

Cite this: *Digital Discovery*, 2025, 4, 1570

# Machine learning-assisted profiling of a kinked ladder polymer structure using scattering†

Lijie Ding,<sup>a</sup> Chi-Huan Tung,<sup>a</sup> Zhiqiang Cao,<sup>a</sup> Zekun Ye,<sup>b</sup> Xiaodan Gu,<sup>c</sup> Yan Xia,<sup>b</sup> Wei-Ren Chen<sup>a</sup> and Changwoo Do<sup>\*,a</sup>

Ladder polymers consisting of fused rings in the backbone have very limited conformational freedom, which results in very different properties from traditional linear polymers. However, accurately determining their size and chain conformations from solution scattering remains a challenge. Their chain conformations of kinked ladder polymers are largely governed by the structures and relative orientations of the repeat units, unlike conventional polymer chains whose bending angles between repeat units follow a unimodal Gaussian distribution. Meanwhile, traditional scattering models for polymer chains do not account for these unique structural features. This work introduces a novel approach that integrates machine learning with Monte Carlo simulations to construct a model that can describe the geometry of a type of kinked CANAL ladder polymers. We first develop a Monte Carlo simulation model for sampling the configuration space of CANAL ladder polymers, where each repeat unit is modeled as a biaxial segment. Then, we establish a machine learning-assisted scattering analysis framework based on Gaussian Process Regression. Finally, we conduct small-angle neutron scattering experiments on a CANAL ladder polymer solution to apply our approach. Our method uncovers structural features of such ladder polymers that conventional methods fail to capture.

Received 3rd February 2025  
Accepted 7th May 2025

DOI: 10.1039/d5dd00051c

rsc.li/digitaldiscovery

## 1 Introduction

Ladder polymers are a unique polymer architecture, consisting of continuously fused, conformationally restrictive rings.<sup>1–3</sup> Such unique backbone structures result in many different properties from traditional linear polymers, which are desired in a range of applications ranging from electronics to membrane separations.<sup>1–3</sup> In particular, non-conjugated ladder polymers are characteristic of their kinked rigid conformations.<sup>2,4</sup> Xia and coworkers have developed a type of kinked ladder polymer *via* catalytic arene norbornene annulation (CANAL) using norbornadiene and aryl dibromides as monomers.<sup>5</sup> The rigid and frequently kinked structures of CANAL polymers result in frustrated packing and high microporosity.<sup>5,6</sup> Certain CANAL polymer films can exhibit strong size sieving effect and remarkable performance in gas separations.<sup>6</sup> Probing the overall conformations and dimensional characteristics of CANAL ladder polymers is the first step toward understanding their macromolecular packing behavior. CANAL reaction results in norbornyl benzocyclobutene structures with exclusive *exo*-

configuration.<sup>5</sup> In CANAL polymerization, the bridge carbon of neighboring norbornyl units can orient to either the same or opposite side of the ladder chain, resulting in *syn* or *anti*-configuration, respectively, as shown in Fig. 1. The sequence and distribution of these configurations determine the overall ladder chain dimensions. Therefore, it is important to first develop a model that can describe the statistical distribution of the *syn* and *anti*-configurations in a CANAL ladder polymer chain.

Small angle scattering experiments,<sup>7</sup> including X-ray scattering<sup>8</sup> and neutron scattering<sup>9,10</sup> are often used to study the characteristics of polymer system, and to unveil the single polymer structure using dilute polymer solutions. The scattering data is often analyzed using various polymer models to extract the polymer parameters, *e.g.* contour length, radius of gyration and persistence length. However, traditional polymer models, such as Gaussian coils<sup>11</sup> or worm-like chains<sup>12</sup> are inadequate for capturing the distinctive features of ladder

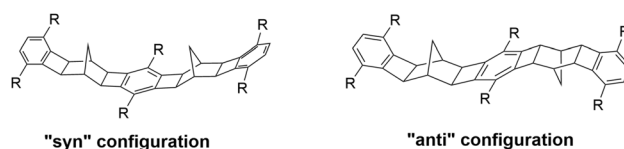


Fig. 1 *Syn* and *anti*-configurations of a representative CANAL ladder dimer.

<sup>a</sup>Neutron Scattering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. E-mail: doc1@ornl.gov

<sup>b</sup>Department of Chemistry, Stanford University, Stanford, CA 94305, USA

<sup>c</sup>School of Polymer Science and Engineering, Center for Optoelectronic Materials and Devices, The University of Southern Mississippi, Hattiesburg, MS 39406, USA

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d5dd00051c>



polymers since they are designed to model the single-stranded polymers and discard bending. These models do not fully represent the inherent rigidity and extended conformation of the ladder polymer, thus fail to provide an accurate depiction of the ladder polymers structure.

To overcome these challenges and provide an accurate description of the ladder polymer structure using scattering data, we build a new model for the simplest CANAL ladder polymer consisting only fused norbornyl and benzocyclobutene units, produced from norbornadiene and dibromo-*p*-xylene. This model accounts for the biaxial nature of its monomer structure and inherent rigidity. Due to the complexity of this model, it is difficult to derive the analytical form of the scattering function, which is typically required for fitting scattering data using traditional approaches. To address this, we leverage the power of Machine Learning (ML)<sup>13</sup> and Monte Carlo (MC)<sup>14</sup> simulations.

The recent advancements in ML have enabled numerous applications in materials science, including the analysis of scattering data<sup>15</sup> without knowing an explicit analytical form. This approach relies on large data sets that include scattering functions and corresponding polymer parameters, allowing ML to learn the relationship between them. Meanwhile, MC can be used to build such data sets. Given a set of polymer parameters, such as contour length and bending rigidity, we can use MC simulation to generate an ensemble of the polymer conformations and calculate the structure factor, or scattering function. This combination of ML and MC provides a powerful framework for analyzing complex polymer systems and has been proven useful for various single-stranded polymer systems<sup>16–19</sup> and other soft matter system.<sup>20</sup> Other works such as SCAN automates structural analysis using predefined particle shape models, while CREASE employs genetic algorithms and surrogate ML to reconstruct 3D features—such as domain size, shape, orientation, and spatial distributions—from scattering profiles.<sup>21–24</sup> Other ML approaches have been used for particle tracking in soft materials<sup>25</sup> and for surface scattering analysis.<sup>26</sup> Nevertheless, these methods do not provide insight for the model-specific parameters for systems like the ladder polymer and can not capture the unique structural nuances of such systems.

In this paper, we present a framework for analyzing the scattering data of ladder polymer using ML. We firstly introduce a model of the ladder polymer where the biaxiality, inherent rigidity and arrangement of successive monomers all play crucial role in determining the polymer conformation. We then carry out MC simulation to generate a large data set of the scattering data and train a ML model of Gaussian process regression<sup>27</sup> (GPR) to obtain the mapping between scattering data and polymer parameters. Finally, we synthesize ladder polymer samples and measure the scattering function using small-angle neutron scattering (SANS) experiment and apply our method to the extract important polymer parameters for the measured sample. In contrast to conventional Gaussian process-based data inversion approaches,<sup>28–30</sup> our approach avoid the potentially large computational cost in posterior sampling and predict each polymer parameters separately.

## 2 Model

To capture the ladder shape and biaxiality of the polymer, we model each monomer unit of the polymer as a rectangular segment whose orientation is specified by two unit vectors,  $\hat{\mathbf{u}}$  and  $\hat{\mathbf{v}}$ , where  $\hat{\mathbf{u}}$  is along the along axis of the segment, or the polymer tangent direction, and  $\hat{\mathbf{v}}$  is along the segment short axis and perpendicular to  $\hat{\mathbf{u}}$ . A polymer is then modeled as a chain of  $L$  segments, where  $L$  is the contour length in unit of monomer length  $B$ .

Unlike conventional polymer, the successive segments on the ladder polymer, *i.e.* catalytic arene-norbornene annulation (CANAL) polymer, tend to form an angle, as shown in Fig. 2(a and b). We introduce another two unit vectors,  $\hat{\mathbf{u}}'$  and  $\hat{\mathbf{v}}'$ , to represent this preferred orientation for the successive segment. For two connecting segments  $i$  and  $j$ , the angle between  $(\hat{\mathbf{u}}, \hat{\mathbf{v}})$  and  $(\hat{\mathbf{u}}', \hat{\mathbf{v}}')$  is the inherent bending and twisting. For this specific model we are concerned of, the  $\hat{\mathbf{v}} = \hat{\mathbf{v}}'$ , and we denote the inherent bending  $\cos(\alpha) = \hat{\mathbf{u}} \cdot \hat{\mathbf{u}}'$ . There is a energy cost when  $(\hat{\mathbf{u}}_{i+1}, \hat{\mathbf{v}}_{i+1})$  tilt away from  $(\hat{\mathbf{u}}'_i, \hat{\mathbf{v}}'_i)$ , given polymer energy  $E = \sum_i \frac{1}{2} K_t \phi_i^2 + \frac{1}{2} K_b \theta_i^2$ , where the bending is  $\cos(\theta_i) = \hat{\mathbf{u}}'_i \cdot \hat{\mathbf{u}}_{i+1}$ , twisting is  $\cos(\phi_i) = \hat{\mathbf{v}}'_i \cdot \hat{\mathbf{v}}_{i+1}$ , and  $K_t$  and  $K_b$  are the twisting and bending modulus, respectively.

Finally, the preferred orientation for successive segment at each segment may not stay on the same side. Comparing Fig. 2(a and b), when they stay on the same side, we call them

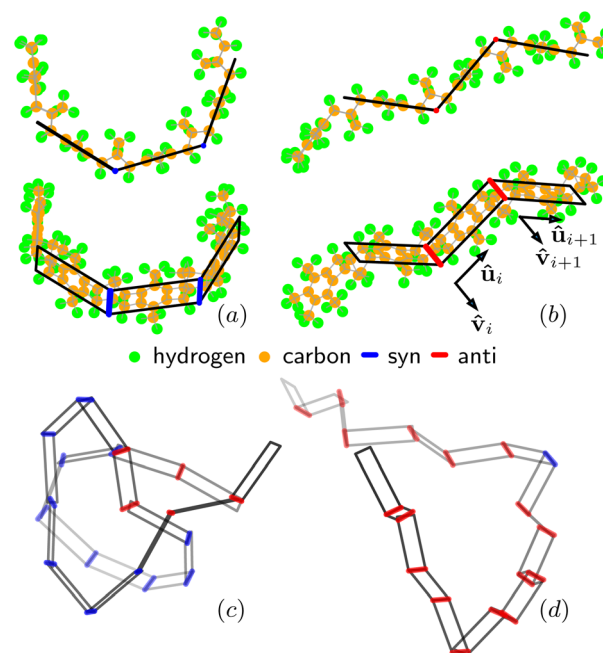


Fig. 2 Illustration of the segmentation model of CANAL ladder polymer. (a) Molecular structure of monomer segments connected through *syn* link, overlapped with rectangle used in our model, top and bottom two are the same polymer with different point of view. (b) Similar to (a), but with segments connected through *anti* link. (c) Monte Carlo generated polymer with low anti rate  $R_a = 0.1$  and (d) high anti rate  $R_a = 0.9$ .



connected by *syn* links, the polymer rolls up and become coil shape as shown in Fig. 2(c). On the contrary, if they flip side, or connected by *anti* links, the polymer tend to extend longer, as shown in Fig. 2(d). We define the probability of a link being a *anti* link as anti rate  $R_a$ .

Given a contour length  $L$ , inherent bending angle  $\alpha$ , anti rate  $R_a$ , bending modulus  $K_t$  and twisting modulus  $K_b$ , the ensemble of ladder polymer configuration is determined. The configuration can be captured by the intra-polymer structure factor, given by:<sup>7,9</sup>

$$S(QB) = \frac{1}{L^2} \sum_{i=1}^L \sum_{j=1}^L \frac{\sin(Q|\vec{r}_i - \vec{r}_j|)}{Q|\vec{r}_i - \vec{r}_j|} \quad (1)$$

where  $Q$  is the scattering vector and  $\vec{r}_i$  is the position vector of segment  $i$  and  $\vec{r}_{i+1} = \vec{r}_i + B\hat{u}_i$ . In addition, we also calculate the radius of gyration  $R_g^2 = \frac{1}{2} \langle (\vec{r}_i - \vec{r}_j)^2 \rangle_{ij}$ , with  $\langle \dots \rangle_{ij}$  denoting the average over all pairs of segments. We will use MC and ML to understand the relationship between structure factor  $S(QB)$  and other polymer parameters ( $R_a$ ,  $\alpha$ ,  $L$ ,  $R_g^2$ ,  $K_t$ ,  $K_b$ ).

## 3 Method

### 3.1 Synthesis of CANAL ladder polymer

To a flame-dried 15 mL glass pressure tube was added 1,4-dibromo-2,5-diethylbenzene<sup>31</sup> (584 mg, 2 mmol), Pd(OAc)<sub>2</sub> (9 mg, 0.04 mmol), PPh<sub>3</sub> (21 mg, 0.08 mmol) and butylated hydroxytoluene (1 mg). The tube was transferred into a nitrogen-filled glove box, and norbornadiene (220  $\mu$ L, 2.2 mmol), Cs<sub>2</sub>CO<sub>3</sub> (1.3 mg, 4 mmol) and THF (2 mL) was added. The tube was sealed with a Teflon cap and removed from the glovebox. The reaction mixture was heated to 150 °C for 24 h. The mixture was then cooled to room temperature and passed through Celite to remove inorganic salts. Chloroform (3  $\times$  5 mL) was used to wash the residue. The filtrate was concentrated and dissolved in a minimum amount of chloroform, which was then precipitated into methanol. The precipitated polymer was collected by centrifugation, washed with methanol, and dried under vacuum. The obtained polymer were fractionated using Soxhlet Extractor to generate low molecular weight polymer fraction (washed down from ethyl acetate) and high molecular weight polymer fraction (washed down from chloroform).

### 3.2 Small-angle neutron scattering

The extended Q-range small-angle neutron diffractometer (EQ-SANS) at the Spallation Neutron Source at the Oak Ridge National Laboratory was used to characterize the conformation of the ladder polymer.<sup>32,33</sup> Low molecular weight CANAL ladder polymer was dissolved in deuterated 1,2-dichlorobenzene at 5 mg mL<sup>-1</sup>. Two sample-to-detector distances (2.5 m and 4 m) were used with two wavelength bands defined by the minimum wavelength of  $\lambda_{\min} = 2.5$  Å and  $\lambda_{\min} = 10$  Å, respectively, to cover scattering wave vectors ranging from 0.006 to 0.5 Å<sup>-1</sup>. The choppers were operated at 60 Hz. The ladder polymer solution was loaded in the quartz cell of 2 mm path length and measured at 25 °C, 75 °C and 125 °C. The measured data were corrected by

detector sensitivity and background scattering from the empty cell and then converted into absolute scale intensities (cm<sup>-1</sup>) using a porous silica standard sample.<sup>34,35</sup> Finally, scattering from the solvent was subtracted before the data is scaled into the unit-less  $QB$  axis using the monomer length  $B$ .

### 3.3 Monte Carlo simulation

To calculate the intra-polymer structure factor of the polymer ensemble at various polymer parameters, we sample the configuration space of the polymer using direct sampling.<sup>14</sup> For a given set of ( $R_a$ ,  $\alpha$ ,  $L$ ,  $K_t$ ,  $K_b$ ), we generate 2000 polymer configurations and calculate the averaged structure factor  $S(QB)$ . The polymer configuration is determined by the link type  $l_i$ , relative bending and twisting angles  $\{(l_i, \theta_i, \phi_i)\}$ , letting  $l_i = 0$  represent *syn* link and  $l_i = 1$  denotes *anti* link, the  $l_i$  follows a Bernoulli distribution with probability  $P(l_i = 1) = R_a$ . In addition  $\theta_i$  and  $\phi_i$  follows the Gaussian distribution  $\theta_i \sim N(0, \sqrt{K_b/k_B T})$  and  $\phi_i \sim N(0, \sqrt{K_t/k_B T})$ , as they are independent in the polymer energy  $E = \sum_i \frac{1}{2} K_t \phi_i^2 + \frac{1}{2} K_b \theta_i^2$ , which follow the Boltzmann distribution  $P(E) \sim e^{-E/k_B T}$ . After sampling  $\{(l_i, \theta_i, \phi_i)\}$  for all segments based on their distribution, we calculate  $(\mathbf{u}_i, \mathbf{v}_i)$  and  $\vec{r}_{i+1} = \vec{r}_i + B\hat{u}_i$  of each polymer segments, then check the self-avoidance criteria  $|\vec{r}_i - \vec{r}_j| < 0.5B$  for all pairs of segments, only configurations satisfying these criteria are kept.

### 3.4 Gaussian process regression

Under the framework of GPR,<sup>27</sup> the goal is to obtain the posterior  $p(\mathbf{Y}_* | \mathbf{X}_*, \mathbf{X}, \mathbf{Y})$  of the function output  $\mathbf{y}$ , where  $\mathbf{X} = \{\ln S(QB)_{\text{train}}\}$ ,  $\mathbf{X}_* = \{\ln S(QB)_{\text{test}}\}$  are the training set and test set,  $\mathbf{Y}$  and  $\mathbf{Y}_*$  are the corresponding polymer parameters ( $R_a$ ,  $\alpha$ ,  $L$ ,  $R_g^2$ ). In our case, we use 70% of the data set  $F = \{\ln S(QB)\}$  as the training set, and the rest 30% as the test set. The joint distribution is for a Gaussian process is given by eqn (2)

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{Y}_* \end{pmatrix} \sim \mathcal{N} \left( \begin{bmatrix} m(\mathbf{X}) \\ m(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} k(\mathbf{X}, \mathbf{X}) & k(\mathbf{X}, \mathbf{X}_*) \\ k(\mathbf{X}_*, \mathbf{X}) & k(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix} \right) \quad (2)$$

where a constant prior mean  $m(\mathbf{x})$  and a linear combination of a Radial basis function (Gaussian) kernel and white noise for the kernel  $k(\mathbf{x}, \mathbf{x}') = e^{-\frac{|\mathbf{x}-\mathbf{x}'|^2}{2l}} + \sigma\delta(\mathbf{x}, \mathbf{x}')$  are used, in which  $l$  is the correlation length,  $\sigma$  is the variance of observational noise and  $\delta$  is the Kronecker delta function.

## 4 Results

We prepare the data set  $F = \{\ln S(QB)\}$  by generating conformations of ladder polymers using MC for 6000 random combination of ( $R_a$ ,  $\alpha$ ,  $L$ ,  $K_t$ ,  $K_b$ ) and calculate the corresponding  $R_g^2$  and  $S(QB)$ . The  $S(QB)$  are calculated for 100 different  $QB \in [0.07, 3]$ , such that the  $\ln QB$  grid is uniformly placed in this interval. The polymer parameters are sampled as  $R_a \sim U(0, 1)$ ,  $\alpha \sim \frac{\pi}{180} U(45, 60)$ ,  $L \sim U(4, 50)$ ,  $K_t \sim U(50, 100)$  and  $K_b \sim U(50, 100)$ , where  $U(a, b)$  is the uniform distribution in interval  $[a, b]$ .



In practice, these simulations are carried out in parallel on different CPUs, and each simulation takes up to half hour to complete. Natural units are used, such that length are in unit of segment or monomer length  $B$ , and energy are measured in unit of thermal noise  $k_B T$ . We firstly study the effect of polymer parameters on the structure factor, then validate the feasibility for ML inversion of each polymer parameter, train a GPR and test it using MC generated data. Finally, we carry out SANS experiment and applied the trained GPR to the experimentally obtained structure factor.

#### 4.1 Intra-polymer structure factor of the ladder polymer

While the polymer energy is only directly related to the bending modulus  $K_b$  and twisting modulus  $K_t$ , these modulus are relative large as the segments are connected by strong chemical bonds, leaving the major conformation change determined by the inherent bending angle  $\alpha$ , and anti rate  $R_a$ . These conformation change is captured by the structure factor. Fig. 3 shows the  $S(QB)$  at various contour length  $L$ , anti rate  $R_a$  and inherent bending angle  $\alpha$ . As shown in Fig. 3(a), increasing the contour length lead to rapid decrease of  $S(QB)$ , resulting from the extension of the polymer that increase the scattering at low  $Q$ . Fig. 3(b) shows that increasing anti rate  $R_a$  has similar effect of increasing  $L$ , as it also make the polymer extending longer. Increasing inherent bending angle  $\alpha$  make the opposite effect and increase the  $S(QB)$  as it effectively make the polymer more straight.

#### 4.2 Feasibility of machine learning inversion

To access the feasibility of using GPR to map the structure factor  $F = \{\ln S(QB)\}$  to polymer parameters  $Y = \{(R_a, \alpha, L, R_g^2, K_t, K_b)\}$ , following the similar ML inversion framework,<sup>15</sup> we carry out principle component analysis of  $6000 \times 100$  matrix  $F$ , by decomposing it into  $F = U\Sigma V^T$  using singular value decomposition (SVD), where  $U$ ,  $\Sigma$ , and  $V$  are matrices of  $6000 \times 6000$ ,  $6000 \times 100$ , and  $100 \times 100$  sizes, respectively.  $V$  is consist of the singular vectors, and the entries of  $\Sigma^2$  are proportional to the variance of the projection of  $F$  onto corresponding principal vectors in  $V$ .

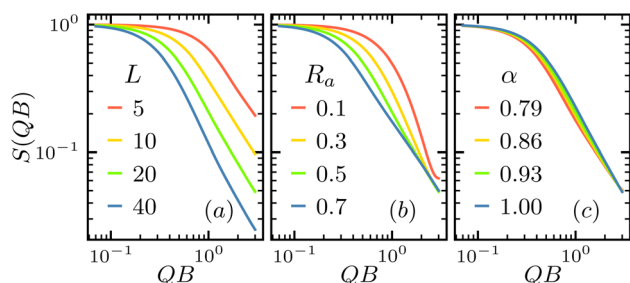


Fig. 3 Examples of simulated structure factor  $S(QB)$  versus scattering vector  $Q$  normalized by monomer length  $B$ , with  $K_t = K_b = 100$  at various contour length  $L$ , anti rate  $R_a$  and inherent bending angle  $\alpha$ . (a)  $S(QB)$  at various  $L$  with  $R_a = 0.5$  and  $\alpha = 0.93$  (or  $53.3^\circ$ ) (b)  $S(QB)$  at various  $R_a$  with  $L = 20$  and  $\alpha = 0.93$ . (c)  $S(QB)$  at various  $\alpha$  with  $L = 20$  and  $R_a = 0.5$ .

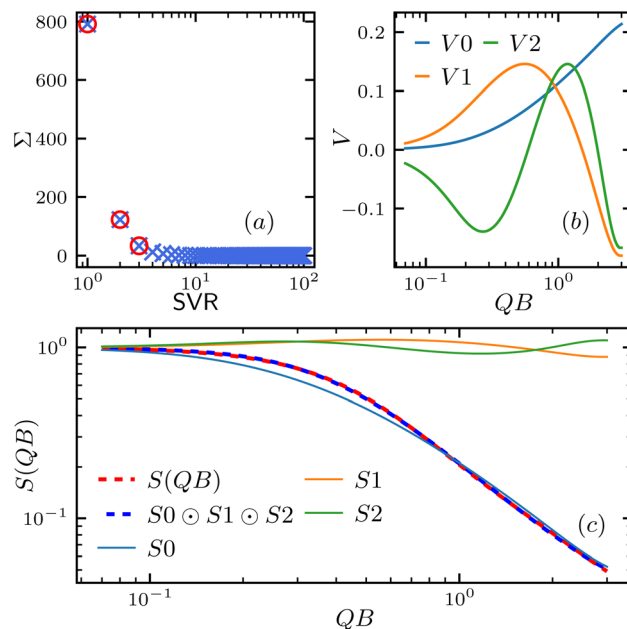


Fig. 4 Singular value decomposition (SVD) of scattering function data set  $F = \{\ln S(QB)\}$ . (a) Singular value  $\Sigma$  versus Singular Value Rank (SVR), value with top 3 rank are highlighted in red circle. (b) First 3 singular vectors  $V_0$ ,  $V_1$  and  $V_2$ . (c) Decomposition of  $\ln S(QB)$  with  $L = 20$ ,  $\alpha = 0.93$ ,  $R_a = 0.5$  and  $K_b = K_t = 100$ ,  $\ln S_0$ ,  $\ln S_1$  and  $\ln S_2$  are the projection of  $\ln S(QB)$  onto  $V_0$ ,  $V_1$  and  $V_2$ , respectively, e.g.  $S_0(QB) = \exp\{V_0(QB) \sum_Q \ln S(Q'B) V_0(Q'B)\}$ , and  $\odot$  denotes the Hadamard, or entrywise, product, i.e.  $(a \odot b)_i = a_i b_i$ .

As shown in Fig. 4(a), the singular value decays rapidly with its rank, suggesting the projecting  $\ln S(QB) \in F$  onto the space spanned by the high rank singular vectors manifest good approximation of the entire  $\ln S(QB)$ . Fig. 4(b) shows the first three singular vectors ( $V_1$ ,  $V_2$ ,  $V_3$ ), and Fig. 4(c) demonstrate the projection of  $\ln S(QB)$  on to these top 3 singular vectors do recover the original  $\ln S(QB)$  very well.

By projecting the  $F = \{\ln S(QB)\}$  onto the singular vector space of ( $V_0$ ,  $V_1$ ,  $V_2$ ), each  $\ln S(QB)$  become a coordinate in the three dimensional space,  $(FV_0, FV_1, FV_2)$ , and the entire set of coordinates provides a good proxy of the raw data set  $F$ . By plotting the distribution of polymer parameters  $Y$  in the  $(FV_0, FV_1, FV_2)$ , Fig. 5 provide insight for the feasibility of ML inversion of each of the polymer parameter, in which the corresponding value are represented by color distribution.

As shown in Fig. 5(a-d), the polymer parameters ( $R_a$ ,  $\alpha$ ,  $L$ ,  $R_g^2$ ) are well spread out in the  $(FV_0, FV_1, FV_2)$  space, indicating a good reversed mapping from  $\ln S(QB)$  to these parameters, indicating they are good inversion targets. On the contrary, Fig. 5(e and f) show that the distribution of the bending and twisting modulus  $K_b$  and  $K_t$  are rather random, suggesting there they can not be easily extract from the  $\ln S(QB)$ . This is in line with our expectation as the conformation of the ladder polymer is not sensitive to the wiggling around the inherent bending angle  $\alpha$  since  $\alpha$  is very large compare to the flexibility of the chemical bond.



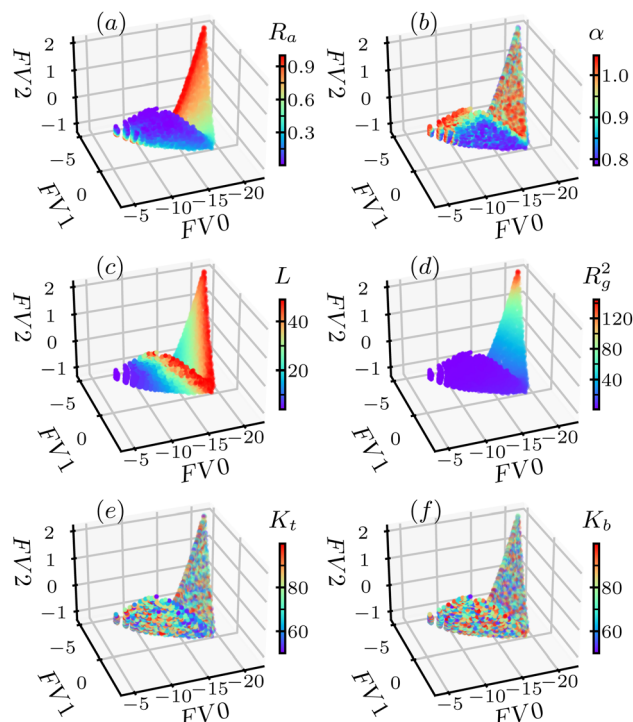


Fig. 5 Distribution of various inversion targets of data set  $F = \{\ln S(QB)\}$  projected into the singular value space ( $FV_0, FV_1, FV_2$ ) described by the first 3 singular vectors ( $V_0, V_1, V_2$ ). (a) Anti rate  $R_a$ . (b) Inherent bending angle  $\alpha$ . (c) Contour length  $L$ . (d) Radius of gyration square  $R_g^2$ . (e) Twisting modulus  $K_t$  and (f) bending modulus  $K_b$ .

Table 1 Optimized hyperparameters for each features, obtained from maximum log marginal likelihood

	$l$	$\sigma$
$R_a$	$6.497 \times 10^{-1}$	$6.301 \times 10^{-4}$
$\alpha$	$3.570 \times 10^{-1}$	$1.585 \times 10^{-2}$
$L$	1.043	$4.442 \times 10^{-3}$
$R_g^2$	3.843	$1.447 \times 10^{-7}$

### 4.3 Machine learning inversion of simulation data

With the feasibility for ML inversion for ( $R_a, \alpha, L, R_g^2$ ) from the  $\ln S(QB)$  established, we test such inversion using simulation data. We divide the data set  $F = \{\ln S(QB)\}$  randomly into two parts, a training set  $\{\ln S(QB)_{\text{train}}\}$  consisting 70% of  $F$  and a testing set  $\{\ln S(QB)_{\text{test}}\}$  made of the rest 30%. We optimize the hyperparameters of the GPR model using the training set for each polymer parameter and then extract the corresponding polymer parameters ( $R_a, \alpha, L, R_g^2$ ) from the  $\ln S(QB) \in \{\ln S(QB)_{\text{test}}\}$ . The scikit-learn Gaussian process library<sup>36</sup> was used for the training. Table 1 shows the optimized hyperparameters for each polymer parameters, obtained by maximizing the log marginal likelihood<sup>27</sup> as shown in Fig. 6.

Fig. 7 shows the comparison between the polymer parameters ( $R_a, \alpha, L, R_g^2$ ) obtained from ML inversion and the

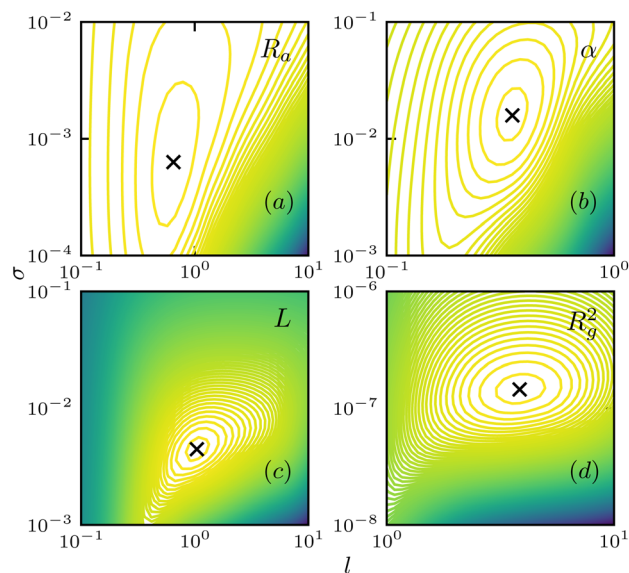


Fig. 6 log marginal likelihood surface of hyperparameters  $l$  and  $\sigma$  for various polymer parameters, with optimized value marked with black cross. (a) Anti rate  $R_a$ . (b) Inherent bending angle  $\alpha$ . (c) Contour length  $L$ . (d) Radius of gyration  $R_g^2$ .

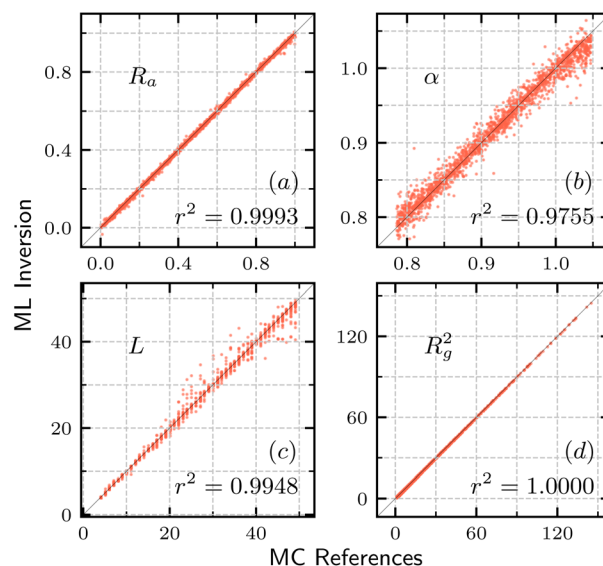


Fig. 7 Comparison of polymer parameters in simulation and inverted by machine learning. (a) Anti rate  $R_a$ . (b) Inherent bending angle  $\alpha$ . (c) Contour length  $L$ . (d) Radius of gyration  $R_g^2$ .

corresponding MC references. The data agree very well, and lie closely to the diagonal line, with coefficient of determination  $r^2$  score close to 1. The high precision highlights the effectiveness of extracting key parameters from the structure factor and further confirms the robustness of our GPR model. These results also indicate that, for our model, these polymer parameter can be extracted from the scattering curve independently, whereas there may need to be additional constraints in some cases, *e.g.* charged polymer.<sup>18</sup>



#### 4.4 Analysis of experimental measurement

To put our ML inversion model into practice, we synthesize CANAL ladder polymer and carry out small-angle neutron scattering (SANS) experiment to measure its form factor.

Fig. 8 shows normalized form factor measured from the SANS experiment and the ML implied curve. The SANS measured  $S(QB)$  shows good flat part in the low  $Q$  region in the log-log plot, allow us to fit for the normalization coefficient using Guinier approximation<sup>7,37</sup>  $S(QB) \sim e^{-(QR_g)^2/3}$ , and the monomer length  $B$  obtained by molecular structure optimization allow us to rescale the horizontal axis. By feeding the normalized experimental  $\ln S(QB)$  to the trained GPR, we obtain the polymer parameters ( $R_a$ ,  $\alpha$ ,  $L$ ,  $R_g^2$ ), as shown in Table 2, and then run MC simulation with these parameters to reconstruct the ML implied  $S(QB)$ . The SANS measured  $S(QB)$  and the ML implied one agree closely. As shown in Fig. 8, the black line, which reproduced using the mean value of the GPR predicted polymer parameter agrees with the experimental data very well, and the gray region are reproduced by taking the extreme of the error bar of each polymer parameter. We note that although the experimentally measured SANS data show high noise at low  $Q$  range, it is a common issue due to low neutron counts and since it is known that the polymer structure factor has the universal

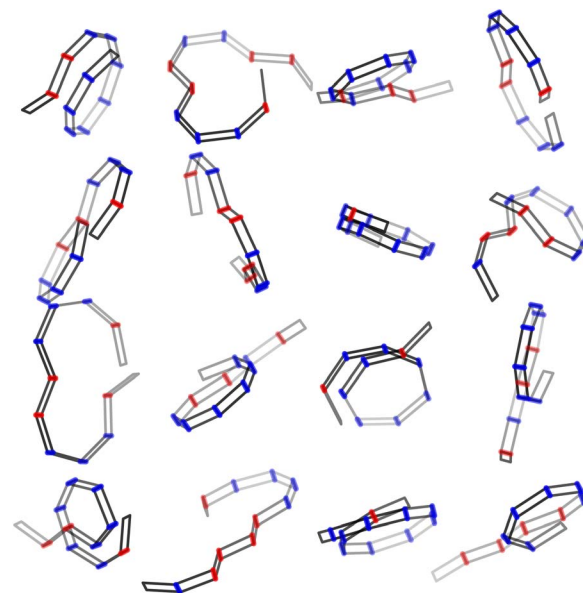


Fig. 9 Sample ladder polymer configurations generated using MC with  $(L, R_a, \alpha) = (12, 0.14, 0.89)$  and  $K_t = K_b = 100$ .

Guinier form at low  $Q$ , we firstly fit for the Guinier region<sup>37</sup> to find the normalization factor of the scattering density and replace the low  $Q$  data with smooth Guinier form before feeding in to the GPR. In addition, while the SANS measured  $S(QB)$  exhibit different noise levels at different  $Q$  due to neutron counting and instrument error, we only used the mean for the inversion as it is conventional for the SANS analysis, such impact can be minimized by taking longer and costly SANS experiment or use higher concentration of the sample to improve the signal. The current method cannot account for the error bar in the experimental data when applying the GPR, which can result in a larger uncertainty for the extracted parameters.

Table 2 shows the four GPR predicted polymer parameters of SANS our synthesized CANAL ladder polymer along with comparison with parameters obtained from other traditional methods. Note that our ML inversion method can extract all parameters simultaneously, and those parameters that the traditional method can extract, ( $\alpha$ ,  $L$ ,  $R_g^2$ ), show strong agreement with its results. Moreover, due to the special monomer structure of the CANAL ladder polymer, the anti rate  $R_a$  is

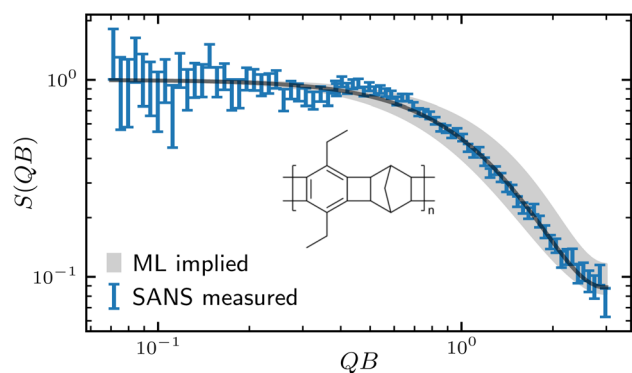


Fig. 8 Experimentally measured ladder polymer structure factor  $S(QB)$  using small-angle neutron scattering with  $Q$  normalized by  $B = 8.12 \text{ \AA}$ , and MC reconstructed  $S(QB)$  based on ML implied polymer parameters. The Dark gray line is calculated using  $(L, R_a, \alpha) = (12, 0.14, 0.89)$ , gray region indicate the uncertainty with upper bound calculated using  $(L, R_a, \alpha) = (9, 0.07, 0.94)$ , and lower bound using  $(L, R_a, \alpha) = (14, 0.21, 0.84)$ . Insert is the chemical structure of the CANAL ladder polymer.

Table 2 Comparison of ladder polymer structure parameters extracted from scattering function using machine learning and other traditional methods

	$R_a$	$\alpha$	$L$	$R_g^2$
Machine learning inversion	$0.14 \pm 0.07$	$0.89 \pm 0.05$	$11.6 \pm 2.8$	$2.07 \pm 0.28$
Molecular structure optimization <sup>a38</sup>	N/A	$0.96 \pm 0.08$	N/A	N/A
Flexible cylinder fitting <sup>12,39</sup>	N/A	N/A	$12.0 \pm 0.8$	N/A
Guinier approximation fitting <sup>7,37</sup>	N/A	N/A	N/A	$1.82 \pm 0.15$

<sup>a</sup> The atomistic structure of the ladder polymer with 4 monomer units were optimized using the Forcite Module with COMPASS force field in Materials Studio 8.0, BIOVIA.



a unique parameter that only can be obtained using the ML inversion method. The ML inversion method suggest our sample is relative short, with only about 12 segments, and the radius of gyration is even just  $R_g^2 \approx 2$ , fairly small for such contour length comparing to semiflexible chains.<sup>40</sup> This discrepancy is explained by the low anti rate  $R_a \approx 0.14$ , which suggesting the monomers are most connected through *syn* link, making the polymer roll up, as shown in Fig. 9. The tendency to have more coiling structure of ladder polymers has also been observed from other systems.<sup>41</sup>

## 5 Conclusions

In this paper, we introduce an biaxial segmentation model for the ladder polymer, we use a ML inversion method to extract polymer parameters from the form factor of the polymer system, and apply such method on real scattering data of the CANAL ladder polymer. The segmentation model represent the ladder polymer as a chain of two dimensional rectangular segments whose orientation is given by two unit vectors corresponding to the long and short axis. We prepare a data set consisting of 6000 structure factor  $F = \{S(QB)\}$  and corresponding polymer parameters  $Y = \{(R_a, \alpha, L, R_g^2)\}$  including anti rate  $R_a$ , inherent bending angle  $\alpha$ , contour length  $L$ , and radius of gyration  $R_g^2$ . We train a GPR using part of data set as training set to achieve the mapping from  $F$  to  $Y$ , and show that the trained GPR achieves excellent mapping when applied on the rest of data set, *i.e.* test set. Given that, we apply the ML inversion analysis on real scattering data. We firstly synthesize a CANAL ladder polymer, and run SANS experiment for a dilute sample. We normalize the SANS measured  $S(QB)$  and feed it into the trained GPR. All four polymer parameters are successfully extracted and the consistent with other traditional method when applicable. The anti rate  $R_a$  is extracted from the scattering data for the first time, providing new insight for the understanding of ladder polymer. While in this work we used the GPR to achieve the inversion from scattering to polymer parameter due to the simplicity and interpretability of GPR, there are alternative approaches utilizing neural network<sup>19,42,43</sup> can also be applied to our system.

Using the ML extracted polymer parameters, we can regenerate sample configurations using MC. It is expected that the CANAL ladder polymer sample we synthesized roll up to a coil or ring shape due to it's low anti rate. Further studies on single polymer imaging using scanning tunneling microscope<sup>44</sup> (STM) or ultra resolution atomic force microscopy<sup>45</sup> (AFM) would be highly beneficial. Moreover, the sample we used in this work only have inherent bending, application of this ML inversion method for other CANAL ladder polymers with both inherent bending and twisting can also be carried out in the future.

We also note that the CANAL ladder polymer structure we studied is dominated by the inherent bending angle and anti rate, the effect of bending modulus  $K_b$  and twisting modulus  $K_t$  are too weak to be extracted for this system. For the study of these  $K_b$  and  $K_t$ , ladder polymer whose monomers are connected in a flat manner are more suitable, as well as conjugated

polymer<sup>46,47</sup> whose twisting can be more significant due to the existent of single bond.

## Data availability

The code and data for this work are available at the GitHub repository [https://github.com/ljding94/Ladder\\_Polymer](https://github.com/ljding94/Ladder_Polymer) (<https://doi.org/10.5281/zenodo.15318836>) and Hugging Face repository [https://huggingface.co/datasets/ljding94/Ladder\\_Polymer](https://huggingface.co/datasets/ljding94/Ladder_Polymer) (<https://doi.org/10.57967/hf/5316>), respectively. The provided code was developed for internal use by the author to generate the results in this study. It is not optimized for general user-friendliness. For support or clarification on its implementation please contact the author.

## Author contributions

L. D. designed the model, carried out the MC simulation and ML analysis and draft the manuscript. C. H. T discussed the model and results, reviewed the manuscript. Z. C measured the SANS spectrum, discussed the results and reviewed the manuscript, Z. Y. synthesized the material, discussed the results and reviewed the manuscript, X. G. discussed the results and reviewed the manuscript. Y. X discussed the results and reviewed the manuscript. W. R. C discussed the results and reviewed the manuscript. C. D. conceived this work, discussed the model and results, and draft the manuscript.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This research was performed at the Spallation Neutron Source and the Center for Nanophase Materials Sciences, which are DOE Office of Science User Facilities operated by Oak Ridge National Laboratory. This research was sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U. S. Department of Energy. Monte Carlo simulations and computations used resources of the Oak Ridge Leadership Computing Facility, which is supported by the DOE Office of Science under Contract DE-AC05-00OR22725. Z. C. and X. G. acknowledge partial support from the U.S. Department of Energy under Award number DE-SC0022050 for the scattering experiments conducted in this work.

## Notes and references

- 1 Y. Xia, M. Yamaguchi and T.-Y. Luh, *Ladder Polymers: Synthesis, Properties, Applications and Perspectives*, John Wiley & Sons, 2023.
- 2 Y. C. Teo, H. W. Lai and Y. Xia, *Chem.-Eur. J.*, 2017, **23**, 14101–14112.
- 3 J. S.-J. Yang and L. Fang, *Chem*, 2024, **10**(6), 1668–1724.



- 4 N. B. McKeown, P. M. Budd, K. J. Msayib, B. S. Ghanem, H. J. Kingston, C. E. Tattershall, S. Makhseed, K. J. Reynolds and D. Fritsch, *Chem.–Eur. J.*, 2005, **11**, 2610–2620.
- 5 S. Liu, Z. Jin, Y. C. Teo and Y. Xia, *J. Am. Chem. Soc.*, 2014, **136**, 17434–17437.
- 6 H. W. Lai, F. M. Benedetti, J. M. Ahn, A. M. Robinson, Y. Wang, I. Pinnau, Z. P. Smith and Y. Xia, *Science*, 2022, **375**, 1390–1392.
- 7 P. Lindner and T. Zemb, *Neutrons, X-rays and Light: Scattering Methods Applied to Soft Condensed Matter*, Elsevier Science, 2002.
- 8 B. Chu and B. S. Hsiao, *Chem. Rev.*, 2001, **101**, 1727–1762.
- 9 S.-H. Chen, *Annu. Rev. Phys. Chem.*, 1986, **37**, 351–399.
- 10 M. Shibayama, *Polym. J.*, 2011, **43**, 18–34.
- 11 P. Debye, *J. Phys. Chem.*, 1947, **51**, 18–32.
- 12 J. S. Pedersen and P. Schurtenberger, *Macromolecules*, 1996, **29**, 7602–7612.
- 13 K. P. Murphy, *Machine learning: a probabilistic perspective*, MIT press, 2012.
- 14 W. Krauth, *Statistical mechanics: algorithms and computations*, OUP Oxford, 2006, vol. 13.
- 15 M.-C. Chang, C.-H. Tung, S.-Y. Chang, J. M. Carrillo, Y. Wang, B. G. Sumpter, G.-R. Huang, C. Do and W.-R. Chen, *Commun. Phys.*, 2022, **5**, 46.
- 16 C.-H. Tung, S.-Y. Chang, H.-L. Chen, Y. Wang, K. Hong, J. M. Carrillo, B. G. Sumpter, Y. Shinohara, C. Do and W.-R. Chen, *J. Chem. Phys.*, 2022, **156**, 13.
- 17 L. Ding, C.-H. Tung, B. G. Sumpter, W.-R. Chen and C. Do, *arXiv*, 2024, preprint, arXiv:2410.05574, DOI: [10.48550/arXiv.2410.05574](https://doi.org/10.48550/arXiv.2410.05574).
- 18 L. Ding, C.-H. Tung, J.-M. Y. Carrillo, W.-R. Chen and C. Do, *arXiv*, 2025, preprint, arXiv:2501.14647, DOI: [10.48550/arXiv.2501.14647](https://doi.org/10.48550/arXiv.2501.14647).
- 19 L. Ding, C.-H. Tung, B. G. Sumpter, W.-R. Chen and C. Do, *J. Chem. Theory Comput.*, 2025, **21**, 4176–4182.
- 20 L. Ding, Y. Chen and C. Do, *arXiv*, 2024, preprint, arXiv:2412.07926, DOI: [10.48550/arXiv.2412.07926](https://doi.org/10.48550/arXiv.2412.07926).
- 21 P. Tomaszewski, S. Yu, M. Borg and J. Rönnols, *2021 Swedish Workshop on Data Science (SweDS)*, 2021, pp. 1–6.
- 22 A. S. Anker, K. T. Butler, R. Selvan and K. M. Jensen, *Chem. Sci.*, 2023, **14**, 14003–14019.
- 23 S. Lu and A. Jayaraman, *Prog. Polym. Sci.*, 2024, 101828.
- 24 S. V. R. Akepati, N. Gupta and A. Jayaraman, *JACS Au*, 2024, **4**, 1570–1582.
- 25 P. S. Clegg, *Soft Matter*, 2021, **17**, 3991–4005.
- 26 A. Hinderhofer, A. Greco, V. Starostin, V. Munteanu, L. Pithan, A. Gerlach and F. Schreiber, *J. Appl. Crystallogr.*, 2023, **56**, 3–11.
- 27 C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, MIT press, Cambridge, MA, 2006, vol. 2.
- 28 M. Gu and L. Wang, *SIAM/ASA J. Uncertain. Quantification*, 2018, **6**, 1555–1583.
- 29 M. C. Kennedy and A. O'Hagan, *J. Roy. Stat. Soc. B Stat. Methodol.*, 2001, **63**, 425–464.
- 30 A. M. Stuart, *Acta Numer.*, 2010, **19**, 451–559.
- 31 K. Sugamata, S. Kobayashi, T. Iihama and M. Minoura, *Eur. J. Inorg. Chem.*, 2021, **2021**, 3185–3190.
- 32 J. K. Zhao, C. Y. Gao and D. Liu, *J. Appl. Crystallogr.*, 2010, **43**, 1068–1077.
- 33 W. T. Heller, M. Cuneo, L. Debeer-Schmitt, C. Do, L. He, L. Heroux, K. Littrell, S. V. Pingali, S. Qian, C. Stanley, V. S. Urban, B. Wu, W. Bras and IUCr, *J. Appl. Crystallogr.*, 2018, **51**, 242–248.
- 34 O. Arnold, J.-C. Bilheux, J. Borreguero, A. Buts, S. I. Campbell, L. Chapon, M. Doucet, N. Draper, R. F. Leal, M. Gigg, *et al.*, *Nucl. Instrum. Methods Phys. Res., Sect. A*, 2014, **764**, 156–166.
- 35 W. T. Heller, J. Hetrick, J. Bilheux, J. M. B. Calvo, W.-R. Chen, L. DeBeer-Schmitt, C. Do, M. Doucet, M. R. Fitzsimmons, W. F. Godoy, G. E. Granroth, S. Hahn, L. He, F. Islam, J. Lin, K. C. Littrell, M. McDonnell, J. McGaha, P. F. Peterson, S. V. Pingali, S. Qian, A. T. Savici, Y. Shang, C. B. Stanley, V. S. Urban, R. E. Whitfield, C. Zhang, W. Zhou, J. J. Billings, M. J. Cuneo, R. M. F. Leal, T. Wang and B. Wu, *SoftwareX*, 2022, **19**, 101101.
- 36 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.
- 37 A. Guinier, G. Fournet, C. B. Walker and K. L. Yudowitch, *Small-angle Scattering of X-rays*, Wiley, New York, 1955.
- 38 U. Shankar, R. Gogoi, S. K. Sethi and A. Verma, in *Forcefields for atomistic-scale simulations: materials and applications*, Springer, 2022, pp. 299–313.
- 39 W.-R. Chen, P. D. Butler and L. J. Magid, *Langmuir*, 2006, **22**, 6539–6548.
- 40 L. Ding, C.-H. Tung, B. G. Sumpter, W.-R. Chen and C. Do, *J. Chem. Theory Comput.*, 2024, **20**, 10697–10702.
- 41 T. Ikai, T. Yoshida, K.-i. Shinohara, T. Taniguchi, Y. Wada and T. M. Swager, *J. Am. Chem. Soc.*, 2019, **141**, 4696–4703.
- 42 C.-H. Tung, L. Ding, G.-R. Huang, L. Porcar, Y. Shinohara, B. G. Sumpter, C. Do and W.-R. Chen, *Appl. Crystallogr.*, 2025, **58**, 523–534.
- 43 C.-H. Tung, L. Ding, M.-C. Chang, G.-R. Huang, L. Porcar, Y. Wang, J.-M. Y. Carrillo, B. G. Sumpter, Y. Shinohara, C. Do and W.-R. Chen, *J. Chem. Phys.*, 2025, **162**(7), 074106.
- 44 G. Binnig and H. Rohrer, *Rev. Mod. Phys.*, 1987, **59**, 615.
- 45 F. J. Giessibl, *Rev. Mod. Phys.*, 2003, **75**, 949.
- 46 X. Yin, K. Zheng, Z. Jin, M. Horst and Y. Xia, *J. Am. Chem. Soc.*, 2022, **144**, 12715–12724.
- 47 Z. Cao, S. A. Tolba, Z. Li, G. T. Mason, Y. Wang, C. Do, S. Rondeau-Gagné, W. Xia and X. Gu, *Adv. Mater.*, 2023, **35**, 2302178.

