Estimation of Molecular Structures of Polymers Using Machine Learning

Sumitomo Chemical Co., Ltd.
Essential Chemicals Research Laboratory
Seiji SHIROMOTO
Hiroyuki KUDO
Akinori BANDO

The molecular structure of polymers was estimated using experimental values of viscoelasticity and molecular-weight distribution. This estimation was achieved through a combination of a viscoelastic simulation and machine learning with Bayesian estimation. The sample utilized was high-pressure low-density polyethylene with branching chains. Using the methods described before, we obtained a molecular structure that accurately reproduced the experimental values. The derived molecular structure was then compared with reports of direct observation of a single molecular chain. The comparison confirmed a general correspondence between the number of branched chains per molecule and the ratio of branching chain length to backbone chain length.

Introduction

Melt viscoelasticity significantly impacts polymer processing, and hence, understanding the relationship between molecular structure and viscoelasticity is crucial. Viscoelasticity depends on the molecular structure, including the chemical species that constitute the polymer, molecular shape, molecular weight, and molecular-weight distribution. The difference in viscoelasticity due to molecular structure is particularly noticeable between linear and branched chains, leading to substantial differences in polymer processability. For instance, linear low-density polyethylene (LLDPE), which has a relatively narrow molecular-weight distribution, is less likely to break during high-speed processing; however, its high melt tension at high strain rates imposes a high load on the processing machine. Furthermore, the absence of strain hardening during flow results in compromising processing stability of film processing. By contrast, the viscoelasticity and processability of high-pressure low-density polyethylene (LDPE), which has long-chain branches, exhibit opposite tendencies to those of LLDPE. Therefore, although LLDPE and LDPE possess the same chemical species, they exhibit substantially different viscoelasticity and processability, which are attributable to their distinct molecular structures.

A considerable number of theoretical and experimental studies on the melt viscoelasticity of branched polymers have been reported[1]–[3]. These theoretical models accurately predict the viscoelasticity of polymers that have been precisely polymerized and have specific molecular shapes. Das et al.[4],[5] developed a simulator called bob-rheology[6] based on the viscoelastic theory. This simulator is a computational tool that enables the swift prediction of viscoelastic behavior from molecular structures with reduced computational requirements. van Gurp and Palmen[7] proposed a method (vGP plot) of using linear viscoelasticity to estimate the presence of branched structures. They showed that the presence of branched chains can be inferred from the relationship between the complex modulus $|G^*|$, which is the response of a polymer to an external strain, and the phase δ.

With recent advances in artificial intelligence and machine learning, the focus in material development has increasingly turned toward the active application of data science in the field of material development. Data science is applied in high-throughput screening based on performance prediction as well as in improving data processing. For example, the structural analysis of powder crystals, with unknown crystal structures and similar X-ray diffraction profiles that are not registered in the database, is challenging even for

1

experienced researchers. Lee *et al.*[8] conducted inverse analysis from a measured X-ray diffraction profile by combining simulation, evolutionary algorithms, and Bayesian optimization; they reported the generation of a crystal structure that accurately reproduced the X-ray diffraction profile.

In this study, we first provide an overview of simulation techniques and machine learning methods related to the viscoelastic properties of polymers. Next, we detail the characteristics of our analysis technique, which combines simulation and Bayesian optimization. Finally, we demonstrate the application of this method in the estimation of the primary structure of polymers with long-chain branching.
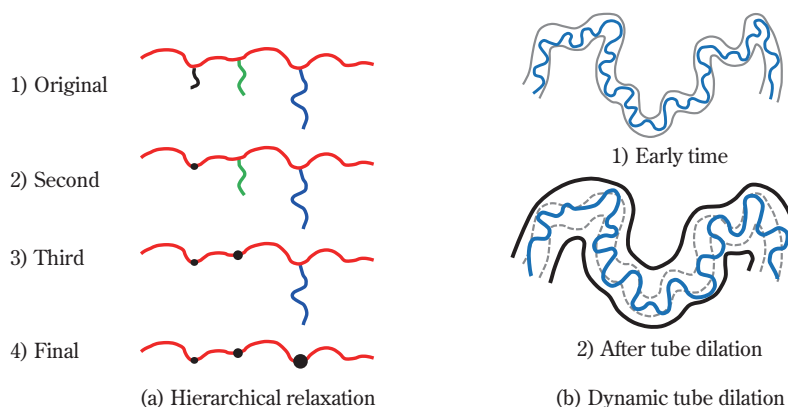
Polymer rheology

1. Relaxation of polymer chains

The molecular dynamics of polymer chains are described by two fundamental models: the Rouse model for nonentangled systems and the tube model for entangled systems. The Rouse model is applied to describe isolated polymer chains, whereas the tube model can be applied to polymer melts with entanglement interactions, where the excluded volume effect predominates and hydrodynamic interactions are comparatively trivial. Since the molecular weight of the sample used in this study is sufficiently higher than that between the entanglement points, the tube model is applicable. The linear polymer relaxation behavior has been established based on the theoretical framework proposed by de Gennes[9] and the tube model developed by Doi and Edwards[10]−[14]. The tube model stipulates that polymer chains are constrained within a surrounding tubular domain, undergoing

conformational adjustments to applied strain through immediate local movements via contraction motions on a short timescale and reptation motion on a long timescale.

Certain polymers, such as LDPE, feature branched chain architecture. Studies suggest that the minimum branching chain length (M_a), at which the rheological effects of branching are observed, is more than twice the molecular weight (M_e) between entanglement points, with the effects becoming pronounced when the length is 5–6 times greater[15]−[17]. It is reported that the molecular weight between the entanglement points of polymer chains depends on the chemical species and is equivalent to that of linear chains, irrespective of the branching polymer chains[18]−[21]. The relaxation of polymer chains with multiple branch points, such as H-branched, pom-pom-branched, and comb-branched structures, can be described by a combination of hierarchical relaxation[1]−[3],[22]−[25] and dynamic tube dilation[26]−[31].

The hierarchical relaxation mechanism for branched polymers is illustrated in **Fig. 1**(a) using a comb-branched polymer with multiple branched chains of different lengths as an example. Initially, the short-length branched chains relax by retraction into the tube surrounding the backbone chain (arm retraction, AR). The relaxed branched chains do not intertwine with other chains, thus acting as friction points at the branch points. AR relaxation occurs sequentially from the shortest branched chains, and in the final stage, the backbone chain relaxes as if it were a straight chain with multiple friction points. This hierarchical relaxation model has been refined to take into account the short-term relaxation of branched chains[2]. Subsequently, Das *et al.* proposed the branch-on-branch

1) Original

2) Second

3) Third

4) Final

(a) Hierarchical relaxation

1) Early time

2) After tube dilation

(b) Dynamic tube dilation

Fig. 1   Conceptualization of algorithm for hierarchical relaxation of a comb-branched polymer

model, which can be used to analyze multistage branching[4),5)] and aligns well with the viscoelastic analysis results of comb-branched polymers[32)]. Similar theoretical models for the dynamics of multibranched chains have been successful in replicating experimental values[3),24),26)].

The concept of dynamic tube dilation proposed by Marrucci[26)] is depicted in **Fig. 1**(b). This process involves the progressive relaxation of the polymer chain of interest, which is accompanied by relaxation of neighboring chains, leading to a decrease in the number of entanglement points with surrounding chains. Consequently, the confining tube around the chain dilates, exhibiting a dynamic tube dilation mechanism. Once the branched chains have relaxed, the backbone chain behaves as a straight chain, with the branched chains acting as friction points, and undergoes reptation relaxation in the dilated tube. The dynamic tube dilation theory has also been extended to branched chains[27),33)] and polydisperse systems[30),34)], with modifications to account for the re-equilibration of the chain tension[31)].

2. Stress relaxation simulation

We conducted stress relaxation simulations to investigate the viscoelasticity of branched polymers. The longest relaxation time of a polymer is orders of magnitude longer than the local motion time of a molecular chain; thus, simulations based on all-atom models are computationally infeasible owing to the extensive computational demand. The motion of the entire polymer chain is defined by two parameters: entanglement length (tube diameter a and molecular weight between entanglement points M_e) and relaxation time or molecular chain friction coefficient ζ.

Even if the chemical species of the polymer chain is the same, the primary structure of the chain (*e.g.*, straight chain/branched chain) significantly impacts the motion of the chain.

The viscoelastic simulator Naples, developed by Masubuchi *et al.*, was used for the large deformation stress relaxation simulation[35)]. Polyisoprene (PI), reported by Kirkwood *et al.*, was used as a reference sample[36)]. Kirkwood *et al.* obtained PI with a narrow molecular-weight distribution and long-chain branches by precise polymerization using macromonomers (PI-C). For comparison, a linear PI (PI-L) with the same molecular weight of the backbone chain was modeled and used. **Table 1** lists the characteristics of the sample, including the molecular weight of the chain (M), molecular weight of the backbone chain (M_b), molecular weight of the branched chain (D_a), molecular-weight dispersity of the backbone chain (D_b), molecular-weight dispersity of the branched chain (D_a), number of branched chains per molecule (q), molecular-weight fraction of the backbone chain (ϕ_b), and molecular-weight fraction of the branched chain (ϕ_a).

The relaxation modulus $G(t,\gamma)$ after applying a step state deformation to a polymer is defined by Eq. (1):

$$G(t,\gamma) = \sigma(t)/\gamma \qquad (1)$$

Hear σ is stress, t is time, and γ is strain. **Fig. 2** exhibits the relaxation modulus obtained when a step state shear strain γ in the range of 0.5–5 is applied; here, τ_R is the Rouse relaxation time, τ_k is the characteristic relaxation time, and τ_d is the longest relaxation time. Specifically, τ_R is associated with the contraction relaxation time of the chains oriented and stretched by deformation, τ_k corresponds to the relaxation time from Rouse relaxation to reptation relaxation, and τ_d represents the relaxation time for the chain morphology to revert to the equilibrium state before the strain is applied. Over time, the relaxation modulus of PI-L decreases with increasing applied strain. In the linear region where $\gamma < 1$, the chains are oriented but not stretched. Thus, almost no reduction in the number of

Table 1 Molecular characteristics of the materials used for the viscoelastic simulation

Code	M	M_b	M_a	D_b	D_a	q	ϕ_b	ϕ_a
PI-L	85,100	–	–	1.05	–	0	1.0	0.0
PI-C	132,000	85,100	10,200	1.05	1.05	4.6	0.64	0.36
PI-Ma2	179,000	85,100	20,400	1.05	1.05	4.6	0.48	0.52
PI-Ma3	226,000	85,100	30,600	1.05	1.05	4.6	0.38	0.62
PI-q8	173,000	85,100	10,200	1.05	1.05	8.6	0.49	0.51
PI-q12	214,000	85,100	10,200	1.05	1.05	12.6	0.40	0.60

     3
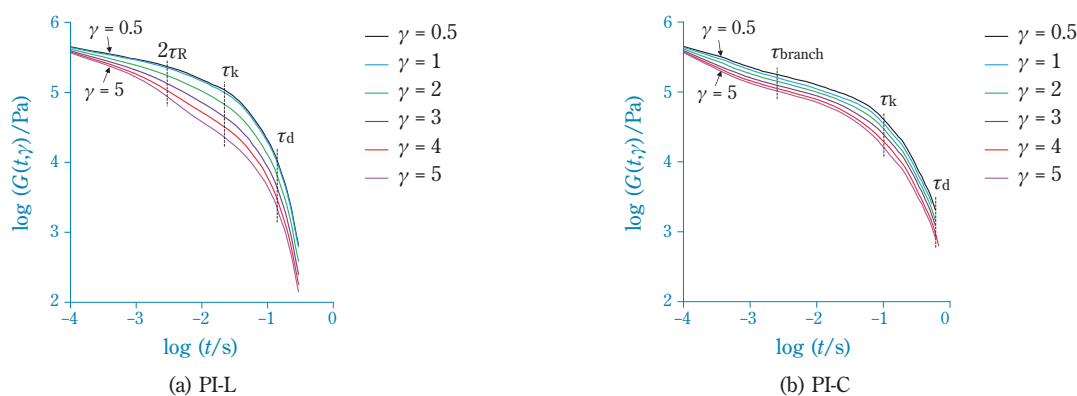
Fig. 2 Relaxation modulus after large-scale step shear strain

entanglements due to chain contraction is observed, and the response remains in the linear region. In the nonlinear region, where $\gamma > 1$, the stretched chains contract due to Rouse relaxation, leading to a decrease in the number of entanglements. The number of entanglements per chain supporting the stress decreases, and consequently, the relaxation modulus decreases with increasing applied strain.

The relaxation modulus of PI-C initially decreases gradually over time and shows a sharp decrease after a certain period, indicating a two-step relaxation behavior. The relaxation behavior on the short timescale corresponds to the contraction relaxation of the branched chains, with τ_{branch} representing the relaxation time of the branched chains. On a longer timescale, the relaxation behavior corresponds to the contraction relaxation of the backbone chain, between branch points, and reptation relaxation. Therefore, the reduction in the relaxation modulus of PI-C is slower than that of PI-L, resulting in a more extended τ_d. Additionally, the extent of the decrease in the modulus due to an increased strain is less pronounced in PI-C.

Fig. 3 depicts the relaxation modulus $G(t,\gamma)/G(t,0)$, which is nondimensionalized by the relaxation modulus $G(t,0)$ in the linear region. In the case of PI-L, the entire chain contracts and relaxes within approximately $2\tau_R$, and the uneven tension is eliminated. At this point, the number of entanglements per chain becomes minimal, and the elastic modulus decreases. Subsequently, by τ_k, constraints around the chain re-establish their equilibrium length. After τ_k, reptation relaxation occurs, and the number of entanglements returns to that in the equilibrium state.

The nondimensional elastic modulus at each strain remains constant, indicating that reptation is the dominant relaxation mechanism. Consequently, the

number of entanglements per chain decreases, and the reduction in the elastic modulus can be attributed to the contraction and relaxation of the chain under a given large deformation.

For PI-C, the nondimensional elastic modulus exhibits a two-stage decrease. The initial stage of stress relaxation occurs by τ_{branch} on the short timescale, during which the branched chains and terminal chains of the backbone chain contract and relax. The proportion of branched chains in the total is low, and the backbone chain between the branch points cannot move at this stage; therefore, there is only a marginal reduction in the elastic modulus. After τ_{branch}, the branched chains undergo arm retraction (AR) relaxation similar to that in star-branched chains, leading to the equilibration of chain length and the dilation of the tube surrounding the entire chain. The branched chains complete their relaxation by τ_k, after which they behave as friction points at the branch points. After τ_k, the backbone chain undergoes reptation relaxation in the dilated tube. Therefore, branched polymers undergo hierarchical relaxation in the two stages, corresponding to the branched chains and backbone chain. The total relaxed stress due to the relaxation of the branched chains and backbone chain of a branched polymer is lower than that of a linear chain because the branched chains are shorter and have fewer entanglement points compared to linear chains. Furthermore, dynamic tube dilation occurs prior to the relaxation of the backbone chain, reducing the number of entanglement points with other chains[37]−[39].

3. vGP plot
(1) Theory

vGP plots are useful tools for understanding the

(a) PI-L

(b) PI-C
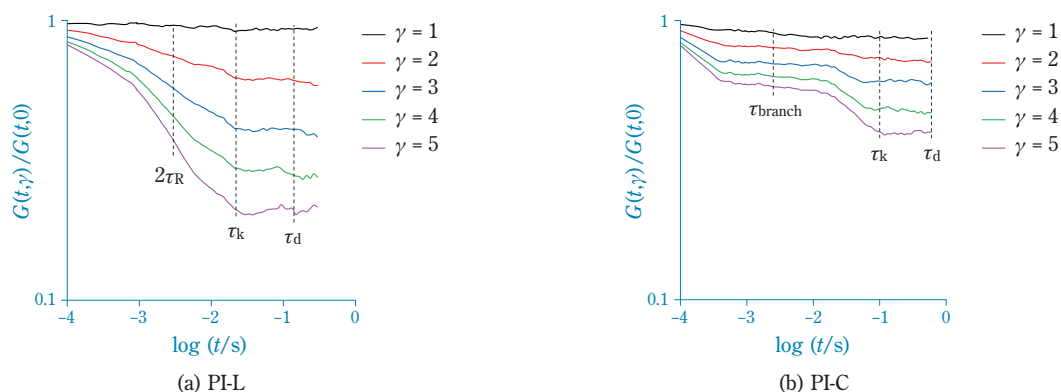
Fig. 3 Relaxation modulus normalized by the linear relaxation modulus

viscoelasticity of polymers. Under sinusoidal vibration, the stress in an elastic body is directly proportional to the strain according to Hooke's law, resulting in a stress in which the phase angle of the stress relative to the strain is 0°. In contrast, a viscous body follows Newton's law; the stress depends on the strain rate, which is the time derivative of the strain and has a lag phase of $\pi/2$ relative to the strain. Therefore, the phase of the stress of a viscous body lags behind the strain by $\pi/2$. In other words, the response phase of an elastic body to an external deformation is 0°, and the response phase of a viscous body is 90°.

van Gurp and Palmen focused on the difference between the responses of an elastic body and a viscous body to an externally applied strain[7]. The relationship between the molecular structure and viscoelasticity was discussed from the relationship between the complex modulus $|G^*|$, which is the response of the material, and the phase δ (vGP plot), with δ and $|G^*|$ given by the following equations:

$$\delta = \tan^{-1}(G''/G') \tag{2}$$

$$|G^*| = (G''^2 + G'^2)^{1/2} \tag{3}$$

Here, G' is the storage modulus, G'' is the loss modulus, and δ is the phase angle of the loss tangent $\tan\delta$. According to the temperature-time superposition principle, the high $|G^*|$ side correlates with a short timescale or low-temperature condition, indicating an elastic polymer response, with δ approaching 0°. Conversely, a low $|G^*|$ side correlates with a long timescale or high-temperature condition, suggesting a viscous polymer response with $\delta \approx 90°$. Consequently, the vGP plot is an index for evaluating the viscoelastic response of polymers. Reports have indicated that the

vGP plot can be employed to assess the influence of branched chains on viscoelastic properties[40]−[42].

(2) vGP plot simulation

The viscoelastic simulator bob-rheology[6] can calculate the dynamic viscoelasticity and uniaxial extensional viscosity of polymers under low loads and at high strain rates using parameters that depend on the chemical species and molecular architecture. There are no restrictions on molecular shape, molecular weight, or molecular-weight distribution, allowing it to model any polymer. However, this simulator is limited to a single chemical species and exclusively models homopolymers.

We investigated the impact of branching structure on viscoelasticity using the samples shown in **Table 1** and conducted viscoelasticity simulations using bob-rheology. Samples with different branching chain lengths (PI-Mb2, PI-Mb3) and different numbers of branched chains (PI-q8, PI-q12) were used for PI-C to confirm the effect of molecular shape on viscoelasticity. **Fig. 4** presents the vGP plots for each sample. The vGP plot for PI-L, a monodisperse linear chain, displayed a minimum value on the high $|G^*|$ side. This corresponds to the high-frequency region and represents the pseudo-equilibrium elastic modulus G_N^0. At the high $|G^*|$ side, $\delta \approx 0°$, reflecting elastic behavior, while lower $|G^*|$ corresponds to δ approaching 90°, where viscous behavior dominates.

Two distinct minimum values were observed in the vGP plot of the monodisperse branched chain PI-C. The minimum value on the high $|G^*|$ side corresponded to relaxation of the branched chain. In contrast, the minimum value G_δ on the low $|G^*|$ side was attributed to relaxation of the backbone chain. Consequently, whereas a single minimum was observed in the vGP
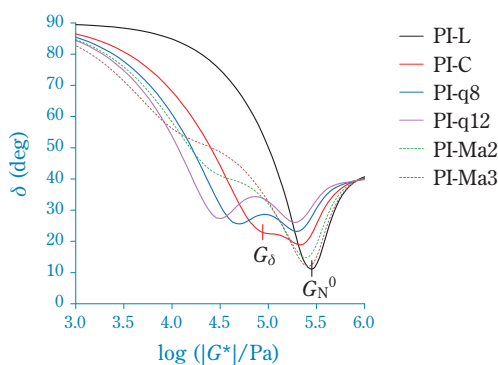
Fig. 4 Relationship between complex modulus $|G^*|$ and phase angle δ

plot for linear chains, multiple minima were observed for the branched chains.

We examined the impact of the number of branched chains q on the vGP plot (PI-C, PI-q8, PI-q12). As the number of branched chains increased, the minimum value G_δ shifted to the lower $|G^*|$ side. This shift may be attributed to the fact that the increased number of branched chains inhibited the relaxation of the entanglements with other chains, causing the minimum value to shift to the longer timescale. Additionally, the minimum value also shifted to the higher phase angle δ, suggesting that the extended relaxation time of the branched chains leads to dynamic tube dilation around the backbone chain, which consequently manifests as viscous behavior.

We then investigated the effect of the branch chain length M_a on the vGP plot (PI-C, PI-Ma2, PI-Ma3). As the branching chain length increased, the minimum value G_δ, which reflects the relaxation of the branched chains, shifted to the lower $|G^*|$ side. The mechanism underlying this shift might be attributed to the increase in the number of branched chains. The minimum value G_δ also shifted significantly to the higher δ side compared with the shift due to the increase in the number of branched chains. This result was attributed to the increase in the relaxation time with an increase in the branching chain length, which resulted in progressive dynamic tube dilation around the backbone chain and subsequently viscous behavior at the later stage where the backbone chain relaxed.

Analysis technology for analytical data

In the analysis of analytical data, structural features are commonly obtained by fitting one-dimensional data, such as spectra, with theoretical or phenomenological equations using the nonlinear least-squares method. Using a theoretical equation, physically meaningful features can be obtained and fed back to material design. However, because actual products have complex structures, few cases where materials can be fitted with theoretical equations are available. On the other hand, with the development of simulation technologies, the number of cases where analytical data can be calculated by simulation has increased. Structural information that can be fed back to material design can be obtained if analytical data can be fitted by simulation.

When fitting analytical data by simulation, computational time and convergence to an appropriate solution are considered critical parameters. The longer the computational time required for a single simulation, the longer it takes to converge to the optimal solution. Therefore, achieving convergence in fewer iterations is essential. In addition, for a complex relationship between the input and output values of the simulation, the values remain trapped in a local solution and fail to converge to an appropriate solution. Compared with the nonlinear least-squares method, fitting using various optimization methods, such as the reverse Monte Carlo method[43], genetic algorithms[44]–[45], and Bayesian optimization[8], allows convergence to an optimal solution more efficiently.

Bayesian optimization, which can converge to an optimal solution in few iterations, has garnered considerable attention in recent years. It predicts the objective variable (output variable) from the explanatory variable (input variable) using machine learning models such as Gaussian process regression, which allows simultaneous evaluation of the standard deviation corresponding to the objective variable's predicted value. **Fig. 5** shows an example of a simple case, where the objective variable is minimized when both the explanatory variable (x) and objective variable (y) are one-dimensional. The standard deviation around the explanatory variable being tested is small, whereas that in the range away from it is large.

Bayesian optimization uses an acquisition function rather than a predicted value. The acquisition function is expressed using a predicted value and its standard deviation, and it is a criterion for selecting the next point to be evaluated while considering uncertainty. **Fig. 5**(a) shows that the predicted value is at its minimum near $x = 0.7$. However, $x = 4.0$, where the acquisition function considering the standard deviation is at its maximum, is selected as the next candidate. As shown
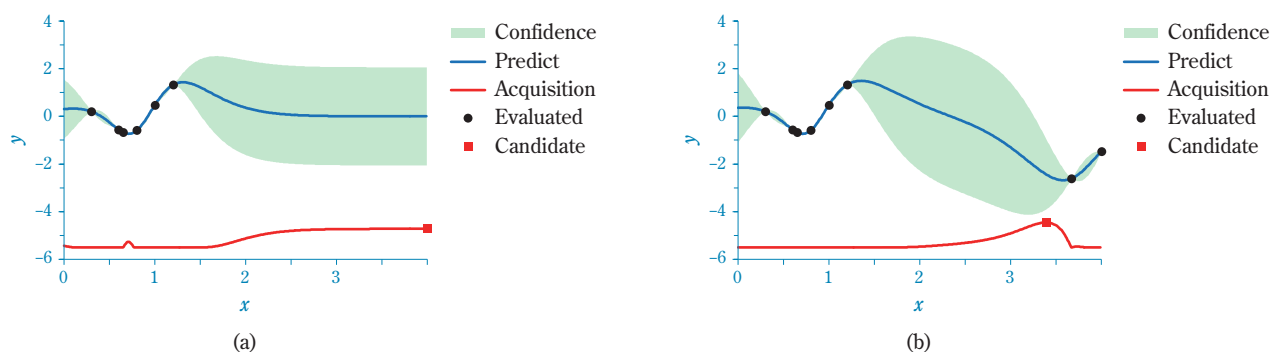
Fig. 5 Bayesian optimization after 6 (a) and 8 (b) iterations

in **Fig. 5**(b), a smaller y value than that in **Fig. 5**(a) was obtained after two additional evaluations. This method is less likely to be trapped in a local solution and allows efficient convergence to the optimal solution.

Samples and experiments

1. Samples

Industrial polyethylene (PE) was used to estimate the molecular structure using machine learning. This PE was an LDPE (F200-0) polymerized by a high-pressure process with a melt flow rate (MFR) of 2.0 g/10 min and density of 924 kg/m^3. We also used LLDPE (PE-A) polymerized by a gas-phase process using a metallocene catalyst with an MFR of 0.1 g/10 min and a density of 926 kg/m^3. **Table 2** shows the characteristics of the samples.

2. Viscoelasticity

For the dynamic viscoelasticity measurement, first, the sample was melted and pressurized at 150 °C using a heat press. Then, it was cooled in a cooling press at a temperature of 30 °C to create a disk-shaped test piece with a thickness of 2 mm and diameter of 25 mm. The ARES G2 manufactured by TA Instruments was used to measure the dynamic viscoelasticity. The test piece was fixed in a parallel plate fixture with a diameter of 25 mm, and measurements were conducted under a nitrogen atmosphere at temperatures of 130–210 °C

and dynamic strain of 5%. A master curve at 190 °C was obtained from the temperature dependence of the obtained viscoelasticity.

3. Molecular-weight distribution

The molecular-weight distribution was measured by gel permeation chromatography (GPC). The sample was dissolved in o-dichlorobenzene at 145 °C and adjusted to a concentration of 1 mg/mL. The sample injection volume was 0.5 mL, and separation was performed using three Tosoh Corp. GMH6-HT columns. The column oven temperature and flow rate were 140 °C and 1 mL/min, respectively, and a differential refractometer was used as the detector. The molecular chain length was calibrated with standard polystyrene, which had a molecular weight of 500–7,000,000 g/mol, and the molecular chain length A obtained by GPC measurements was multiplied by a Q factor of 17.7 g/mol·Å to convert it to a molecular weight M.

Analysis method

We used bob-rheology (ver.2.5) to conduct viscoelasticity simulations. The number of segments between the entanglement points at 190 °C was 42, the entanglement time was 2.5×10^{-8} s, and the dynamic dilation exponent was 1.0. The simulation assumed a three-component system, with the low-molecular-weight component A being a linear chain, and the medium-molecular-weight component B and high-molecular-weight component C being comb-shaped chains. A log-normal distribution was assumed for the molecular-weight distribution of the backbone chain and branched chains of each component. We used 13 input variables: the weight fraction of each component ϕ_X, molecular weight of the backbone chain M_{Xbb}, molecular weight of the branched chains M_{Xa} and its number of branches
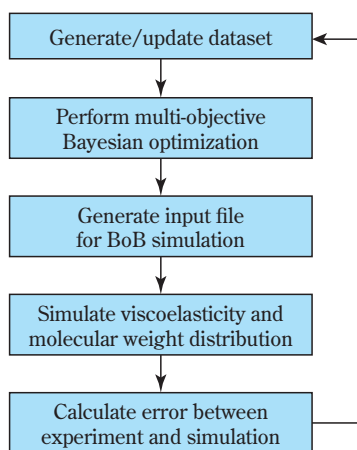
Table 2 Molecular characteristics of F200-0 and PE-A

Code	MFR g/10 min	Density kg/m^3	M_n g/mol	M_w g/mol	M_w/M_n
F200-0	2.0	924	18,700	72,200	3.8
PE-A	0.1	926	16,400	129,600	7.9

    7

q_X, and molecular weight dispersion D_X (where X represents the name of the component).

Fig. 6 shows the overall algorithm. First, a training dataset was generated by sampling various combinations for the 13 variables representing the molecular structure. The viscoelasticity and molecular-weight distribution for these combinations were then simulated. The error E_{VE} between the experimental and simulated values of G' and G'' in viscoelasticity was calculated. Similarly, the error E_{MWD} in the molecular-weight distribution was calculated, resulting in a dataset where the molecular structure was the explanatory variable and E_{VE} and E_{MWD} were the objective variables.

The general-purpose programming language Python and the multi-objective Bayesian optimization package MOBOpt[46] were used for multi-objective optimization. The kernel function in Gaussian process regression was selected from several candidates through grid search, and the one with the highest prediction accuracy for the objective variable was selected. Subsequently, Bayesian optimization was performed using the dataset to estimate the molecular structure that minimized the objective variable.

In Bayesian optimization, constraints were set such that the sum of the weight fractions of components A–C was 1.00±0.01. Simulations were conducted for the molecular structures predicted to have small errors, and viscoelasticity and molecular-weight distributions were obtained. E_{VE} and E_{MWD} were calculated from the simulation results and added to the dataset. The optimal solution was searched by repeatedly performing Bayesian optimization using the updated dataset.

Results and discussion

1. Molecular structure estimation by machine learning-1

(1) Comparison with experimental values

Fig. 7 shows the viscoelasticity of the sample obtained by this analysis method. The complex viscosity η^* was calculated by the following equation:

$$\eta^* = [(G'/\omega)^2 + (G''/\omega)^2]^{1/2} \qquad (4)$$

Here, ω is the angular frequency. The analytical results were in good agreement with the experimental data of G', G'', and η^*. Fig. 8 also represents the vGP plot. The experimentally obtained vGP plot showed a local minimum value, suggesting the presence of long chain branches. This result supports previous studies[39]−[41]. The analytical values reproduced the experimental data within the measured range, confirming quantitative validity. Fig. 9 compares the molecular-weight distribution obtained from the analysis and
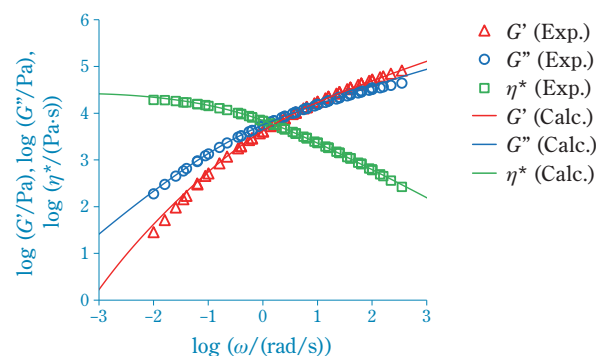
Fig. 7 Viscoelasticity of F200-0 obtained with combination of bob-rheology and machine learning methods with experimental data

Fig. 6 Computational algorithm using multi-objective Bayesian optimization
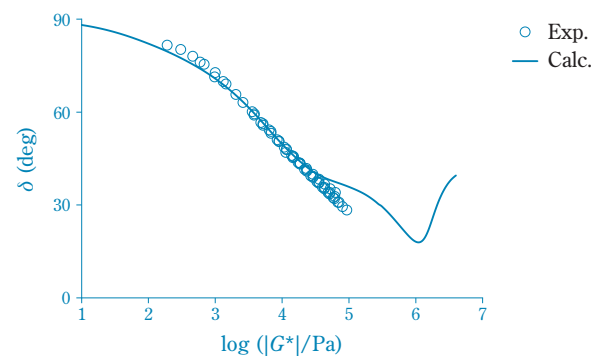
Fig. 8 vGP plot of F200-0 obtained with combination of bob-rheology and machine learning methods with experimental data

the values obtained by GPC. The molecular-weight distribution M_w/M_n of the sample was 3.8 and was unimodal. In this analysis, the number of components in the sample was three, and the determined molecular-weight distribution reproduced the experimental values.

(2) Comparison of results obtained by machine learning with experimental values

Table 3 presents the molecular structure of viscoelasticity simulated using machine learning based on Bayesian optimization. For comparison, the atomic force microscopy (AFM) results of Shinohara *et al.* are also shown[47].

The molecular components were designated A, B, and C from the low-molecular-weight side. The respective molecular weights of the components were 17,000, 69,000, and 232,000 g/mol, respectively. The per-chain numbers of the branched chains q of the components B and C were 4.9 and 7.2, respectively. The predicted ratios A_a/A_b of the branching chain length A_a to the backbone chain length A_b of the components B and C were 0.23 and 0.09, respectively. The weighted average q value of the entire molecule based on the ratio of
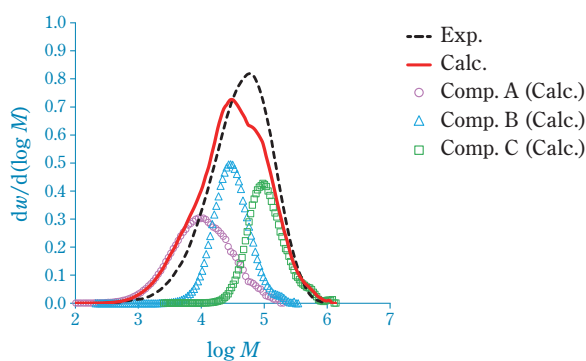
Fig. 9 Comparison of molecular weight distribution between simulated and experimental data

each component was 3.8, and A_a/A_b was 0.11.

For the AFM observation, the F200-0 sample, which was the same as that used for machine learning, was used. However, after dissolution in a solvent, the high-molecular-weight components were separated. Consequently, a direct comparison of the molecular weight and molecular-weight distribution was not possible. Therefore, q and A_a/A_b were compared. The average q value derived from machine learning was 3.8, while the value obtained by AFM observation was 3. Additionally, the A_a/A_b value predicted by machine learning was 0.11, while the observed result was 0.12. The number of specimens for the observed result was 1, and thus, a quantitative comparison was challenging. Nevertheless, the two results were in good agreement.

2. Molecular structure estimation by machine learning-2

We attempted to estimate the molecular structure of PE-A using machine learning with bob-rheology and Bayesian optimization. We set the objective variable as E_{VE}, which was calculated from the accumulated error between the viscoelasticity obtained by bob-rheology and the experimental value. Furthermore, the accumulated error E_{MWD} between the molecular-weight distribution obtained by GPC measurement and the bob-rheology simulation was used. In the analysis, the number of components in the sample was assumed to be 3, with the low-molecular-weight chains defined as linear chains, and the medium- and high-molecular-weight chains defined as branched chains.

In constructing the regression model in machine learning, 80% of the total data was used for training, while the remaining 20% was used for testing. **Fig. 10** shows the coefficient of determination of the regression model and test data in optimization. The coefficients of determination of the training data and test

**Table 3** Comparison of the molecular structure obtained from the machine learning estimations and AFM imaging

	Simulation			AFM observation[43]		
	Comp. A	Comp. B	Comp. C	Arm 1	Arm 2	Arm 3
M_w (g/mol)	17,000	69,000	232,000		275,000	
A_b (nm)	94	182	779		162	
A_a (nm)		43	74	10	18	31
Number of arms		4.9	7.2		3	
Average number of arms		3.8			3	
A_a/A_b		0.23	0.09	0.06	0.11	0.19
Average A_a/A_b		0.11			0.12	

(a) Error of viscoelasticity

(b) Error of molecular weight distribution
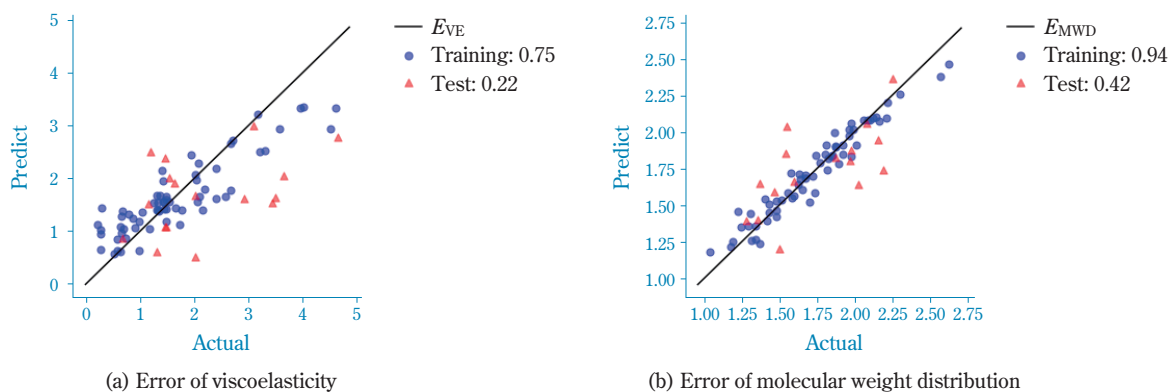
Fig. 10   Predictive machine learning modeling using viscoelasticity and molecular weight distribution

data for the viscoelasticity error E_{VE} were 0.75 and 0.22, respectively. The coefficients of determination of the training data and test data for the molecular-weight distribution error E_{MWD} were 0.94 and 0.42, respectively. These results indicate that the reliability of the regression model using the training data was high, but that of the test data was insufficient.

Fig. 11 shows the Pareto solutions obtained by multi-objective Bayesian optimization. Initially, a model
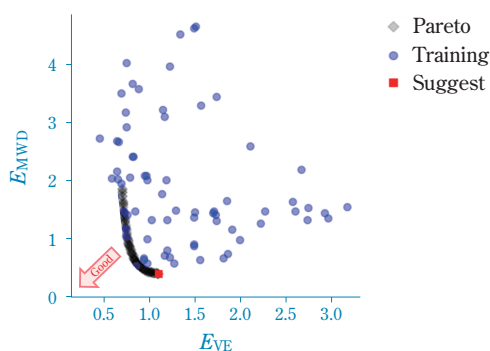
Fig. 11   Relationship between objective variables obtained from Bayesian optimization

was constructed using 80% of the training data, and its performance was evaluated based on the coefficient of determination. Subsequently, a comprehensive model was constructed utilizing the entire data set as training data. The lower left region in **Fig. 11** indicates low errors in both objective variables, and an optimal Pareto solution is proposed based on this analysis model.

One of the molecular structures obtained by machine learning was selected, and its viscoelasticity is shown in **Fig. 12**(a). The Pareto solution with the smallest viscoelasticity error was selected. The viscoelasticity of the molecular structure obtained by machine learning was in good agreement with the experimental data within the measured range. **Fig. 12**(b) shows the molecular-weight distribution of the molecular structure obtained by machine learning. Although the molecular weight was slightly lower than the experimental value, the overall distribution pattern of the molecular-weight distribution was generally consistent. However, no molecular structure was obtained that could simultaneously reproduce both the experimental values of viscoelasticity and molecular-weight

(a) Linear viscoelasticity

(b) Molecular weight distribution
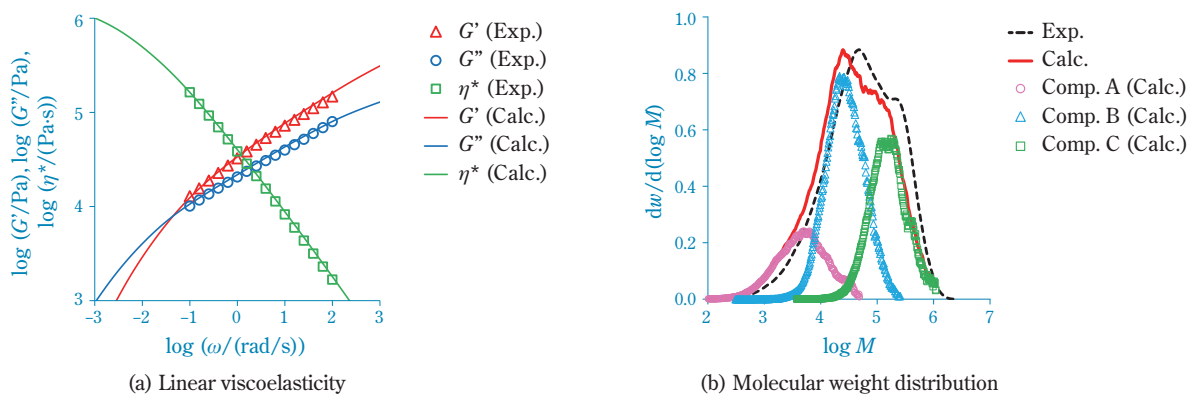
Fig. 12   Linear viscoelasticity and molecular weight distribution of PE-A obtained with experiment and molecular simulation applying the multi-objective variables Bayesian optimization method

     10

distribution with high accuracy. The analytically obtained molecular-weight distribution differed from the experimental value obtained by GPC measurements. This discrepancy may arise because the bob-rheology relaxation algorithm, based on the hierarchical relaxation mechanism and dynamic tube dilation, may not perfectly represent the actual relaxation behavior of polymer chains. Consequently, an error occurs in the molecular-weight distribution when a molecular structure is searched to accurately reproduce viscoelasticity.

Conclusion

A viscoelasticity simulation using the rheology simulator bob-rheology was combined with Bayesian optimization to obtain the viscoelasticity and molecular architecture of the primary structure of a resin. The discrepancies between the simulated and experimental data were quantified. A machine learning analysis was performed using the molecular structure as the explanatory variable and the calculated error as the objective variable to estimate the molecular shape.

The results confirmed that polymer chains with multiple branched chains undergo relaxation through hierarchical relaxation and dynamic tube dilation after deformation. In addition, the presence of branched chains was confirmed from vGP plots. The molecular structure of LDPE predicted by a combination of viscoelasticity simulation and machine learning was compared with the observation results of the molecular chain. The simulated branched chain structure was generally in agreement with the experimental one, and one molecular chain had several branched chains. The branching chain length was approximately 10% of the backbone chain length. For PE-A, which has an uncharacterized molecular structure, machine learning effectively reproduced the experimental values of viscoelasticity and molecular-weight distribution. Therefore, the validity of the estimated molecular structure was quantitatively confirmed.

Reference

1) R. G. Larson, Macromolecules, 34(13), 4556 (2001).
2) S. J. Park et al., Rheol. Acta, 44(3), 319 (2005).
3) N. J. Inkson et al., Macromolecules, 39(12), 4217 (2006).
4) C. Das et al., J. Rheol., 50(2), 207 (2006).
5) C. Das et al., Phys. Rev. E, 74(1), 011404 (2006).
6) C. Das and D. J. Read, "bob-rheology", http://sourceforge.net/projects/bob-rheology (referred on 2024/3/12).
7) M. van Gurp and J. Palmen, Rheol. Bull., 67(1), 5 (1998).
8) J. Lee et al., npj Comput. Mater., 9, 135 (2023).
9) P. G. de Gennes, J. Chem. Phys., 55(2), 572 (1971).
10) M. Doi and S. F. Edwards, J. Chem. Soc., Faraday Trans. 2, 74, 1789 (1978).
11) M. Doi and S. F. Edwards, J. Chem. Soc., Faraday Trans. 2, 74, 1802 (1978).
12) M. Doi and S. F. Edwards, J. Chem. Soc., Faraday Trans. 2, 74, 1818 (1978).
13) M. Doi and S. F. Edwards, J. Chem. Soc., Faraday Trans. 2, 75, 38 (1979).
14) M. Doi and S. F. Edwards, "The Theory of Polymer Dynamics", Oxford University Press (1986).
15) M. J. Struglinski and W. W. Graessley, Macromolecules, 18(12), 2630 (1985).
16) Y. H. Zang et al., Polymer, 28, 1577 (1987).
17) M. Takahashi, Kobunshi, 47(11), 820 (1998).
18) J. Roovers and W. W. Graessley, Macromolecules, 14(3), 766 (1981).
19) J. Roovers, Macromolecules, 17(6), 1196 (1984).
20) J. Roovers, Polymer, 26(7), 1091 (1985).
21) W. W. Graessley and J. Roovers, Macromolecules, 12(5), 959 (1979).
22) T. C. B. McLeish, Europhys. Lett., 6(6), 511 (1988).
23) D. R. Daniels et al., Macromolecules, 34(20), 7025 (2001).
24) S. J. Park and R. G. Larson, J. Rheol., 50(1), 21 (2006).
25) M. Kapnistos et al., Macromolecules, 38(18), 7852 (2005).
26) G. Marrucci, J. Polym. Sci., Polym. Phys. Ed., 23(1), 159 (1985).
27) R. C. Ball and T. C. B. McLeish, Macromolecules, 22(4), 1911 (1989).
28) A. E. Likhtman and T. C. B. McLeish, Macromolecules, 35(16), 6332 (2002).
29) H. Watanabe et al., Macromolecules, 35(23), 8802 (2002).
30) S. J. Park and R. G. Larson, Macromolecules, 37(2), 597 (2004).
31) E. van Ruymbeke et al., Macromolecules, 45(4), 2085 (2012).
32) C. Das et al., Rheol. Acta, 58(3–4), 159 (2019).
33) T. C. B. McLeish, Adv. Phys., 51(6), 1379 (2002).

34) E. van Ruymbeke *et al.*, Macromolecules, 47(21), 7653 (2014).

35) Y. Masubuchi *et al.*, J. Chem. Phys., 115(9), 4387 (2001).

36) K. M. Kirkwood *et al.*, Macromolecules, 42(24), 9592 (2009).

37) Y. Masubuchi *et al.*, Rheol. Acta, 51(3), 193 (2012).

38) S. Shiromoto *et al.*, Seikei-kakou symposium 2010, Japanese Society of Polymer Processing, p.353 (2010).

39) S. Shiromoto *et al.*, Reoroji-Toronkai, The Society of Rheology, Japan, Presentation No. 3B07 (2011).

40) D. J. Lohse *et al.*, Macromolecules, 35(8), 3066 (2002).

41) S. Trinkle *et al.*, Rheol. Acta, 41(1–2), 103 (2002).

42) K. M. Zentel *et al.*, RSC Adv., 11(52), 33114 (2021).

43) K. Hagita *et al.*, Rheol. Acta, 47(5–6), 537 (2008).

44) M. Wormington *et al.*, Philos. Trans. R. Soc., A, 357(1761), 2827 (1999).

45) A. Ulyanenkov *et al.*, physica B, 283(1–3), 237 (2000).

46) P. P. Galuzio *et al.*, SoftwareX, 12, 100520 (2020).
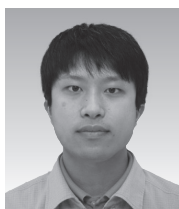
47) K. Shinohara *et al.*, Sci. Rep., 9, 9791 (2019).

PROFILE

Seiji SHIROMOTO

Sumitomo Chemical Co., Ltd.
Essential Chemicals Research
Laboratory
Senior Research Associate, Ph.D.

Akinori BANDO

Sumitomo Chemical Co., Ltd.
Essential Chemicals Research
Laboratory
Research Associate, Ph.D.

Hiroyuki KUDO

Sumitomo Chemical Co., Ltd.
Essential Chemicals Research
Laboratory
Researcher