Digital Discovery

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: R. Shibata, N. Shioda, H. Imai, Y. Igarashi and Y. Oaki, *Digital Discovery*, 2025, DOI: 10.1039/D5DD00442J.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/digitaldiscovery

View Article Online
DOI: 10.1039/D5DD00442J

ARTICLE

A data-driven approach to control stimulus responsivity of functional polymer materials: Predicting thermoresponsive color-change properties of polydiacetylene

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Risako Shibata,[a] Nano Shioda,[a] Hiroaki Imai,[a] Yasuhiko Igarashi,[b] Yuya Oaki*,[a]

Sensing devices are fabricated using stimuli-responsive materials. In general, the responsivity is controlled by designing molecules and materials based on professional experience. If predictors are constructed for the responsivity control, the number of experiments can be reduced without consumption of time, cost, and effort. However, such dynamic properties of functional polymer materials are not easily predicted because of the small data and complex structure-function relationship. How to prepare dataset and train small data remain significant challenges. The present work shows construction and application of a prediction model for controlling thermoresponsive color-changing properties of layered polydiacetylenes (PDAs). The responsivity was changed by the intercalated guest molecules. The training dataset was prepared from a series of the photographs representing the color at each temperature. The prediction model of the thermoresponsivity, namely color-changing temperature, was constructed by combining machine learning and our chemical insight on the small experimental data. The thermoresponsivity of the newly synthesized layered PDAs was predicted by the model. The modeling methods can be applied to predicting various dynamic properties of functional polymer materials.

1. Introduction

Stimuli-responsive molecules and materials have various applications, such as sensors and actuators.[1–11] An input external stimulus, such as thermal, chemical, optical, and mechanical stimuli, is converted to a specific output responsiveness by molecules and materials. Visible and/or fluorescent colors are used for detection of the applied stresses. The responsivity, such as sensitivity, chromaticity, and reversibility, is tuned by design and synthesis of chromophores and their organization states. In general, such molecules and materials are designed by professional experiences and synthesized with trial and error. If data-driven approach is effectively applied to these processes, the responsivity can be efficiently tuned without consumption of time, cost, and effort. Although this motivation is rational, data-driven approaches are not easily applied to designing such functional soft materials because of the following reasons. One of the problems is insufficient data size of typical experimental works for the use of conventional machine-learning (ML) algorithms. Another problem is how to prepare the dataset suitable for ML based on the non-numerical experimental results. The targeted problem, *i.e.* stimuli responsivity, should be translated to a ML-solvable problem. For example, photographs, spectra, and their changes are needed to be converted to ML-applicable data. Moreover, as dynamic properties are related to the complex structural hierarchy from molecules to organized states, the factors related to the properties as the features are not easily prepared only using conventional tools for automatic generation of descriptors. In this manner, ML has not been fully applied to experimental studies for controlling the properties of stimuli-responsive materials because of the issues in data size, data curation for the targets, and preparation of the descriptors. Although ML has been applied to design polymers and soft materials, such as polymers, gels, liquid crystals, and bubbles, in recent years,[12–22] the further methodological advances are required to develop the methods for small experimental data. The present work shows construction of a prediction model for thermoresponsive color-changing properties of layered PDA (Fig. 1). Based on the photographic data, a straightforward linear regression model was prepared by sparse modeling for small data (SpM-S). The data acquisition, curation, and modeling methods can be applied to construct the predictors of the other stimuli-responsive functional materials with small data.

[a] *Department of Applied Chemistry, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan.*
[b] *Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8573, Japan.*

Supplementary Information available: [Experimental methods, List of guest molecules, Datasets, XRD patterns and FT-IR spectra, Thermoresponsive color-changing properties]. See DOI: 10.1039/x0xx00000x

View Article Online
DOI: 10.1039/D5DD00442J

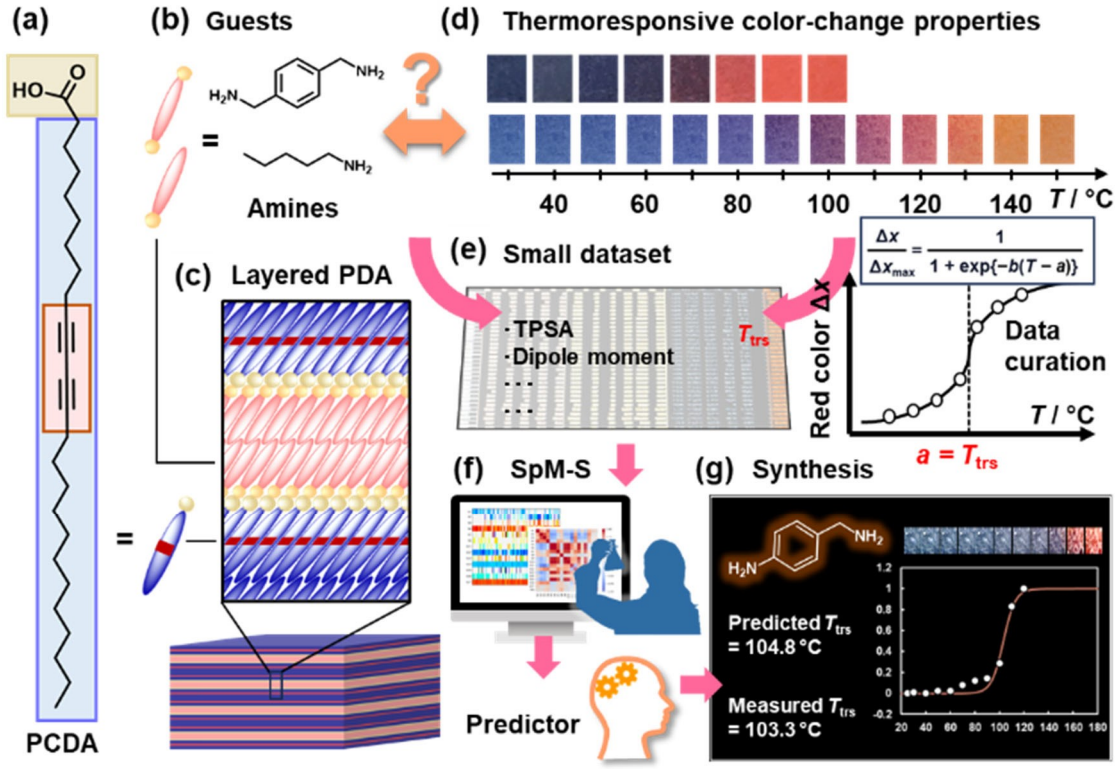

Fig. 1. Schematic illustration for predicting thermoresponsive color-changing properties of layered PDA. (a) PCDA monomer. (b) Guest amines. (c) Guest-intercalated layered PDA with the topochemical polymerization. (d) Photographs for preparing thermoresponsive color-changing properties. (e) Small dataset containing the color transition temperature ($T_{trs} = y$, objective variables) and physicochemical parameter of the guests (x_{Gn}, explanatory variables). (f) SpM-S for extraction of the descriptors and model construction. (g) Synthesis of layered PDA with intercalating new guests based on the predicted T_{trs}.

PDA, a conjugated polymer, exhibits color changes in response to external stimuli, such as thermal, chemical, and frictional stresses.[23–30] A wide variety of sensing devices and systems were fabricated using PDA. The stresses induce the molecular motion and subsequent shortening the effective conjugation length of PDA main chain. The stimuli responsivity has been controlled by the molecular design.[31–43] Our group has studied intercalation approach to control the responsivity of the layered PDAs (Fig. 1a–d).[45–52] The layered structure of an amphiphilic diacetylene (DA) monomer, 10,12-pentacosadiynoic acid (PCDA), is topochemically polymerized in the solid crystalline state (Fig. 1a,c). Whereas the layered PDA derived from PCDA shows the blue-to-red color transition around 65 °C with heating (Fig. 1d), the color-changing temperature varied in the range of −0.2 to 146 °C depending on the intercalated guests, such as metal ions and alkyl amines, in the interlayer space (Fig. 1b,d).[44–52] However, the responsivity control based on the experience and intuition meets the limitations for both the molecular-design and intercalation approaches. Here we used ML to construct the predictor for controlling the responsivity.

In recent years, ML has been widely used in general materials science.[53–60] Most of the algorithms are suitable for big data. Sufficient size of data is not prepared for all the experimental works. In small data, conventional ML algorithms cause problems, such as overtraining. Recent studies have proposed the specific methods for small data, such as transfer and active learnings.[61–67] In our group, sparse modeling for small data (SpM-S) has been studied; the approach combining ML and our chemical insight provides straightforward and interpretable predictors.[68–73] SpM-S was used for the simple numerical data, such as yield, size, and capacity, directly obtained from the experimental works. The approach was not applied to small experimental data requiring the curation prior to the use, such as photograph and graph. If the targeted data and problem are converted to SpM-S applicable formats in an appropriate manner, the approach can be applied to small data in the broader fields. In the present work, a prediction model for the thermoresponsive color-changing properties of the layered PDA was constructed by SpM-S (Fig. 1d–g). After the training dataset was prepared from a series of the photographic data (Fig. 1d,e), combination of ML and chemical insight provides the predictors based on the small data (Fig. 1f). The model successfully predicted the thermoresponsivity of the layered PDAs with the intercalation of new unknown guests (Fig. 1g). Based on these results, if the targeted stimuli responsivity

is converted to the appropriate dataset, the predictors can be constructed by SpM-S even on small experimental data.

2. Results and discussion

2. 1. Preparation of Datasets from Photographs

The data about the thermoresponsive color-changing properties were extracted from our previous works and newly added in the present work (Fig. 1b–d and Figs. S1–S3 and Tables S1–S3 in the Electronic Supplementary Information (ESI)).[44–51] The layered PDAs show the color changes from blue to red with increasing temperature (T / °C) (Fig. 1d). Whereas the layered PDA without the guest showed the blue-to-red color transition around 65 °C, for example, the color transition was observed around 80 and 100 °C with the intercalation of tetradecylamine (C_{14}-NH_2) and *p*-xylylenediamine (*p*-Xy), respectively (Fig. 2a,b).

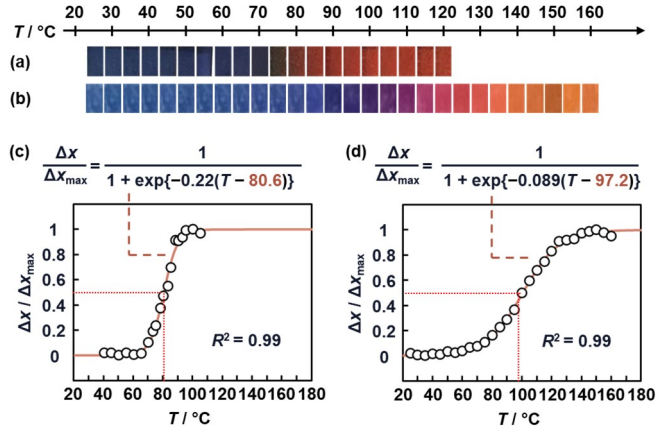

Fig. 2. Data curation of the thermoresponsive color-changing properties. (a,b) Photographs representing the relationship between T and color of layered PDA with the intercalation of C_{14}-NH_2 (a) and *p*-Xy (b). (c,d) Approximation of the relationship between T and Δx / Δx_{max} (each plot) generated from the photographic data using sigmoidal function (red curve) for PDA-C_{14}-NH_2 (c) and PDA-*p*-Xy. The same data for the other amines were summarized in Fig. S3 in the ESI.

The original experimental data is a series of the photographs exhibiting the color of the sample at different temperature. The color of the photographs was converted to the red-color intensity (x) based on the RGB values using an international standard (See Experimental Method in the ESI). An increment of x ($\Delta x = x - x_0$) was calculated at each T in reference to the initial state before heating (x_0). Then, Δx was normalized by dividing the maximum Δx_{max} (Δx / Δx_{max}). The relationship between T and Δx / Δx_{max} was prepared for each layered PDA containing 75 different guests (Fig. 2c,d and Fig. S3 in the ESI). The T–(Δx / Δx_{max}) curve was approximated to sigmoidal function Eq. (1) using two constants a and b with the coefficient of determination (R^2).

$$\frac{\Delta x}{\Delta x_{max}} = \frac{1}{1+\exp\{-b(T-a)\}} \quad \cdots \quad \text{Eq. (1)}$$

In the curve, the constant a corresponds to the color-change temperature reaching $0.5\Delta x$ / Δx_{max} (Fig. 1e). The constant b corresponds to the slope representing the temperature responsivity. The observed T–Δx plots were approximated using Eq. (1) with optimizing a and b to minimize R^2 value. The average R^2 values were 0.952 ± 0.066 for all the data (Table S4 in the ESI). The fitting function was prepared with the specific a and b for each sample (Fig. 2c,d and Fig. S3 in the ESI). The color-transition temperature (T_{trs}) was defined as T to reach $0.5\Delta x$ / Δx_{max}, namely a in Eq. (1). Sigmoidal function is a suitable approximation to describe the temperature-responsive chromaticity change. Sigmoidal function generally exhibits the following trend of the increase in y: gentle increases in the initial and final stages and steep increase in the middle range. The same trend was experimentally observed for the thermoresponsive color-changing behavior of PDAs (Fig. 2c,d). As the variation of the chromaticity change, Δx / Δx_{max}, is limited in the specific range, this behavior is also represented by sigmoidal function. In this manner, the thermoresponsive color-changing properties based on a series of photographs were converted to the numerical data.

T_{trs} as objective variable (y) was calculated for the layered PDA with the intercalated 75 different guests, such as alkyl amines, cationic polymers, and aromatic amines (Table S1 in the ESI). These guests are numbered as **S005**, **S006**, …, **S161** in our group. Table 1 summarizes the explanatory variables (x_{Gn}: n = 1–17) as the potential descriptors corresponding to the physicochemical parameters of the guest molecules, such as boiling point (x_{G6}) and dipole moment (x_{G16}).

Table 1. List of explanatory variables.

x_{Gn}	Parameter	Unit	[a]Method
x_{G1}	HSP dispersion	$MPa^{1/2}$	H
x_{G2}	HSP polarity	$MPa^{1/2}$	H
x_{G3}	HSP hydrogen bonding	$MPa^{1/2}$	H
x_{G4}	Density	g cm^{-3}	H
x_{G5}	Molecular volume	cm^3 mol^{-1}	H
x_{G6}	Molecular surface area	cm^2 mol^{-1}	H
x_{G7}	Boiling point (1.01×10^5 Pa)	°C	H
x_{G8}	Melting point (1.01×10^5 Pa)	°C	H
x_{G9}	Ovality	–	H
x_{G10}	Molecular weight	g mol^{-1}	R
x_{G11}	[b]Log P	–	R
x_{G12}	[c]TPSA	$(10^{-10}$ m$)^2$	R
x_{G13}	Molecular length	10^{-10} m	R
x_{G14}	Minimum of partial charge density	C m^{-3}	G
x_{G15}	Polarizability	$(10^{-10}$ m$)^3$	G
x_{G16}	Dipole moment	Debye	G
x_{G17}	Valence of amine	–	–

[a] Softwares: HSP-ip (H), RDKit (R), Gaussian (G) (See ESI). [b] Logarithm of partition coefficient to water and octanol. [c] Topological polar surface area.

View Article Online
DOI: 10.1039/D5DD00443J

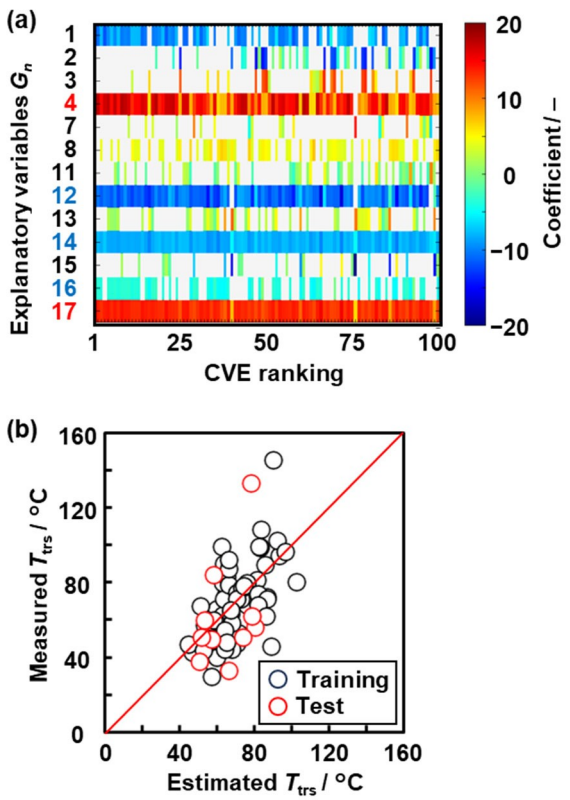

Fig. 3. Construction of the T_{trs} prediction model. (a) Weight diagram of ES-LiR. (b) Relationship between the estimated and measured T_{trs} for the training (black, 65 *y*) and test (red, 10 *y*) datasets using the model Eq. (2) comprising of five x_{Gn} (x_{G4}, x_{G12}, x_{G14}, x_{G16}, and x_{G17}).

2. 2. Construction of T_{trs} Predictor

The original 75 data was divided into the training and test datasets containing 65 *y* and 10 *y*, respectively (Tables S2 and S3 in the ESI). The variable selection was carried out using the training dataset (Fig. 3 and Table S2 in the ESI). As the correlation coefficients of x_{Gn} (n = 5, 6, 9, 10) were larger than 0.9 (Fig. S4 in the ESI), these four x_{Gn} were removed to avoid the multicollinearity. In SpM-S,[68] the significance of x_{Gn} was visualized in the weight diagram of exhaustive search with linear regression (ES-LiR) (Fig. 3a).[69,73] Linear regression models were prepared using the training dataset in all the possible combinations of 13 x_{Gn} (n = 1–4, 7, 8, 11–17), 2^{13}–1 (= 8.2×10^3) combinations, with five-fold cross validation (CV). The constructed models were sorted in the ascending order of the cross-validation error (CVE) values. The positive and negative coefficients of x_{Gn} were represented by the warm and cool colors in the diagram, respectively (Fig. 3a). The more frequently used x_{Gn} was displayed by the more densely colored bar in the vertical axis. Based on the weight diagram, we selected five x_{Gn} (x_{G4}, x_{G12}, x_{G14}, x_{G16}, and x_{G17}) as the descriptors to prepare the linear regression model Eq. (2).

$$y = 8.85x_{G4} - 6.66x_{G12} - 7.38x_{G14} - 5.76x_{G16} + 12.04x_{G17} + 68.67 \ldots \text{Eq. (2)}$$

As the coefficients of this linear regression are normalized in frequency distribution with mean 0 and standard deviation 1, the contribution of each x_{Gn} can be comparable. The relationship between the estimated and measured T_{trs} had root mean squared error (RMSE) 15.9 °C for the training dataset (65 *y*, black circles) and 25.3 °C for the test dataset (10 *y*, red circles) (Fig. 3b). The coefficient of determination (R^2) was 0.388 for the training dataset and 0.139 for the test dataset. Five-fold CV was carried out using the merged dataset of the training and test ones (75 *y*) (Fig. S5 in the ESI). In the five-fold CV, the regression equations had the same positive and negative coefficients as those in Eq. (2). RMSE was 15.6 ± 1.2 °C for training data and 17.3 ± 4.1 °C for test data. The RMSE values in the five-fold CV imply that the model Eq. (2) is not overtrained in the training dataset but generalizable. The RMSE and R^2 values are not so high to predict T_{trs} precisely. On the other hand, the relationship between the estimated and predicted values in Fig. 3b indicates that the overall trends of the higher and lower T_{trs} are roughly described by the model. In general, a model with high prediction accuracy (e.g. RMSE and R^2 values) is not easily constructed based on small data because of the lack of data size. In addition to these quantitative metrics, whether the overall trend is described by the model or not is qualitatively evaluated by the plots representing the relationship between the estimated and predicted values.[68,69] If the model describing the overall trend is constructed based on small data, the next experiments can be accelerated with reducing the number of trials. In this manner, a linear regression model for predicting T_{trs} of the layered PDA was constructed by SpM-S.

The contribution of each descriptor to T_{trs} was studied based on the positive and negative of the coefficients with our chemical insight. Density (x_{G4}) and valence (x_{G17}) of the guests have the positive correlation to T_{trs}. If the guest molecule with the higher density (larger x_{G4}) is intercalated in the interlayer space, the more densely packed guests can provide the rigid layered structure. The divalent guest amines also form the rigid layered structure with anchoring the layers by both the terminals. The minimum value of the partial charge density (x_{G14}) shows the negative correlation. The guests with heteroatoms in the alkyl chain or aromatic ring, such as secondary amines and pyridine rings, have the larger x_{G14} values compared with the other primary amines (Tables S2 and S3 in the ESI). These guest molecules direct formation of the soft layered structures with the weakened interlayer interaction leading to a decrease in T_{trs} because of the loose packing. Topological polar surface area (TPSA, x_{G4}) and dipole moment (x_{G16}) have the negative correlation to T_{trs}. As the guest molecules with the smaller x_{G4} and x_{G16} are organized and packed in the interlayer space with van der Waals interaction, an increase in T_{trs} is derived by the more rigid layered structure. In this manner, the contribution of each x_{Gn} can be interpretable by our experience and chemical insight. In the present study, the gest monoamines and diamines with the alkyl chain, cycloalkane, aromatic and heteroaromatic rings, and alcohol were mainly used for the training (Table S1 in the ESI). The other amines, such as those containing branched alkyl chains and ethers, were not fully trained in the present model. In addition,

View Article Online
DOI: 10.1039/D5DD00443A

the more different types, such as amino acids, macromolecules, and polyamines, need to train with adding the more data. The more accurate model can be constructed with adding the data.

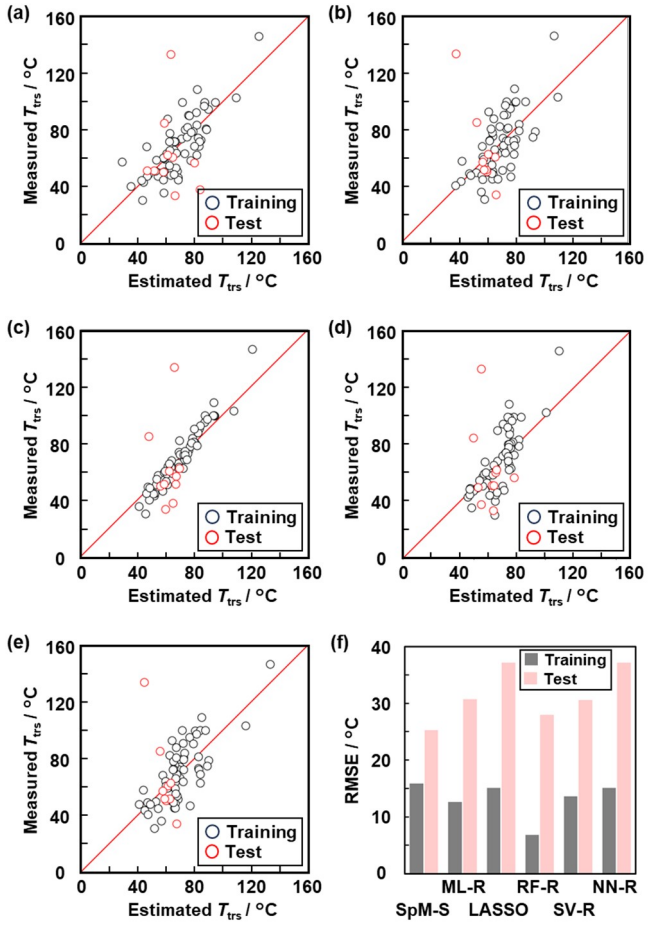

Fig. 4. Model construction using the other ML algorithms. (a–e) Relationship between the estimated and measured T_{trs} for the training dataset (black, 65 *y*) and test dataset (red, 10 *y*) for ML-R (a), LASSO (b), RF-R (c), SV-R (d), and NN-R (e). (f) RMSE values of the constructed models for the training (gray) and test (pink) datasets.

2. 3. Model construction using other ML algorithms

The model construction was carried out based on the same training and test datasets using the following other linear and nonlinear ML algorithms (Fig. 4): multiple linear regression (MLR) without variable selection, least absolute shrinkage and selection operate (LASSO), random forest regression (RF-R), support vector regression (SV-R), and neural network regression (NN-R). Whereas the relationship between the estimated and observed T_{trs} were approached to the diagonal line for the training dataset, the large error was found for the test dataset (Fig. 4a–e). Fig. 4f summarizes the RMSE values to the original training and test datasets. The other algorithms provide the slightly smaller RMSE values to the training dataset compared with that of the model constructed by SpM-S. In contrast, the larger RMSE values were obtained to the test dataset. In the case of LASSO, the larger number of the descriptors x_n (2, 5–8, 10, 12, 13, 15) was used for the model. These results imply that the other algorithms lead to the overfitting to the training dataset and lowering the generalizability to the test dataset. The similar results were obtained in our previous works using small data.[69,70,74,75] SpM-S provides the accurate model comparable to the other models even though the number of the selected descriptors is limited. The linear regression model comprised of the selected descriptor and coefficient has interpretability and straightforwardness. The smallest RMSE value to the test dataset means the generalizability to unknown test data. In this manner, SpM-S is a suitable approach to small data in terms of the accuracy, interpretability, and generalizability.

2. 4. Prediction-Based Syntheses of New Layered PDA

Nine commercial amines (**S901**–**S909**) were selected as the new guests (Fig. 5a and Table S5 in the ESI). These molecules were not used in the previous works and datasets for the model construction. Prior to the experiments, the predicted T_{trs} was calculated using the model Eq. (2). These guest-intercalated precursor layered crystals were synthesized by self-organization from the solution containing the host PCDA and guest amine with the evaporation of solvent.[45] The layered PDA with the intercalated guests was obtained by the UV-light irradiation. The intercalated structures were analyzed by X-ray diffraction (XRD) and Fourier-transform infrared (FT-IR) spectroscopy (Fig. S6 in the ESI). The thermoresponsive color-changing properties were observed by heating the samples (Fig. 5b). The actual T_{trs} was calculated from the T–(Δx / Δx_{max}) relationship by approximation using Eq. (1) (Fig. S7 in the ESI). The predicted and actual T_{trs} values were summarized using the red circles in Fig. 5c. The RMSE value of the actual T_{trs} was 15.6 °C comparable to that of the training data (15.9 °C, black circles in Fig. 5c). The results indicate that T_{trs} of the guest-intercalated layered PDA can be predicted using the model. Although the guests have been selected based on our experience and intuition, the more efficient selection can be achieved using the predictor in the future. Moreover, the similar approach can be applied to the molecular design of DA monomers to achieve tailored responsivity.

View Article Online
DOI: 10.1039/D5DD00442J

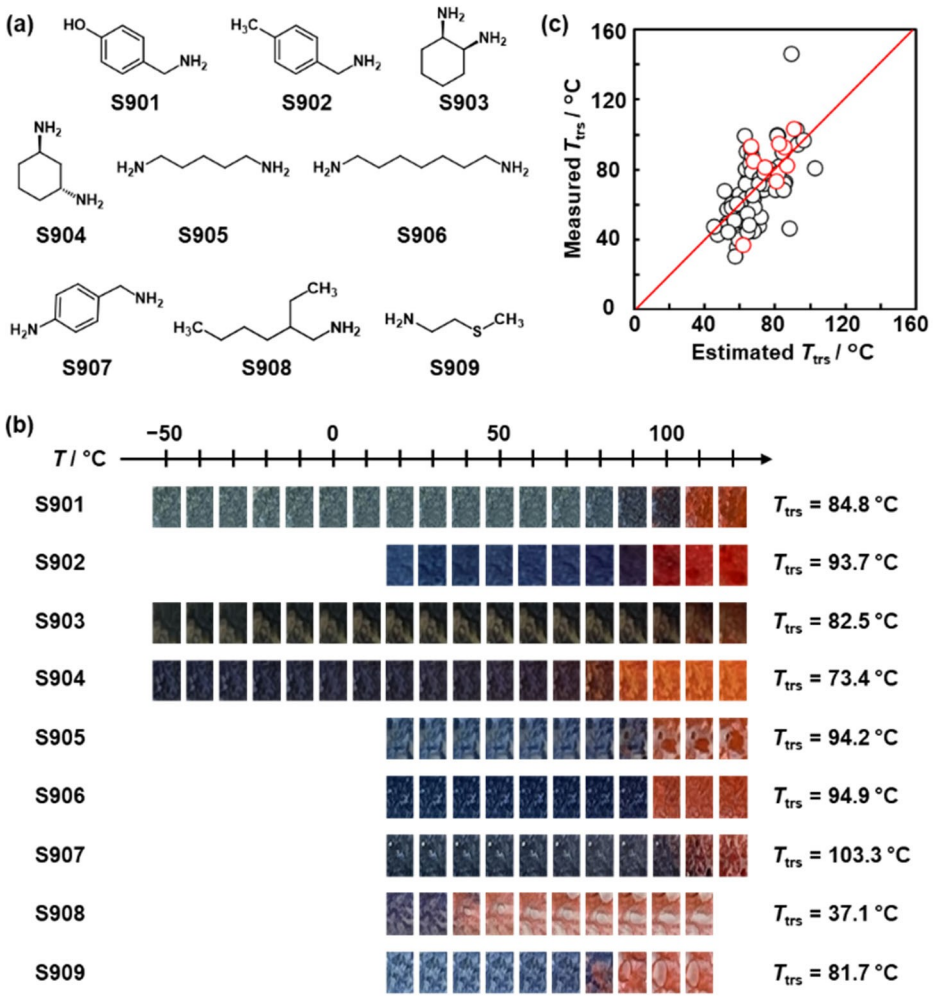

Fig. 5. Prediction-based synthesis of the layered PDA with the intercalation of new guests. (a) Molecular structure of new guests **S901**–**S909**. (b) Photographs representing the relationship between *T* and color of the layered PDA with the intercalation of the new guests **S901**–**S909**. The measured T_{trs} was noted with the photographs. (c) Relationship between the estimated and measured T_{trs} for the training dataset (black, 65 *y*) and newly synthesized ones (red, 9 *y*).

3. Conclusions

A dynamic function of polymer material has been predicted based on small experimental data with assistance of ML, SpM-S. Layered PDAs show the thermoresponsive color-change properties depending on the intercalated guest molecules. As the color-changing properties were represented by a series of photographs in our experimental studies, the ML-applicable data was prepared at the initial data-curation step. The relationship between *T* and Δx / Δx_{max} representing the thermoresponsivity was approximated by sigmoidal function to conversion of the photographic data to the numerical data for ML. A constant in the fitting function was set as *y*. The descriptors were extracted from the weight diagram representing the contribution of each x_n. The straightforward and interpretable linear model was constructed for predicting T_{trs} by SpM-S. Furthermore, T_{trs} of the layered PDA with intercalating the new guest was predicted prior to the experiment. The actual T_{trs} was consistent with the predicted one. These results indicate the successful construction of the T_{trs} predictor for the layered PDA. In general, dynamic functions of polymer materials are not easily predicted because of the complex structural hierarchy. The present approach combining data curation and SpM-S can be applied to the other stimuli-responsive materials with input triggers (e.g. temperature, concentration, pH, and light) and output signals (e.g. color change, phase and structural transitions, morphology change, current). Moreover, designing the other functional polymer materials can be achieved using the present method based on small data.

Conflicts of interest

There are no conflicts to declare.

View Article Online
DOI: 10.1039/D5DD00442J

Data availability

The data supporting this article have been included as part of the ESI. The following datasets are available at GitHub of our group (https://github.com/Oaki-Group/202511_PDA) (DOI: https://doi.org/10.5281/zenodo.18060256): Tables S1 and S5 as csv file with simplified molecular input line entry system (SMILES) and mol2 files; Table S2 as csv file for the training dataset; Table S3 as csv file for the test dataset. The following codes are available at the above website: ES-LiR; Drawing weight diagram; The other ML algorithms as the reference.

Acknowledgements

This work was partially supported by JST PRESTO (Y.O., JPMJPR16N2 and Y. I. JPMJPR17N2), JST CREST (Y.I., JPMJCR21O1), JSPS-KAKENHI (Y. O. JP22H02148, JP23K23416, JP24K01550, JP25H00425), AMED (JP25ym0126819).

Notes and references

1 Y. Osada and J. P. Gong, *Prog. Polym. Sci.*, 1993, **18**, 187.
2 M. M. Caruso, D. A. Davis, Q. Shen, S. A. Odom, N. R. Sottos, S. R. White and J. S. Moore, *Chem. Rev.*, 2009, **109**, 5755.
3 Y. Sagara, S. Yamane, M. Mitani, C. Weder and T. Kato, *Adv. Mater.*, 2016, **28**, 1073.
4 C. Isapour and M. Lattuada, *Adv. Mater.*, 2018, **30**, 1707069.
5 M. I. Khazi, W. Jeong and J. M. Kim, *Adv. Mater.*, 2018, **30**, 1705310.
6 K. Tanaka and Y. Chujo, *Polym. J.*, 2023, **55**, 353.
7 T. Watabe and H. Ohtsuka, *Macromolecules*, 2024, **57**, 425.
8 Y. Oaki and S. Fujii, *Chem. Commun.*, 2024, **60**, 9163.
9 X. Tian, Y. Guo, J. Zhang, O. M. Ivasishin, J. Jia and J. Yan, *Small*, 2024, **20**, 2306952.
10 B. Yoon, T. Oh, Y. J. Chang and J. Suhr, *Small*, 2025, **21**, 2310682.
11 S. Dolui, B. Sahu and S. Banerjee, *Macromol. Chem. Phys.*, 2025, **226**, 2400472.
12 T. Bereau, D. Andrienko and K. Kremer, *APL Mater.*, 2016, **4**, 053101.
13 D. J. Audus and J. J. de Pablo, *ACS Macro Lett.*, 2017, **6**, 1078.
14 J. S. Peerless, N. J. B. Milliken, T. J. Oweida, M. D. Manning and Y. G. Yingling, *Adv. Theory Simul.*, 2019, **2**, 1800129.
15 K. Sattari, Y. Xie and J. Lin, *Soft. Mater.*, 2021, **17**, 7607.
16 K. Zhang, X. Gong and Y. Jiang, *Adv. Funct. Mater.*, 2024, **34**, 2315177.
17 K. Hatakeyama-Sato, *Polym. J.* 2023, **55**, 117.
18 T. Orlova, A. Piven, D. Darmoroz, T. Aliev, T. M. T. A. Razik, A. Boitsev, N. Grafeeva and E. Skorb, *Digital Discovery*, 2023, **2**, 298.
19 K. Wang, H. Shi, T. Li, L. Zhao, H. Zhai, D. Korani and J. Yeo, *Digital Discovery*, 2023, **2**, 1660.
20 C. Yan and G. Li, *Adv. Intell. Syst.*, 2023, **5**, 2200243.
21 T. B. Martin and D. J. Audus, *ACS Polym. Au*, 2023, **3**, 239.
22 W. Ge, R. De Silva, Y. Fan, S. A. Sisson and M. H. Stenzei, *Adv. Mater.*, 2025, **37**, 2413695.
23 S. Okada, S. Peng, W. Spevak and D. Charych. *Acc. Chem. Res.*, 1998, **31**, 229.
24 D. J. Ahn and J. M. Kim, *Acc. Chem. Res.*, 2008, **41**, 805.
25 X. Sun, T. Chen, S. Huang, L. Li and H. Peng, *Chem. Soc. Rev.*, 2010, **39**, 4244.
26 X. Qian and B. Städler, *Chem. Mater.*, 2019, **31**, 1196.
27 Y. Oaki, *Chem. Commun.*, 2020, **56**, 13069.
28 B. Das, S. Jo, J. Zheng, J. Chen and K. Sugihara, *Nanoscale*, 2022, **14**, 1670.
29 Y. Kim, K. Iimura and N. Tamaoki, *Bull. Chem. Soc. Jpn.*, 2024, **97**, uoae034.
30 C. C. Revadekar, A. A. Patil, J. M. Kim and B. J. Park, *J. Photochem. Photobiol. C, Photochem. Rev.*, 2025, **63**, 100699.
31 Z. Huo, Q. Deng, T. Fan, G. He, X. Hu, X. Hong, H. Chen, S. Luo, Z. Wang and D. Chen, *Polym. Chem.*, 2017, **8**, 7438.
32 Z. Yu, C. MuYu, H. Xu, J. Zhao and G. Yang, *Polym. Chem.*, 2023, **14**, 2266.
33 A. Thakuri, M. Banerjee and A. Chatterjee, *Chem. Asian J.*, 2025, **20**, e202500219.
34 S. Dei, M. Matsumoto and A. Matsumoto, *Macromolecules*, 2008, **41**, 2467.
35 S. Ampornpun, S. Montha, G. Tumcharern, V. Vchirawongkwin, M. Sukwattanasinitt and S. Wacharasindhu, *Macromolecules.*, 2012, **45**, 9038.
36 J. Seo, C. Kantha, J. F. Joung, S. Park, R. Jelinek and J. M. Kim, *Small*, 2019, **15**, 1901342.
37 B. Hu and P. Wu, *Giant.*, 2020, **3**, 100025.
38 B. S. Kim, M. I, Khazi and J. M. Kim, *Macromolecules*, 2021, **54**, 8220.
39 R. Saymung, N. Traiphol and R. Traiphol, *Colloids Surf. A Physicochem. Eng. Asp.*, 2021, **626**, 120746.
40 S. Baek, J. M. Heo, K. Bae, M. I. Khazi, S. Lee, K. Kim and J. M. Kim, *Langmuir.*, 2024, **40**, 18272.
41 N. N. Kadamannil, D. Jang, H. Lee, J. M. Kim and R. Jelinek, *Small Methods*, 2024, **8**, 2301286.
42 J. Zheng, J. Chen, M. Galluzzi, Y. Hou and K. Sugihara, *Nano Lett.*, 2025, **25**, 7307.
43 Y. Takeuchi, H. Imai and Y. Oaki, *J. Mater. Chem. C*, 2025, **13**, 3209.
44 Y. Ishijima, H. Imai and Y. Oaki, *Chem.*, 2017, **3**, 509.
45 H. Terada, H. Imai and Y. Oaki, *Adv. Mater.*, 2018, **30**, 1801121.
46 Y. Oaki, Y. Ishijima and H. Imai, *Polym. J.*, 2018, **50**, 319.
47 K. Watanabe, H. Imai and Y. Oaki, *Small.*, 2020, **16**, 20045.
48 K. Watanabe, H. Imai and Y. Oaki, *J. Mater. Chem. C*, 2020, **8**, 1265.
49 Y. Mochizuki, H. Imai and Y. Oaki, *ChemPlusChem*, 2021, **86**, 1563.
50 A. Edagawa, S. Matsuda, H. Kawakubo, H. Imai and Y. Oaki, *ACS Appl. Mater. Interfaces.*, 2022, **14**, 43792.
51 N. Shioda, J. M. Heo, B. Kim, H. Imai, J. M. Kim and Y. Oaki, *Sens. Diagn.*, 2022, **1**, 163.
52 R. Shibata, S. Matsuda, H. Kawakubo, H. Imai and Y. Oaki, *J. Mater. Chem. B*, 2024, **12**, 10886.
53 S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito and O. Levy, *Nat. Mater.*, 2013, **12**, 191.
54 K. Rajan, *Annu. Rev. Mater. Res.*, 2015, **45**, 153.
55 K. T. Butler, J. M. Frost, J. M. Skelton, K. L. Svanea and A. Walsh, *Chem. Soc. Rev.*, 2016, **45**, 6138.
56 B. Sanchez-Lengeling and A. Aspuru-Guzik, *Science*, 2018, **361**, 360.
57 A. Agrawal and A. Choudhary, *MRS Commun.*, 2019, **9**, 779.
58 L. Himanen, A. Geurts, A. S. Foster and P. Rinke, *Adv. Sci.*, 2019, **6**, 1900808.
59 Y. Miyake and A. Saeki, *J. Phys. Chem. Lett.*, 2021, **12**, 12391.
60 S. C. Smith, C. S. Horbaczewskyj, T. F. N. Tanner, J. J. Walder and J. S. Fairlamb, *Digital Discovery*, 2024, **3**, 1467.
61 A. D. Sendek, B. Ransom, E. D. Cubuk, L. A. Pellouchoud, J. Nanda and E. J. Reed, *Adv. Energy Mater.*, 2022, **12**, 2200553.
62 Y. Zhang and C. Ling, *npj Comp. Mater.*, 2018, **4**, 25.
63 Y. Liu, B. Guo, X. Zou, Y. Li and S. Shi, *Energy Storage Mater.*, 2020, **31**, 434.
64 B. Dou, Z. Zhu, E. Merkurjev, L. Ke, L. Chen, J. Jian, Y. Zhu, J. Liu, B. Zhang and G. W. Wei, *Chem. Rev.*, 2023, **123**, 8736.

View Article Online
DOI: 10.1039/D5DD00442J

65 P. Xu, X. Ji, M. Li and W. Lu, *npj Comp. Mater.*, 2023, **9**, 42.
66 S. J. Pan and Q. Yang. *IEEE Trans. Knowl. Data Eng.*, 2010, **22**, 1345.
67 A. U. Mahmood, M. M. Ghelardini, J. B. Tracy and Y. G. Yingling, *Chem. Mater.*, 2024, **36**, 9330.
68 Y. Oaki and Y. Igarashi, *Bull. Chem. Soc. Jpn.*, 2021, **94**, 2410.
69 Y. Haraguchi, Y. Igarashi, H. Imai and Y. Oaki, *Digital Discovery*, 2022, **1**, 26.
70 H. Tobita, Y. Namiuchi, T. Komura, H. Imai, K. Obinata, M. Okada, Y. Igarashi and Y. Oaki, *Energy Adv.*, 2023, **2**, 1014.
71 W. Hamada, M. Hishida, R. Sugiura, H. Tobita, H. Imai, Y. Igarashi and Y. Oaki, *J. Mater. Chem. A*, 2024, **12**, 3294.
72 D. Suzuki, H. Minato, Y. Sato, R. Namioka, Y. Igarashi, R. Shibata and Y. Oaki, *Chem. Commun.*, 2024, **60**, 13678.
73 Y. Kitamura, Y. Namiuchi, H. Imai, Y. Igarashi and Y. Oaki, *Nanoscale Adv.*, 2025, **7**, 4620.
74 K. Sakano, Y. Igarashi, H. Imai, S. Miyakawa, T. Saito, Y. Takayanagi, K. Nishiyama and Y. Oaki, *ACS Appl. Energy Mater.*, 2022, **5**, 2074.
75 R. Yamamoto, Y. Igarashi, H. Imai, T. Sakata, S. Miyakawa, S. Yoshizaki, T. Saito and Y. Oaki, *Batteries & Supercaps*, 2025, **8**, e202500288.

Data Availability Statement

View Article Online
DOI: 10.1039/D5DD00442J

A data-driven approach to control stimulus responsivity of functional polymer materials: Predicting thermoresponsive color-change properties of polydiacetylene

Risako Shibata,[a] Nano Shioda,[a] Hiroaki Imai,[a] Yasuhiko Igarashi,[b] Yuya Oaki*,[a]

[a] Department of Applied Chemistry, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan.

[b] Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8573, Japan

E-mail: oakiyuya@applc.keio.ac.jp

The data supporting this article have been included as part of the ESI. The following datasets are available at GitHub of our group (https://github.com/Oaki-Group/202511_PDA) (DOI: https://doi.org/10.5281/zenodo.18060256): Tables S1 and S5 as csv file with simplified molecular input line entry system (SMILES) and mol2 files; Table S2 as csv file for the training dataset; Table S3 as csv file for the test dataset. The following codes are available at the above website: ES-LiR; Drawing weight diagram; The other ML algorithms as the reference.