Multivariate statistical analyses of ^{1}H NMR data for binary and ternary copolymer blends to determine the chemical compositions and blending fractions of the components

Ryota Kamiike[a,b], Tomohiro Hirano[a], Koichi Ute[a]

[a] Department of Applied Chemistry, Tokushima University, 2-1 Minamijosanjima, Tokushima 770-8506, Japan

[b] R&D Center, Nippon A&L Inc., 2-10-2 Kikumoto, Niihama, Ehime 792-0801, Japan

Correspondence:

Tomohiro Hirano

hirano@tokushima-u.ac.jp

Abstract

Statistical ^{1}H nuclear magnetic resonance (NMR) analyses were conducted with ternary copolymer blends. Two out of the three monomers, acrylonitrile, styrene, and α-methylstyrene, were subjected to radical copolymerization to synthesize three kinds of copolymers that were mixed to prepare binary and ternary copolymer blends. The ^{1}H NMR spectral matrix for the copolymers and their blends (explanatory variables) was combined with the blending parameter matrix (objective variables). Cross-validation with the least absolute shrinkage and selection operator regression confirmed that excellent regression models were constructed with a dataset composed of data for eight copolymers and forty-five binary blends; these were used to predict the blending parameters for the binary blends, such as the chemical compositions and mole fractions of the component copolymers. Accordingly, the models were then used to predict the blending parameters for the ternary blends, which resulted in successful and highly accurate predictions. Other regularized regression models, such as Ridge regression and Elastic Net, were also examined.

Introduction

Since chemometrics was established early in the 1970s, a wide range of multivariate analyses, particularly principal component analysis (PCA) [1, 2] and partial least-squares regression (PLS) [3-5], have been applied to measured data to extract chemical information [6, 7]. The first application of chemometrics to nuclear magnetic resonance (NMR) spectra, which was used to cluster the substituents on monosubstituted benzenes from the chemical shifts of the aryl carbons, was reported in 1983 [8]. In the early 1990s, PCA was introduced to classify the ^{1}H NMR spectra of urine [9]. Now, multivariate analyses of biological NMR data have become recognized as metabolomics [10].

NMR spectroscopy is also a powerful tool for analyzing polymer microstructures, including their stereochemistries, chemical compositions, and monomer sequences [11, 12]. The chemometric approach has been successfully applied to the NMR spectra of synthetic polymers [13-25]. Recently, we reported multivariate statistical analyses of the ^{1}H NMR spectra for binary blends (*bi*-blends) of copolymers prepared by emulsion copolymerization with two of the three monomers, acrylonitrile (AN), styrene (ST), and α-methylstyrene (MS) [24]. Two linear regression models, PLS and least absolute shrinkage and selection operator regression (LASSO) [26], were applied to the NMR spectral data as explanatory variables and the blending parameters (BPs) as objective variables. If the component copolymers in *bi*-blends were defined as "polymer-**A**" and "polymer-**B**" with priority given in the order poly(AN-*co*-ST) (AN/ST), poly(AN-*co*-MS) (AN/MS), and poly(MS-*co*-ST) (MS/ST), the predictions of BPs, such as the chemical compositions and mole fractions of the component copolymers, failed. A new parameter was introduced, and the objective variables were redefined based on the amounts of the

component copolymers in the blend: the component copolymers with larger and smaller amounts were classified as "polymer-**L**" and "polymer-**S**", respectively. Furthermore, the addition of squared spectral data to the explanatory variables was required to predict the BPs with linear regression models based on PLS and LASSO because the newly defined objective variables had nonlinear relationships with the ^{1}H NMR spectral data. After optimizing both the objective and explanatory variables, the BPs were successfully predicted. Note that LASSO showed slightly better accuracy than PLS.

In the present study, we investigated the extent to which multivariate statistical analyses of ^{1}H NMR spectral data are useful for extracting the BPs for ternary blends (*ter*-blends). As with our previous work [24], three kinds of copolymers, AN/ST, AN/MS, and MS/ST, were synthesized by emulsion copolymerization to prepare one series of *ter*-blends and three series of *bi*-blends. These copolymers were simply defined as "polymer-**a**", "polymer-**b**", and "polymer-**c**", respectively. The BPs for the *ter*-blends were successfully predicted with regularized regression analyses, such as LASSO, Ridge [27], and Elastic Net [28], by using the dataset for the copolymers and the *bi*-blends as a training dataset and the ^{1}H NMR spectral data of the *ter*-blends as a test dataset. Note that no optimizations were needed for either the objective or explanatory variables.

Experimental

Multivariate statistical analyses of ^{1}H NMR spectral data

Eight copolymers were prepared by changing the combinations of comonomers and chemical compositions (Table 1). The polymer preparation details were described in our previous paper [24].

（**Table 1**）

Forty-five *bi*-blends and six *ter*-blends were prepared by varying the species and fractions of the component copolymers. Fig. 1 shows a ternary composition diagram for the *bi*-blends and *ter*-blends, and additional details are given in Table S1. The codes for the *bi*-blends, "A_S/M", "S_A/M", and "M_A/S", were assigned for the combinations of "AN/ST+AN/MS", "AN/ST+MS/ST", and "AN/MS+MS/ST", respectively, in which the first symbol means the common monomeric unit. Furthermore, identification numbers were given with the format S_A/M_25/40, which was based on the run-numbers for the component copolymers in Table 1 (the *bi*-blends of Run 2 and Run 5 in this example) and the weight fraction of the first component copolymer (40% for Run 2 in this example). The code "ASM" was used for the *ter*-blends by adding both the run-numbers of the component copolymers and the ratio of their weight fractions in the format ASM_138_343 (the *ter*-blends of Run 1, Run 3, and Run 8 with a weight ratio of 3:4:3 in this example). The mole fractions for the copolymer blends were calculated as in a previous report [24].

(**Fig. 1**)

^{1}H NMR spectra of the copolymers and their blends were recorded with an ECZ400 spectrometer (JEOL Ltd., Tokyo, Japan) equipped with a 5-mm ROYAL probe under the following conditions: 2 wt/vol% in deuterated pyridine, 100 °C, 45° pulse (3.925 µs), spectral width of 7,423 Hz, 128 scans, repetition time of 9.4145 s and 32,768

data points. The free induction decays were exponentially apodized with 0.2 Hz line-broadening and zero-filled to 65,536 points prior to the Fourier transformations. The signal for H_2O overlapped those of the main-chain CH_2 and CH groups when $CDCl_3$ was used as the solvent. Therefore, deuterated pyridine was chosen as the solvent to avoid overlap of the H_2O signal. The signal for the nondeuterated *para*-proton of the solvent (7.52 ppm) was used as an internal reference for the chemical shift.

Bucket integration of the resonance regions for the α-CH_3 group in the MS units and the main-chain CH_2 and CH groups (0.25–3.40 ppm) was performed with intervals of 0.01 ppm to express each spectrum with 314 variables (JEOL Delta NMR ver. 5.2 software). The integrated intensities were normalized so that the sum was 100. The spectral data were combined with the structural data to conduct regularized regressions with R software ver. 3.4.4 packaged in the Visual R Platform ver. 2.0 (NTT DATA Mathematical Systems Inc., Tokyo, Japan).

Regularized regressions, such as LASSO, Ridge and Elastic Net, are extensions of ordinary least-squares regressions [26-28]. The prediction accuracy is improved by minimizing the objective function (E), which is expressed as:

$$\text{LASSO}: E = \|y - xb\|^2 + \lambda \sum_{i=1}^{m} |b_i| \quad (1)$$

$$\text{Ridge}: E = \|y - xb\|^2 + \lambda \sum_{i=1}^{m} b_i^2 \quad (2)$$

$$\text{Elastic Net}: E = \|y - xb\|^2 + \lambda \left[\alpha \sum_{i=1}^{m} |b_i| + (1 - \alpha) \sum_{i=1}^{m} b_i^2 \right] \quad (3)$$

where y denotes a dataset of material data (the chemical compositions and mole fractions of the component copolymers in this study); x denotes a dataset of bucket integral values;

b denotes a dataset of regression coefficients; b_i denotes the *i*th regression coefficient; λ denotes a regularization parameter; and α denotes a tuning parameter. The second term on the right-hand side is a penalty for increasing the coefficient values. The sum of the absolute values of the *i*th regression coefficients was used for LASSO, whereas the sum of the squared values of the *i*th regression coefficients was used for Ridge. The Elastic Net penalty was composed of a combination of the LASSO and Ridge penalties. The contributions were determined by the tuning parameter α: Elastic Net provides the same result as LASSO when α is unity and the same as Ridge when α is zero. Note that in LASSO, meaningful variables are selected from many predictors by setting b_i to zero, which results in a decreased number of variables used in the regression (df) with an increase in λ value. A fivefold cross-validation (CV) was used to optimize both the λ and α values by minimizing the mean absolute error.

Results and Discussion

^{1}H NMR spectra of the component copolymers, *bi*-blends and *ter*-blends

The signals observed at 0.25–3.40 ppm in the ^{1}H NMR spectra of the component copolymers were assigned to the α-CH$_3$ groups in the MS units and the main-chain CH$_2$ and CH groups (Fig. 2 (a)–(c); Runs 2, 6, and 7 in Table 1). The combinations of the monomers used significantly affected the spectral patterns. However, the chemical composition could not be determined because the broadened and overlapped signals made it difficult to assign and quantify the signals for each monomeric unit. The spectra of the *bi*-blends were broadened more than those of their component copolymers [Fig. 2(d)–(f)], likely because the spectra for those two copolymers were superimposed. In addition, the spectrum of the *ter*-blend was so broadened [Fig. 2 (g)] that it was indistinguishable from

those of the *bi*-blends. Therefore, it was practically impossible to determine which component copolymers were contained and how much of each copolymer was added just by looking at the spectra.

(Fig. 2)

Definition of the BPs for the *ter*-blends

In our previous study [24], the three kinds of copolymers were categorized either as "polymer-**A**" or "polymer-**B**" to determine the BPs for the *bi*-blends. This awkward definition induced a contradiction in the blending fractions depending on the combination of copolymers, and optimizations of the explanatory and objective variables were required to predict the BPs. In the present study, however, the three kinds of copolymers, AN/ST, AN/MS, and MS/ST, are simply defined as "polymer-**a**", "polymer-**b**" and "polymer-**c**", respectively. The BPs defined for the *bi*-blends were extended to those for the *ter*-blends. Multiplication of the chemical composition and mole fraction of a certain component copolymer corresponds to the chemical composition derived from the component copolymer in the *ter*-blends. Accordingly, the averaged chemical composition in the *ter*-blends can be defined with Equations (4) and (5):

$$Comp = (Comp_a \times f_a) + (Comp_b \times f_b) + (Comp_c \times f_c) \qquad (4)$$

$$f_a + f_b + f_c = 1 \qquad (5)$$

where *Comp* denotes the averaged chemical compositions of the AN, MS, or ST monomeric units in the copolymer blends; $Comp_a$, $Comp_b$, and $Comp_c$ denote the

chemical compositions of the component copolymers named "polymer-**a**", "polymer-**b**" and "polymer-**c**", respectively; and f_a, f_b, and f_c denote the mole fractions of polymer-**a**, polymer-**b** and polymer-**c**, respectively, in the blends (Scheme 1).

(Scheme 1)

Prediction of the BPs for the *bi*-blends by LASSO

The *Comp* values for the *bi*-blends can be expressed with the abovementioned definitions, in which one of the mole fractions is set to zero. At first, therefore, fivefold CVs of LASSO were conducted with the dataset combining the bucket-integral values and the *Comp* values of the eight copolymers and the forty-five *bi*-blends. Note that the *Comp* values of the eight copolymers were the same as the chemical compositions. The predicted values agreed well with the theoretical values calculated from the weight fractions of the component copolymers with high coefficient-of-determination (R^2) values over 0.98 (Fig. S1). This was not surprising, because the *Comp* values should be the same as the values calculated with the previous definition, "polymer-**A**" and "polymer-**B**" [24].

Then, five CVs were performed with LASSO to predict the BPs for the *bi*-blends. As mentioned above, for example, the composition of AN derived from polymer-**a** in the *bi*-blends can be expressed as $AN_a \times f_a$. Therefore, the chemical compositions derived from the component copolymers ($Comp_a \times f_a$, $Comp_b \times f_b$, and $Comp_c \times f_c$) and the mole fractions (f_a, f_b and f_c) were used as the objective variables. Fig. 3 shows the relationships between the predicted and theoretical values for the chemical compositions derived from the individual component copolymers and the mole fractions. Excellent relationships

were observed with R^2 values over 0.98, regardless of the BPs. It was therefore assumed that simple categorization of the three kinds of copolymers into "polymer-**a**", "polymer-**b**", and "polymer-**c**" resulted in the construction of straightforward regression models with which to predict the BPs with high accuracy. Consequently, optimizations of neither the objective variables nor explanatory variables were needed.

(Fig. 3)

Prediction of the BPs for the *ter*-blends with LASSO

As with the *bi*-blends, the theoretical values for the *ter*-blends were calculated from the weight fractions of the three component copolymers. First, the *Comp* values were predicted by LASSO. Considering that excellent regression models were constructed with the dataset composed of eight copolymers and forty-five *bi*-blends, the prediction was carried out by using the dataset for the copolymers and the *bi*-blends as the training dataset. The *Comp* values were successfully predicted with R^2 values over 0.95, regardless of the monomeric unit (Fig. 4).

(Fig. 4)

Next, the chemical compositions derived from the component copolymers ($Comp_a \times f_a$, $Comp_b \times f_b$, $Comp_c \times f_c$) and the mole fractions (f_a, f_b and f_c) for the *ter*-blends were predicted with LASSO by using the dataset for the copolymers and the *bi*-blends as the training dataset. Excellent relationships were obtained with R^2 values over

0.93, regardless of the BPs (Fig. 5), indicating that the BPs for the *ter*-blends were also predicted with practical accuracies. It should be noted that the predictions of the BPs for the *ter*-blends were carried out without adding the data for the *ter*-blends in the training dataset.

(Fig. 5)

Prediction of the BPs for the *ter*-blends with Ridge regression and Elastic Net

Instead of LASSO, Ridge and Elastic Net were used to predict the BPs for the *bi*-blends. As with LASSO, good relationships were obtained with high R^2 values, regardless of the regression used (Figs. S2 and S3). Therefore, the BPs for the *ter*-blends were also predicted with Ridge and Elastic Net by using the datasets for the copolymers and the *bi*-blends as the training dataset. Table 2 summarizes the R^2 values obtained with Ridge and Elastic Net together with those obtained with LASSO. The R^2 values obtained with Ridge were lower than those obtained with LASSO, except for the chemical composition of the MS unit derived from polymer-**b** (MS**b**) in the *ter*-blends. The R^2 values obtained with Elastic Net were comparable to those obtained with LASSO. These results indicated that the use of LASSO or Elastic Net gave better predictions of the BPs for the *ter*-blends from the ^{1}H NMR spectral data for at least the copolymer blends examined in this study.

(Table 2)

Conclusions

Statistical structural analyses of ternary copolymer blends were conducted by

LASSO with the dataset comprising the ^{1}H NMR spectral data as explanatory variables and the BPs as objective variables. Definitions of the three kinds of copolymers as "polymer-**a**", "polymer-**b**" and "polymer-**c**" resulted in the construction of straightforward regression models with which to predict the chemical compositions derived from the component copolymers and the mole fractions. This indicated that the use of LASSO could be expanded to include structural analyses of ternary copolymer blends. Note that the predictions were conducted with the data for eight copolymers and forty-five *bi*-blends used as the training dataset: the data for the *ter*-blends were not necessary for the predictions. In addition to LASSO, Ridge and Elastic Net were employed to predict the BPs. Elastic Net predicted the BPs with accuracies comparable to those of LASSO, whereas Ridge exhibited poor prediction accuracies. Further work is now underway to examine the extent to which regularized regression is useful in extracting the BPs for copolymer blends containing terpolymers by using ^{1}H and/or ^{13}C NMR spectral data as the explanatory variables.

Conflict of interest The authors declare no competing interests.

References

1. Kowalski BR, Bender CF. Pattern recognition. Powerful approach to interpreting chemical data. J Am Chem Soc. 1972;94:5632-9. https://doi.org/10.1021/ja00771a016.

2. Kowalski BR, Bender CF. Pattern recognition. II. Linear and nonlinear methods for displaying chemical data. J Am Chem Soc. 1973;95:686-93. https://doi.org/10.1021/ja00784a007.

3. Wold S. Spline functions in data analysis. Technometrics. 1974;16:1-11. https://doi.org/10.1080/00401706.1974.10489142.

4. Wold S. Nonlinear partial least squares modelling ii. Spline inner relation. Chemom Intell Lab Syst. 1992;14:71-84. https://doi.org/10.1016/0169-7439(92)80093-J.

5. Wold S, Kettaneh-Wold N, Skagerberg B. Nonlinear pls modeling. Chemom Intell Lab Syst. 1989;7:53-65. https://doi.org/10.1016/0169-7439(89)80111-X.

6. Lavine BK. Chemometrics. Anal Chem. 1998;70:209-28. https://doi.org/10.1021/a19800085.

7. Lavine BK, Brown SD, Booksh KS. 40 years of chemometrics – from bruce kowalski to the future. Washington, DC: Oxford University Press; 2015.

8. Johnels D, Edlund U, Grahn H, Hellberg S, Sjöström M, Wold S, Clementi S, Dunn WJ. Clustering of aryl carbon-13 nuclear magnetic resonance substituent chemical shifts. A multivariate data analysis using principal components. J Chem Soc, Parkin Trans 2. 1983:863-71. https://doi.org/10.1039/P29830000863.

9. Gartland KP, Beddell CR, Lindon JC, Nicholson JK. Application of pattern recognition methods to the analysis and classification of toxicological data derived from proton nuclear magnetic resonance spectroscopy of urine. Mol Pharmacol. 1991;39:629-42.

10. Lindon JC, Nicholson JK, Holmes E, Everett JR. Metabonomics: Metabolic processes studied by NMR spectroscopy of biofluids. Concepts Magn Reson. 2000;12:289-320. https://doi.org/10.1002/1099-0534(2000)12:5<289::AID-

CMR3>3.0.CO;2-W.

11. Bovey FA. High resolution NMR of macromolecules. New York: Academic Press; 1972.

12. Hatada K, Kitayama T. NMR spectroscopy of polymers. Berlin: Springer; 2004.

13. Sardashti M, Gislason JJ, Lai X, Stewart CA, O'Donnell DJ. Determination of total styrene in styrene/butadiene block copolymers by process NMR and chemometrics modeling. Appl Spectrosc. 2001;55:467-71. https://doi.org/10.1366/0003702011951984.

14. Momose H, Hattori K, Hirano T, Ute K. Multivariate analysis of ^{13}C NMR spectra of methacrylate copolymers and homopolymer blends. Polymer. 2009;50:3819-21. https://doi.org/10.1016/j.polymer.2009.05.048.

15. Momose H, Maeda T, Hattori K, Hirano T, Ute K. Statistical determination of chemical composition and monomer sequence distribution of poly(methyl methacrylate-*co-tert*-butyl methacrylate)s by multivariate analysis of ^{13}C NMR spectra. Polym J. 2012;44:808-14. https://doi.org/10.1038/pj.2012.110.

16. Hirano T, Anmoto T, Umezawa N, Momose H, Katsumoto Y, Oshimura M, Ute K. Application of multivariate analysis of NMR spectra of poly(*n*-isopropylacrylamide) to assignment of stereostructures and prediction of tacticity distribution. Polym J. 2012;44:815-20. https://doi.org/10.1038/pj.2012.61.

17. Mayer BP, Lewicki JP, Chinn SC, Overturf GE, Maxwell RS. Nuclear magnetic resonance and principal component analysis for investigating the degradation of poly[chlorotrifluoroethylene-*co*-(vinylidene fluoride)] by ionizing radiation. Polym Degrad Stab. 2012;97:1151-7. https://doi.org/10.1016/j.polymdegradstab.2012.04.001.

18. Hirano T, Kamiike R, Hsu Y, Momose H, Ute K. Multivariate analysis of ^{13}C NMR spectra of branched copolymers prepared by initiator-fragment incorporation radical copolymerization of ethylene glycol dimethacrylate and *tert*-butyl methacrylate. Polym J. 2016;48:793-800. https://doi.org/10.1038/pj.2016.16.

19. Kukackova O, Dung NV, Abbrent S, Urbanova M, Kotek J, Brus J. A novel insight into the origin of toughness in polypropylene–calcium carbonate microcomposites: Multivariate analysis of ss-NMR spectra. Polymer. 2017;132:106-13. https://doi.org/10.1016/j.polymer.2017.10.071.

20. Hsu Y, Chuang M, Hirano T, Ute K. Multivariate analysis of ^{13}C NMR spectra to extract information about monomer sequences in poly(methyl methacrylate-*co*-benzyl methacrylate)s prepared by various polymer reactions. Polym J. 2018;50:355-63. https://doi.org/10.1038/s41428-018-0027-9.

21. Kusaka Y, Hasegawa T, Kaji H. Noise reduction in solid-state NMR spectra using principal component analysis. J Phys Chem A. 2019;123:10333-8. https://doi.org/10.1021/acs.jpca.9b04437.

22. Fernando LD, Dickwella Widanage MC, Penfield J, Lipton AS, Washton N, Latgé J-P, Wang P, Zhang L, Wang T. Structural polymorphism of chitin and chitosan in fungal cell walls from solid-state NMR and principal component analysis. Frontiers in Molecular Biosciences. 2021;8. https://doi.org/10.3389/fmolb.2021.727053.

23. Hirano T, Kamiike R, Yuki T, Matsumoto D, Ute K. Determination of monomer reactivity ratios from a single sample using multivariate analysis of the ^{1}H NMR spectra of poly[(methyl methacrylate)-*co*-(benzyl methacrylate)]. Polym J. 2022;54:623-31. https://doi.org/10.1038/s41428-022-00618-y.

24. Kamiike R, Hirano T, Ute K. Statistical determination of chemical composition and

blending fraction of copolymers by multivariate analysis of ^{1}H NMR spectra of binary blends of the copolymers. Polymer. 2022;256:125207. https://doi.org/10.1016/j.polymer.2022.125207.

25. Novotny EH, Garcia RHS, deAzevedo ER. Pulse sequence induced variability combined with multivariate analysis as a potential tool for ^{13}C solid-state NMR signals separation, quantification, and classification. J Magn Reson Open. 2023;14-15:100089. https://doi.org/10.1016/j.jmro.2022.100089.

26. Tibshirani R. Regression shrinkage and selection via the lasso. J Roy Statist Soc B. 1996;58:267-88. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

27. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics. 1970;12:55-67. https://doi.org/10.1080/00401706.1970.10488634.

28. Zou H, Hastie T. Regularization and variable selection via the elastic net. J Roy Statist Soc B. 2005;67:301-20. https://doi.org/10.1111/j.1467-9868.2005.00503.x.
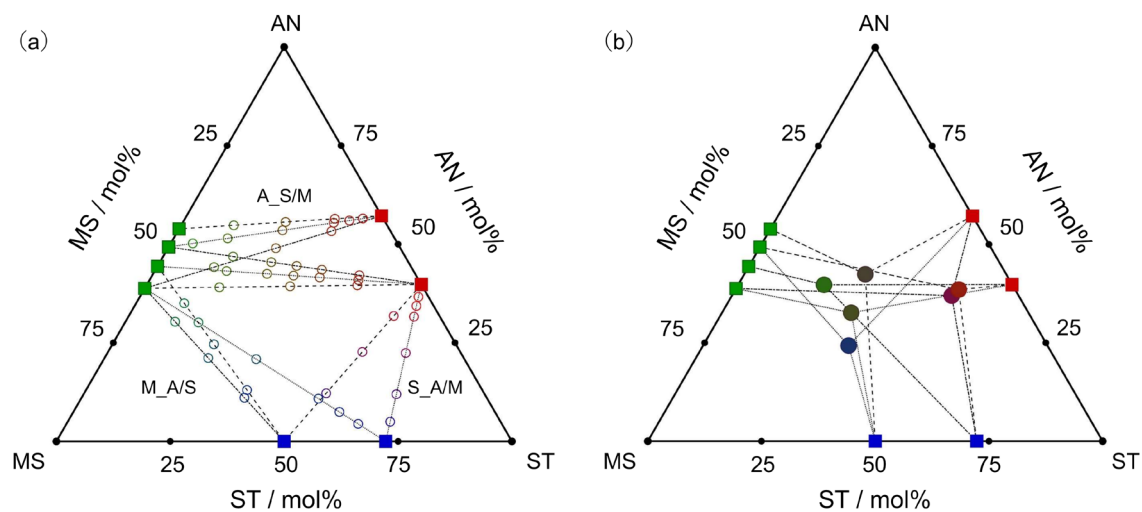
Fig. 6. Ternary composition diagrams for the copolymers and their blends. The symbols ■, ■, and ■ denote the plots of AN/ST, AN/MS, and MS/ST, respectively. (a) The symbol ○ denotes plots of the *bi*-blend series. (b) The symbol ● denotes plots of the *ter*-blends.
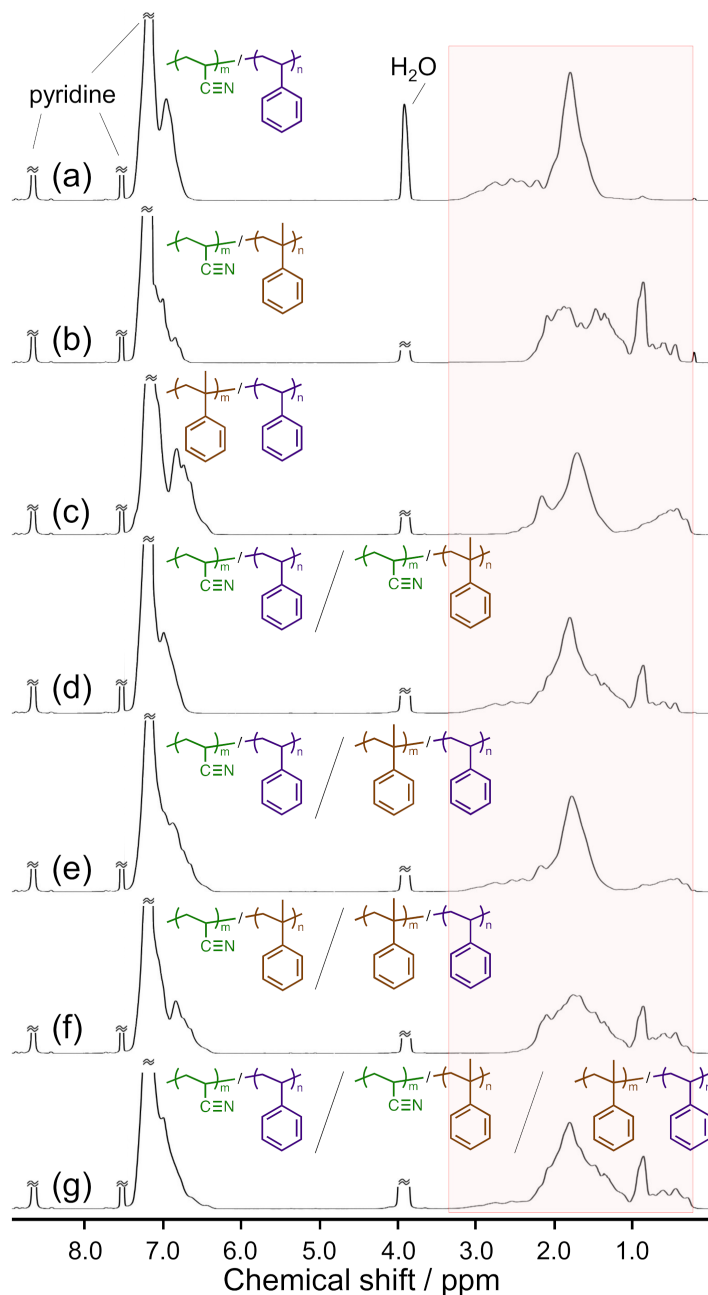
Fig. 7. ^{1}H NMR spectra of the component copolymers (a) AN/ST (Run 2), (b) AN/MS (Run 6), and (c) MS/ST (Run 7); the *bi*-blends (d) A_SM_26/50, (e) S_A/M_27/50, and (f) M_A/S_67/50; and the *ter*-blend (g) ASM_268_352, as measured in deuterated pyridine at 100 °C. The resonance region for the α-CH$_3$ groups in the MS units and the main-chain CH$_2$ and CH groups (0.25–3.40 ppm), emphasized in red, was used to conduct the chemometric analyses.
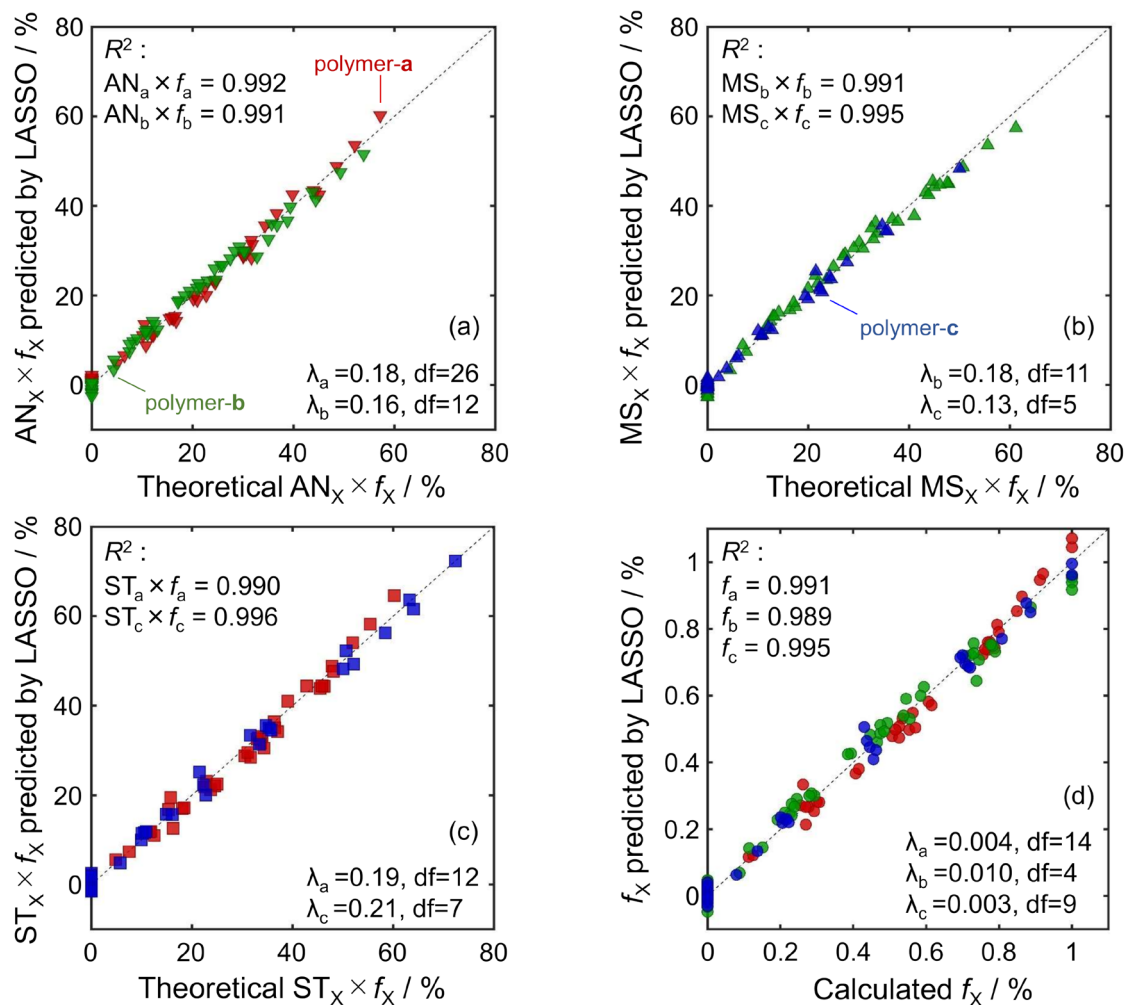
18
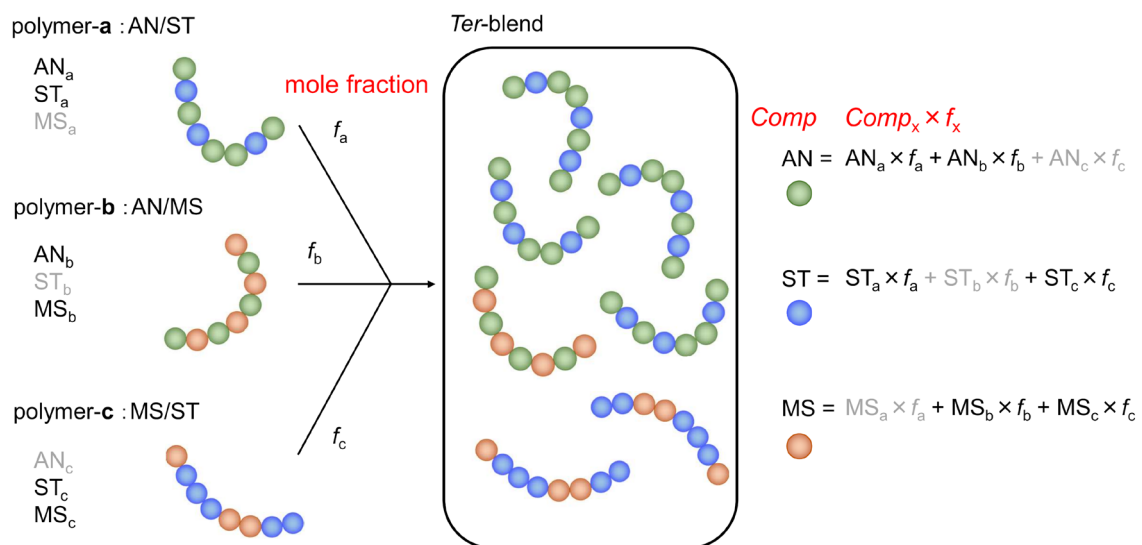
Scheme 2. Definition of the BPs in the *ter*-blends.

Fig. 8. Relationships of (a) the AN composition, (b) the MS composition, and (c) the ST composition derived from polymer-**a** (▼, ■), polymer-**b** (▼, ▲), and polymer-**c** (■, ▲) and (d) the mole fraction of polymer-**a** (●), polymer-**b** (●) and polymer-**c** (●), respectively, in the *bi*-blends predicted by LASSO with theoretical values.
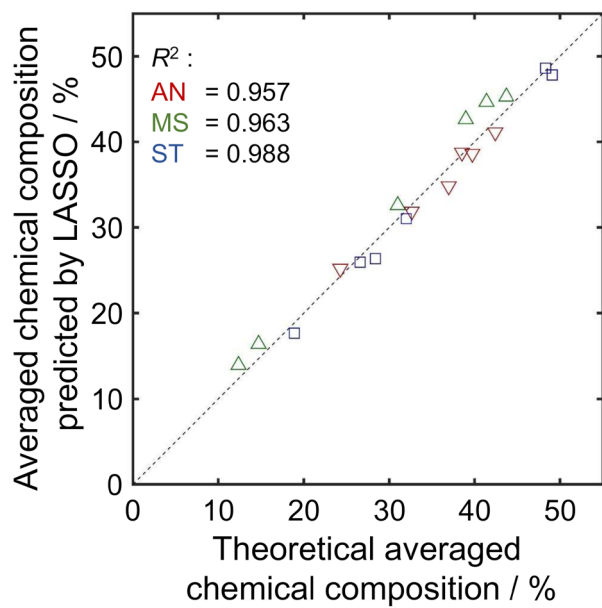
Fig. 9. Relationships of the *Comp* values for AN (∇), MS ($\triangle$), and ST ($\square$) in the *ter*-blends predicted by LASSO with the theoretical values.
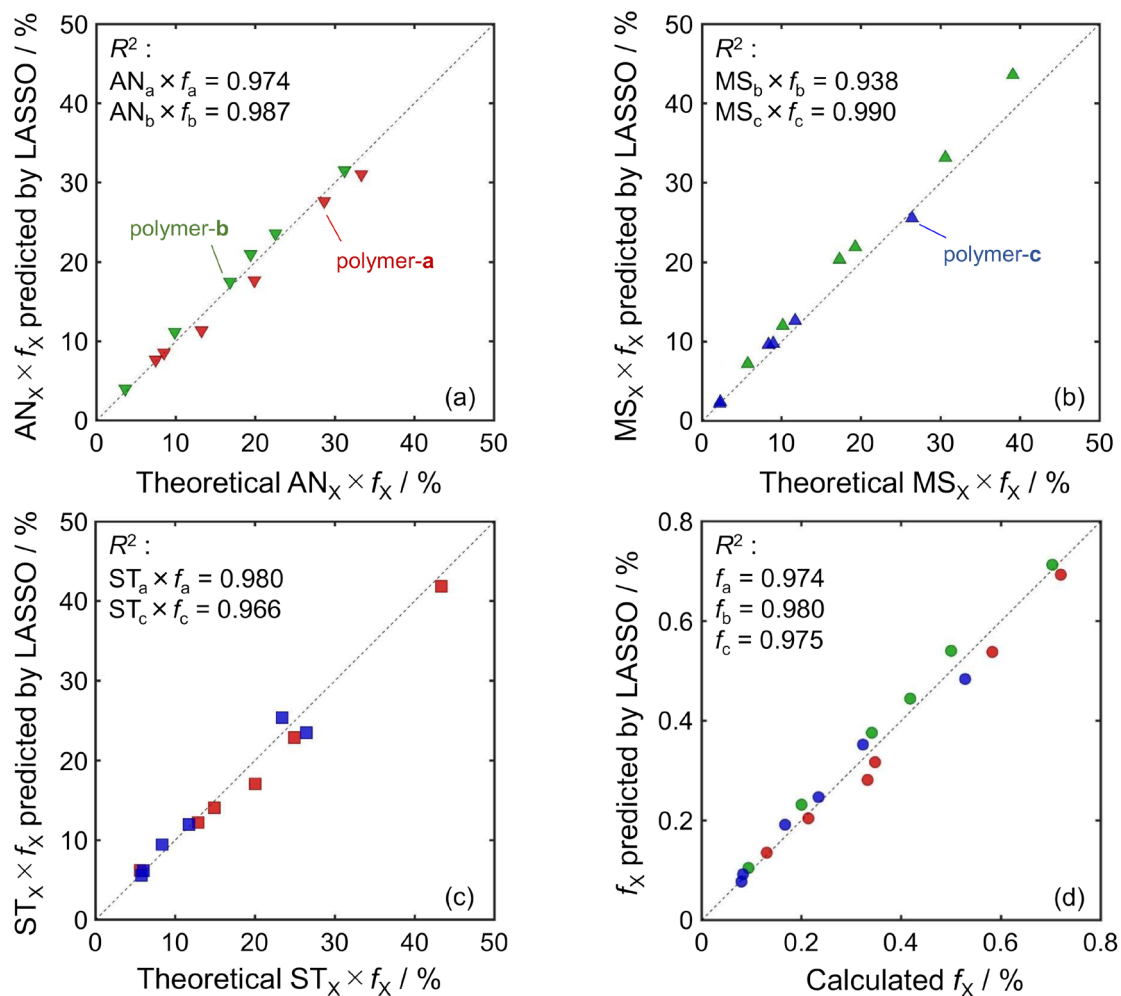
Fig. 10. Relationships of (a) the AN composition, (b) the MS composition, and (c) the ST composition derived from polymer-**a** (▼, ■), polymer-**b** (▼, ▲), and polymer-**c** (■, ▲), and (d) the mole fractions of polymer-**a** (●), polymer-**b** (●) and polymer-**c** (●), respectively, in the *ter*-blends predicted by LASSO with the theoretical values.

Table 3. Copolymer samples used as component copolymers.

Run	Chemical composition/mol%						$M_n^b \times 10^{-4}$	M_w/M_n^b
	in the feed			in the copolymer[a]				
	AN	MS	ST	AN	MS	ST		
1	60.0		40.0	57.2		42.8	4.6	2.8
2	40.0		60.0	39.8		60.2	4.0	7.0
3	61.0	39.0		53.9	46.1		5.6	2.7
4	50.0	50.0		49.3	50.7		5.6	2.4
5	45.9	54.1		44.4	55.6		4.0	2.6
6	35.8	64.2		37.4	62.6		3.9	2.4
7		27.4	72.6		27.7	72.3	3.4	2.2
8		55.0	45.0		50.0	50.0	2.1	2.5

[a] Determined by gas chromatography.

[b] Determined by size-exclusion chromatography (with THF, polystyrene standards).

Table 4. R^2 values for predictions of the BPs for the *ter*-blends with Ridge, Elastic Net, and LASSO using a dataset for the copolymers and the *bi*-blends as the training dataset.

	R^2 values											
	Ridge	Elastic Net										LASSO
	$\alpha=0$	0.19	0.20	0.32	0.57	0.84	0.86	0.95	0.96	1.00		$\alpha=1$
$AN_a \times f_a$	0.969		0.973									0.974
$AN_b \times f_b$	0.966	0.984										0.987
$MS_b \times f_b$	0.966						0.968					0.938
$MS_c \times f_c$	0.931				0.989							0.990
$ST_a \times f_a$	0.955					0.981						0.980
$ST_c \times f_c$	0.895							0.966				0.966
f_a	0.966			0.975								0.974
f_b	0.974								0.980			0.980
f_c	0.928							0.975				0.975