waseda University, Tokyo 169-8555, Japan; Research Organization for Nano & Life Innovation, Waseda University, Tokyo 162-0041, Japan*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c07859

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by JSPS Grant-in-Aid for Scientific Research B (17H03107) for H.K., JSPS Research Fellowship for Young Scientists (22J22384) for S.H, and the Early Bird Program of Waseda University (BA080Z003500) for S.H. S.H. thanks the Graduate Program for Power Energy Professionals, Waseda University, from the WISE Program of MEXT.

REFERENCES

(1) Qin, M.; Huang, Y.; Li, F.; Song, Y. photochromic sensors: a versatile approach for recognition and discrimination. *J. Mater. Chem. C* **2015**, *3*, 9265−9275.

(2) Irie, M. *Diarylethene Molecular Photoswitches*; Wiley-VCH: Weinheim, Germany, 2021.

(3) Koshima, H.; Hasebe, S.; Hagiwara, Y.; Asahi, T. Mechanically Responsive Organic Crystals by Light. *Isr. J. Chem.* **2021**, *61*, 683−696.

(4) Cohen, M. D.; Schmidt, G. M. J. Photochromy and thermochromy of anils. *J. Phys. Chem.* **1962**, *66*, 2442−2445.

(5) Hadjoudis, E. Chapter 17: Tautomerism by hydrogen transfer in anils. In *Photochromism: Molecules and Systems*; Dürr, H.; Bouas-Laurent, Eds.; Elsevier: Amsterdam, 1990; pp 685−712.

(6) Hadjoudis, E.; Marvidis, I. M. Photochromism and thermochromism of Schiff bases in the solid state: structural aspects. *Chem. Soc. Rev.* **2004**, *33*, 579−588.

(7) Amimoto, K.; Kawato, T. photochromism of organic compounds in the crystal state. *J. Photochem. Photobiol. C* **2005**, *6*, 207−226.

(8) Sliwa, M.; Mouton, N.; Ruckebusch, C.; Aloïse, S.; Poizat, O.; Buntinx, G.; Métivier, R.; Nakatani, K.; Masuhara, H.; Asahi, T. Comparative Investigation of Ultrafast Photoinduced Processes in Salicylidene-Aminopyridine in Solution and Solid State. *J. Phys. Chem. C* **2009**, *113*, 11959−11968.

(9) Spörkel, L.; Cui, G.; Thiel, W. Photodynamics of Schiff Base Salicylideneaniline: Trajectory Surface-Hopping Simulations. *J. Phys. Chem. A* **2013**, *117*, 4574−4583.

(10) Hasebe, S.; Hagiwara, Y.; Komiya, J.; Ryu, M.; Fujisawa, H.; Morikawa, J.; Katayama, T.; Yamanaka, D.; Furube, A.; Sato, H.; et al. Photothermally Driven High-Speed Crystal Actuation and Its Simulation. *J. Am. Chem. Soc.* **2021**, *143*, 8866−8877.

(11) Hasebe, S.; Hagiwara, Y.; Takechi, K.; Katayama, T.; Furube, A.; Asahi, T.; Koshima, H. Polymorph-Derived Diversification of Crystal Actuation by Photoisomerization and the Photothermal Effect. *Chem. Mater.* **2022**, *34*, 1315−1324.

(12) Harada, J.; Uekusa, H.; Ohashi, Y. X-ray Analysis of Structural Changes in photochromic Salicylideneaniline Crystals. Solid-State Reaction Induced by Two-Photon Excitation. *J. Am. Chem. Soc.* **1999**, *121*, 5809−5810.

(13) Johmoto, K.; Ishida, T.; Sekine, A.; Uekusa, H.; Ohashi, Y. Relation between photochromic properties and molecular structures in salicylideneaniline crystals. *Acta Crystallogr., Sect. B: Struct. Sci.* **2012**, *68*, 297−304.

(14) Carletta, A.; Buol, X.; Leyssens, T.; Champagne, B.; Wouters, J. Polymorphic and Isomorphic cocrystals of a *N*-Salicylidene-3-aminopyridine with Dicarboxylic Acids: Tuning of Solid-State Photo- and Thermochromism. *J. Phys. Chem. C* **2016**, *120*, 10001−10008.

(15) Sugiyama, H.; Uekusa, H. Relationship between crystal structures and photochromic properties of *N*-salicylideneaminopyridine derivatives. *CrystEngComm* **2018**, *20*, 2144−2151.

(16) Mercier, G. M.; Robeyns, K.; Tumanov, N.; Champagne, B.; Wouters, J.; Leyssens, T. New Insights into photochromic Properties of *N*-Salicylideneaniline Derivatives Using a Cocrystal Engineering Approach. *Cryst. Growth Des.* **2019**, *19*, 5544−5556.

(17) Arcovito, G.; Bonamico, M.; Domenicano, A.; Vaciago, A. Crystal and Molecular Structure of Salicylaldehyde Azine. *J. Chem. Soc. B* **1969**, 733−741.

(18) Destro, R.; Gavezzotti, A.; Simonetta, M. Salicylideneaniline. *Acta Crystallogr., Sect. B: Struct. Sci.* **1978**, *34*, 2867−2869.

(19) Moustakali-Mavridis, I.; Hadjoudis, E.; Mavridis, A. Crystal and Molecular Structure of Some Thermochromic Schiff Bases. *Acta Crystallogr., Sect. B: Struct. Sci.* **1978**, *34*, 3709−3715.

(20) Moustakali-Mavridis, I.; Hadjoudis, E.; Mavridis, A. Structure of Thermochromic Schiff Bases. II. Structures of *N*-Salicylidene-3-aminopyridine and *N*-(5-Methoxysalicylidene)-3- aminopyridine. *Acta Crystallogr., Sect. B: Struct. Sci.* **1980**, *36*, 1126−1130.

(21) Hadjoudis, E.; Vittorakis, M.; Moustakali-Mavridis, I. photochromism and Thermochromism of Schiff Bases in the Solid State and In Rigid Glasses. *Tetrahedron* **1987**, *43*, 1345−1360.

(22) Moustakali-Mavridis, I.; Terzis, A.; Hadjoudis, E. Structure of *N*-(4-Methoxysalicylidene)-2-thienylmethylamine. *Acta Crystallogr., Sect. C: Struct. chem.* **1987**, *43*, 1389−1391.

(23) Hadjoudis, E. photochromic and Thermochromic Anils. *Mol. Eng.* **1995**, *5*, 301−337.

(24) Burgess, J.; Fawcett, J.; Russell, D. R.; Gilani, S. R.; Palma, V. Four *N*-(2-hydroxybenzylidene)aniline derivatives. *Acta Crystallogr., Sect. C: Struct. chem.* **1999**, *55*, 1707−1710.

(25) Fukuda, H.; Amimoto, K.; Koyama, H.; Kawato, T. Crystalline photochromism of *N*-salicylidene-2,6-dialkylanilines: advantage of 2,6-dialkyl substituents of aniline for preparation of photochromic Schiff base crystals. *Org. Biomol. Chem.* **2003**, *1*, 1578−1583.

https://doi.org/10.1021/acsomega.3c07859
ACS Omega 2024, 9, 1463−1471

(26) Tooke, D. M.; Song, Y.; van Albada, G. A.; Reedijk, J.; Spek, A. L. 1,4-Bis(3,5-di-*tert*-butyl-2-hydroxybenzylidene-aminomethyl)-benzene. *Acta Crystallogr.* **2004**, *E60*, o718−o719.
(27) Arod, F.; Gardon, M.; Pattison, P.; Chapuis, G. The α_2-polymorph of salicylideneaniline[1]. *Acta Crystallogr., Sect C. Struct. Chem.* **2005**, *61*, o317−o320.
(28) Sliwa, M.; Létard, S.; Malfant, I.; Nierlich, M.; Lacroix, P. G.; Asahi, T.; Masuhara, H.; Yu, P.; Nakatani, K. Design, Synthesis, Structural and Nonlinear Optical Properties of photochromic Crystals: Toward Reversible Molecular Switches. *Chem. Mater.* **2005**, *17*, 4727−4735.
(29) Chatziefthimiou, S. D.; Lazarou, Y. G.; Hadjoudis, E.; Dziembowska, T.; Mavridis, I. M. Keto Forms of Salicylaldehyde Schiff Bases: Structural and Theoretical Aspects. *J. Phys. Chem. B* **2006**, *110*, 23701−23709.
(30) Arod, F.; Pattison, P.; Schenk, K. J.; Chapuis, G. Polymorphism in *N*-Salicylideneaniline Reconsidered. *Cryst. Growth Des.* **2007**, *7*, 1679−1685.
(31) Fukuda, H.; Amimoto, K.; Koyama, H.; Kawato, T. A different photo-sensitivity of isostructural crystals of *N*-(3,5-dihalosalicylidene)-2,6-dimethylaniline analogues: search for the definite reaction room in the crystal to exhibit photochromism. *Tetrahedron Lett.* **2009**, *50*, 5376−5378.
(32) Johmoto, K.; Sekine, A.; Uekusa, H.; Ohashi, Y. Elongated Lifetime of Unstable Colored Species by Intermolecular Hydrogen Bond Formation in photochromic Crystals. *Bull. Chem. Soc. Jpn.* **2009**, *82*, 50−57.
(33) Robert, F.; Naik, A. D.; Tinant, B.; Robiette, R.; Garcia, Y. Insights into the Origin of Solid-State photochromism and Thermochromism of *N*-Salicylideneanils: The Intriguing Case of Aminopyridines. *Chem.—Eur. J.* **2009**, *15*, 4327−4342.
(34) Wang, Y.; Yu, X.; Sun, Y.; Wang, Y.; Lu, L. Synthesis, vibrational spectral and nonlinear optical studies of N-(4-hydroxyphenyl)-2-hydroxybenzaldehyde-imine: A combined experimental and theoretical investigation. *Spectrochim. Acta, Part A* **2011**, *79*, 1475−1482.
(35) Johmoto, K.; Sekine, A.; Uekusa, H. photochromism Control of Salicylideneaniline Derivatives by Acid—Base Co-Crystallization. *Cryst. Growth Des.* **2012**, *12*, 4779−4786.
(36) Avadanei, M.; Cozan, V.; Shova, S.; Paixão, J. A. Solid state photochromism and thermochromism of two related N-salicylidene anilines. *Chem. Phys.* **2014**, *444*, 43−51.
(37) Jacquemin, P.; Robeyns, K.; Devillers, M.; Garcia, Y. Reversible photochromism of an *N*-salicylideneaniline anion. *Chem. Commun.* **2014**, *50*, 649−651.
(38) Hutchins, K. M.; Dutta, S.; Loren, B. P.; MacGillivray, L. R. Co-Crystals of a Salicylideneaniline: photochromism Involving Planar Dihedral Angles. *Chem. Mater.* **2014**, *26*, 3042−3044.
(39) Safin, D. A.; Bolte, M.; Garcia, Y. Solid-state photochromism and thermochromism of *N*-salicylidene pyrene derivatives. *CrystEngComm* **2014**, *16*, 8786−8793.
(40) Jacquemin, P.; Robeyns, K.; Devillers, M.; Garcia, Y. Photochromism Emergence in *N*-Salicylidene *p*-Aminobenzenesulfonate Diallylammonium Salts. *Chem.—Eur. J.* **2015**, *21*, 6832−6845.
(41) Carletta, A.; Dubois, J.; Tilborga, A.; Wouters, J. Solid-state investigation on a new dimorphic substituted *N*-salicylidene compound: insights into its thermochromic behaviour. *CrystEngComm* **2015**, *17*, 3509−3518.
(42) Zbačnik, M.; Nogalo, I.; Cinčića, D.; Kaitner, B. Polymorphism control in the mechanochemical and solution-based synthesis of a thermochromic Schiff base. *CrystEngComm* **2015**, *17*, 7870−7877.
(43) Safin, D. A.; Babashkina, M. G.; Robeyns, K.; Garcia, Y. C−H—Br−C *vs.* C−Br—Br−C *vs.* C−Br—N bonding in molecular self-assembly of pyridine-containing dyes. *RSC Adv.* **2016**, *6*, 53669−53678.
(44) Sugiyama, H. Molecular Planarity and Crystal Structures of *N*-Salicylideaminopyrazine Derivatives. *X-ray Structure Analysis Online* **2018**, *34*, 57−58.
(45) Taniguchi, T.; Sato, H.; Hagiwara, Y.; Asahi, T.; Koshima, H. Photo-triggered phase transition of a crystal. *Commun. Chem.* **2019**, *2*, 19.
(46) Houjou, H.; Kato, T.; Huang, H.; Suzuki, Y.; Yoshikawa, I.; Mutai, T. Re-evaluation of the *tert*-Butyl Method in Crystal Engineering of Salicylideneanilines by Simultaneous Observation of photochromism and Thermochromism in Single Crystals. *Cryst. Growth Des.* **2019**, *19*, 1384−1390.
(47) Hagiwara, Y.; Taniguchi, T.; Asahi, T.; Koshima, H. Crystal actuator based on a thermal phase transition and photothermal effect. *J. Mater. Chem. C* **2020**, *8*, 4876−4884.
(48) Sugiyama, H. Crystal structures and thermochromic and fluorescence properties of *N*-salicylideneaniline derivatives having a naphthyl-isoquinoline group. *J. Mol. Str.* **2021**, *1229*, No. 129603.
(49) Ishizaki, K.; Sugimoto, R.; Hagiwara, Y.; Koshima, H.; Taniguchi, T.; Asahi, T. Actuation performance of a photo-bending crystal modeled by machine learning-based regression. *CrystEngComm* **2021**, *23*, 5839−5847.
(50) Tang, Y.; Liu, J.; Zeng, Y.; Peng, H.; Huang, X.; Yang, M.; Xiong, R. Optical Control of Polarization Switching in a Single-Component Organic Ferroelectric Crystal. *J. Am. Chem. Soc.* **2021**, *143*, 13816−13823.
(51) Liao, W.; Zeng, Y.; Tang, Y.; Peng, H.; Liu, J.; Xiong, R. Multichannel Control of Multiferroicity in Single-Component Homochiral Organic Crystals. *J. Am. Chem. Soc.* **2021**, *143*, 21685−21693.
(52) Wang, Z.; Huang, C.; Liu, J.; Zeng, Y.; Xiong, R. Salicylideneaniline is a Photoswitchable Ferroelectric Crystal. *Chem.—Eur. J.* **2021**, *27*, 14831−14835.
(53) Liao, W.; Deng, B.; Wang, Z.; Cheng, T.; Hu, Y.; Cheng, S.; Xiong, R. Optically Induced Ferroelectric Polarization Switching in a Molecular Ferroelectric with Reversible Photoisomerization. *Adv. Sci.* **2021**, *8*, No. 2102614.
(54) Hasebe, S.; Hagiwara, Y.; Hirata, K.; Asahi, T.; Koshima, H. Crystal actuation switching by crystal thickness and light wavelength. *Mater. Adv.* **2022**, *3*, 7098−7106.
(55) Demšar, J.; Curk, T.; Erjavec, A.; Gorup, Č.; Hočevar, T.; Milutinovič, M.; Možina, M.; Polajnar, M.; Toplak, M.; Starič, A.; et al. Orange: Data Mining Toolbox in Python. *J. Mach. Learn. Res.* **2013**, *14*, 2349−2353.
(56) Kawato, T.; Koyama, H.; Kanamori, H.; Tagawa, H.; Iga, K. Photoisomerization and thermoisomerization III. *tert*-Butyl method of selective preparation of photochromic crystalline salicylideneanilines and elucidation of substituent effects on the kinetics of the colour-change process. *J. Photochem. Photobiol., A* **1994**, *78*, 71−77.
(57) Chen, T.; Guestrin, C. *XGBoost: A Scalable Tree Boosting System*, 2016.
(58) Zou, M.; Jiang, W.-G.; Qin, Q.-H.; Liu, Y.-C.; Li, M.-L. Optimized XGBoost Model with Small Dataset for Predicting Relative Density of Ti-6Al-4V Parts Manufactured by Selective Laser Melting. *Materials* **2022**, *15*, 5298.
(59) Chaikittisilp, W.; Yamauchi, Y.; Ariga, K. Material Evolution with Nanotechnology, Nanoarchitectonics, and Materials Informatics: What will be the Next Paradigm Shift in Nanoporous Materials? *Adv. Mater.* **2022**, *34*, No. 2107212.
(60) Li, C.; Zheng, K. Methods, progresses, and opportunities of materials informatics. *InfoMat* **2023**, *5*, No. e12425.
(61) Lundberg, S. M.; Lee, S. A Unified Approach to Interpret Model Predictions. In *Advances in Neural Information Processing Systems*; NIPS, 2017; Vol. *30*.
(62) Spackman, P. R.; Turner, M. J.; McKinnon, J. J.; Wolff, S. K.; Grimwood, D. J.; Jayatilaka, D.; Spackman, M. A. *CrystalExplorer*: a program for Hirshfeld surface analysis, visualization and quantitative analysis of molecular crystals. *J. Appl. Crystallogr.* **2021**, *54*, 1006−1011.
(63) *rdkit: Open-Source Cheminformatics*, http://www.rdkit.org.

https://doi.org/10.1021/acsomega.3c07859
ACS Omega 2024, 9, 1463−1471