Support Information for

Thermogravimetry-Synchronized, Reference-Free Quantitative Mass Spectrometry for Accurate Compositional Analysis of Polymer Systems Without Prior Knowledge of Constituents

Yusuke Hibi[1]*, Shiho Uesaka[1] and Masanobu Naito[1]*

[1]Data-driven Polymer Design Group, Research Center for Macromolecules and Biomaterials, National Institute for Materials Science (NIMS); 1-2-1, Sengen, Tsukuba, Ibaraki 305-0047, Japan.

*Corresponding authors. Emails: hibi.yusuke@nims.go.jp and naito.masanobu@nims.go.jp

This PDF file includes:

Materials and Methods

Supplementary Figures S1-S2

Supplementary Table 1-2

Captions for Data S1 to S3

Reference list cited in this support information file

Other Supplementary Materials for this manuscript include the following:

Data S1 to S3 are attached in csv format.

Materials and Methods

Materials

Poly(ethyl methacrylate) (E, M_w=515,000), Poly(methyl methacrylate) (M, M_w=667,00) and polystyrene (S, M_w=18,100) were purchased from Polymer Source, Shodex and Tosoh (M_n=102,00), respectively. Diglycidyl ether of bisphenol A-based epoxy (Gly, M_w=1,650) was purchased from Mitsubishi Chemical. Poly(propylene glycol) diamine (Jeff, M_w=2,000) and polydimethylsiloxane diamine (Silox, M_w=2,500) were purchased from Sigma Aldrich. Perfluorotributylamine (PFTBA) and 4,4'-di-*tert*-butylbiphenyl (dtBbph) were purchased from Tokyo Chemical Industry. TG-MS samples were prepared as described in main text. Random copolymer of methyl methacrylate and styrene with the monomer composition of 1:1 was synthesized following the reported procedure[1].

Methods

JMS-T2000GC (JEOL) with EI ion source connected to TG-DTA8122 (Rigaku) was used in this study. Ion source was kept at 300 ºC. Unless otherwise specified, the ionization voltage was set at 70 eV. The connecting tube between MS and TG was kept at 350 ºC. Helium gas was used for transporting the pyrolyzed gas from TG combustion chamber to MS (200 mL/min). Before the measurements, TOFMS was calibrated along m/z axis using PFTBA. MS peak drift correction was conducted using internal standard peak of dtBbph (266.2029 m/z) evaporated around 200-250 ºC. The MS profile spectra were converted into centroid spectra using msAxel (JEOL, noise threshold: 100, peak merge torrelance: 2 mDa). Pyrolysis was conducted at elevated temperature from 50–600 ºC at heating rate of 25 ºC/min

(22 min) after preheating period (40–50 ºC at heating rate of 2 ºC/min). Both TG and MS were recorded at every seconds.

The output MS spectra were imported into Python3.9 using netCDF4.0 package. After MS intensity and delay correction as described in the main text, the spectra recorded at 250–600 ºC (8-22 min) were extracted and divided into 10 spectra. This formatted spectral dataset right before being analyzed by (TG-) RQMS is available in attached Data S1 with all sample information and the corresponding TG data. TG-RQMS use TG data to convert spectral fragmentation abundances (FA) into weight-base FA. Despite this conversion and different ionization/MS methods, the hyperparameters used for TG-RQMS was set almost same as those for previously reported RQMS (Table S1), suggesting the robustness for hyperparameter selections in (TG-) RQMS.

Supplementary Figures and Tables

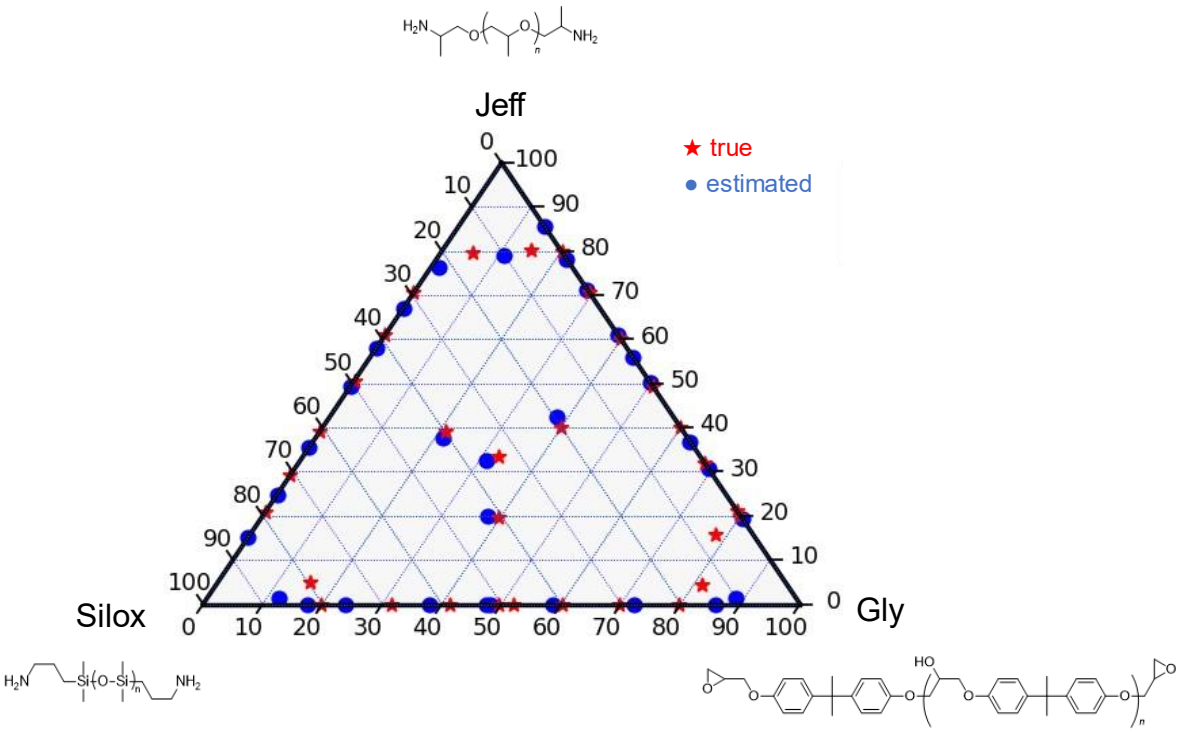

Fig. S1. CA for a strongly interacted system via TG-RQMS. Glycidyl group is highly reactive with primary or secondary amines, especially at elevated temperature. Here, glycidyl groups of Gly would react with primary amines of Jeff and Silox during sample preparation and/or pyrolysis, which may generate new peaks of cross products. Therefore, the linear mixing assumption (Eq. 1) did not hold for this reactive polymer system, compromising the CA accuracy (RMSE=0.067) as compared to CA for non-interacted system (RMSE=0.013, see Fig. 3B).

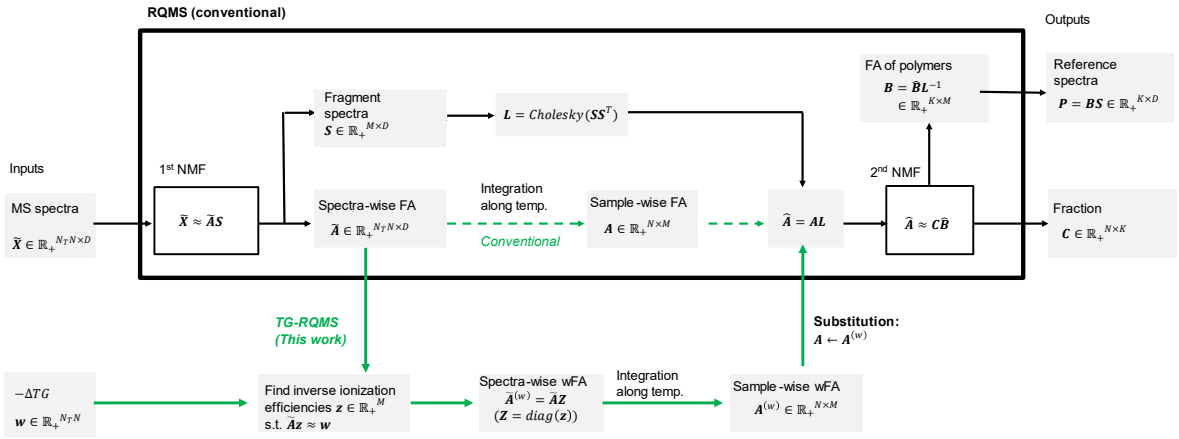

Fig. S2. Flow chart of TG-RQMS as compared to the reported RQMS. The modified parts are indicated in green.

Table S1.

Numerical data of CA results for E/M/S ternary system of Fig. 3.

	Inferred Composition			Known Composition		
sample Name	M	S	E	M	S	E
E10M10S80	0.10	0.80	0.10	0.10	0.79	0.12
E10M10S80_re	0.07	0.80	0.13	0.07	0.79	0.14
E10M80S10	0.82	0.09	0.09	0.80	0.09	0.11
E10M80S10_re	0.80	0.11	0.09	0.78	0.12	0.10
E20M40S40	0.41	0.39	0.20	0.40	0.40	0.19
E20S80_re	0.01	0.82	0.17	0.00	0.80	0.20
E30M30S30	0.32	0.34	0.34	0.33	0.34	0.33
E30M70	0.73	0.00	0.27	0.72	0.00	0.28
E30S70_re	0.01	0.72	0.27	0.00	0.71	0.29
E40M20S40	0.20	0.39	0.41	0.21	0.39	0.41
E40M40S20	0.38	0.21	0.42	0.39	0.20	0.40
E40M60	0.61	0.00	0.39	0.60	0.00	0.40
E40S60_re	0.01	0.57	0.42	0.00	0.58	0.43
E50M50	0.49	0.00	0.51	0.50	0.00	0.50
E50S50_re	0.01	0.50	0.50	0.00	0.50	0.50
E60M40	0.38	0.00	0.62	0.40	0.00	0.60
E60S40_re	0.00	0.40	0.60	0.00	0.41	0.59
E70M30	0.28	0.00	0.72	0.30	0.00	0.70
E70S30_re	0.00	0.27	0.73	0.00	0.29	0.71
E80M10S10	0.09	0.14	0.77	0.10	0.15	0.75
E80M10S10_re	0.16	0.09	0.75	0.18	0.10	0.73
E80M20_re	0.17	0.00	0.83	0.19	0.00	0.81
E80S20	0.00	0.20	0.80	0.00	0.21	0.80
E80S20_re	0.00	0.20	0.80	0.00	0.22	0.78
M20S80	0.22	0.78	0.00	0.20	0.80	0.00
M30S70	0.32	0.68	0.00	0.30	0.70	0.00
M40S60	0.41	0.59	0.00	0.40	0.60	0.00
M50S50_re	0.52	0.48	0.00	0.50	0.50	0.00
M60S40_re	0.62	0.38	0.00	0.61	0.39	0.00

M70S30_re	0.72	0.28	0.00	0.71	0.29	0.00
M80S20_re	0.82	0.17	0.00	0.81	0.20	0.00

Table S2.

Numerical data of CA results for E/M/R/S quaternary system of Fig. 4.

	Inferred Composition				Known Composition			
Sample Names	R	M	S	E	R	M	S	E
M10S10E80	0.01	0.06	0.10	0.83	0.00	0.10	0.10	0.80
M10S80E10	0.00	0.08	0.82	0.11	0.00	0.10	0.79	0.11
M20E80	0.00	0.13	0.00	0.87	0.00	0.19	0.00	0.81
M20S80	0.02	0.16	0.83	0.00	0.00	0.19	0.81	0.00
M30E70	0.00	0.22	0.00	0.78	0.00	0.29	0.00	0.71
M30S70	0.03	0.22	0.76	0.00	0.00	0.20	0.81	0.00
M40E60	0.00	0.36	0.00	0.64	0.00	0.41	0.00	0.59
M40S60	0.00	0.30	0.70	0.00	0.00	0.29	0.71	0.00
M50E50	0.00	0.47	0.00	0.53	0.00	0.50	0.00	0.50
M50S50	0.00	0.39	0.60	0.00	0.00	0.39	0.61	0.00
M60E40	0.00	0.57	0.00	0.43	0.00	0.59	0.00	0.41
M60S40	0.00	0.52	0.48	0.00	0.00	0.48	0.52	0.00
M70E30	0.00	0.69	0.00	0.31	0.00	0.70	0.00	0.30
M70S30	0.04	0.60	0.36	0.00	0.00	0.61	0.39	0.00
M80E20	0.00	0.79	0.00	0.21	0.00	0.79	0.00	0.21
M80S10E10	0.03	0.83	0.10	0.04	0.00	0.81	0.10	0.09
M80S20	0.03	0.69	0.28	0.00	0.00	0.70	0.30	0.00
R10M10E80	0.12	0.07	0.00	0.81	0.12	0.10	0.00	0.78
R10M10S80	0.18	0.06	0.76	0.00	0.12	0.09	0.79	0.00
R10M80E10	0.13	0.81	0.00	0.06	0.11	0.79	0.00	0.11
R10M80S10	0.17	0.75	0.08	0.00	0.11	0.79	0.10	0.00
R10S10E80	0.13	0.00	0.10	0.77	0.12	0.00	0.10	0.78
R10S80E10	0.11	0.00	0.82	0.08	0.10	0.00	0.80	0.10
R20E80	0.21	0.00	0.00	0.79	0.21	0.00	0.00	0.79
R20M80	0.20	0.80	0.00	0.00	0.22	0.78	0.00	0.00
R20S80	0.17	0.00	0.83	0.00	0.20	0.00	0.80	0.00
R30E70	0.32	0.00	0.00	0.68	0.30	0.00	0.00	0.70
R30M70	0.35	0.65	0.00	0.00	0.32	0.68	0.00	0.00
R30S70	0.28	0.00	0.72	0.00	0.30	0.00	0.70	0.00

R40E60	0.44	0.00	0.00	0.56	0.41	0.00	0.00	0.59
R40M60	0.44	0.56	0.00	0.00	0.40	0.60	0.00	0.00
R40S60	0.41	0.00	0.59	0.00	0.40	0.00	0.60	0.00
R50E50	0.57	0.00	0.00	0.43	0.52	0.00	0.00	0.48
R50M50	0.49	0.45	0.00	0.06	0.50	0.50	0.00	0.00
R50S50	0.52	0.00	0.48	0.00	0.50	0.00	0.50	0.00
R60E40	0.66	0.00	0.00	0.34	0.60	0.00	0.00	0.40
R60M40	0.61	0.34	0.01	0.04	0.61	0.39	0.00	0.00
R60S40	0.60	0.00	0.40	0.00	0.61	0.00	0.39	0.00
R70E30	0.75	0.01	0.00	0.24	0.70	0.00	0.00	0.30
R70M30	0.68	0.27	0.02	0.04	0.67	0.33	0.00	0.00
R70S30	0.71	0.00	0.29	0.00	0.70	0.00	0.30	0.00
R80E20	0.85	0.02	0.00	0.14	0.81	0.00	0.00	0.19
R80M10E10	0.85	0.10	0.00	0.06	0.81	0.10	0.00	0.09
R80M10S10	0.89	0.11	0.00	0.00	0.80	0.09	0.11	0.00
R80M20	0.82	0.18	0.00	0.00	0.79	0.21	0.00	0.00
R80S10E10	0.88	0.00	0.07	0.05	0.82	0.00	0.09	0.09
R80S20	0.77	0.00	0.23	0.00	0.79	0.00	0.21	0.00
S20E80	0.02	0.00	0.19	0.78	0.00	0.00	0.21	0.79
S30E70	0.04	0.00	0.32	0.64	0.00	0.00	0.32	0.68
S40E60	0.04	0.00	0.40	0.56	0.00	0.00	0.41	0.59
S50E50	0.02	0.00	0.52	0.46	0.00	0.00	0.50	0.50
S60E40	0.05	0.00	0.60	0.36	0.00	0.00	0.60	0.40
S70E30	0.00	0.00	0.73	0.27	0.00	0.00	0.70	0.30
S80E20	0.06	0.00	0.78	0.16	0.00	0.00	0.80	0.20

Table S3.

Hyperparameters for benchmark CA tests via RQMS[1].

Dataset	Sample number N	First NMF				Second NMF		
		w_o	Merging threshold	Initial M	iteration	K	$\alpha = \beta$	ρ
Previously reported benchmark test[1]	24	0.1/0.2/0.3	0.99	30	3000	3	0.1	1/1.5
E/M/S (Fig. 3)	32	0.1	0.99	30	3000	3	0.1	1.5
E/M/S/R (Fig. 4)	58	0.1	0.99	30	3000	4	0.1	1.5

The meaning of parameters for the first NMF for consolidating the peaks into basis spectra are as follows.

w_o: The stringent level of orthogonality.[2] The higher w_o means less shared peaks among the basis spectra. w_o should be given between 0 and 1, corresponding to no orthogonality constraint and most strict constraint.

Merging threshold and initial M: These parameters are related to automatic relevance determination (ARD) mechanism. The initial basis number should be set larger than the appropriate number, and the ARD mechanism reduces the number of basis spectra by merging, when the cosine similarity of spectra goes beyond "merging threshold".[2]

The meaning of parameters for the second NMF for calculating the final output of composition are as follows.

K: The number of system constituents.

α: The regularization hyper parameter controlling the minimum volume constraint for the simplex spanned by the reference spectra. The simplex would be shrunk more with higher α.[3]

β: The regularization hyper parameter controlling the orthogonality among the reference spectra. The simplex would be expanded more with the higher β. [1]

ρ: The hyper parameter controlling the robustness of outlier should be specified between [0.5, 2]. Higher ρ makes the second NMF more sensitive to the data variation but less robust to the outliers.[3]

Caption for Data S1-S3.

Formatted ready-to-use spectral datasets with TG and sample information for non-interacted ternary M/S/E (Data S1), quaternary M/S/E/R (Data S2), interacted Gly/Jeff/Silox systems (Data S3) are attached in csv format.

Reference

(1) Hibi, Y.; Uesaka, S.; Naito, M. A Data-Driven Sequencer That Unveils Latent "Codons" in Synthetic Copolymers. *Chem. Sci.* **2023**, *14*, 5619-5626.

(2) Shiga, M.; Tatsumi, K.; Muto, S.; Tsuda, K.; Yamamoto, Y.; Mori, T.; Tanji, T. Sparse Modeling of EELS and EDX Spectral Imaging Data by Nonnegative Matrix Factorization. *Ultramicroscopy*, **2016**, *170*, 43–59.

(3) Fu, X.; Huang, K.; Yang, B.; Ma, W. K.; Sidiropoulos, N. D. Robust Volume Minimization-Based Matrix Factorization for Remote Sensing and Document Clustering. *IEEE Transactions on Signal Processing* **2016**, *64* (23), 6254–6268.