ORIGINAL ARTICLE

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

Hikaru Momose[1,2], Tomoya Maeda[1], Kosuke Hattori[1], Tomohiro Hirano[1] and Koichi Ute[1]

To evaluate chemical compositions and heterogeneity of comonomer sequences in methyl methacrylate (MMA)-*tert*-butyl methacrylate (TBMA) copolymers, multivariate analysis was applied to the ^{13}C nuclear magnetic resonance (NMR) spectra of the carbonyl, backbone quaternary and α-methyl carbons of the copolymers. A better linear relationship was found between the first principal component score and the chemical composition in copolymers than was found between the results from spectra of the carbonyl and backbone quaternary carbons. The chemical compositions of 16 copolymers were successfully predicted by partial least squares regression (PLSR). The second principal component was found to reflect the fraction of MMA-TBMA hetero dyad sequence. Dyad sequence distributions of copolymers that were obtained at higher conversions were successfully determined by PLSR with those of copolymers obtained at low conversions as a training set. Multivariate analysis using properly prepared samples provided us with quantitative information of chemical compositions and comonomer sequence distributions, without assignment of the ^{13}C NMR resonance peaks.

INTRODUCTION

Multivariate analysis is a powerful tool that can transform complex information into more useful sets of information, and that can extract vital differences from data that may look similar by other methods. Application of multivariate analysis to metabolite evaluations, so-called 'metabolomics', is well known. Principal component analysis (PCA) and partial least squares regression (PLSR), developed by Kowalski[1,2] and Wold,[3] have been successfully applied to many applications such as the study of the impact of stress conditions on the plant metabolome,[4] the evaluation of neurological disease progression[5] and studies on toxicological mechanisms.[6] An interesting application is the quality classification of Japanese green tea[7] or curative plants[8] through ^{1}H nuclear magnetic resonance (NMR) spectra of their extracts. Multivariate analysis has also been applied to characterization of polymer materials. PCA and PLSR were utilized for the discrimination of ethylene-vinyl acetate copolymers with different compositions and the prediction of the content of vinyl acetate in the copolymers by infrared emission,[9] Raman[10] or near-infrared[11] spectroscopy. PCA was also applied to analyze the Fourier transform infrared spectroscopy and differential scanning calorimetry

data of various types of polyethylenes to classify chain-branching types, and chain-branching content and distribution.[12]

Recently, we have reported the PCA and PLSR of ^{13}C NMR spectra of methyl methacrylate (MMA) and *tert*-butyl methacrylate (TBMA) copolymers, homopolymers and different blends of two methacrylates (poly MMA (PMMA) and poly TBMA (PTBMA)).[13] The first principal component (PC1) scores contained most of the information of the chemical compositions of polymers, and PLSR (without assigning resonance peaks) predicted the chemical compositions accurately and precisely. The second principal component (PC2) scores reflected the heterogeneity of comonomer sequences. In that case, PCA and PLSR were performed on the spectrum regions of the carbonyl and quaternary backbone carbons. Maximization of the spectral information for entire spectrum regions requires further examination.

In this paper, the shapes of PCA score plots in each spectral region were compared to determine suitable spectral regions that maximized the complete spectral information. The relationship between PC1 scores and the chemical compositions of the samples, as well as the influence on the precision and accuracy of PLSR from the difference

[1]Department of Chemical Science and Technology, The University of Tokushima, Tokushima, Japan and [2]Corporate Research Laboratories, Mitsubishi Rayon Co. Ltd, Otake, Hiroshima, Japan
Correspondence: Professor K Ute, Department of Chemical Science and Technology, The University of Tokushima, 2-1 Minami-Josanjima, Tokushima 770-8506, Japan.
E-mail: ute@chem.tokushima-u.ac.jp

of the spectrum regions set, were examined. Furthermore, we focused on quantitative analysis of the heterogeneity of comonomer sequences of MMA-TBMA copolymers. The fractions of comonomer sequences at dyad and triad levels for two blends of copolymers and seven copolymers obtained at the end stage of polymerization by PLSR were determined.

EXPERIMENTAL PROCEDURE
Materials
MMA and TBMA (supplied by Mitsubishi Rayon, Tokyo, Japan) were purified by distillation under reduced pressure. 2,2′-Azobisisobutyronitrile (Wako Pure Chemical Industries, Osaka, Japan) was recrystallized from methanol. Ethyl lactate and methanol (Kishida Chemical, Osaka, Japan) were used without purification.

Polymerization
A mixture of the monomer (20 wt%) and 2,2′-azobisisobutyronitrile (0.5 mol% of the monomer mixture) were dissolved in ethyl lactate. Polymerization was carried out at 80 °C under a nitrogen atmosphere. After 3–7 min, the polymerization mixture was cooled to room temperature and poured into a large volume of a methanol/water mixture (3/7 vol/vol). The polymer precipitated and was collected by filtration and dried over night at 60 °C *in vacuo*. Table 1 summarizes the feed ratio and chemical composition (mol% in TBMA units), M_n, M_w/M_n and the yield of the polymer samples together with those of the homopolymers and the copolymers obtained at higher yields.[13]

Measurements and procedures
The homopolymers, copolymers and homopolymer blends were dissolved in chloroform-d (8% wt/vol). ^{1}H and ^{13}C NMR spectra of the sample solutions were measured at 55 °C on a JNM-ECX400 spectrometer (JEOL, Tokyo, Japan)

Table 1 Preparation of PMMA, PTBMA and poly(MMA-*co*-TBMA)s[a]

| Code | TBMA (mol%) | | M_n^c (10^{-3}) | M_w/M_n^c | Yield (%) |
	Feed	Copolymer[b]			
H-0	0	0	9.22	2.22	83
H-24	20.0	24.0	8.33	2.46	83
H-43	40.0	43.1	8.50	2.47	85
H-65	60.0	64.5	8.47	2.51	85
H-83	80.0	82.7	8.47	2.54	87
H-100	100	100	9.18	2.39	86
M-29	25.0	29.2	6.34	1.39	35
M-54	50.0	53.8	5.70	1.56	45
M-78	75.0	78.0	4.60	1.69	55
L-6	5.0	6.0	15.9	1.79	7
L-18	15.0	18.3	15.6	1.90	2
L-28	25.0	28.4	14.3	1.87	6
L-41	35.0	40.9	15.6	1.89	6
L-56	49.9	55.5	16.0	2.13	4
L-71	65.0	70.8	14.0	2.12	5
L-79	75.0	78.6	13.9	2.10	9
L-88	85.0	88.1	14.4	2.11	4
L-93	91.9	93.2	13.0	2.25	5

Abbreviations: AIBN, 2,2′-azobisisobutyronitrile; MMA, methyl methacrylate (MMA); NMR, nuclear magnetic resonance; PMMA, poly MMA; PTBMA, poly *tert*-butyl methacrylate; SEC, size exclusion chromatography; THF, tetrahydrofuran.
[a](AIBN)$_0$ = 0.5 mol% of monomer (H-100–H-0 and L-93–L-8). (AIBN)$_0$ = 20 mol% of monomer (M-78–M-29).
[b]Determined by ^{1}H NMR.
[c]Determined by SEC in THF calibrated with standard PMMA samples.

equipped with a 10 mm multinuclear probe (^{1}H: 45° pulse (8.5 µs), pulse repetition 8.90 s, 16 scans; ^{13}C: 45° pulse (7.5 µs), pulse repetition 2.73 s, 10000 scans, with ^{1}H broadband decoupling). Chemical composition was determined from the integral intensities of ^{1}H NMR signals of the ester groups of MMA and TBMA units.

Each ^{13}C NMR spectrum was stored into 32 768 complex data points covering a spectral width 31250 Hz, and zero-filled to 13 1072 points before Fourier transformation. An exponential apodization function was applied to free induction decays, which corresponded to a line-broadening factor of 3.0 Hz. The ^{13}C NMR chemical shift of chloroform-d was set to 77.0 p.p.m. as an internal standard. Bucket integration at an interval of 0.25 p.p.m. was performed with the JEOL Alice2 ver.5 for metabolome ver.1.6 software for the following six spectral regions; 15.1–23.1 p.p.m. (α-methyl carbons), 26.0–29.0 p.p.m. (methyl carbons of *tert*-butyl groups in TBMA units), 44.1–48.1 p.p.m. (backbone quaternary carbons), 48.2–58.2 p.p.m. (ester methyl carbons of MMA units and backbone methylene carbons), 79.5–83.0 p.p.m. (quaternary carbons of *tert*-butyl groups in TBMA units) and 175–179 p.p.m. (carbonyl carbons). The sum of integral intensities in each resonance region was normalized to 100. The average integral intensity was subtracted from each integral intensity. The data set therefore consisted of mean-centered bucket integral values. PCA and PLSR of the data sets were conducted using the Sirius ver. 7.0 software (Pattern Recognition Systems, Bergen, Norway).

RESULTS AND DISCUSSION
PCA
Figure 1 shows a ^{13}C NMR spectrum of poly(MMA-*co*-TBMA) with 53.8 mol% in TBMA units (M-54 in Table 1). The spectra of α-methyl, backbone quaternary and carbonyl carbons of the copolymer were broad and complicated because of configurational and comonomer sequences. Therefore, the assignment of the individual peaks is troublesome.

The individual PCA for 27 samples (2 homopolymers, 9 blends and 16 copolymers) was performed on the six spectral regions described above. Figure 2 shows the PC1–PC2 score plots for these regions. The number in parentheses denotes the contribution rate of variance for each principal component. Spectral information in each region could mostly be explained by the PC1–PC2 score plots. The cumulative contribution rate was over 96% in each case. As shown Figures 2a–c, their score plots reflected on the spectral information of the carbonyl, backbone quaternary and α-methyl carbons were the same shapes, because their carbons were the common structures in MMA and TBMA units.

A new data set was constructed by combining bucket integral intensities of the three spectral regions of the carbonyl, backbone quaternary and α-methyl carbons to improve reliability of the PCA. Figure 3 shows the PC1–PC2 score plots for the combined data set. The contribution rates of PC1 and PC2 were 77.1% and 20.7%, respectively. Further investigation described below was made on this data set.

Statistical determination of chemical composition
Figure 4 shows ^{13}C NMR extended spectra with the carbonyl, backbone quaternary and α-methyl carbons of PMMA, PTBMA and poly(MMA-*co*-TBMA) of L-54, together with the corresponding PCA loading histogram. The loading histograms indicate the pseudo-signal reflecting each principal component in the bucket integral intensities for the three resonance regions of all sample polymers. Positive loadings of PC1 were observed at the same chemical shift owing to the signals of PMMA, whereas negative ones were observed from PTBMA. Their loading histograms indicated that PC1 reflected the chemical compositions of MMA and TBMA units in polymer samples, as described in the previous section. The relationship

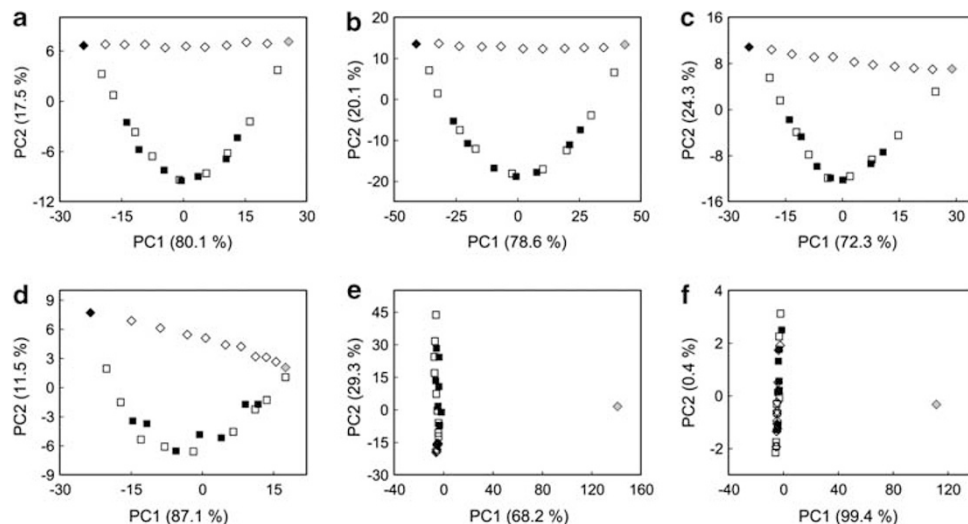**Figure 1** 100 MHz ^{13}C NMR spectrum of poly(MMA-*co*-TBMA) with 55.5 mol% in TBMA units (L-56),as measured in chloroform-d at 55 °C. (**a**) Whole, (**b**) carbonyl carbon, (**c**) backbone quaternary carbon and (**d**) α-methyl carbon regions.

Figure 2 First (PC1) and second (PC2) principal component score plots for (**a**) the carbonyl, (**b**) the backbone quaternary, (**c**) the α-methyl, (**d**) the backbone methylene and the methyl group of MMA pendant, (**e**) the quaternary of TBMA pendant and (**f**) the methyl groups of TBMA pendant ^{13}C NMR signals of PMMA (◆), PTBMA (◆), homopolymer blends (◇), copolymers obtained at high conversions (■) and low conversions (□). The number in parentheses is the contribution rate of variance for each principal component.
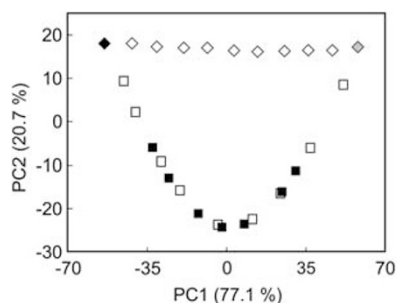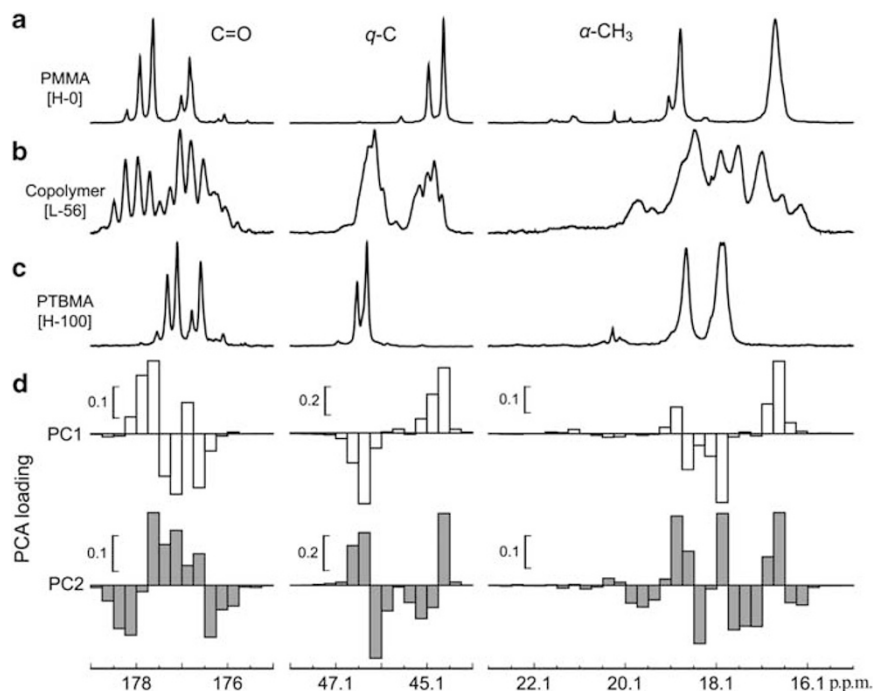

Figure 3 First (PC1) and second (PC2) principal component score plots for the combined ^{13}C NMR signals in the carbonyl, backbone quaternary and α-methyl groups of two homopolymers and 16 copolymers. See Figure 2 for each symbol. The number in parentheses is the contribution rate of variance for each principal component.

between the PC1 score (*y*) and the TBMA unit composition in mol% (*x*) was derived as the following equation with a correlation coefficient (R^2) of 0.998:

$$y = -1.12x + +58.7$$

The R^2 is slightly better than that reported in our previous paper ($R^2 = 0.997$),[13] where the two spectral regions due to the carbonyl and backbone quaternary carbons were used for the PCA.

To predict TBMA unit composition in polymer samples, a PLSR model was constructed with bucket-integral data sets of the three spectral regions and the TBMA unit composition data of the two homopolymers and nine their blends as a training (reference) set. The first latent variable (LV1) was used for the regression model because 99.5% of the spectral information was explained with only LV1. The LV1 loading histograms, shown in Figure 5, were the same shape as the PC1 loading. Therefore, this model was reasonable to predict

Figure 4 ^{13}C NMR spectra of the carbonyl, backbone quaternary and α-methyl carbons of (**a**) PMMA (H-0), (**b**) poly(MMA-*co*-TBMA) (L-56) and (**c**) PTBMA (H-100). (**d**) The corresponding loading histograms for the first (PC1) and second (PC2) principal component of the ^{13}C NMR spectra.
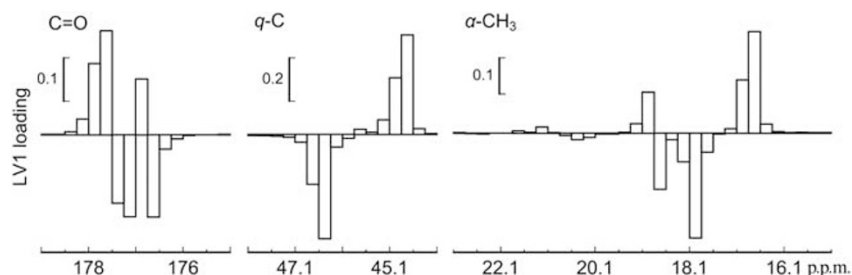
Figure 5 The loading histograms of the PLSR model for predicting chemical compositions in samples, constructed in the ^{13}C NMR spectra and TBMA unit compositions with the PMMA, PTBMA and nine of their blends.
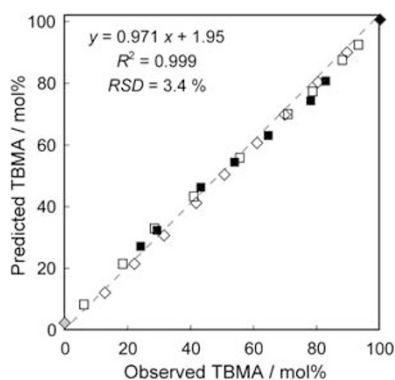
Figure 6 Correlation between predicted TBMA compositions by partial least square regression and observed by 1H NMR for PMMA (♦), PTBMA (♦), nine homopolymer blends (◇) and 16 copolymers (□, ■). The solid line indicates the ideal correlation. Training (reference) set model: PMMA, PTBMA and nine homopolymer blends.

chemical compositions in samples. Figure 6 shows the relationship between the TBMA unit compositions determined by ^{1}H NMR and those predicted by PLSR. An excellent relationship was obtained with an R^2 of 0.999, which is also slightly better than that in the previous paper ($R^2 = 0.998$).[13] The accuracy of predicted TBMA unit compositions by PLSR was evaluated with relative standard deviation, *RSD*. The *RSD* was calculated to be 3.4%, which is slightly worse than that of the 2.8% (this *RSD* of 2.8% is as same sense as the absolute error of 3.4% indicated in our previous report) in the previous paper.[13] This is because of the two methyl groups in the 2,2′-azobisisobutyronitrile fragment,[14] which were observed in the same region as the α-methyl carbons.

Relationship between the PC2 score and the heterogeneity of monomer sequences

The PC2-positive loadings in the PCA loading histograms, as shown in Figure 4, appeared at 176.50–177.75 p.p.m. in the resonance region of carbonyl carbon, at 44.60–44.85 p.p.m. and 46.35–46.85 p.p.m. in that of backbone quaternary carbon, 16.60–17.10 17.85–18.10 and

18.60–19.10 p.p.m. in that of α-methyl carbon, and corresponded to the signals of the two homopolymers. Conversely, the characteristic signals of copolymers, especially at 175.75–176.50, 178.00–179.00 and 44.85–46.35 p.p.m., arose on the negative side in PC2 loadings. The prime characteristic of copolymers is the presence of different monomeric sequences. Therefore, PC2 should reflect the heterogeneity of monomer sequences, as the stereoregularity of the copolymers is not expected to vary greatly with the feed monomer ratio in this copolymerization.[15]

To obtain quantitative information about the heterogeneity of monomer sequences from PC2, the fraction of hetero dyad sequence f_{12}, consisting of M_1 and M_2 units was used as an index. The radical copolymerization of methacrylate monomers generally proceeds according to the terminal model. The f_{12} in the copolymer is expressed as Equation (1),[16]

$$f_{12} = \frac{2P_{12}P_{21}}{P_{12} + P_{21}} \tag{1}$$

where P_{12} and P_{21} denote the probability of addition to M_2 monomer by the M_1-ended radical and *vice versa*, respectively. P_{12} and P_{21} are defined by Equations (2) and (3) in the case of low conversions,[17]

$$P_{12} = \frac{2F_2}{1 + \sqrt{1 + 4(r_1 r_2 - 1)F_1 F_2}} \tag{2}$$

$$P_{21} = \frac{2F_1}{1 + \sqrt{1 + 4(r_1 r_2 - 1)F_1 F_2}} \tag{3}$$

where r_1 and r_2 denote reactivity ratios of monomers M_1 and M_2, respectively, and F_1 and F_2 denote chemical composition of M_1 and M_2 units in the copolymer obtained at an early stage of polymerization ($F_1 + F_2 = 1$). Thus, the f_{12} in the copolymer obtained at low conversions is calculated by Equation (4):

$$f_{12} = \frac{4F_1 F_2}{1 + \sqrt{1 + 4(r_1 r_2 - 1)F_1 F_2}} \tag{4}$$

The f_{12} depends on comonomer composition. A copolymer with equimolar composition, $F_1 = F_2 = 0.5$, maximizes f_{12}. A completely random copolymer whose comonomer sequence distribution obeys Bernoullian statistics ($r_1 = r_2 = 1$) gives $f_{12} = 0.5$ at $F_1 = 0.5$. Homopolymer and completely alternating copolymers exhibit extreme values of f_{12}, 0 and 1, respectively. The f_{12} is equivalent to the RN defined by Harwood *et al.*[18] (run number = 100 f_{12}).

The monomer reactivity ratios of MMA, r_M and TBMA, r_T were calculated to be 0.81 ± 0.06 and 1.26 ± 0.03 with an R^2 of 0.997, respectively, by the Kelen-Tüdõs method[19] using nine copolymers obtained at early stages of copolymerization (L-6–L-93, Table 1). These were comparable to previously reported values ($r_M = 0.96$ and $r_T = 1.35$ by Yuki *et al.*,[15] $r_M = 0.68$ and $r_T = 1.29$ by Zhao *et al.*[20]). The f_{MT} (M_1: MMA, M_2: TBMA) for L-6–L-93 were calculated with Equation (4). The good relationship between the PC2 score (w) and the f_{MT} (z) of L-6–L-93 and the two homopolymers was obtained as the following equation with an R^2 of 0.996:

$$w = -83.6z + 18.6$$

PC2 scores succeeded in reducing the quantitative information of the fraction of hetero dyad sequences, f_{MT}.

Statistical determination of dyad sequence distribution by PLSR

Dyad sequence distributions (dyad sequence distributions mean, strictly speaking, the distributions of sequence of two monomeric units in each copolymer chain. In this report, we simply define them as the fractions of each dyad sequence in measured samples) of

copolymers should be determined by PLSR with a training set of copolymers obtained at low conversions, because the information of the fraction of two homo dyads, M_1–M_1: f_{11} and M_2–M_2: f_{22} were also included in the three spectral regions. The f_{11} in the copolymers obtained at an early stage of polymerization is expressed as Equation (5),[16]

$$f_{11} = \frac{P_{21}P_{11}}{P_{12} + P_{21}} = F_1 \left(1 - \frac{2F_2}{1 + \sqrt{1 + 4(r_1 r_2 - 1)F_1 F_2}}\right) \tag{5}$$

where P_{11} denotes the probabilities of addition to the M_1 monomer by the M_1-ended radical ($P_{11} = 1 - P_{12}$).[17] f_{22} is also expressed by replacing suffix number 1 with 2 in Equation (5). The fractions of three dyad sequences, f_{MM}, f_{MT}, f_{TT} for PMMA, PTBMA and the copolymers of L-6–L-93 were calculated from Equations (4) and (5), as summarized in Table 2. Chemical composition was adopted as f_{MM} and f_{TT} of the blends of two homopolymers with $f_{MT} = 0$. Leave-one-out cross-validation[21] was performed by PLSR to validate the training set models of f_{MM}, f_{MT} and f_{TT}, which were constructed with 20 ^{13}C NMR spectra and theoretical fractions in Table 2. Each cumulative contribution rate of three PLSR models, as summarized in Table 3, was over 99% with two latent variables, LV1 and LV2. Figure 7 shows

Table 2 Theoretical fractions of dyad sequences for training set models[a]

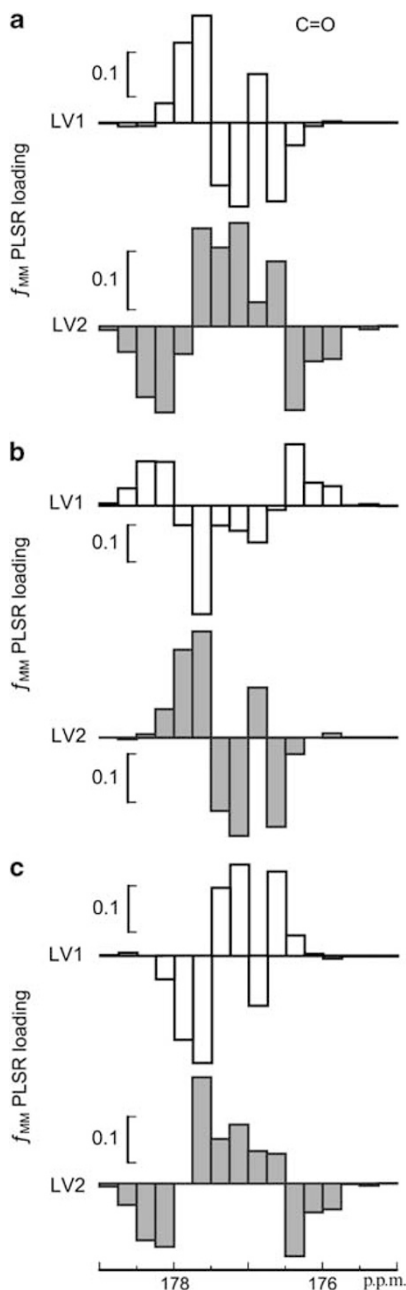| Code	f_{MM} (%)	f_{MT} (%)	f_{TT} (%)
H-0	100	0	0
B-13[b]	87.2	0	12.8
B-22[b]	77.9	0	22.1
B-32[b]	68.4	0	31.6
B-42[b]	58.2	0	41.8
B-51[b]	49.3	0	50.7
B-61[b]	38.9	0	61.1
B-70[b]	30.1	0	69.9
B-80[b]	19.6	0	80.4
B-90[b]	13.1	0	86.9
H-100	0	0	100
L-6	88.4	11.2	0.4
L-18	66.7	29.9	3.4
L-28	51.3	40.6	8.1
L-41	35.0	48.2	16.8
L-56	19.9	49.2	30.9
L-71	8.6	41.3	50.1
L-79	4.6	33.5	61.8
L-88	1.4	21.0	77.5
L-93	0.5	12.7	86.8

[a]Calculated from Equations (4) and (5).
[b]Data from Momose *et al.*[13]

Table 3 Contribution rates of PLSR models to predict the fractions of dyad sequences

LV	f_{MM} (%)	f_{MT} (%)	f_{TT} (%)
LV1	94.7	97.4	93.4
LV2	4.9	2.3	6.4
Total	99.6	99.7	99.8

Abbreviation: PLSR, partial least squares regression.

Figure 7 The loading histograms of the PLSR model, which predicts the fractions of three dyad sequences in samples, constructed with the ^{13}C NMR spectra and theoretical fractions of dyad sequences with the two homopolymers, and nine copolymers and their blends obtained at early stages of copolymerization.

the spectra regions of the quaternary and α-methyl carbons showed the same results as those of the carbonyl carbon. Therefore their PLSR models were reasonable to predict the three fractions of dyad sequences in copolymer samples. Figure 8 (see Supplementary Table S1 and Figure S1) shows the PLSR models predicting f_{MM}, f_{MT} and f_{TT}. This indicates that a good fit between theoretical and predicted fractions was obtained, with an R^2 of 0.999–0.997 and an *RSD* of 3.5–7.1%. Thus, their PLSR models should be able to predict the fractions of three dyad sequences with good accuracy and precision from spectra of unknown samples.

Verifying adaptability to a wide variety of MMA-TBMA copolymers, the prediction of f_{MM}, f_{MT} and f_{TT} of the two copolymers, which were blended with different large copolymers of dyad sequence distributions, were initially tried on the PLSR models in Figure 8. Table 4 shows the amount of the original copolymers and the TBMA unit compositions of the copolymer blends.

The fractions of dyad sequences were determined by PLSR models in Figure 8. Table 2 summarizes the fractions of three dyad sequences calculated from the blend molar ratio of the original copolymers and those predicted by PLSR. Calculation procedures are described in Supplementary information. The predicted fractions of dyad sequences were excellent in relation to their predicted values, with an R^2 of 0.995 and an *RSD* of 2.6% for six of the fractions of three dyad sequences. It was confirmed that the fractions of dyad sequences of copolymer samples were predicted with good accuracy and precision by using those training set models.

Figure 9 shows the predicted distribution of seven copolymers obtained at higher conversions, H-83–H-24 and M-78–M-29, using the training set model. These plots deviated slightly from the simulation for the copolymers obtained at early stages of copolymerization owing to the difference of conversions. The average fractions of each dyad sequence at full conversions calculated by the Spinner method[22] showed the same tendency as Figure 8 (see Supplementary Table S1 and Supplementary Figure S1).
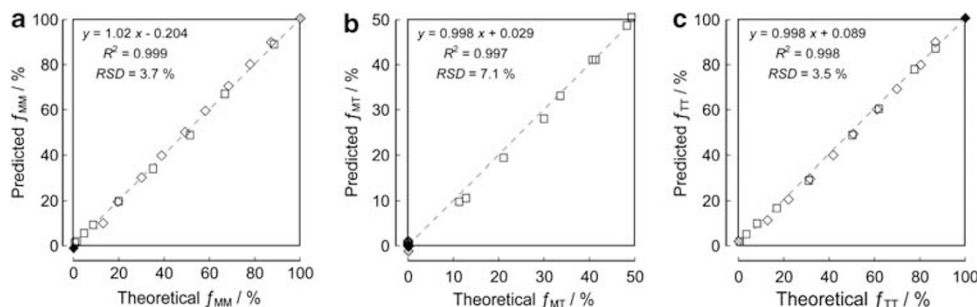
PCA for side chain structures
PCA was also performed for the quaternary carbons (Figure 2e) and the methyl carbons (Figure 2f) of the *tert*-butyl ester groups in TBMA units. Only the plot of the PMMA homopolymer deviated far from the other plots, indicating that PMMA homopolymer was easily distinguished from other samples. This was because no signals were observed in these resonance regions of the spectrum of PMMA.

The PCA score plots for the backbone methylene carbons and the methyl carbons of methyl ester groups in MMA units (Figure 2d) were a similar shape to those for α-methyl, backbone quaternary and carbonyl carbons (Figures 2a–c), although the PC1 plot distances of the blends of homopolymers narrowed with increasing PC1 scores. The narrow plot distance can be attributed to incorporation of the signals from the methyl ester groups, which only belong to MMA monomeric units. If the signals of the methyl carbons of methyl ester groups in MMA units could be separated from those of the methylene carbons, the PCA score plots for the backbone methylene groups would be similar to those for the α-methyl, backbone quaternary and carbonyl carbons (Figures 2a–c).

CONCLUSIONS
We applied multivariate analysis of ^{13}C NMR spectra of PMMA, PTBMA, their blends and poly(MMA-*co*-TBMA)s to compositional and sequential characterization. The two principal components of PCA explained the spectra of the majority of samples tested. PC1 and PC2 indicated the chemical composition and the fraction of hetero

the LV1 and LV2 loading histograms of their PLSR models on the spectra regions of carbonyl carbons. The shapes of the LV1 and LV2 loading histograms of the f_{MM} model coincided with those of PC1 and PC2 on PCA loadings, and indicated that the LV1 and LV2 of f_{MM} model caught information on the homo and hetero dyad sequences in the polymer samples. The LV1 of the f_{MT} model were the same as LV2 of the f_{MM} model reflected the hetero dyad sequences. The positive and negative LV1 loading of the f_{TT} model showed the same shape of PTBMA and PMMA spectra, which were the opposite sign of those of f_{MM} model. The loading histograms on

Figure 8 PLSR models for predicting fractions of (**a**) MMA homo (f_{MM}), (**b**) MMA-TBMA hetero (f_{MT}) and (**c**) TBMA homo (f_{TT}) dyad sequences from ^{13}C NMR signals of the carbonyl, backbone quaternary and α-methyl carbons of PMMA (♦), PTBMA (♦), their blends (◇) and poly(MMA-*co*-TBMA)s obtained at early stages of copolymerization (□).

Table 4 Calculated and predicted fractions of dyad sequences for the blends of copolymers obtained at low conversions

|	Original copolymer		TBMA	f_{MM} (%)		f_{MT} (%)		f_{TT} (%)	
Code	I	II	(mol%)[a]	Calc.[b]	Pred.[c]	Calc.[b]	Pred.[c]	Calc.[b]	Pred.[c]
CB-1	L-71	L-28	47.0	32.5	32.2	40.9	42.1	26.6	25.7
CB-2	L-6	L-93	48.4	45.6	46.3	11.9	12.6	42.4	41.7
		Total R^2				0.995			
		Total RSD (%)				2.6			

Abbreviations: Calc., calculated; NMR, nuclear magnetic resonance; PMMA, poly methyl methacrylate; pred., predicted; PTBMA, poly *tert*-butyl methacrylate.
[a]Determined by ^{1}H NMR.
[b]Calculation procedure of fractions of dyad sequences is described in the Supplementary Information.
[c]Predicted condition: 2LVs. Training set models: the resonance regions owing to the carbonyl, quaternary and α-methyl carbons, and the samples owing to PMMA (H-0), PTBMA (H-100), blends of homopolymers (B-13–B-90) and copolymers obtained at low conversions (L-6–L-93).
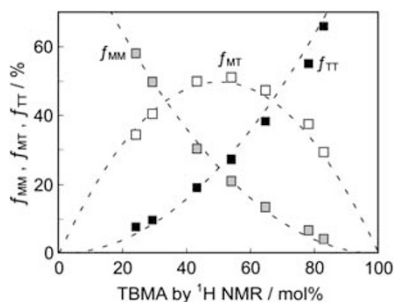
Figure 9 Dyad sequence distributions of the copolymer obtained at higher conversions. Each dashed line indicates the dyad sequence calculated from Equations (4) and (5) obtained at low conversions.

dyad sequence, respectively. Also, the dyad sequence distributions of copolymers obtained at higher conversions were successfully determined by PLSR using homopolymers and the copolymers at early stages of copolymerization as a training data set, without assignment of ^{13}C NMR resonance peaks. We are currently investigating a quantitative determination method of stereoregularity in addition to chemical composition and comonomer sequence distribution of poly(MMA-*co*-TBMA).

1 Kowalski, B. R. & Bender, C. F. Pattern recognition. Powerful approach to interpreting chemical data. *J. Am. Chem. Soc.* **94,** 5632–5639 (1972).
2 Kowalski, B. R. & Bender, C. F. Pattern recognition. II. Linear and nonlinear methods for displaying chemical data. *J. Am. Chem. Soc.* **95,** 686–693 (1973).
3 Wold, S. Spline functions in data analysis. *Technometrics* **16,** 1–11 (1974).
4 Krishnan, P., Kruger, N. J. & Ratcliffe, R. G. Metabolite fingerprinting and profiling in plants using NMR. *J. Exp. Bot.* **56,** 255–265 (2005).
5 Holmes, E., Tsang, T. M. & Tabrizi, S. J. The application of NMR-based metabonomics in neurological disorders. *NeuroRx* **3,** 358–372 (2006).
6 Coen, M., Holmes, E., Lindon, J. C. & Nicholson, J. K. NMR-based metabolic profiling and metabonomic approaches to problems in molecular toxicology. *Chem. Res. Toxicol.* **21,** 9–27 (2008).
7 Tarachiwin, L., Ute, K., Kobayashi, A. & Fukusaki, E. ^{1}H NMR based metabolic profiling in the evaluation of Japanese green tea quality. *J. Agric. Food Chem.* **55,** 9330–9336 (2007).
8 Tarachiwin, L., Katoh, A., Ute, K. & Fukusaki, E. Quality evaluation of Angelica acutiloba Kitagawa roots by ^{1}H NMR-based metabolic fingerprinting. *J. Pharm. Biomed. Anal.* **48,** 42–48 (2008).
9 Pell, R. J., Erickson, B. C., Hannah, R. W., Callis, J. B. & Kowalski, B. R. Quantitative infrared emission spectroscopy using multivariate calibration. *Anal. Chem.* **60,** 2824–2827 (1988).
10 Shimoyama, M., Maeda, H., Matsukawa, K., Inoue, H., Ninomiya, T. & Ozaki, Y. Discrimination of ethylene/vinyl acetate copolymers with different composition and prediction of the vinyl acetate content in the copolymers using Fourier-transform Raman spectroscopy and multivariate data analysis. *Vib. Spectrosc.* **14,** 253–259 (1997).
11 Shimoyama, M., Hayano, S., Matsukawa, K., Ninomiya, T. & Ozaki, Y. Discrimination of ethylene/vinyl acetate copolymers with different composition and prediction of the content of vinyl acetate in the copolymers and their melting points by near-infrared spectroscopy and chemometrics. *J. Polym. Sci. Part B Polym. Phys.* **36,** 1529–1537 (1998).
12 Hughes, J., Shanks, R. & Cerezo, F. Characterisation of the comonomer composition and distribution of copolymers using chemometric techniques. *J. Therm. Anal. Calorim.* **76,** 1069–1078 (2004).
13 Momose, H., Hattori, K., Hirano, T. & Ute, K. Multivariate analysis of ^{13}C NMR spectra of methacrylate copolymers and homopolymer blends. *Polymer* **50,** 3819–3821 (2009).
14 Hatada, K., Kitayama, T., Terawaki, Y. & Chujo, R. Effect of NMR signal from initiator fragment on triad tacticities of poly(methyl methacrylate) formed by 2, 2-azobisiso-butyronitrile and benzoyl peroxide. *Polym. J.* **19,** 1127–1129 (1987).
15 Yuki, H., Okamoto, Y., Shimada, Y., Ohta, K. & Hatada, K. Microstructure of the copolymers of methyl methacrylate with other methacrylates obtained by radical and anionic copolymerizations in tetrahydrofuran. *J. Polym. Sci. Polym. Chem. Ed.* **17,** 1215–1225 (1979).
16 Ito, K. & Yamashita, Y. Copolymer composition and microstructure. *J. Polym Sci. Part A General Papers* **3,** 2165–2187 (1965).
17 Tosi, C. Graphical representation of composition and sequence distribution in copolymers. *Makromol. Chem.* **108,** 307–311 (1967).
18 Harwood, H. J. & Ritchey, W. M. The characterization of sequence distribution in copolymers. *J. Polym. Sci. Part B Polym. Lett.* **2,** 601–607 (1964).
19 Kelen, T. & Tüdős, F. Analysis of the linear methods for determining copolymerization reactivity ratios. I. A new improved linear graphic method. *J. Macromol. Sci. Chem.* **9,** 1–27 (1975).
20 Zhao, Y., Luo, Y. W., Ye, C., Li, B. G. & Zhu, S. Model-based design and synthesis of gradient MMA/*t*BMA copolymers by computer-programmed semibatch atom transfer radical copolymerization. *J.Polym. Sci. Part A Polym. Chem.* **47,** 69–79 (2009).
21 Lorber, A. & Kowalski, B. R. Alternatives to cross-validatory estimation of the number of factors in multivariate calibration. *Appl. Spectrosc.* **44,** 1464–1470 (1990).
22 Spinner, I. H., Lu, B. C. Y. & Graydon, W. F. Copolymer composition distribution. *J. Am. Chem. Soc.* **77,** 2198–2200 (1955).

Supplementary Information accompanies the paper on Polymer Journal website (http://www.nature.com/pj)