Electronic Supplementary Information

Machine-Learning-Assisted Prediction of the Size of Microgels Prepared by Aqueous Precipitation Polymerization

Daisuke Suzuki,*[a][b] Haruka Minato,[a][b] Yuji Sato,[a][b] Ryuji Namioka,[b] Yasuhiko Igarashi, [c] Risako Shibata,[d] and Yuya Oaki*[d]

[a]Graduate School of Environmental, Life, Natural Science and Technology, Okayama University, 3-1-1 Tsushimanaka, Kita-ku, Okayama, 700-8530, Japan.
E-mail: d_suzuki@okayama-u.ac.jp (D.S.)

[b]Graduate School of Textile Science & Technology, Shinshu University, 3-15-1 Tokida, Ueda, Nagano 386-8567, Japan

[c]Faculty of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8573, Japan

[d]Department of Applied Chemistry, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan.
E-mail: oakiyuya@applc.keio.ac.jp (Y.O.)

Table of Contents

Experimental Details

Materials

N-isopropyl acrylamide (NIPAm, purity 98%), *N*,*N*′-methylenebis(acrylamide) (BIS, 97%), potassium peroxodisulfate (KPS, 95%), sodium dodecyl sulfate (SDS, 95%), hydrochloric-acid solution (HClaq), hydroxide solution (NaOHaq), and sodium chloride (NaCl, 99%) were purchased from FUJIFILM Wako Pure Chemical Corporation (Japan) and used as received. Acrylic acid (AAc, 99%) was purchased from Sigma-Aldrich and used as received. Distilled and ion-exchanged water was used for all experiments, including the preparation of solutions and dispersions (EYELA, SA-2100E1).

Microgel synthesis

Aqueous free radical precipitation polymerization was used to synthesize 66 types of microgels according to the conditions shown in Table S1. Initially, an aqueous solution containing the monomers, NIPAm (and AAc), and the crosslinker BIS was poured into a four-necked round-bottom flask equipped with a mechanical stirrer and a reflux condenser. Subsequently, the monomer solution was heated to the constant polymerization temperature (Table S1) in an oil bath. Simultaneously, the dissolved oxygen in the solution was removed by sparging with nitrogen for at least 0.5 h. Then, the surfactant SDS and initiator KPS dissolved in water were added, respectively, to the solution to initiate the polymerization. The polymerization was allowed to proceed for 4 h, and the resultant microgel dispersion was cooled to room temperature to stop the polymerization. The obtained microgels were purified using at least two rounds of centrifugation and redispersion in water and/or by dialysis with pure water for several days.

Scanning electron microscopy (SEM)

The morphology and size uniformity of the 70 types of microgels were evaluated using SEM (JEOL Ltd., JCM-7000) and field emission (FE)-SEM (JEOL Ltd., JSM-IT800SHL). For this purpose, droplets of diluted microgel dispersions were dropped and dried on polystyrene substrates at room temperature. Subsequently, the samples were sputtered with Pt/Pd (15 mA, 6 Pa, 80-320 s) prior to observation.

Dynamic light scattering (DLS)

The hydrodynamic diameter (D_h) of the 70 types of microgels was measured using DLS (Zetasizer Nano S, Malvern Instrument Ltd.). For the preparation of the samples, 1 mL of diluted microgel dispersions with 1 mM ionic strength was used after adjusting to pH = 3. The samples were measured at a constant temperature (25 °C or 40 °C); equilibration was ensured at each temperature by applying a resting period (10 min) prior to the DLS measurements. The correlation function of the scattering intensity $g_2(\tau)$ and the correlation function of the scattering electric field $g_1(\tau)$ are given by the following equations:

$$g_2(\tau) - 1 = \beta |g_1(\tau)|^2 \qquad \text{(Eq. S1)}$$

$$g_1(\tau) = \exp(-Dq^2\tau) \qquad \text{(Eq. S2)}$$

where τ is the decay time, β is the extrapolated value at $\tau = 0$, and D is the diffusion coefficient; q is the scattering vector. The vacuum wavelength λ of the irradiating laser beam was 633 nm, and the total scattering angle θ was 173°. All D_h values were calculated using the Stokes–Einstein equation (Eq. S3), based on the average of three measurement cycles.

$$D_{\mathrm{h}} = \frac{kT}{3\pi\eta D} \qquad \text{(Eq. S3)}$$

Here, k is the Boltzmann constant, T is the absolute temperature (25 °C = 298 K, 40 °C = 318 K), and η is the viscosity of water at 298 K or 318 K. The D_{h} values were calculated using software (Malvern, Zetasizer software v. 7.12) or manually using the slope of the graph in Eq. S2 when the accuracy of the automatic software calculation seemed to be poor based on the correlation function data.

Construction of the prediction model [ref. S1]

The dataset in Table S1 was used for machine learning. Using ES-LiR, linear regression models were prepared for all the combinations of x_n (2^n-1 patterns), i.e., {x_1 only}, {x_1, x_2}, {x_1, x_3}, {x_1, x_4}, … ,{x_1, x_n}, {x_2, x_3}, …,{x_{n-1}, x_n}, {x_1, x_2, x_3}, {x_1, x_2, x_4}, … {x_1, x_2, …, x_n}. Each x_n has an option whether it is used or not. The total 2^n combinations of the multiple linear regression models are constructed exhaustively. As the case without the use of any x_n was removed, total 2^n-1 models were constructed. The prediction accuracy of the models was verified by cross-validation error (CVE). After the constructed models were sorted in the ascending order of the CVE values, the coefficients were summarized in the weight diagram. This algorithm was implemented in Python. The descriptors were first extracted from the weight diagram visually. Then, the further selection, including the addition and removal, was carried out based on our chemical insight. After the descriptors were selected based on the weight diagram, five-fold cross-validation was carried out to validate the model (Figs. S4 and S5).

Results and Discussion

Table S1. List of the experimental values of the explanatory variables (x_n: n = 1–10) and the objective variables (y)

Code	x_1 NIPAm [mM]	x_2 BIS [mM]	x_3 AAc [mM]	x_4 Monomer [mM]	x_5 SDS [mM]	x_6 NaCl [mM]	x_7 KPS [mM]	x_8 Water [mL]	x_9 stirring speed [rpm]	x_{10} T [°C]	y_1 D_h (40 °C, pH = 3) [nm]	y_2 D_h (25 °C, pH = 3) [nm]	References
1	147.8	0.8	1.5	150	0	0	2	100	250	70	369	1380	
2	147.8	1.5	0.8	150	0	0	2	100	250	70	332	1780	
3	147.0	1.5	1.5	150	0	0	2	100	250	70	324	1163	[ref. S2]
4	141.0	1.5	7.5	150	0	0	2	100	250	70	383	1175	[ref. S2]
5	133.5	1.5	15.0	150	0	0	2	200	250	70	567	1522	[ref. S2]
6	103.5	1.5	45.0	150	0	0	2	200	250	70	491	990	
7	103.5	1.5	45.0	150	0	0	2	500	250	70	329	1312	
8	73.5	1.5	75.0	150	0	0	2	100	250	70	338	586	[ref. S2]
9	141.0	7.5	1.5	150	0	0	2	100	250	70	415	707	
10	135.0	7.5	7.5	150	0	0	2	100	250	70	405	768	
11	127.5	7.5	15.0	150	0	0	2	200	250	70	410	966	
12	97.5	7.5	45.0	150	0	0	2	200	250	70	381	621	
13	120.0	15.0	15.0	150	0	0	2	200	250	70	435	829	
14	90.0	15.0	45.0	150	0	0	2	200	250	70	400	597	
15	94.7	5.0	0.0	100	0	0	3.6	56	250	60	320	566	
16	141.0	1.5	7.5	150	1	0	2	200	250	70	146	320	
17	141.0	1.5	7.5	150	5	0	2	100	250	70	89	133	
18	133.5	1.5	15.0	150	1	0	2	200	250	70	171	511	
19	120.0	15.0	15.0	150	1	0	2	200	250	70	204	304	
20	118.5	1.5	30.0	150	5	0	2	100	250	70	64	265	
21	115.5	4.5	30.0	150	4	0	2	100	250	70	79	161	
22	105.0	15.0	30.0	150	1	0	2	100	250	70	249	358	
23	112.5	22.5	30.0	150	1	0	2	100	250	70	221	576	
24	97.5	7.5	30.0	150	1	0	2	100	250	70	375	618	
25	88.5	16.5	45.0	150	1	0	2	100	250	70	385	607	
26	87.0	18.0	45.0	150	1	0	2	100	250	70	299	489	
27	85.5	19.5	45.0	150	1	0	2	100	250	70	361	662	
28	104.3	0.8	45.0	150	1	0	2	100	250	70	181	243	
29	100.5	4.5	45.0	150	4	0	2	100	250	70	99	228	
30	82.5	22.5	45.0	150	5	0	2	100	250	70	103	149	
31	78.0	27.0	45.0	150	5	0	2	100	250	70	151	193	
32	90.0	15.0	45.0	150	5	0	2	100	250	70	93	141	
33	75.0	30.0	45.0	150	7	0	2	100	250	70	230	127	
34	94.0	5.0	1.0	100	0	0	2	100	250	70	369	739	
35	94.0	5.0	10.0	100	1	0	2	100	250	70	157	316	
36	188.0	10.0	10.0	200	0	0	2	100	250	70	663	2844	
37	282.0	3.0	15.0	300	1	0	2	100	250	70	160	334	
38	130.5	4.5	15.0	150	1	10	2	100	250	70	269	609	
39	130.5	4.5	15.0	150	1	30	2	100	250	70	717	4758	
40	144.0	1.5	4.5	150	0	0	2	500	400	70	382	1412	
41	147.0	1.5	1.5	150	0	0	0.5	100	250	60	397	2211	
42	139.5	3.0	7.5	150	0	0	1	100	250	60	442	1875	
43	138.0	4.5	7.5	150	0	0	2	100	250	60	469	2244	
44	99.0	1.0	0.0	100	0	0	2	100	250	70	457	1433	
45	294.0	3.0	3.0	300	0	0	2	100	250	70	811	3080	
46	144.0	4.5	1.5	150	0	0.1	1	100	250	70	449	1132	
47	138.0	4.5	7.5	150	0	1	2	100	250	70	422	1803	
48	130.5	4.5	15.0	150	0	5	1	100	250	70	626	5184	
49	138.0	4.5	7.5	150	0	10	2	100	250	70	735	N.A.	
50	144.0	1.5	4.5	150	0	0	2	100	0	70	453	1685	
51	138.0	7.5	4.5	150	0	0	2	100	100	70	442	1744	
52	144.0	1.5	4.5	150	1	0	2	500	100	70	132	311	
53	138.0	7.5	4.5	150	0	0	2	100	400	70	498	1272	
54	147.0	1.5	1.5	150	0	0	0.5	100	250	80	370	2238	
55	145.5	3.0	1.5	150	0	0	1	100	250	80	296	633	
56	130.5	4.5	15.0	150	0	0	2	100	250	80	373	707	
57	123.0	12.0	15.0	150	0	0	4	100	250	80	437	700	
58	131.0	4.5	15.0	150	1	20	2	100	250	70	536	1134	
59	148.5	1.5	0.0	150	0.1	0	2	100	250	70	245	579	
60	148.5	1.5	0.0	150	1	0	2	100	250	70	147	257	
61	148.5	1.5	0.0	150	8	0	2	100	250	70	56	97	
62	148.5	1.5	0.0	150	16	0	2	100	250	70	47	54	
63	131.0	4.5	15.0	150	0	0	2	50	250	70	551	1604	
64	131.0	4.5	15.0	150	0	0	2	750	250	70	415	1886	
65	131.0	4.5	15.0	150	0	0	2	1000	250	70	431	2806	
66	131.0	4.5	15.0	150	0.1	10	2	100	250	70	454	2691	
67	131.0	4.5	15.0	150	0	0	2	500	600	70	544	N.A.	
68	131.0	4.5	15.0	150	0	0	2	500	800	70	412	1670	
69	131.0	4.5	15.0	150	0	0	2	500	1000	70	545	N.A.	
70	131.0	4.5	15.0	150	0	0	2	500	1200	70	627	N.A.	

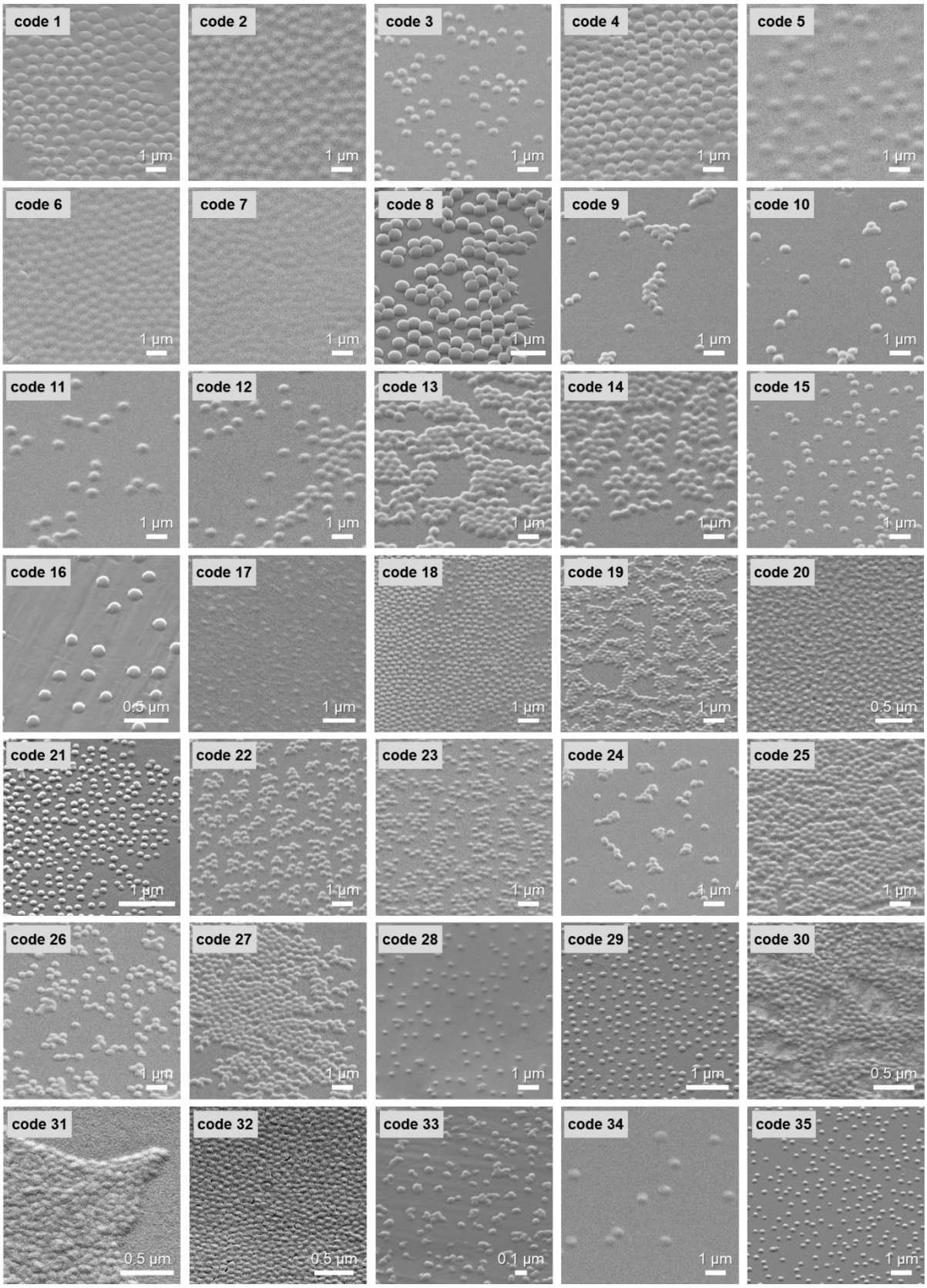

Figure S1 (continues) Representative SEM and FE-SEM images of the microgels.

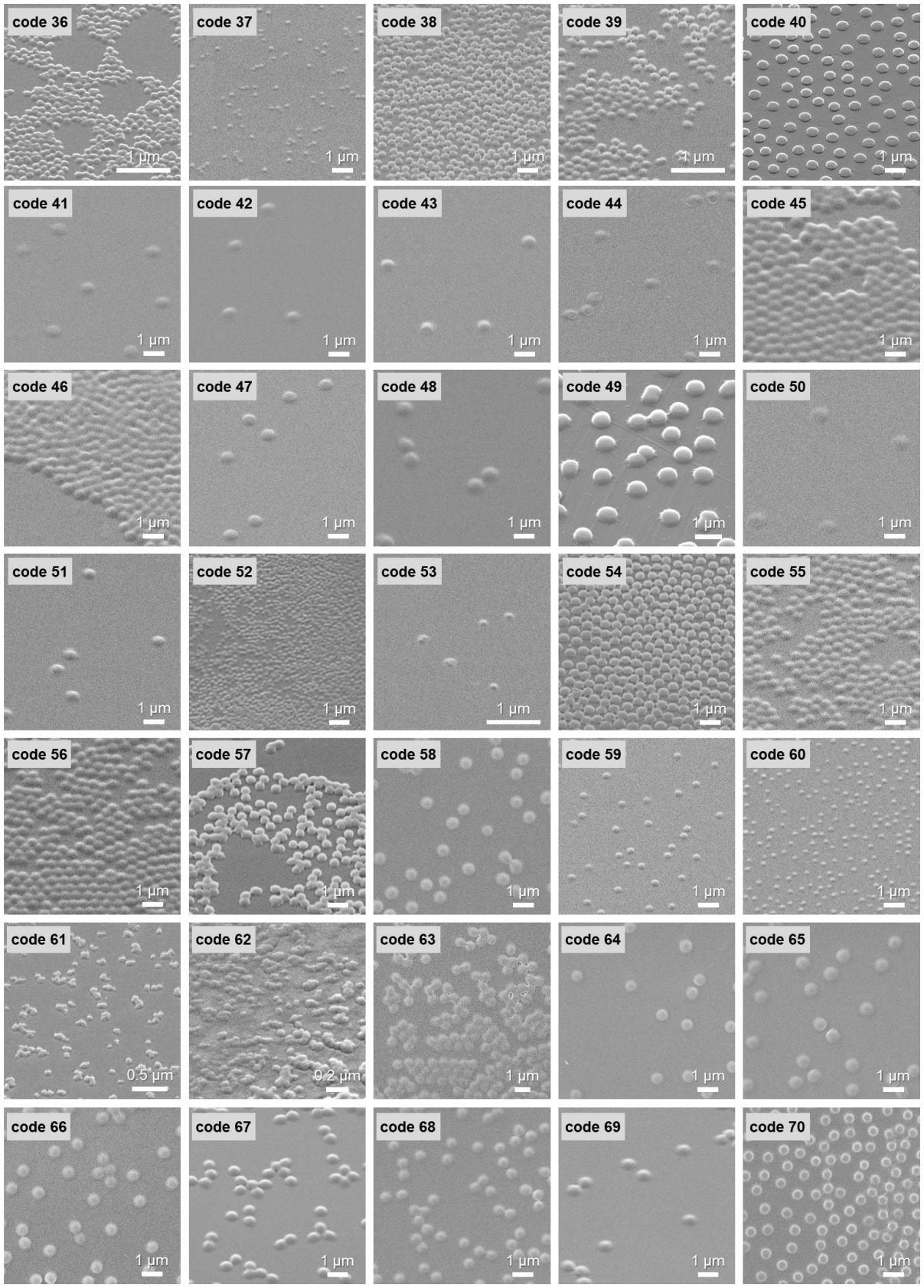

Figure S1 (continued) Representative SEM and FE-SEM images of the microgels.

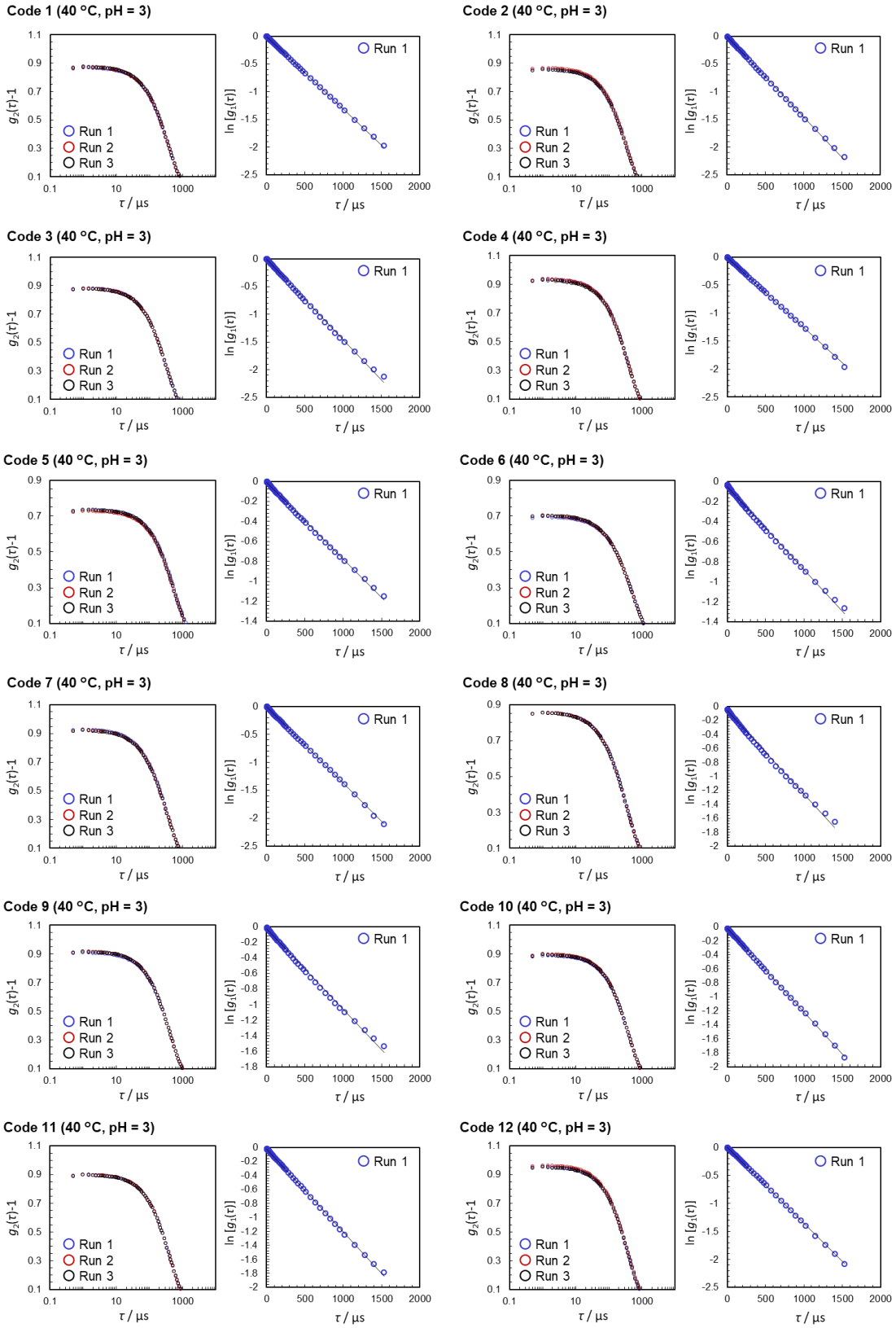

Figure S2 (continues) Time–correlation function of the scattering intensity, $g_2(\tau)-1$, and calculated time-correlation function of the scattering electric field, $\ln[g_1(\tau)]$, of each microgel (y_1).

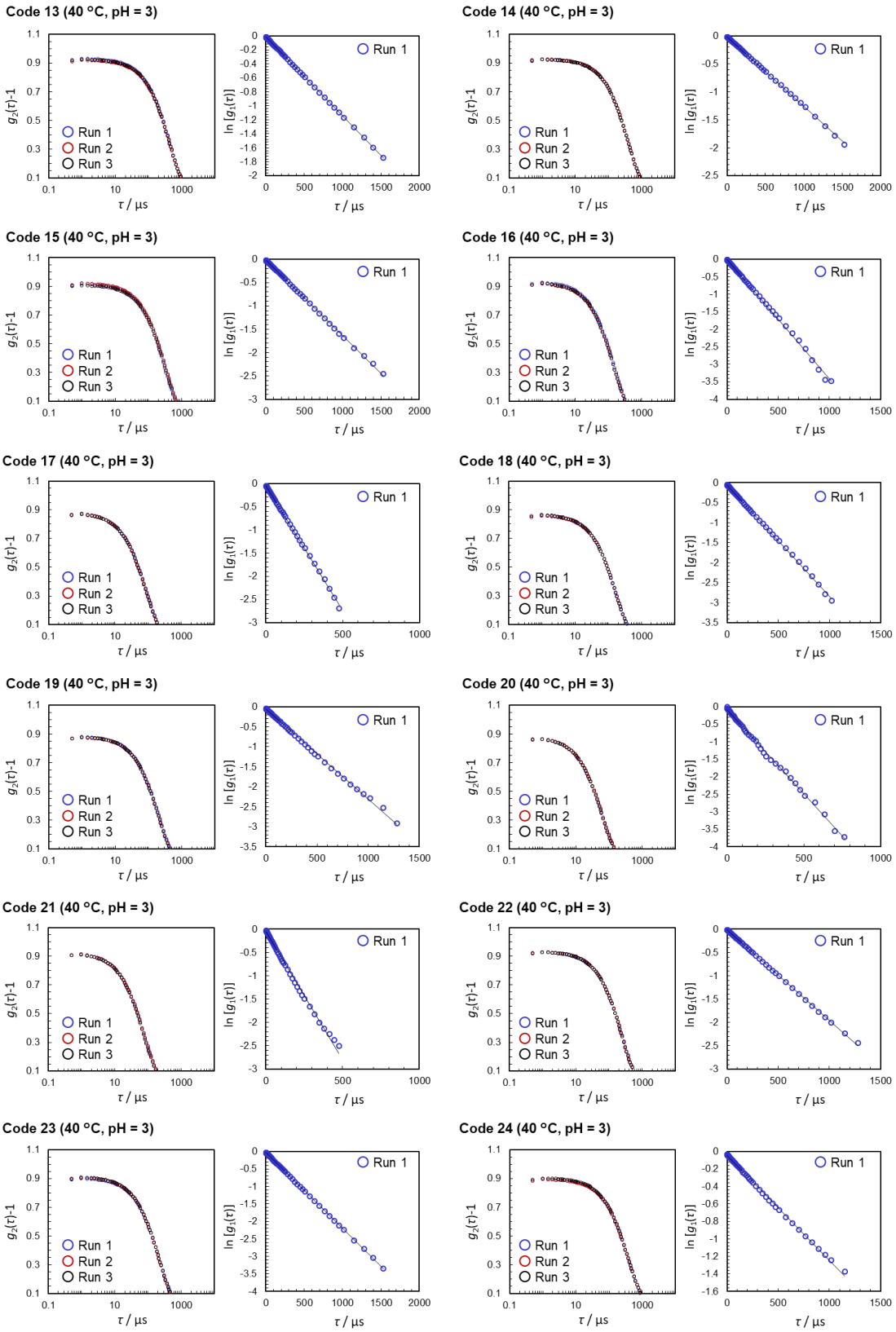

Figure S2 (continues) Time–correlation function of the scattering intensity, $g_2(\tau)-1$, and calculated time-correlation function of the scattering electric field, $\ln[g_1(\tau)]$, of each microgel (y_1).

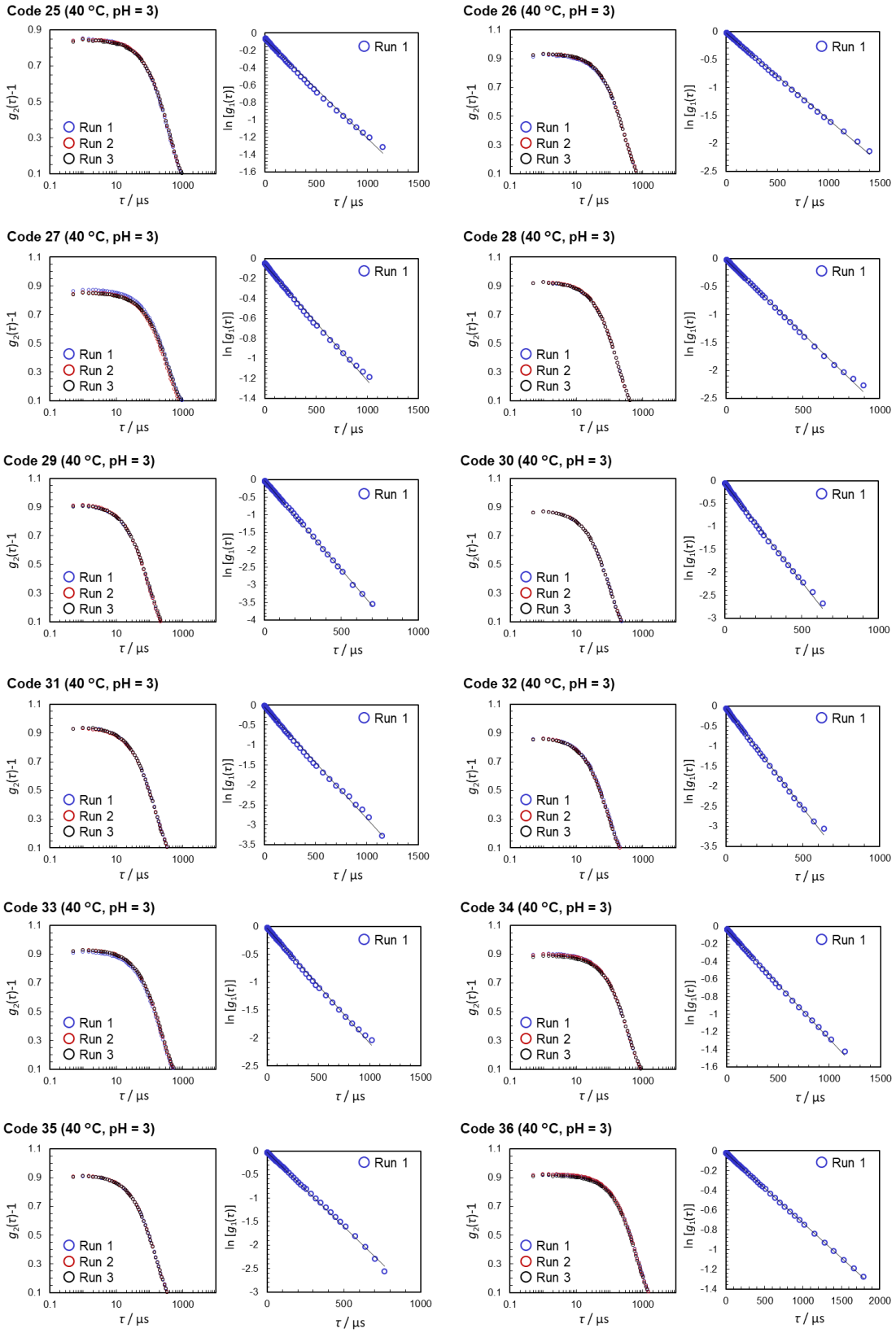

Figure S2 (continues) Time–correlation function of the scattering intensity, $g_2(\tau)-1$, and calculated time-correlation function of the scattering electric field, $\ln[g_1(\tau)]$, of each microgel (y_1).

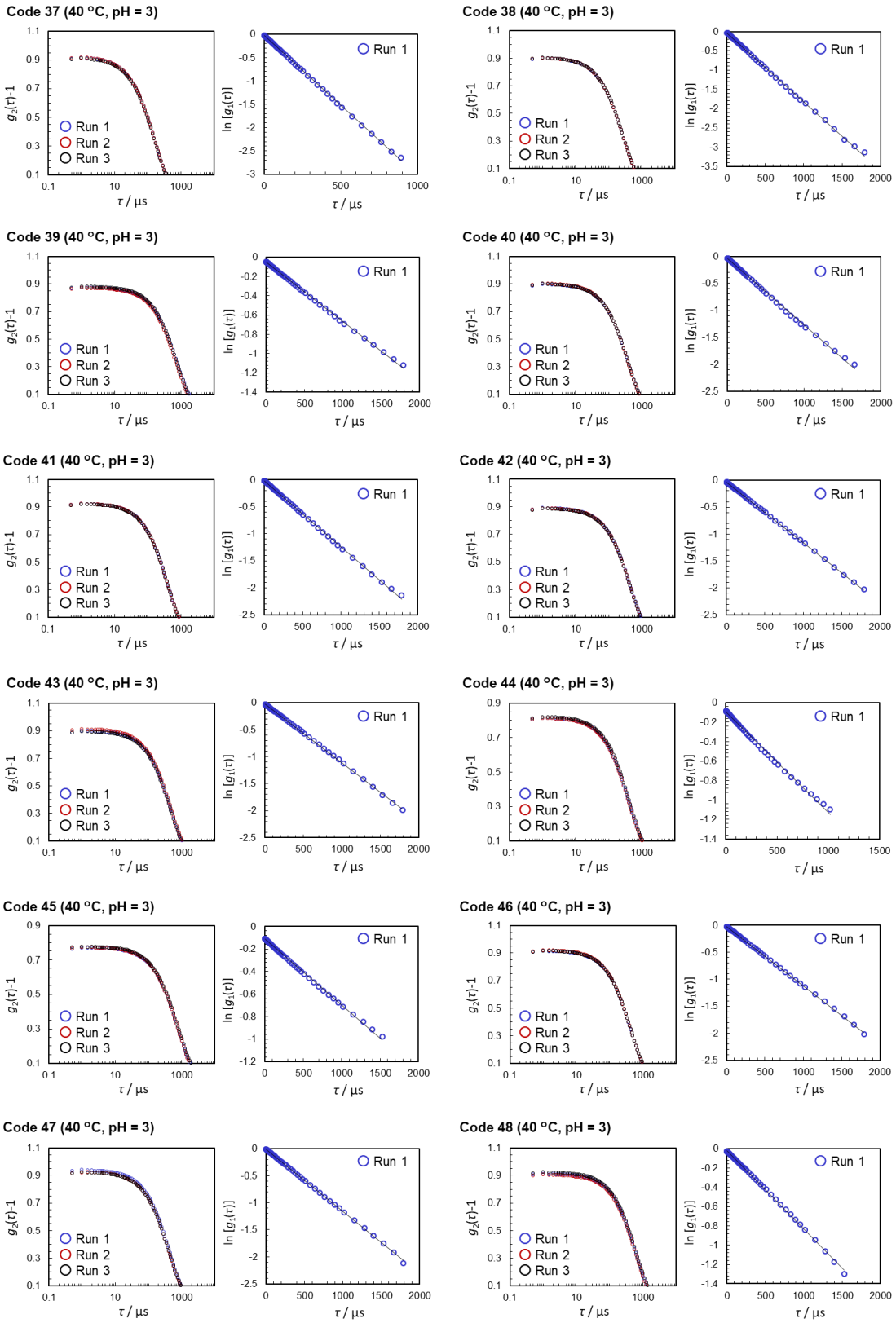

Figure S2 (continues) Time–correlation function of the scattering intensity, $g_2(\tau)-1$, and calculated time-correlation function of the scattering electric field, $\ln[g_1(\tau)]$, of each microgel (y_1).

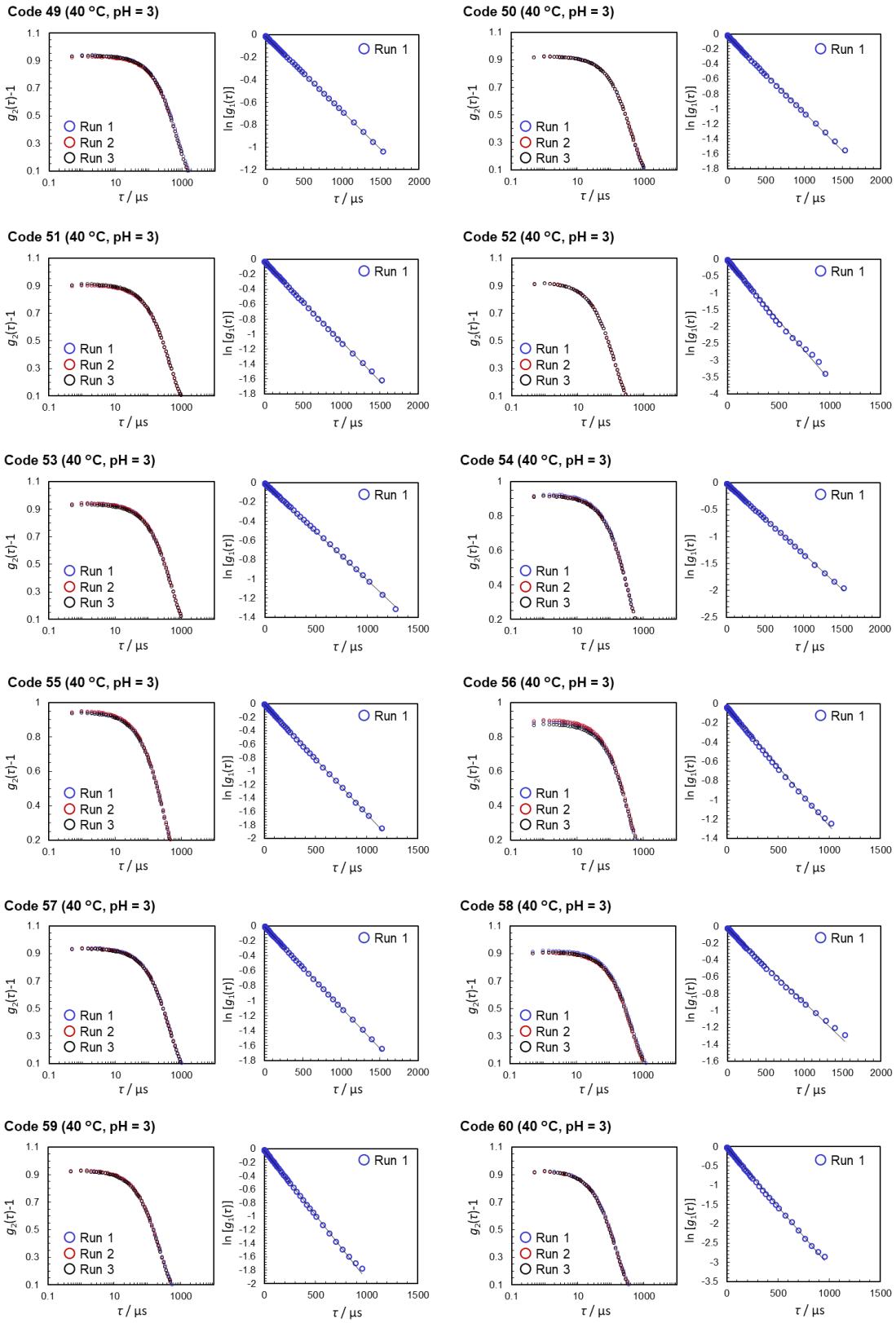

Figure S2 (continues) Time–correlation function of the scattering intensity, $g_2(\tau)-1$, and calculated time-correlation function of the scattering electric field, $\ln[g_1(\tau)]$, of each microgel (y_1).

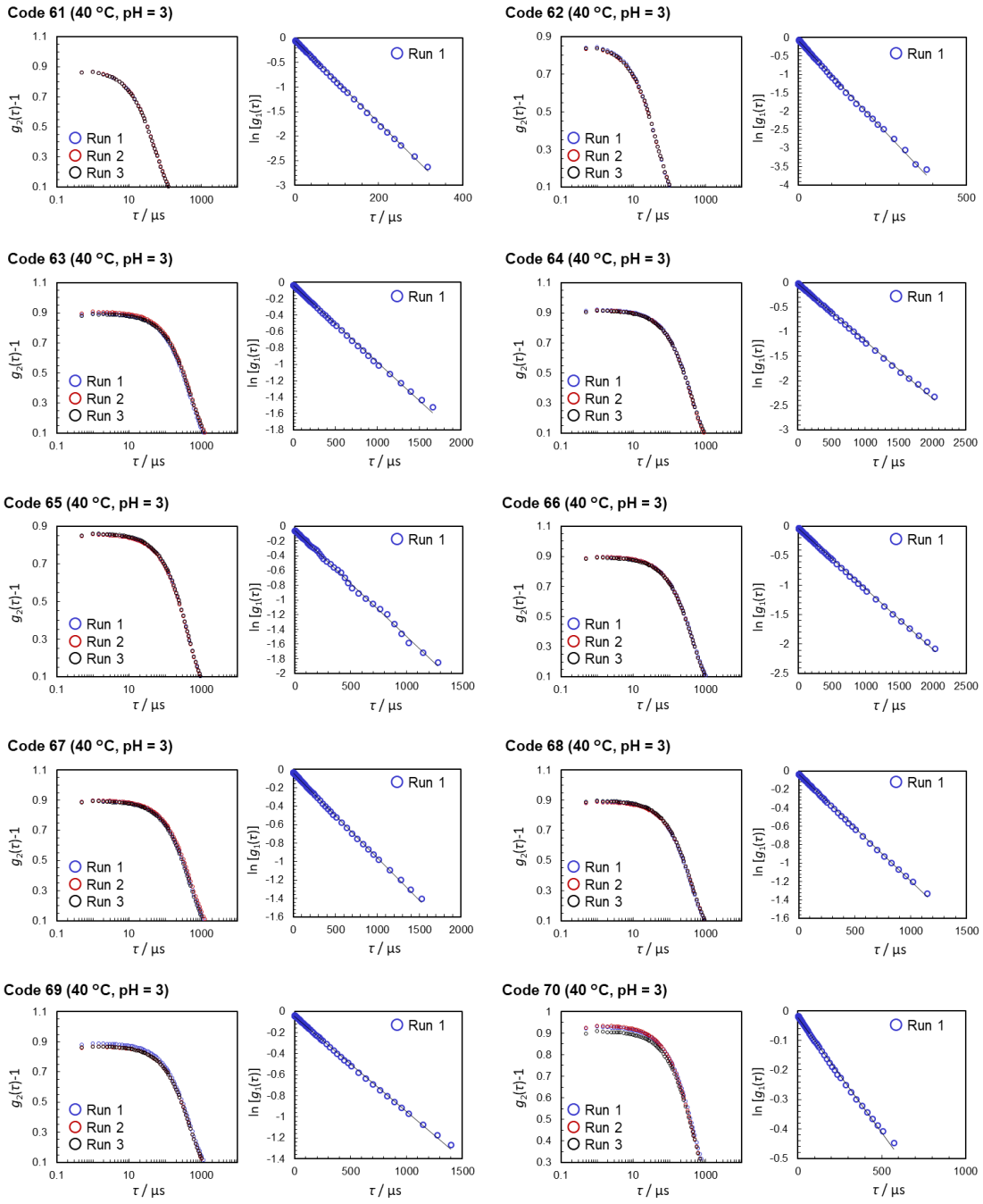

Figure S2 (continued) Time–correlation function of the scattering intensity, $g_2(\tau)-1$, and calculated time-correlation function of the scattering electric field, $\ln[g_1(\tau)]$, of each microgel (y_1).

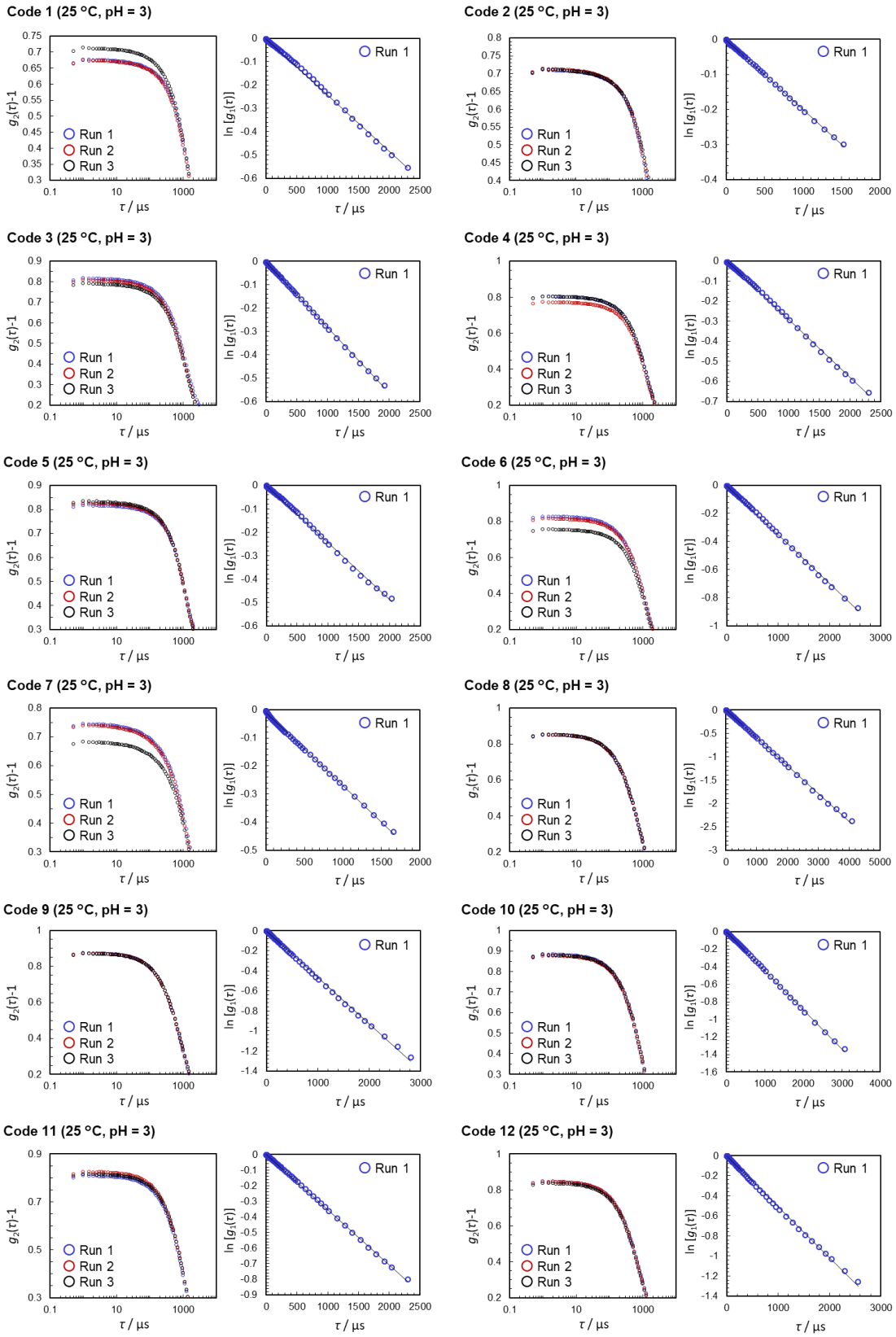

Figure S3 (continues) Time–correlation function of the scattering intensity, $g_2(\tau)-1$, and calculated time-correlation function of the scattering electric field, $\ln[g_1(\tau)]$, of each microgel (y_2).

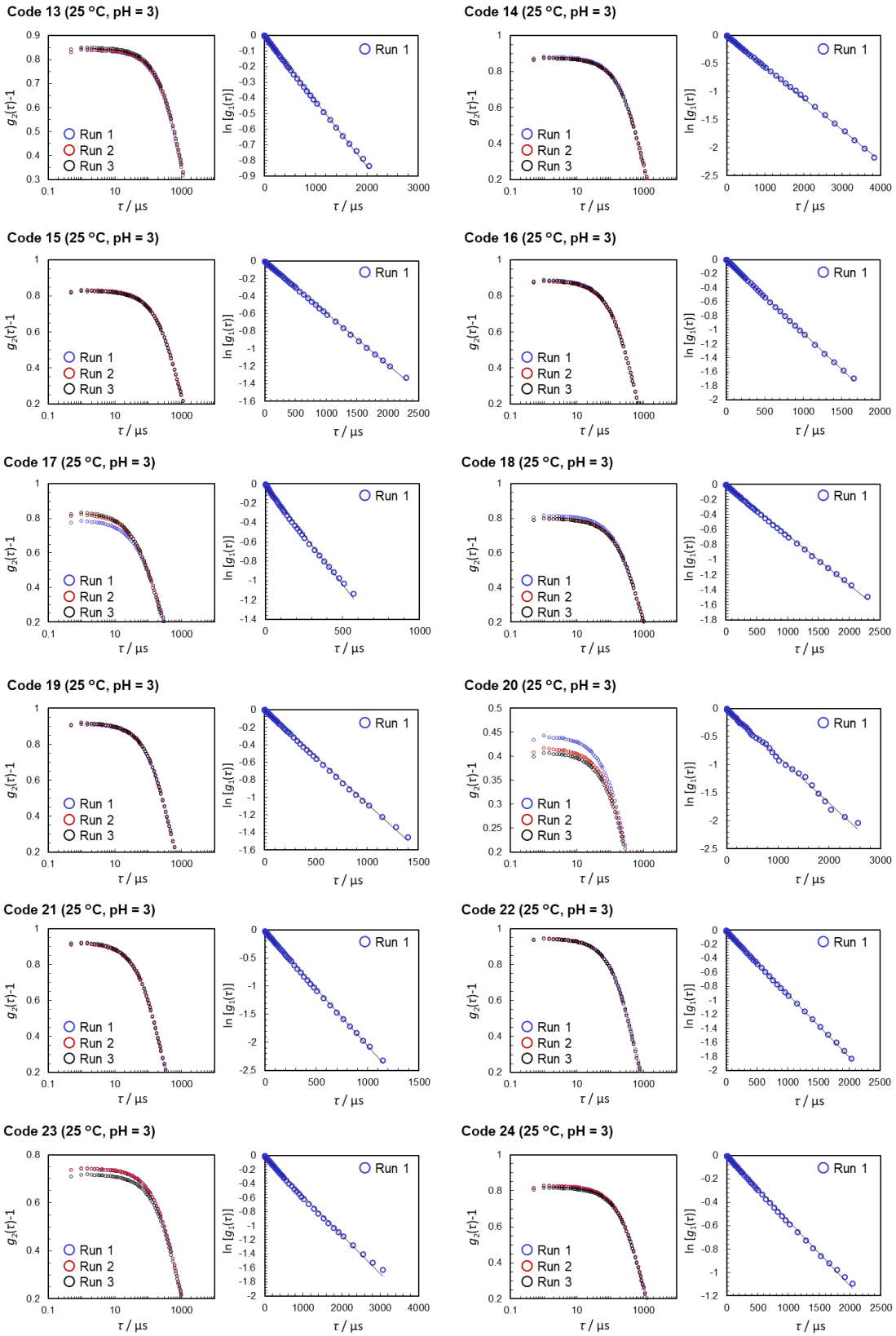

Figure S3 (continues) Time–correlation function of the scattering intensity, $g_2(\tau)-1$, and calculated time-correlation function of the scattering electric field, $\ln[g_1(\tau)]$, of each microgel (y_2).

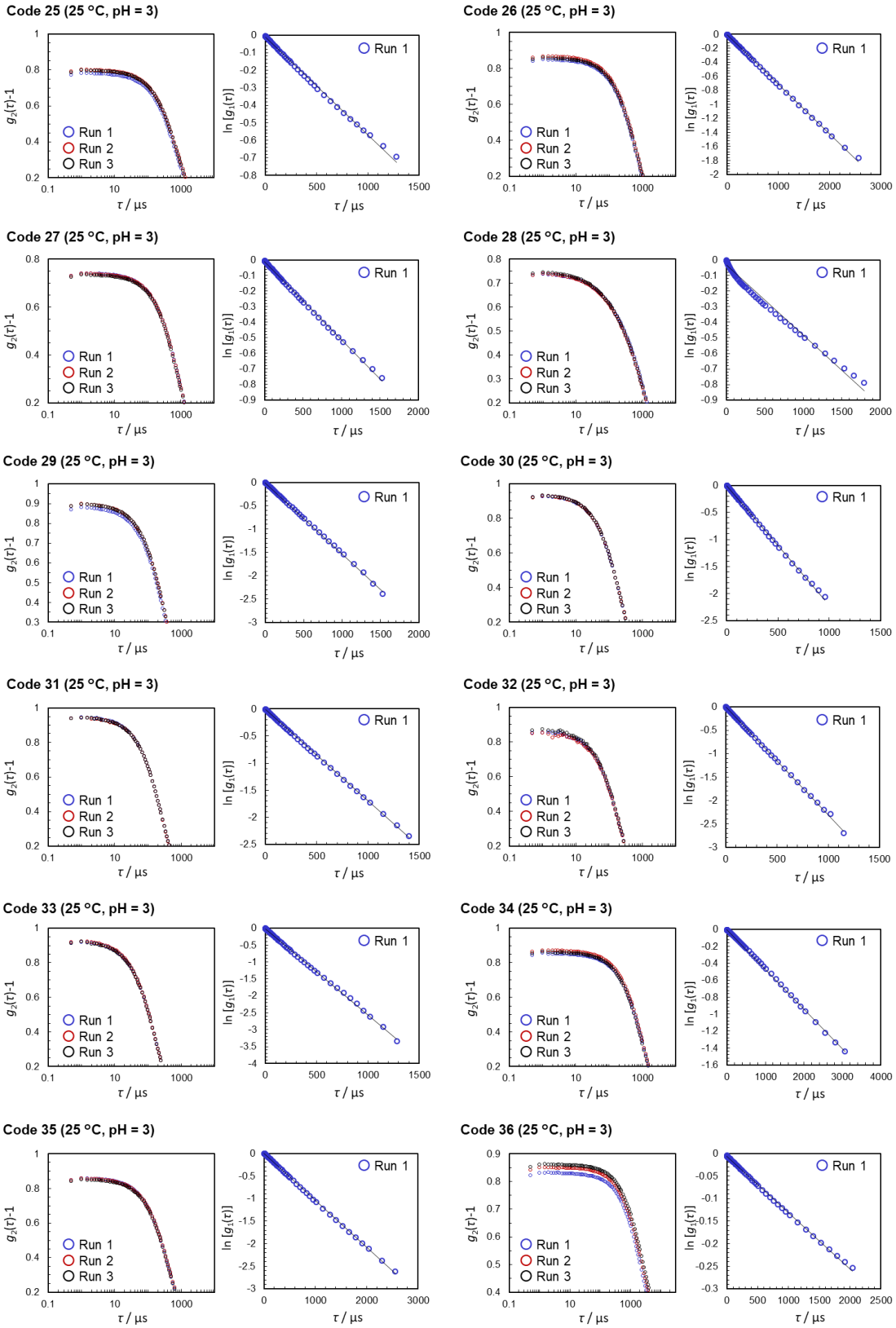

Figure S3 (continues) Time–correlation function of the scattering intensity, $g_2(\tau)-1$, and calculated time-correlation function of the scattering electric field, $\ln[g_1(\tau)]$, of each microgel (y_2).

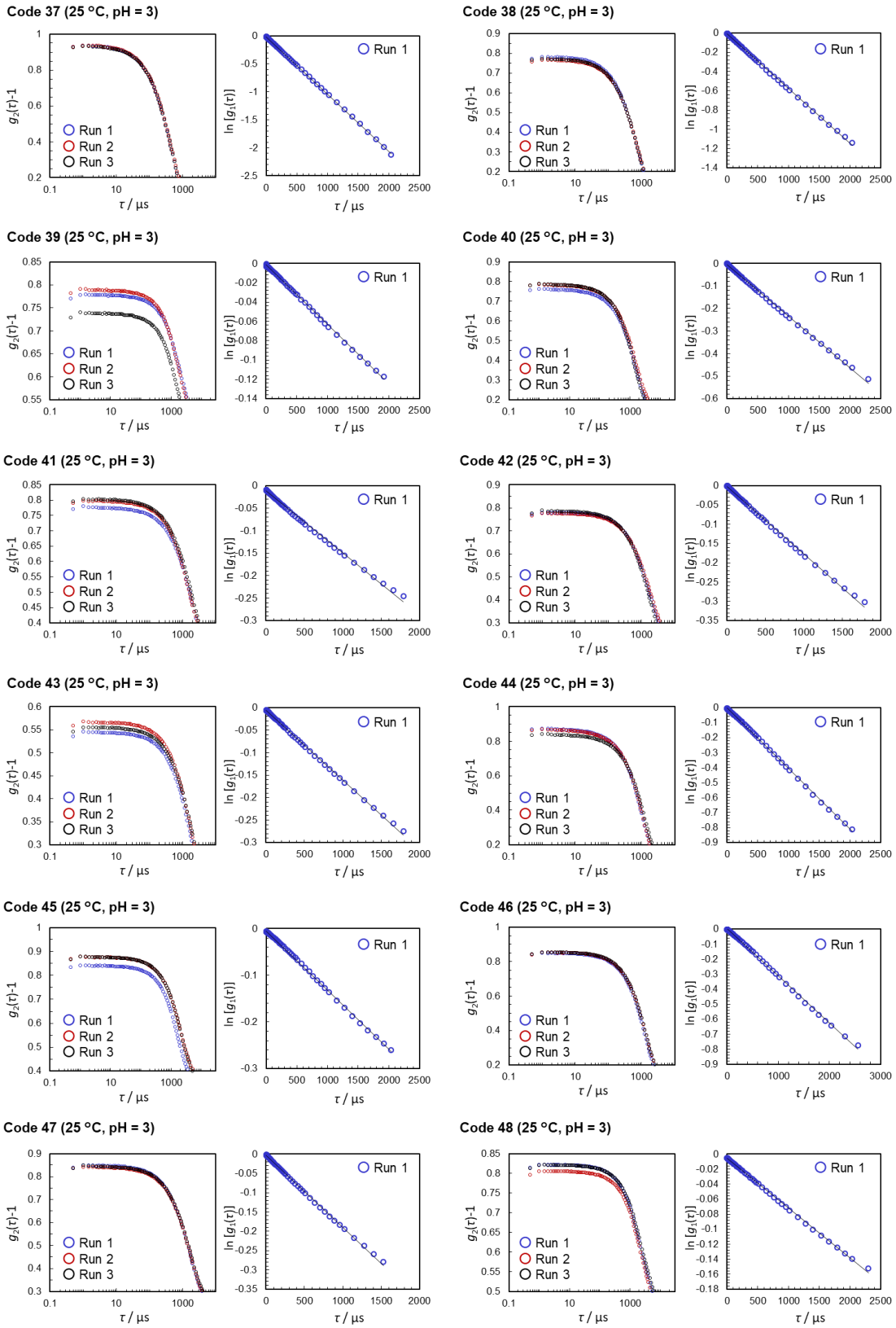

Figure S3 (continued) Time–correlation function of the scattering intensity, $g_2(\tau)-1$, and calculated time-correlation function of the scattering electric field, $\ln[g_1(\tau)]$, of each microgel (y_2).

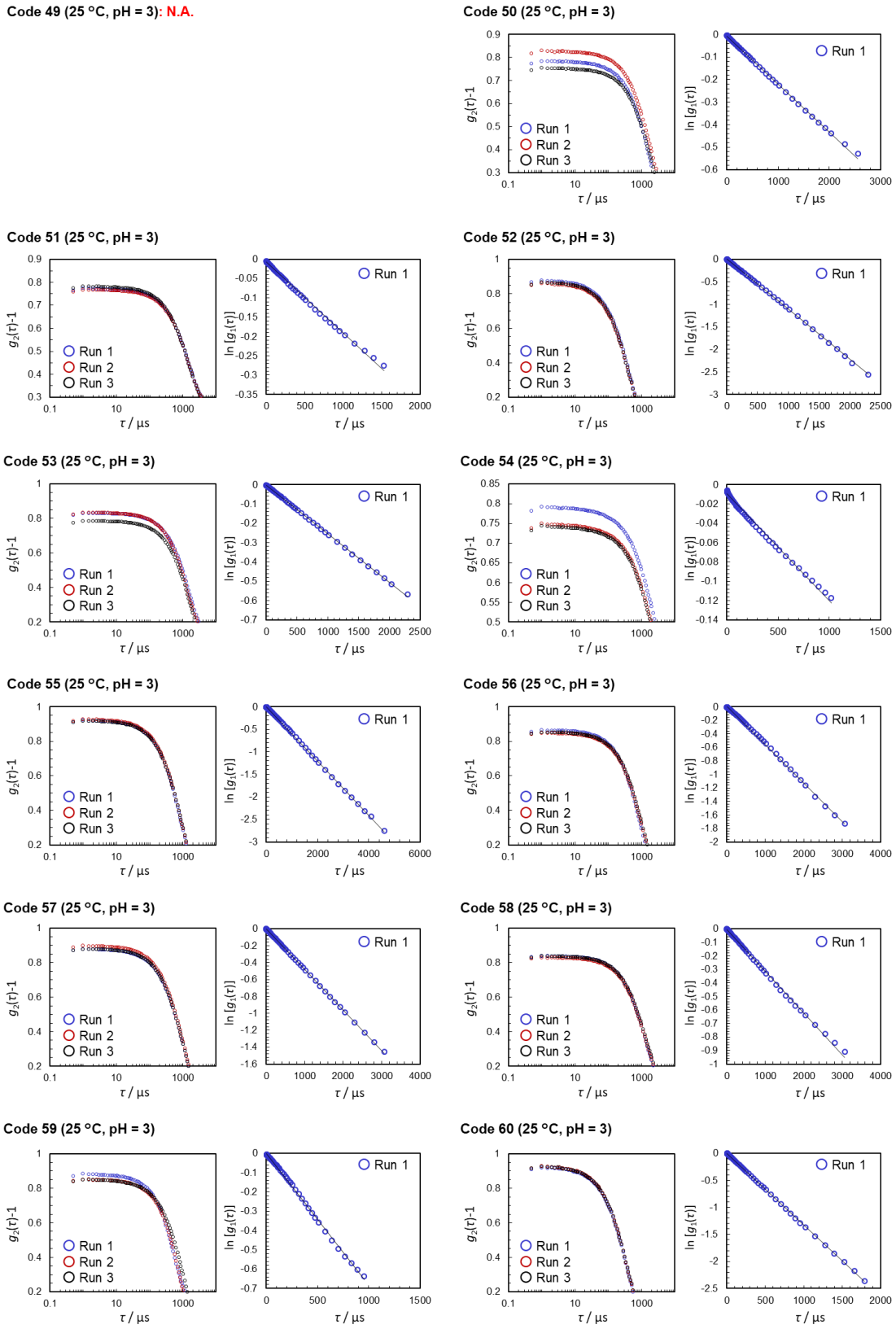

Figure S3 (continued) Time–correlation function of the scattering intensity, $g_2(\tau)-1$, and calculated time-correlation function of the scattering electric field, $\ln[g_1(\tau)]$, of each microgel (y_2).

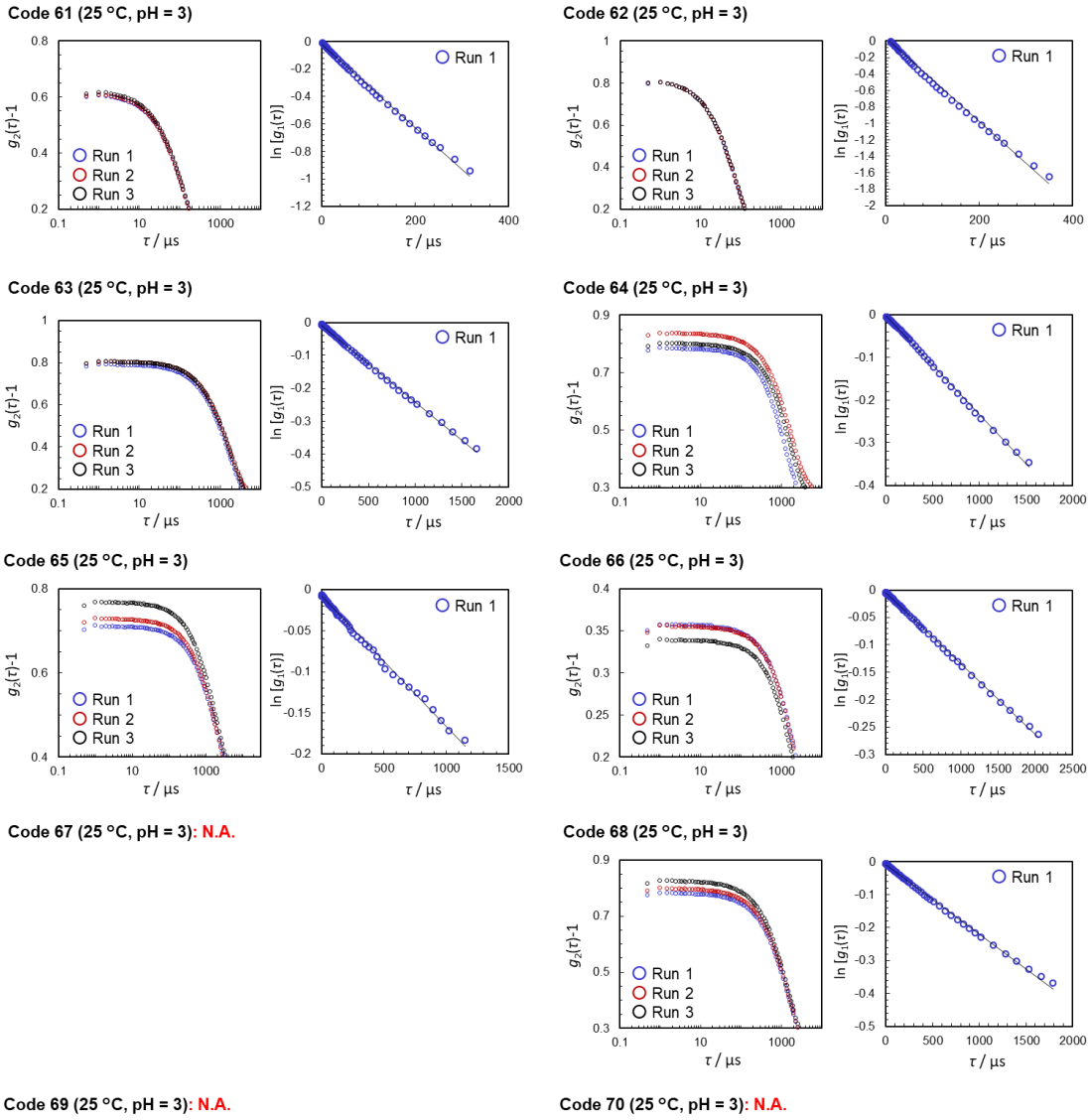

Figure S3 (continues) Time–correlation function of the scattering intensity, $g_2(\tau)-1$, and calculated time-correlation function of the scattering electric field, $\ln[g_1(\tau)]$, of each microgel (y_2).

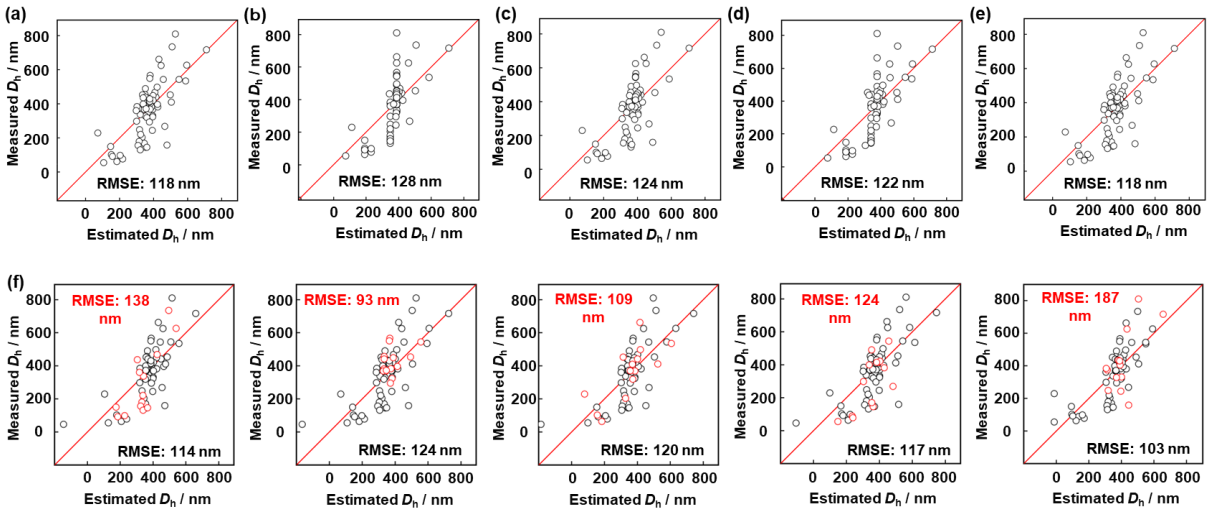

Figure S4. Model construction processes for y_1. (a–e) Relationship between the estimated and measured D_h for the constructed models with the different descriptors: (a) x_1, x_3, x_5, x_6, x_7, x_9, and x_{10}. (b) x_5, x_6, x_7, and x_{10}. (c) x_1, x_5, x_6, x_7, and x_{10}. (d) x_5, x_6, x_7, x_9, and x_{10}. (e) x_1, x_5, x_6, x_7, x_9, and x_{10}. (f) Relationship between the estimated and measured D_h in five-fold cross-validation of y_1 predictor (x_1, x_5, x_6, x_7, x_9, and x_{10}) using the five different training (black) and test (red) datasets.

As mentioned in the main text, the descriptors were selected three steps combining the weight diagram of ES-LiR and our chemical insight. The descriptors were finally determined based on the following factors: the relationship between the measured and estimated D_h, RMSE values, and interpretability of the selected x_n in the chemical insight.

The descriptors were extracted from the weight diagram in the first step. For y_1, the descriptors x_1, x_3, x_5, x_6, x_7, x_9, and x_{10} were extracted based on the color intensity and density from the weight diagram (Fig. 3a and Fig. S4a). Then, the descriptors x_5, x_6, x_7, and x_{10} were selected based on our chemical insight considering the mechanism of precipitation polymerization (Fig. S4b). However, the prediction accuracy lowered with decreasing the number of the descriptors. The descriptors x_1 and/or x_9 were added to x_5, x_6, x_7, and x_{10} (Fig. S4c–e) for improving the prediction accuracy. In this manner, we finally determined the descriptors x_1, x_5, x_6, x_7, x_9, and x_{10} for y_1 (Fig. 3c).

The training dataset was randomly divided into five groups. Four groups were assigned to the training data, and the remaining group was used as the test data for validation. After construction of the model, the root-mean-squared error (RMSE) was calculated for both the training and test data. This validation was carried out in five patterns, changing the assignment of the test data. The average RMSE values were 116 ± 7.0 nm for the training data and 130 ± 32 nm for the test data (Fig. S4f). The positive and negative correlations of the descriptors did not change depending on the model. Therefore, the extracted descriptor is suitable for predicting y_1.

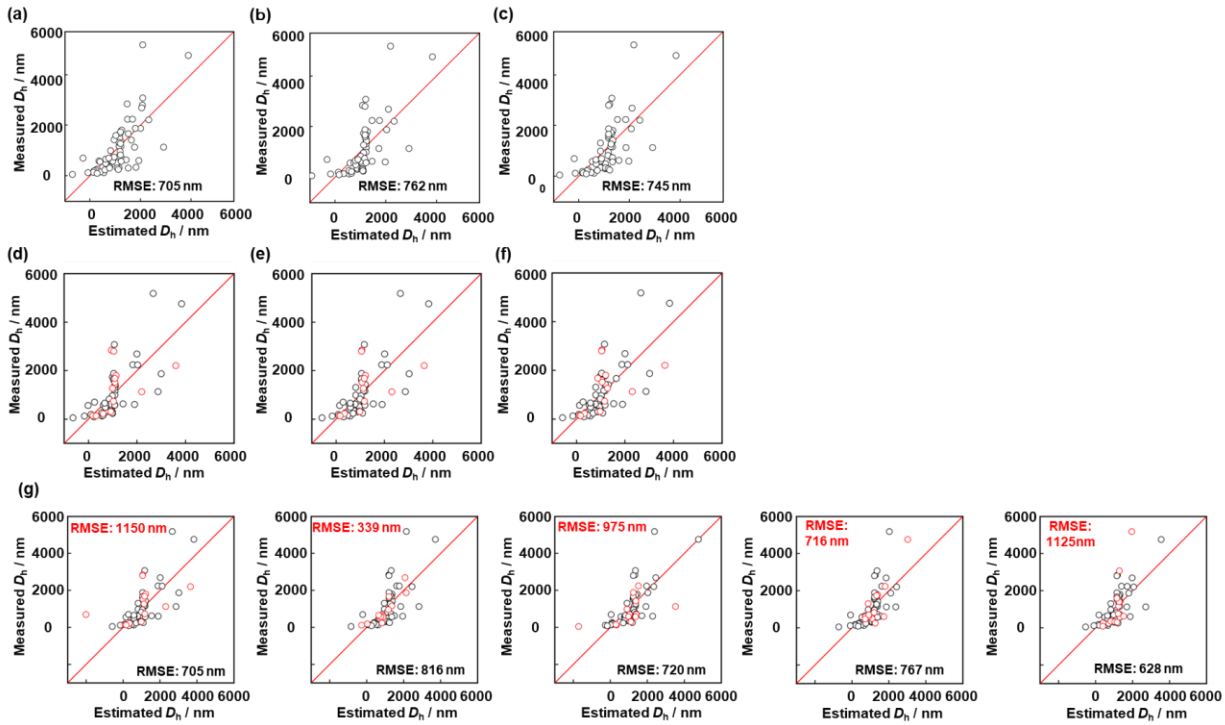

Figure S5. Model construction processes for y_2. (a–c) Relationship between the estimated and measured D_h for the constructed models with the different descriptors: (a) x_1, x_2, x_3, x_5, x_6, x_7, x_8, and x_{10}. (b) x_2, x_5, x_6, x_7, and x_{10}. (c) x_2, x_3, x_5, x_6, x_7, and x_{10}. (d–f) Relationship between the estimated and measured D_h for the model based on x_2, x_5, x_6, x_7, and x_{10} (d) with only x_3 (e) and both x_3 and x_9 (f) in the five-fold cross validation. (g) Relationship between estimated and measured D_h in five-fold cross validation of y_2 predictor (x_2, x_3, x_5, x_6, x_7, and x_{10}) using the five different training (black) and test (red) datasets.

For y_2, the descriptors x_1, x_2, x_3, x_5, x_6, x_7, x_8, and x_{10} were extracted based on the weight diagram (Fig. 3b and Fig. S5a). Then, the descriptors x_2, x_5, x_6, x_7, and x_{10} were selected based on our chemical insight considering the mechanism of precipitation polymerization (Fig. S5b). The descriptor x_3 was added to x_2, x_5, x_6, x_7, and x_{10} (Fig. S5c). The five-fold cross validation was performed in the following combinations to select the additional preferred descriptors (x_3, x_9): x_2, x_5, x_6, x_7, and x_{10} (Fig. S5d), x_2, x_3, x_5, x_6, x_7, and x_{10} (Fig. S5e), and x_2, x_3, x_5, x_6, x_7, x_9, and x_{10} (Fig. S5f). The average RMSE values of the training and test datasets were 744 and 904 nm for the model based on x_2, x_5, x_6, x_7, and x_{10} (Fig. S5d), 727 and 861 nm for the model based on x_2, x_3, x_5, x_6, x_7, and x_{10} (Fig. S5e), and 725 and 865 nm for the model based on x_2, x_3, x_5, x_6, x_7, x_9, and x_{10} (Fig. S5f), respectively. The model based on x_2, x_3, x_5, x_6, x_7, and x_{10} showed the minimum average RSEM value for the test data (Fig. S5e). Therefore, we finally determined the descriptors x_2, x_3, x_5, x_6, x_7, and x_{10} for y_2 (Fig. 3d).

The five-fold cross validation was carried out using the same method as for y_2. The average RMSE values were 727 ± 63 nm for the training data and 861 ± 303 nm for the test data (Fig. S5g). The positive and negative correlations of the descriptors did not change depending on the model. Therefore, the extracted descriptor is suitable for predicting y_2.

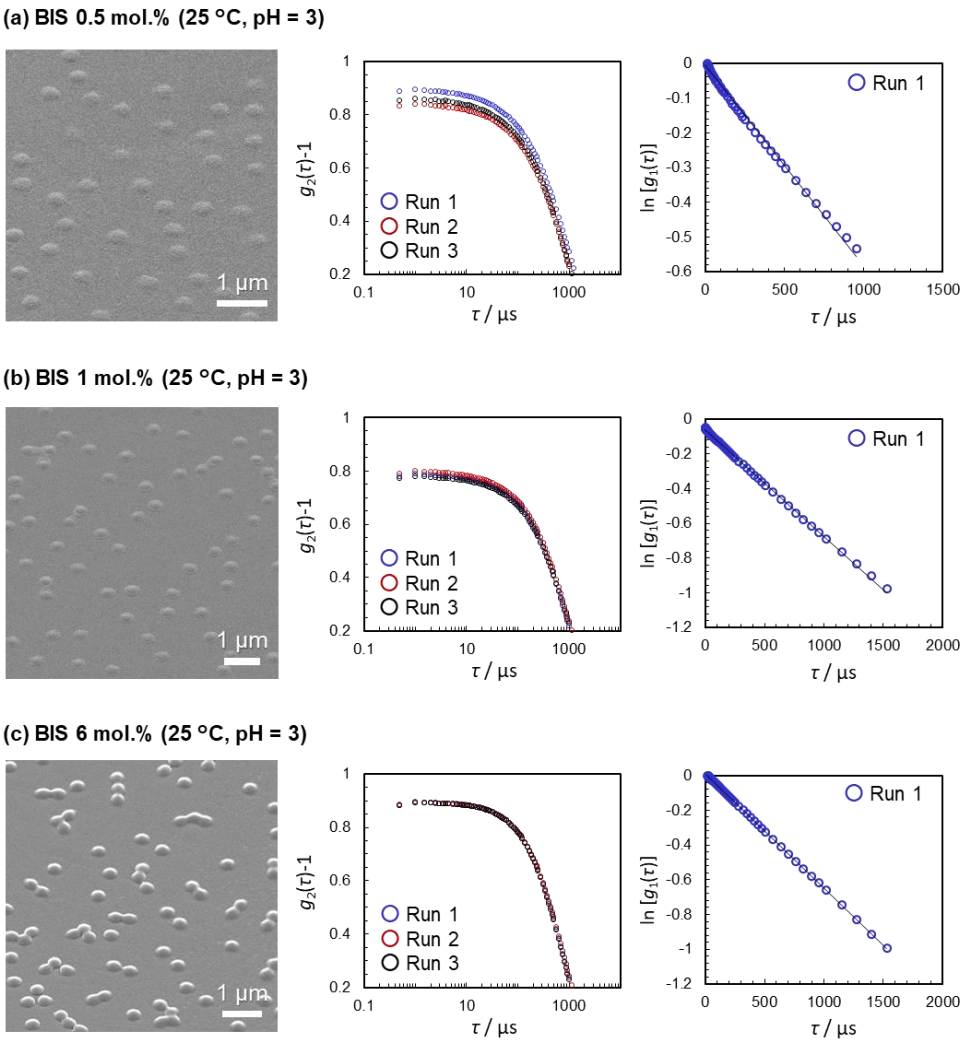

Figure S6 In order to confirm the prediction model for y_2, additional experiments were conducted targeting a swollen microgel size (y_2) of 500 nm. Representative SEM and FE-SEM images, time–correlation function of the scattering intensity, $g_2(\tau)-1$, and calculated time-correlation function of the scattering electric field, $\ln[g_1(\tau)]$, for each microgel (y_2).

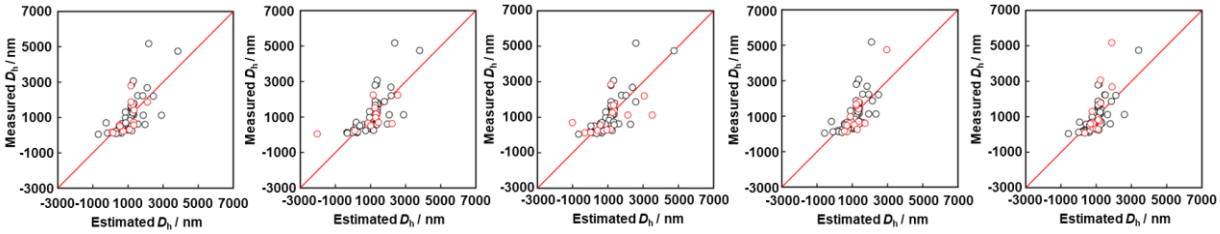

Fig. S7. Relationship between estimated and measured D_h in five-fold cross validation of y_2 predictor using the dataset with adding the test data in Table 2.

The particles with the predicted y_2 = 500 nm were synthesized in three different conditions (Table 2). The validation of the constructed models was carried out by five-fold cross validation, a common method in data science, with the addition of the resultant data (Fig. S6). The test data was added to the original training data. The dataset was randomly divided into the five groups to perform five-fold cross validation. The average RMSE value was 712 ± 59 nm for the training data (black) and 868 ± 196 nm for the test data (red). These average RMSE values were close to 727 ± 63 nm for the training data and 861 ± 303 nm for the test data using original training dataset (Fig. S5g). Therefore, the model has general applicability to predict the particle size.

References

S1. Y. Igarashi, H. Takenaka, Y. Nakanishi-Ohno, M. Uemura, S. Ikeda and M. Okada, *J. Phys. Soc. Jpn.*, 2018, **87**, 044802.

S2. H. Minato, M. Takizawa, S. Hiroshige and D. Suzuki, *Langmuir*, 2019, **35**, 10412-10423.