Nanoscale

# Lateral-Size Control of Exfoliated Transition-Metal-Oxide 2D Materials by Machine Learning on Small Data



| | |
|---|---|
| Journal: | *Nanoscale* |
| Manuscript ID | NR-ART-12-2020-008684.R1 |
| Article Type: | Paper |
| Date Submitted by the Author: | 19-Jan-2021 |
| Complete List of Authors: | Mizuguchi, Ryosuke; Keio University, Department of Applied Chemistry, Faculty of Science and Technology<br>Igarashi, Yasuhiko; Tsukuba Daigaku<br>Imai, Hiroaki; Keio University, Department of Applied Chemistry, Faculty of Science and Technology<br>Oaki, Yuya; Keio University, Department of Applied Chemistry, Faculty of Science and Technology |
| | |

ARTICLE

# Lateral-Size Control of Exfoliated Transition-Metal-Oxide 2D Materials by Machine Learning on Small Data

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

DOI: 10.1039/x0xx00000x

Ryosuke Mizuguchi,[a] Yasuhiko Igarashi,[b,c] Hiroaki Imai,[a] Yuya Oaki*[a,c]

A wide variety of nanosheets including monolayers and few-layers have attracted much interest as two-dimensional (2D) materials for the unique anisotropic structures and properties. On the other hand, one of the significant remaining and challenging issues is the lateral-size control. Since 2D materials are generally synthesized by exfoliation of layered materials, the lateral size is not easily controlled in the breaking-down uncontrollable processes. The experimental factors have not been found for the control and prediction. In the present work, lateral sizes of exfoliated transition-metal-oxide nanosheets were predicted and controlled by assistance of machine learning. Layered composites of host inorganic layers and guest organic molecules were exfoliated into the nanosheets in organic dispersion media. The lateral size of the nanosheets was estimated by dynamic light scattering (DLS), instead of microscopy methods, to achieve high-throughput analyses. Factors governing the lateral size are explored on the small experimental data by assistance of sparse modeling, a method of machine learning. The eight physicochemical parameters of the organic guests and dispersion media were extracted by sparse modeling for construction of the size-prediction model. The size-prediction model accelerated the selective syntheses of nanosheets with large and small lateral sizes in a limited number of the experiments. The results indicate that the prediction model is a guideline to find the suitable exfoliation conditions for the size control. Size-selective syntheses of a variety of 2D materials can be achieved by the similar methods using sparse modeling on small experimental data. Moreover, sparse modeling is applicable to control the design and exploration of the other materials and their processing based on small data.

## Introduction

Two-dimensional (2D) materials attract much interest as one of the most promising nanostructures in recent years.[1–11] The characteristic properties and functions are extracted from the ultrathin, anisotropic, and flexible structures. Nanosheets including monolayers and few-layers are synthesized by exfoliation of precursor layered materials in liquid phase.[12–20] However, it is not easy to obtain the nanosheets with tailored size, yield, and surface chemistry through exfoliation processes. The lateral-size control is an important challenge for the applications, such as assembly of the nanosheets and tuning of the properties. In general, the lateral size is not easily predicted and controlled because the precursor layered materials are broken down into nanosheets through exfoliation coupled with random fracture in lateral direction. In addition, measurement of the lateral size generally requires time- and effort-consuming microscopy analyses. Therefore, the lateral-size control of 2D materials is still a challenging target for chemists and materials scientists. New approaches are required for the lateral-size prediction and control in the exfoliation processes. Here we applied machine learning to change the breaking-down uncontrollable processes to controllable ones (Fig. 1).

Exfoliation methods and processes are different for layered materials, such as those consisting of van der Waals and electrostatic interactions.[12–20] Recently, exfoliated nanosheets are observed on not only classical layered inorganic compounds but also a wide variety of materials.[21–25] In previous works, variation of the lateral size was observed by changes in the exfoliation conditions,[26–35] such as the type and concentration of the exfoliating agents,[26–30,32] exfoliation time,[31–33] and centrifugation conditions during purification.[34,35] The larger and smaller nanosheets were obtained by size control of the precursor layered compounds.[36–39] Here we show a new general strategy to predict and control the lateral size of 2D materials. The layered composites of inorganic host and organic guest were exfoliated into the surface-modified nanosheets in organic media. The exfoliation proceeds on the basis of the affinity between the guests and media through swelling.[40–44] In this method, the exfoliation behavior can be changed by combinations of the host, guest, and dispersion medium.[43,44] The present work shows the lateral-size prediction and control of the transition-metal-oxide nanosheets by an assistance of sparse modeling, a recent successful machine-learning method.

[a] Department of Applied Chemistry, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan. E-mail: oakiyuya@applc.keio.ac.jp
[b] Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8573, Japan.
[c] JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan.
Electronic Supplementary Information (ESI) available: [Experimental methods; Training dataset for the sparse modeling, Electron microscopy images; Structures and compositions of the precursor layered composites; Test dataset; DLS charts of the resultant nanosheets in the dispersion media]. See DOI: 10.1039/x0xx00000x

Sparse modeling is a machine-learning method for explanation of whole data by a simple model using a limited number of significant descriptors on the assumption of sparseness in high-dimensional data.[45–47] The modeling method is applied to a variety of phenomena in physical sciences, such as compression of image data. Our group has studied applications of sparse modeling to small-scale our own experimental data.[43,44,48] The method contributes to extract the descriptors overlooked by researchers. A simple interpretable prediction model using a limited number of descriptors can be constructed by sparse modeling. The approach facilitates accelerated finding of molecules, materials, and processes by experimental scientists.[43,44,48–50] In the present work, lateral-size prediction and control of nanosheets were achieved by sparse modeling on our own small data. The construction method of the size-prediction model can be applied to achieve controlled syntheses of the other 2D materials.

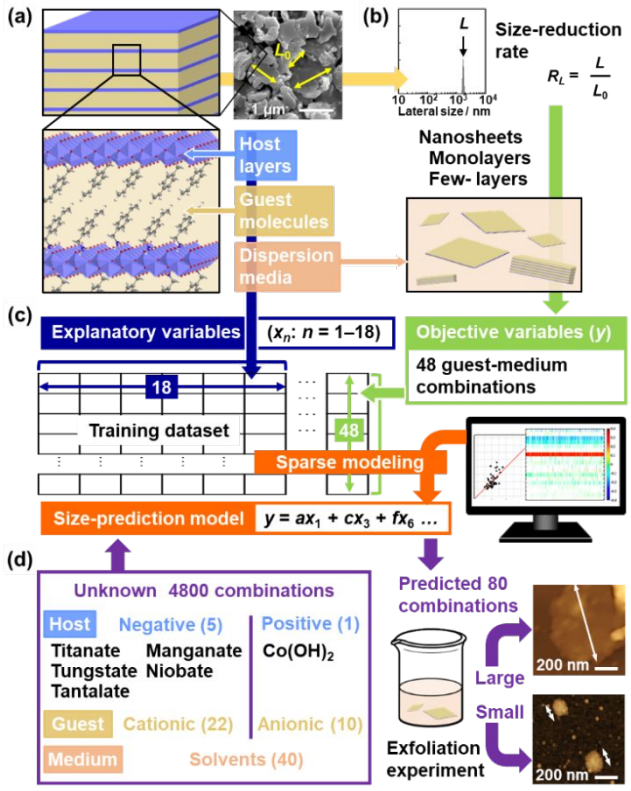


**Fig. 1**. Overview of the present work. (a) Precursor layered composites consisting of the host layers and guest molecules with the original lateral size ($L_0$). (b) Exfoliation of the layered composites into the surface-functionalized nanosheets in organic media and their DLS chart to estimate the size-reduction rate ($R_L = L\,L_0^{-1}$). (c) Training dataset consisting of 48 $R_L$ values as objective variables ($y = R_L$) and 18 potential descriptors as explanatory variables ($x_n$: $n$ = 1–18) for construction of the size-prediction model using sparse modeling. (d) Exfoliation experiments for the tailored syntheses of large and small nanosheets on the predicted 80 host-guest-medium combinations (right) extracted from the total unknown 4800 combinations using the prediction model (left).

## Results and Discussion

The nanosheets were synthesized by exfoliation of the layered composites consisting of the host inorganic layers and interlayer organic guests in organic dispersion media (Fig. 1a,b). A simple lateral-size-prediction model was constructed by sparse modeling on the small experimental data (Fig. 1c). Finally, the lateral-size prediction and control of the exfoliated nanosheets were achieved on unknown host, guest, and medium combinations in a limited number of experiments (Fig. 1d). The detailed methods for the syntheses, characterization, and data analysis were described in the Electronic Supplementary Information (ESI).

### Construction of size-prediction model from small experimental data

The layered composites of host titanate and 8 different interlayer guests were exfoliated in 13 dispersion media (Fig. 1a).[43] The nanosheets were synthesized in the total 104 guest-medium combinations. The particle-size distribution of the dispersion liquid containing the nanosheets was measured by dynamic light scattering (DLS) after removal of the unexfoliated bulky particles by filtration (Fig. 1b). These data were extracted from our previous work.[43] In previous works,[23,43,51] the lateral size was roughly estimated by DLS measurement to without time-consuming microscopy observations. In the present work, the average size ($L$) of the DLS measurement was regarded as the average lateral size of the nanosheets to achieve the high-throughput analysis and data collection (Fig. 1b). The thickness of the transition-metal-oxide nanosheets is only measured by time-consuming AFM, even though that of some layered compounds based on van der Waals interaction can be estimated from the spectroscopic analyses. According to our previous works, the nanosheets include the monolayers and few-layers. In the present work, we focused on the lateral size of the nanosheets measured by DLS. The pristine layered compounds were prepared by solid-state and solution syntheses. Then, the precursor layered composites were obtained by the intercalation of the guest organic molecules. Our previous works suggest that the particle sizes of the pristine layered compounds were not changed after the intercalation of the guests.[44,45] Therefore, the average lateral size ($L_0$) of the pristine layered compounds was measured by scanning electron microscopy (SEM) (Fig. 1a and Fig. S1 in the ESI). The size-reduction rate ($R_L = L\,L_0^{-1}$) was used as a parameter of the lateral size because $L$ depends on $L_0$ of the host layered compounds. The $R_L$ values for the 48 guest-medium combinations with the yield of the nanosheets higher than 3 % were extracted from the original 104 combinations (Fig. 1c and Table S1 in the ESI). The average yield of the nanosheets was 23.9 ± 21.7 % for the 48 guest-medium combinations in the training dataset (Table S1 in the ESI). The yield was measured by collection of the nanosheets through the filtration of the dispersion liquid.[43,44] These 48 $R_L$ values in the range of 0.120–0.576 were set as the objective variables ($y$). Then, 18 potential factors related to $R_L$ were selected on the basis of our perspective and listed as the explanatory variables ($x_n$: $n$ = 1–18) (Table 1 and Fig. 1c). The explanatory variables are physicochemical parameters for the potential descriptors controlling the lateral size, such as molecular weight, polarizability, and Hansen-solubility (-similarity) parameters regarding the dispersion media ($x_n$: $n$ = 1–9), guests ($x_n$: $n$ = 10–15), and guest-medium combinations ($x_n$:

$n$ = 16–18). In this manner, the small-scale training dataset containing the 48 objective variables ($y$) and 18 explanatory variables ($x$) was prepared for sparse modeling (Fig. 1c and Table S1 in the ESI).

Our intention here is to construct a simple and interpretable prediction model for control of $R_L$ by the exfoliation experiments. The important factors to estimate $R_L$ ($y$) were extracted from the training dataset by multiple regression model with penalty based on minimax concave plus (MCP), a method for sparse modeling (Fig. 2).[52,53] This sparse modeling extracted 10 descriptors from the original 18 explanatory variables (the colored $x_n$ in Table 1). The $R_L$ is represented by the descriptors $x_n$ ($n$ = 3, 4, 6, 7, 8, 10, 12, 13, 14, 18). The modified prediction model using these 10 descriptors had the relationship between the measured ($y$) and estimated $R_L$ ($y'$) with cross validation error (CVE) 0.130. However, the descriptors $x_6$ and $x_{12}$ require time-consuming calculation by density functional theory (DFT) (the gray-colored cells in Table 1). The prediction model using 8 descriptors except $x_6$ and $x_{12}$, namely $x_n$ ($n$ = 3, 4, 7, 8, 10, 13, 14, 18), is represented by (eq. 1) with root mean squared error (RMSE) 0.116 (Fig. 2a), where the descriptors $x_n$ are converted to the normalized frequency distribution such that the mean is 0 and standard deviation is 1.

**Table 1**. List of the explanatory variables ($x_n$: $n$ = 1–18).

| Dispersion media | | Guest molecules | |
|---|---|---|---|
| $x_n$ | Parameters | 10 | Molecular weight |
| 1 | Molecular weight | 11 | [b] Polarizability |
| 2 | [a] Melting point | 12 | [b] Dipole moment |
| 3 | [a] Boiling point | 13 | [c] HSP-dispersion |
| 4 | [a] Density | 14 | [c] HSP-polarity |
| 5 | [b] Dipole moment | 15 | [c] HSP-hydrogen bonding |
| 6 | [b] Polarizability | Guest-medium combinations | |
| 7 | [c] HSP-dispersion | 16 | [b] Δ Polarizability |
| 8 | [c] HSP-polarity | 17 | [b] Δ Dipole moment |
| 9 | [c] HSP-hydrogen bonding | 18 | [c] HSP distance |

a literature data, b Calculation data by DFT, c Calculation data by HSPiP software. The red and blue descriptors have positive and negative correlations, respectively. The gray descriptors were not used in the prediction model (eq. 1).

$$y = -0.159x_3 - 0.096x_4 + 0.257x_7 - 0.017x_8 - 0.018x_{10} + 0.028x_{13} - 0.050x_{14} + 0.061x_{18} + 0.267 \ldots \text{(eq. 1)}$$

The relationship between the $y$ and $y'$ indicates that the prediction model enables not accurate but rough estimation of $R_L$ using only 8 descriptors (Fig. 2a).

The additional sparse modeling, exhaustive search with linear regression (ES-LiR),[47] supported the correlation of the extracted 8 descriptors (Fig. 2b). In sparse modeling, linear regression model is represented by a limited number of descriptors with non-zero coefficients. In ES-LiR,[47] which coefficients are zero (or non-zero) is exhaustively searched in all the possible combinations of the explanatory variables, namely $2^N$–1 combinations, where $N$ is the total number of explanatory variables. In the present work, after calculation of the CVE values for total $2^{18}$–1 (= 2.62 × $10^5$) combinations, the combinations were sorted in the ascending order of the CVE values. The weight diagram shows the $10^3$ combinations with the smallest CVE values. The colors represent the positive and negative correlations in the linear regression models (Fig. 2b). In the weigh diagram, the descriptors $x_3$, $x_4$, $x_8$, $x_{10}$, and $x_{14}$ exhibited blue cool colors corresponding to the negative correlation. In contrast, the descriptors $x_7$, $x_{13}$, and $x_{18}$ had red warm color corresponding to the positive coefficients. The trends of the positive and negative correlations are consistent with the coefficients in the prediction model (eq. 1). Therefore, the ES-LiR supports the prediction model (eq. 1).

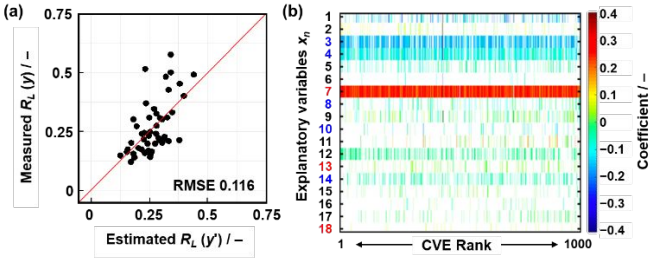


**Fig. 2**. Construction of the size-prediction model using the descriptors extracted by sparce modeling and chemical perspective. (a) Relationship between the measured and estimated $R_L$ values for the 48 samples in the training dataset according to the size-prediction model (eq. 1) using the extracted 8 descriptors. (b) Weight diagram of the ES-LiR analysis representing the coefficients of the non-zero explanatory variables with the warm and cool colors corresponding to the positive and negative values, respectively.

According to the correlations, the larger nanosheets are obtained by the smaller $x_3$ (boiling point of dispersion media), $x_8$ (Hansen-solubility (or -similarity) parameter (HSP) polarity term of dispersion media), and $x_{14}$ (HSP polarity term of guests). The larger $x_7$ (HSP dispersion term of dispersion media) and $x_{13}$ (HSP dispersion term of guest molecules) provide the larger nanosheets. These correlations of the descriptors imply that the larger nanosheets are formed with weaker intermolecular interaction between the nonpolar interlayer guest and dispersion medium. If the guest-medium interaction is weak, the smooth exfoliation without fracture is induced without stress to the layer. In contrast, the stronger intermolecular interaction between the polar guest and medium causes the larger stress to the host metal-oxide layer leading to the fracture in the lateral direction. In this manner, the positive and negative correlations in the model are consistent with the chemical perspective, although the correlation of the other descriptors, such as $x_4$ (density of dispersion media), $x_{10}$ (molecular weight of guests), and $x_{18}$ (HSP distance of the guests and dispersion media) are not clearly explained. The size-prediction model (eq. 1) is a potential guideline for syntheses of the large and small nanosheets.

### Sparse modelling for small experimental data

Sparse linear regression is applied in the present work on the assumption of a linear relationship between the explanatory and objective variables, because the number of data is small for the normal machine-learning (ML) algorithm. Although the

degree of freedom of the model is narrowed by assuming a linear relationship, the model can be trained stably even on small data. The problem for predictions using ML algorithms is not the size of the trained data, but rather the similarity of the trained data to the objective data. If the objective data is similar to the trained data, we can extrapolate from the training data and predict the better results. In the present study, the extrapolation can be achieved because the experimental system of the exfoliation is same in the trained and test data. Among machine learning algorithms, the linear model with sparsity has some degree of reliability to the extrapolation results on the assumption that the correlation between the explanatory and objective variables is constant outside the training data. In addition, a sparse linear regression model is easily interpretable for experimental scientists. Therefore, we believe that the construction and application of the linear model with sparsity is a suitable screening method of the experimental conditions based on small data. The correlations between explanatory and objective variables can be changed by the accumulation of the experimental data. The modifications in the coefficients are required to construct the more accurate model with addition of the data. The other chemical-related features may have correlation with the lateral-size of the nanosheets as the descriptors. In our group, the explanatory variables as the potential descriptors are selected prior to sparse modeling on the basis of our chemical perspective and experience.[43,44] The selection of the explanatory variables as the prior knowledge corresponds to the assumption of the sparseness itself. If the sufficient model is not constructed, reconsideration of the explanatory variables is needed. We first provided the potential factors to avoid the overlooking of the descriptors, as much as we can. If the provided explanatory variables have little correlation, these are not extracted as the descriptors by machine learning. Listing all the potential explanatory variables is important for our sparse modeling on small experimental data. In fact, design of new organic anode for lithium-ion battery was performed by the similar approach in our previous work.[53]

### Predicted syntheses of large and small nanosheets

The size-selective syntheses of metal-oxide nanosheets were demonstrated on unknown host-guest-medium combinations (Fig. 1d). According to the prediction model, the nanosheets with large and small lateral sizes were synthesized only on the predicted 80 combinations in the total 4800 candidates (Fig. 1d and Fig. S2 in the ESI). The number of experiments reduced by 98.3 % using the prediction model. The unknown combinations of 22 guest molecules and 40 dispersion media were prepared for 5 host layered compounds with negatively charged layers, such as layered titanate, manganate, tungstate, niobate, and tantalate (Fig. 1d and Fig. S2 in the ESI).[29,44,54-57] The unknown combinations of 10 guest molecules and 40 dispersion media were prepared for the host containing the positively charged layer, such as α-type layered cobalt hydroxide ($Co(OH)_2$) (Fig. 1d and Fig. S2 in the ESI).[42]

The yield- and size-prediction models recommended the guest-medium combinations for syntheses of large and small nanosheets in high yield. Prior to the size prediction, 468 guest-medium combinations with the predicted yield higher than 30 % were extracted from unknown 880 combinations for the negatively charged host layers using the yield-prediction model in our previous report.[44] Then, the lateral size was predicted using (eq. 1) for the extracted 468 combinations. The exfoliation experiments were only performed on the predicted largest and smallest 6 or 10 guest-medium combinations for each host layer (Table 2 and Table S2 in the ESI). When the target guest molecule was not experimentally intercalated in the layered compounds, the next one was moved up in the rank. In this manner, the ranks of the predicted large and small $R_L$ values were prepared for each layered compound (Table 2 and Tables S2 in the ESI). Table 2 shows the ranks of the predicted 10 guest-medium combinations providing large (L-01–10) and small (S-01–10) titanate nanosheets. The ranks L-01–06 and S-01–06 were prepared for the other host layers (Table S2 in the ESI). The required precursor layered composites were synthesized for the exfoliation experiments (Figs. S3–S8 and Table S3 in the ESI).

**Table 2**. The predicted 10 combinations for syntheses large (L-01–10) and small (S-01–10) titanate nanosheets.

| Large-size nanosheets | | | | |
|---|---|---|---|---|
| Rank | Guest | Medium | Predicted $R_L$ / – | Measured $R_L$ / – |
| L-01 | Vinyl-BA | THF | 0.473 | 1.130 |
| L-02 | H-BA | Benzaldehyde | 0.469 | 0.182 |
| L-03 | Vinyl-BA | Ethylbenzene | 0.465 | 0.233 |
| L-04 | NEA | 1,3-Dioxolane | 0.461 | 0.308 |
| L-05 | Tp-EA | THF | 0.461 | 0.352 |
| L-06 | Tp-MA | Benzaldehyde | 0.453 | 0.434 |
| L-07 | $CH_3$-BA | Benzaldehyde | 0.451 | 0.334 |
| L-08 | DOA | THF | 0.449 | 0.264 |
| L-09 | H-BA | 1,3-Dioxolane | 0.445 | 0.354 |
| L-10 | $NH_2$-BA | 1,3-Dioxolane | 0.392 | 0.592 |
| Average | | | 0.452 | 0.418 |
| Standard deviation | | | 0.023 | 0.275 |
| Small-size nanosheets | | | | |
| S-01 | DAMN | Formamide | −0.567 | 0.179 |
| S-02 | API | DEG | −0.476 | 0.163 |
| S-03 | BPO | DEG | −0.460 | 0.310 |
| S-04 | $C_{18}$-$NH_2$ | DEG | −0.402 | 0.312 |
| S-05 | $C_6$-$NH_2$ | DEG | −0.394 | 0.189 |
| S-06 | MeO-BA | DEG | −0.386 | 0.194 |
| S-07 | F-BA | DEG | −0.382 | 0.117 |
| S-08 | DOA | DEG | −0.368 | 0.200 |
| S-09 | API | DEG | −0.337 | 0.118 |
| S-10 | DEA | Formamide | −0.262 | 0.292 |
| Average | | | −0.403 | 0.207 |
| Standard deviation | | | 0.083 | 0.073 |

The actual $R_L$ values were measured to demonstrate formation of the large and small-size nanosheets for the total 80 host-guest-medium conditions (Table 2 and Table S2 in the ESI). The $R_L$ values were estimated from the DLS measurements to achieve the high-throughput study (Figs. S9–S14 in the ESI).

The average measured $R_L$ values were calculated for the large- (L-01–06) and small-size (S-01–06) conditions of each host layered compound. The average yield of the size-controlled large and small nanosheets was 23.0 ± 23.2 % for all the 80 host-guest-medium combinations (Table S2 in the Supporting Information). Since the yield prediction model was applied according to our previous work,[45] the nanosheets were efficiently obtained in the limited number of the experiments. Table 3 summarizes the average predicted $R_L$, measured $R_L$, and $L$. Fig. 3 shows the average measured $R_L$ values for the large and small nanosheets derived from each layered compound. The average measured $R_L$ values are significantly different for the predicted conditions providing the large and small nanosheets except the layered composites based on tungstate (Table 3 and Fig. 3). According to our previous report,[29] the exfoliation of the layered composite based on tungstate provided the nanoflakes with smaller lateral size through frequent fracture in the lateral direction because the surface of the layer was not fully modified by the guest molecules. Therefore, the lateral sizes of the tungstate nanosheets are not controlled in the predicted conditions.

**Table 3**. Average predicted $R_L$, measured $L$, and measured $R_L$ of the predicted conditions providing the large and small nanosheets for each host layered compound.

| Host layers | $L_0$ / μm | Samples | Predicted $R_L$ / – | Measured $L$ / μm | Measured $R_L$ / – |
|---|---|---|---|---|---|
| Titanate | 1.52 | L-01–10 | 0.452 | 0.635 | 0.418 |
| | | S-01–10 | −0.403 | 0.315 | 0.207 |
| Manganate | 3.66 | L-01–06 | 0.459 | 2.11 | 0.577 |
| | | S-01–06 | −0.362 | 0.344 | 0.094 |
| Tungstate | 4.78 | L-01–06 | 0.458 | 0.392 | 0.082 |
| | | S-01–06 | −0.386 | 0.521 | 0.110 |
| Niobate | 4.60 | L-01–06 | 0.435 | 1.38 | 0.301 |
| | | S-01–06 | −0.399 | 0.451 | 0.098 |
| Tantalate | 2.75 | L-01–06 | 0.435 | 1.76 | 0.639 |
| | | S-01–06 | −0.416 | 0.490 | 0.178 |
| $Co(OH)_2$ | 3.72 | L-01–06 | 0.378 | 1.15 | 0.308 |
| | | S-01–06 | −0.456 | 0.480 | 0.129 |

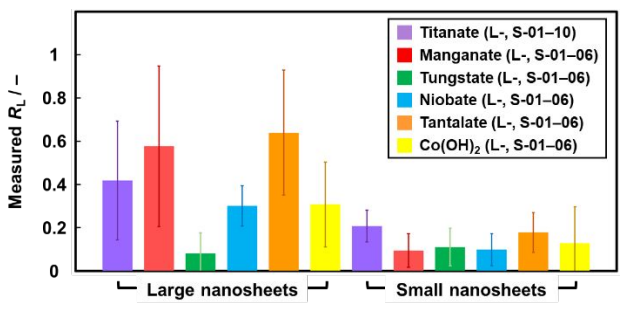


**Fig. 3**. Average measured $R_L$ values and their standard deviations of the large and small nanosheets on the predicted conditions for each host layer (Table 2 and Table S3 in the ESI).

The large and small nanosheets were actually observed on the transmission electron microscopy (TEM) images (Fig. 4 and Fig. S15 in the ESI). The layered titanate with intercalation of 4-vinylbenzylamine (Vinyl-BA) was exfoliated into the large nanosheets in tetrahydrofuran (THF) (Table 2 and Fig. 4a–c). The DLS chart showed the peak around 1.72 ± 0.079 μm (Fig. 4a). The large nanosheets were actually observed on the TEM images (Fig. 4b). The average size was estimated to be 0.489 ± 0.308 μm from the TEM images on the histogram (Fig. 4c). The average size estimated from the DLS chart is larger than that from the TEM image because the nanosheets as the primary particle form the aggregates in the dispersion liquid. In contrast, the smaller nanosheets were obtained by exfoliation of the layered titanate with intercalation of hexylamine ($C_6$-$NH_2$) in diethylene glycol (DEG) (Table 2 and Fig. 4d–f). The DLS measurement and TEM image showed formation of the smaller nanosheets 0.288 ± 0.022 μm and 0.213 ± 0.103 μm in size, respectively. The similar size differences in the large and small nanosheets were observed on the other hosts (Fig. S15 in the ESI). In this manner, the large- and small-size nanosheets were synthesized by exfoliation of the layered composites according to the prediction model constructed by machine learning on small experimental data. The constructed model is not a precise accurate predictor but a rough guideline to estimate the lateral size for reduction in the number of the experiments. The inaccuracy of the predicted values is ascribed to the linear regression model based on small-scale data using a limited number of the descriptors. In addition, the experimental error is included in the trained data. The extrapolation of the linear regression model causes the overestimation, such as the negative predicted values in small-size nanosheets in Table 2. Nevertheless, the nanosheets with the large and small lateral size were selectively prepared using the prediction model in a limited number of the experiments. The prediction model can be regarded as a guideline for controlled syntheses of the nanosheets. Our next challenge is prediction and control of the size distribution, as well as the lateral size itself. Moreover, the size-prediction model and its construction method can be applied to a variety of layered materials for controlled syntheses of 2D materials.

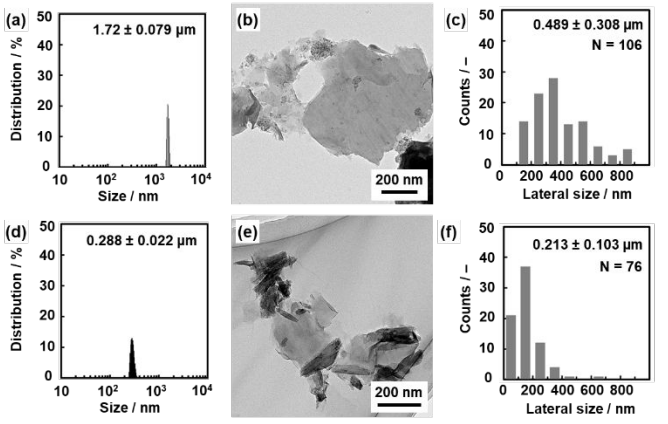


**Fig. 4**. Examples of the large (a–c) and small (d–f) nanosheets obtained by exfoliation of the layered titanate with intercalation of Vinyl-BA and $C_6$-$NH_2$ dispersed in THF (a–c) and DEG (d–f), respectively. (a,d) Size distribution by DLS. (b,e) TEM images. (c,f) Histograms prepared from the TEM images. TEM images of the other nanosheets were listed in Fig. S15 in the ESI.

## Conclusions

Lateral-size control of the exfoliated nanosheets was achieved by an assistance of machine learning. The size-prediction model was constructed from the small training dataset based on our own experiments by sparse modeling. Combination of the yield- and size-prediction models extracted unknown 80 host-guest-medium combinations providing large- and small-size nanosheets from the total 4800 candidates. The large- and small-size nanosheets were actually obtained in a limited number of the exfoliation experiments. The present size-prediction model can be regarded as a guideline to explore the exfoliation conditions. Moreover, the yield and size-prediction models can be applied to controlled syntheses of the other 2D materials. The present methods with combination of high-throughput analysis and ML on small data are applied to similar experimental works.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work was supported by JST PRESTO (Y.O., JPMJPR16N2 and Y. I. JPMJPR17N2).

## Notes and references

1 M. Osada and T. Sasaki, *Adv. Mater.*, 2012, **24**, 210.
2 V. Nicolosi, M. Chhowalla, M. G. Kanatzidis, M. S. Strano and J. N. Coleman, *Science*, 2013, **340**, 1226419.
3 H. P. Cong, J. F. Chen and S. H. Yu, *Chem. Soc. Rev.*, 2014, **43**, 7295.
4 R. Ma and T. Sasaki, *Acc. Chem. Res.*, 2015, **48**, 136.
5 J. E. ten Elshof, H. Yuan and P. G. Rodriguez, *Adv. Energy Mater.*, 2016, **6**, 1600355.
6 J. R. Brent, N. Savjani and P. O'Brien, *Prog. Mater. Sci.*, 2017, **89**, 411.
7 M. Servalli and A. D. Schlüter, *Annu. Rev. Mater. Res.*, 2017, **47**, 361.
8 K. Ariga, S. Watanabe, T. Mori and J. Takeya, *NPG Asia Mater.*, 2018, **10**, 90.
9 C. N. R. Rao and K. Paramoda, *Bull. Chem. Soc. Jpn.*, 2019, **92**, 441.
10 P. Ganter and B. V. Lotsch, *Mol. Syst. Des. Eng.*, 2019, **4**, 566.
11 M. A. Timmerman, R. Xia, R. T. P. Le, Y. Wang and J. E. ten Elshof, *Chem. Eur. J.*, 2020, **27**, 9084.
12 T. Sasaki, M. Watanabe, H. Hashizume, H. Yamada and H. Nakazawa, *J. Am. Chem. Soc.*, 1996, **118**, 8329.
13 R. Ma, Z. Liu, K. Takada, N. Iyi, Y. Bando and T. Sasaki, *J. Am. Chem. Soc.*, 2007, **129**, 5257.
14 Q. Wang and D. O'Hare, *Chem. Rev.*, 2012, **112**, 4124.
15 A. Clesielski and P. Samori, *Chem. Soc. Rev.*, 2014, **43**, 381.
16 A. Ambrosi and M. Pumera, *Chem. Soc. Rev.*, 2018, **47**, 7213.
17 Y. Yang, H. Hou, G. Zou, W. Shi, H. Shuai, J. Li and X, Ji. *Nanoscale*, 2019, **11**, 16.
18 Z. Sun, Q. Fan, M. Zhang, S. Liu, H. Tao and J. Texter, *Adv. Sci.*, 2019, **6**, 1901084.
19 P. Tao, S. Yao, F. Liu, B. Wang, F. Huang and M. Wang, *J. Mater. Chem. A*, 2019, **7**, 23512.
20 C. Chen, L. Tao, S. Du, W. Chen, Y. Wang, Y. Zou and S. Wang, *Adv. Funct. Mater.*, 2020, **30**, 1909832.
21 X. Zhuang, Y. Mai, D. Wu, F. Zhang and X. Feng, *Adv. Mater.*, 2015, **27**, 403.
22 Y. Ishijima, M. Okaniwa, Y. Oaki and H. Imai, *Chem. Sci.*, 2017, **8**, 647.
23 S, Yano, K. Sato, J. Suzuki, H. Imai and Y. Oaki, *Commun. Chem.*, 2019, **2**, 97.
24 D. X. Feng and D. A. Schlüter, *Angew. Chem. Int. Ed.*, 2018, **57**, 13748.
25 D. Rodríguez-San-Miguel, C. Montoro and F. Zamora, *Chem. Soc. Rev.*, 2020, **49**, 2291.
26 C. Backes, R. J. Smith, N. McEvoy, N. C. Berner, D. McCloskey, H. C. Nerl, A. O'Neill, P. J. King, T. Higgins, D. Hanlon, N. Scheuschner, J. Maultzsch, L. Houben, G. S. Duesberg, J. F. Donegan, V. Nicolosi and J. N. Coleman, *Nat. Commun.*, 2014, **5**, 4576.
27 L. J. Ji, Y. Qin, D. Gui, W. Li, Y. LI, X. Li and P. Lu, *Chem. Mater.*, 2018, **30**, 8732.
28 C. Backes, D. Campi, B. M. Szydlowska, K. Synnatschke, E. Ojala, F. Rashvand, A. Harvey, A. Griffin, Z. Sofer, N. Marzari, J. N. Coleman and D. D. O'Regan, *ACS Nano*, 2019**, 13**, 7050.
29 M. Honda, Y. Oaki and H. Imai, *Chem. Commun.*, 2015, **51**, 10046.
30 E. Varrla, C. Backes, K. R. Paton, A. Harvey, Z. Gholamvand, J. McCauley and J. N. Coleman, *Chem. Mater.*, 2015, **27**, 1129.
31 N. Miyamoto and T. Nakato, *J. Phys. Chem. B*, 2004, **108**, 6152.
32 T. Maluangnont, K. Matsuba, F. Geng, R. Ma, Y. Yamauchi and T. Sasaki, *Chem. Mater.*, 2013, **25**, 3137.
33 R. Mizuguchi, H. Imai and Y. Oaki, *Nanoscale Adv.*, 2020, **2**, 1168.
34 U. Khan, A. O'Neill, H. Porwal, P. May, K. Nawaz and J. N. Coleman, *Carbon*, 2012, **50**, 470.
35 C. Backes, B. M. Szydlowska, A. Harvey, S. Yuan, V. Vega-Mayoral, B. R. Davies, P. Zhao, D. Hanlon, E. J. G. Santos, M. I. Katsnelson, W. J. Blau, C. Gademaier and J. N. Coleman, *ACS Nano*, 2016, **10**, 1589.
36 N. Miyamoto, H. Yamamoto, R. Kaito and K. Kuroda, *Chem. Commun.*, 2002, 2378.
37 T. Tanaka, Y. Ebina, K. Takada, K. Kurashima and T. Sasaki, *Chem. Mater.*, 2003, **15**, 3564.
38 J. Peng, J. Wu, X. Li, Y. Zhou, Z. Yu, Y. Guo, J. Wu, Y. Lin, Z. Li, X. Wu, C. Wu and Y. Xie, *J. Am. Chem. Soc.*, 2017, **139**, 9019.
39 K. Nakamura, Y. Oaki and H. Imai, *J. Am. Chem. Soc.*, 2013, **135**, 4501.
40 M. Honda, Y. Oaki and H. Imai, *Chem. Mater.*, 2014, **26**, 3579.
41 G. Nakada, H. Imai and Y. Oaki, *Chem. Commun.*, 2018, **54**, 244.
42 Y. Yamamoto, H. Imai and Y. Oaki, *Bull. Chem. Soc. Jpn.*, 2019, **92**, 779.
43 G. Nakada, Y. Igarashi, H. Imai and Y. Oaki, *Adv. Theory Simul.*, 2019, **2**, 1800180.
44 K. Noda, Y. Igarashi, H. Imai and Y. Oaki, *Adv. Theory Simul.*, 2020, **3**, 2000084.
45 R. Tibshirani, M. Wainwright and T. Hastie, S*tatistical Learning with Sparsity: The Lasso and Generalizations*, Chapman and Hall/CRC, Philadelphia, PA 2015.
46 Y. Igarashi, K. Nagata, T. Kuwatani, T. Omori, Y. Nakanishi-Ohno and M. Okada, *J. Phys. Conf. Ser.*, 2016, **699**, 012001.
47 Y. Igarashi, H. Takenaka, Y. Nakanishi-Ohno, M. Uemura, S. Ikeda and M. Okada, *J. Phys. Soc. Jpn.*, 2018, **87**, 044802.
48 H. Numazawa, Y. Igarashi, K. Sato, H. Imai and Y. Oaki, *Adv. Theory Simul.*, 2019, **2**, 1900130.
49 K. Sodeyama, Y. Igarashi, T. Nakayama, Y. Tateyama and M. Okada, *Phys. Chem. Chem. Phys.*, 2018, **20**, 22585.
50 A. Ishikawa, K. Sodeyama, Y. Igarashi, T. Nakayama, Y. Tateyama and M. Okada, *Phys. Chem. Chem. Phys.*, 2019, **21**, 26399.

 

51 M. Lotya, A. Rakovich, J. F. Donegan and J. N. Coleman, *Nanotechnology*, 2013, **24**, 265703.
52 C. H. Zhang, *Ann. Stat.*, 2010, **38**, 894.
53 P. Breheny and J. Huang, *Ann. Appl. Stat.*, 2011, **5**, 232.
54 K. Nassau, J. W. Shiever and J. L. Bernstein, *J. Electrochem. Soc.*, 1969, **116**, 348.
55 N. Kimura, Y. Kato, R. Suizuki, A. Shimada, S. Tahara, T. Nakato, K. Matsukawa, P. H. Mutin and Y. Sugahara, *Langmuir*, 2014, **30**, 1169.
56 Y. Tsunoda, M, Shirata, W. Sugimoto, Z. Liu, O. Terasaki, K. Kuroda and Y. Sugahara, *Inorg. Chem.*, 2001, **40**, 5768.
57 Y. Tsunoda, W. Sugimoto and Y. Sugahara, *Chem. Mater.*, 2003, **15**, 632.