Employing an R Software Package rsm for Optimizing of Genistein, Daidzein, and Glycitein Separation and Its Application for Soy Milk Analysis by HPLC Method

Florentinus Dika Octa Riswanto[1,2], Alni Desra[2], Rinjani Mustika Sari[2], Valentino Thomas[2], Abdul Rohman[2], Suwidjiyo Pramono[3], and Sudibyo Martono[2,*]

[1]*Department of Pharmacy, Faculty of Pharmacy, Sanata Dharma University, Campus III Paingan, Maguwoharjo, Depok, Sleman, Yogyakarta 55282, Indonesia*

[2]*Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia*

[3]*Department of Pharmaceutical Biology, Faculty of Pharmacy, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia*

* ***Corresponding author:***

tel: +62-8122729826
email: sudibyo_martono@ugm.ac.id

Received: November 20, 2019
Accepted: January 16, 2020

DOI: 10.22146/ijc.51669

Abstract: *Soy milk, one of the soybean products, become more popular in recent years due to its benefit for human health. Biological activities of soybean products have been widely studied according to the presence of isoflavone aglycones content, including genistein, daidzein, and glycitein. Hence, it is important to develop an effective and efficient analytical method to provide guidance regarding appropriate isoflavone intake levels for soy milk. A reversed-phase high performance liquid chromatography (HPLC) method was developed and optimized in this study employed by rsm package in R statistical software. A C_{18} column was used for HPLC separation with the detection at 260 nm. Optimized condition for HPLC separation has been achieved with the mobile phase of methanol: water (63.26:36.74), a flow rate of 0.81 mL/min, and a column temperature of 45.31 °C. These conditions were applied in the HPLC system and successfully tested for the system suitability. Quantitative estimation was performed and resulted that the genistein, daidzein, and glycitein content in soy milk samples were 6.372, 6.273, and 2.853 µg/mL, respectively.*

Keywords: *daidzein; genistein; glycitein; HPLC; R software*

■ INTRODUCTION

Functional food from soybeans (*Glycine max* L. Merrill) has attracted much attention in recent years due to their beneficial effects on immune function, brain function, and neurochemistry towards healthy people [1]. It was stated that soy foods were not only providing high-quality protein and healthful fat but also having health benefits such as reducing the risk of coronary heart disease, performing an anticancer activity for breast and prostate cancer, improving renal function, brain smartness, and skin health [2-3]. In the previous studies [4-6], several biological activities of soy foods indicated the presence of isoflavones content as bioactive compounds.

Soy milk, one of the most popular beverages in Asian countries, is commonly produced by the grinding process of soybeans with water to achieve soybean water extract [7]. As one of soy foods products, soy milk was reported to have isoflavone aglycones content, including genistein, daidzein, and glycitein [8]. Fig. 1 presents the chemical structure of genistein, daidzein, and glycitein as the three major isoflavone aglycones in soy milk.

More than two thousand soy-related studies have been annually published since there was evidence of the activity from its nutrients content [2]. Hence, it is necessary to provide guidance regarding appropriate isoflavone intake levels for soybeans food products [9]. High performance liquid chromatography (HPLC) has been widely used as an analytical method to analyze isoflavones content in several soy foods [10-12]. As an analytical method, analytes separation by HPLC was affected by several experimental conditions, including column conditions, solvent purity, column temperature,

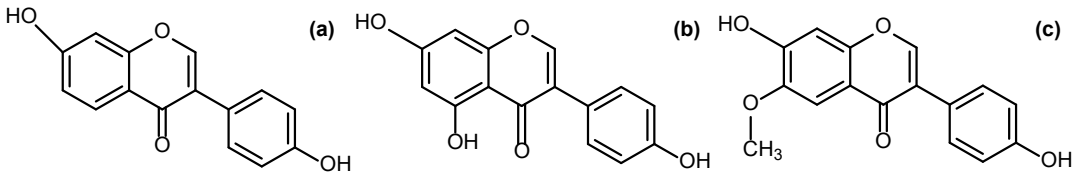

Fig 1. Chemical structure of genistein (a), daidzein (b), and glycitein (c)

flow rate, and mobile phase composition [13]. Observation of the optimum conditions for the HPLC method can be performed using response surface methodology (RSM) as a mathematical-statistical model for optimization [14-18]. RSM can be employed in the optimization of independent variables in HPLC separation [18].

RSM, as one of the statistical techniques, can be applied in optimization research using commercial software such as Minitab® [19] and Design-Expert® [20]. However, this kind of software is expensive and protected by software licensing [21]. On the other hand, R, as open-source software with the GNU General Public License (GPL), offers a broad range of statistical technique by installing software packages for each statistical need [22]. The interest in implementing RSM for the HPLC method has increased since there were many reports on its application in the field of chromatographic method optimization over the last decade [23-28]. Nevertheless, there was limited publication reporting the employment of R software for HPLC optimization. The aim of this study was to optimize the HPLC separation of genistein, daidzein, and glycitein in soy milk analysis employed by R software using the rsm package. This package supported the application of Box-Behnken Design (BBD) and provided response surface estimation with contour and perspective plots in order to present a sophisticated visualization of optimization [29]. In this paper, the R codes function, and their implementation in the rsm package was discussed. Parameters, namely methanol percentage, flowrate, and column temperature, were stated as independent variables, whereas several chromatographic responses, namely retention time, resolution, tailing factor, and theoretical plate number, were chosen to be evaluated as dependent variables. The system suitability test of the HPLC method, followed by the quantitative estimation of the three isoflavone aglycones, was also performed in this study.

■ EXPERIMENTAL SECTION

Materials

All the reference standard including genistein, daidzein, and glycitein were purchased from Sigma-Aldrich. The solvents used in this study were methanol gradient grade for liquid chromatography (Merck Millipore), petroleum ether, ethyl acetate (Smart Lab), and redistilled water (PT. Ikapharmindo Putramas). Sodium sulfate anhydrous was purchased from Merck Millipore. Soymilk sample was produced by the local brand and obtained in Yogyakarta, Indonesia.

Instrumentation and Software

Instrumentation used in this study were ultra-micro analytical balance RADWAG® series of UYA 2.3Y (max: 2.1 g, min 0.01 mg), HPLC system of Shimadzu® LC-2010 CHT with UV/Vis detector, a C_{18} column of Hibar® 250-4,6 Purospher® STAR RP-18 endcapped (5 µm), a system of Buchi Rotary Evaporator (Interface I-300, Rotavapor R-300, Heating Bath B-300), Retsch® T460 ultrasonicator, Gast® vacuum pump model DOA-P504-BN, sterile syringe filter with a 0.2 µm pore size hydrophilic PTFE membrane (Merck Millipore), and a set of Socorex® micropipettes. Statistical technique and response surface application was carried out using R Studio software version 1.1.456 with an rsm package.

Procedure

Standard and sample preparation

Accurate weight of 4.976 mg of genistein, 5.003 mg of daidzein, and 1.068 mg of glycitein were transferred into 10 mL volumetric flask for genistein and daidzein and 5 mL volumetric flask for glycitein. The contents of each volumetric flask were diluted in methanol to achieve a stock solution for each standard. The mixture solution containing all standards was prepared by transferring 0.2 mL, 0.2 mL, and 0.4 mL of genistein, daidzein, and

glycitein stock solution into a 5 mL volumetric flask, respectively, followed by dilution to volume. This solution was filtered using a sterile syringe filter membrane before injection into the HPLC system.

Fifty milliliters of soy milk samples were transferred into beaker glass and macerated in 50 mL petroleum ether for 40 min (150 rpm) followed by the removal of the petroleum ether. The hydrophilic solution was then fractionated using ethyl acetate and water. The ethyl acetate fraction was subsequently separated and filtered to remove the solid phase residue. Sodium sulfate anhydrous was added into the filtrate follow by filtration to achieve the yellowish solution of ethyl acetate. This yellowish filtrate solution was evaporated using a rotary evaporator. The residue obtained from the evaporation stage was diluted in 10 mL methanol. The working sample solution was prepared by transferring a 5 mL sample diluted solution into a 10 mL volumetric flask followed by dilution to volume. This solution was filtered using a sterile syringe filter membrane before injection into the HPLC system.

Experimental design

RSM as an experimental design approach was implemented in this study to optimize chromatographic separation. The BBD [30], one of the RSM technique, has been applied as experimental design since there were three observed factors with three levels for each [15]. Preliminary HPLC optimization for condition screening was performed by one variable at time (OVAT) observation. The percentage of methanol (X1), flowrate (X2), and column temperature (X3) were selected as independent variables, whereas HPLC separation responses namely retention time (Y1), resolution (Y2), tailing factor (Y3), and theoretical plates number (Y4) was selected as dependent variables. The BBD design for this study was generated using R software using the following formula:

```
> library(rsm)
> bbd(3, randomize=FALSE)
```

Every independent variable was coded as -1, 0, and +1 for low, medium, and high levels, respectively. The coded levels of variables and their values are presented in Table 1. After successfully generated a BBD model with 16 runs, HPLC separation was performed for each run by HPLC system of Shimadzu® LC-2010 CHT with UV/Vis detector and a C_{18} column of Hibar® 250-4,6 Purospher® STAR RP-18 endcapped (5 μm). Wavelength detection was set to 260 nm in which three analytes were resulting in a good response. The volume injection for each run was 10 μL.

Table 1. Coded levels of variables and their observed values

Variables	Coded levels		
	-1	0	+1
X1: methanol (%)	60	65	70
X2: flowrate (mL/min)	0.6	0.8	1.0
X3: column temperature (°C)	30	40	50

RSM observation

The results of 16 BBD runs were recorded and listed for each response and each compound to build RSM coded model. After storing coded data for each compound, the effect of independent variables toward each dependent variable was fitted using the RSM function following the second-order model since it provided flexibility, simplicity for parameter estimation, and the ability for solving response surface problems [31]. The statistical analysis resulted from an analysis of variance (ANOVA) table to observe the significant differences among independent variables. Estimated coefficients for each RSM equation [28] were also generated by the software. The second-order polynomials are considered to estimate the response of the dependent variable:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{33} X_3^2 \quad (1)$$

where,

Y = predicted response

X_1, X_2, X_3 = independent variables

β_0 = intercept

β_1, β_1, β_3 = linear effect

β_{12}, β_{13}, β_{23} = interaction effect

β_1^2, β_1^2, β_3^2 = squared effect

Further, statistical analysis using R software provided stationary points along with the eigenvalues, which practically assist the optimization process. The

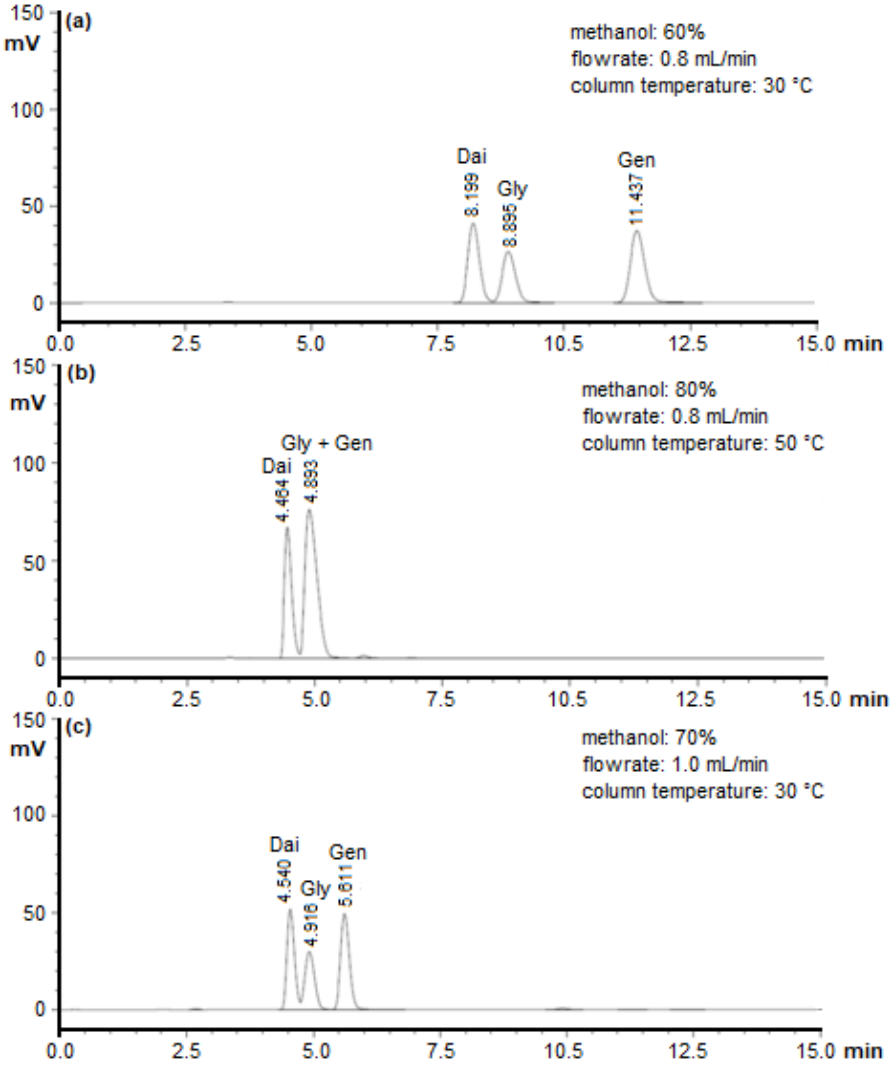

Fig 2. Preliminary HPLC optimization using three variations separation condition of OVAT1 (a), OVAT2 (b), and OVAT3 (c) for mixture standard solution containing genistein (Gen), daidzein (Dai), and glycitein (Gly). Column: C_{18} column of Hibar® 250-4,6 Purospher® STAR RP-18 endcapped (5 µm). Wavelength detection at 260 nm. Volume injection: 10 µL

perspective plots for each dependent variable were also generated to depict the effective visualization of RSM models.

System suitability test

The system suitability test was performed by injecting a mixture solution containing all standards of 10 µL with a concentration of 9.951, 10.005, and 8.544 µg/mL for genistein, daidzein, and glycitein, respectively. These injections were replicated five times.

Quantitative estimation

Qualitative estimation for genistein, daidzein, and glycitein was performed using a single-point calibration method [27,32]. The concentration of genistein, daidzein, and glycitein was determined by the peak area sample and standards.

■ RESULTS AND DISCUSSION

BBD, as one of the three-level fractional factorial design classes, providing an effective and economic statistical model to solve the optimization problem [33]. The percentage of methanol in combination with water, flow rate, and the column temperature was stated as independent variables due to their contribution for

achieving a sophisticated HPLC separation. The selection of mobile phase composition played an important role in liquid chromatography after considering stationary phase selection [34] and supported with an optimized flow rate to minimize the peak broadening [35]. The column temperature was also considered in HPLC optimization due to its contribution toward the elution strength of the mobile phase [36]. Preliminary HPLC optimization has been performed by the conventional OVAT approach using three condition variations (OVAT1, OVAT2, and OVAT3) of the mobile phase, flow rate, and column temperature, as presented in Fig. 2. This approach has been considered for preliminary screening of various independent variables influencing HPLC separation responses for further statistical optimization using the BBD technique.

The BBD model has been generated using R software with the formula of bbd(3, randomize=FALSE). This formula resulted in 16 runs for RSM experimental model, including four central point replications agreed with BBD number of experiments formula: 2k(k–1)+Cp, where k represents the number of factors and Cp represents the replicate number of the central point [15]. The BBD and the experimental responses of genistein, daidzein, and glycitein separation were presented in Table 2.

RSM Optimization of Genistein, Daidzein, and Glycitein Separation

A total of 16 runs of experiments were performed with different combinations of variables X1, X2, and X3 using BBD which also includes four central point replication of second order response surface. These runs allowed to collect responses data of independent variables Y1, Y2, Y3, and Y4. RSM statistical analysis was performed to observe each response for genistein, daidzein, and glycitein. It should be noted that setting up the coded formula of independent variables was the important step for generating RSM model in R software. The medium level of each factor and its interval to low and high level should be clearly defined. RSM coded model for each compound was input in R software using the

Table 2. The Box-Behnken Design and the experimental responses of dependent variables of genistein, daidzein, and glycitein separation

Run	Independent variables			Dependent variables							
				Genistein				Daidzein			
	met	flow	temp	ret	res	tai	n	ret	res	tai	n
1	60	0.6	40	12.699	4.292	1.238	8821.962	9.553	13.294	1.221	6475.064
2	70	0.6	40	8.346	1.738	1.234	5465.418	6.962	3.397	0.000	5110.210
3	60	1.0	40	7.796	4.020	1.201	7802.569	5.848	8.875	1.188	5865.786
4	70	1.0	40	5.090	1.649	1.208	5199.550	4.253	7.057	1.184	4941.509
5	60	0.8	30	11.437	5.089	1.184	7808.821	8.199	8.960	1.194	5986.763
6	70	0.8	30	6.833	2.297	1.270	5269.899	5.562	4.435	0.000	4636.857
7	60	0.8	50	8.800	2.779	1.209	8139.261	6.820	11.540	1.093	4931.067
8	70	0.8	50	5.923	1.336	1.177	5335.514	5.032	5.102	1.455	5121.019
9	65	0.6	30	10.776	3.213	1.266	6168.342	8.352	7.573	0.000	4939.654
10	65	1.0	30	6.687	3.275	1.198	5568.114	5.128	6.449	0.000	4376.574
11	65	0.6	50	9.019	2.109	1.247	6399.267	7.382	3.737	1.252	5135.400
12	65	1.0	50	5.335	1.874	1.227	5634.854	4.390	7.679	1.243	4743.974
13	65	0.8	40	7.476	2.699	1.233	6716.002	5.959	6.112	1.314	4995.631
14	65	0.8	40	7.482	2.688	1.233	6665.227	5.961	6.310	1.312	5021.040
15	65	0.8	40	7.451	2.679	1.232	6614.460	5.935	6.238	1.312	5010.788
16	65	0.8	40	7.460	2.672	1.228	6549.274	5.956	4.319	1.309	5122.781

Table 2. The Box-Behnken Design and the experimental responses of dependent variables of genistein, daidzein, and glycitein separation (*Continued*)

Run	Independent variables			Dependent variables			
				Glycitein			
	met	flow	temp	ret	res	tai	n
1	60	0.6	40	10.450	1.825	1.208	6778.520
2	70	0.6	40	7.526	1.282	0.000	3757.794
3	60	1.0	40	6.413	1.764	1.178	5863.593
4	70	1.0	40	4.603	1.270	0.000	3558.951
5	60	0.8	30	8.895	1.537	1.168	5464.280
6	70	0.8	30	5.939	1.035	0.000	3496.609
7	60	0.8	50	7.689	2.179	1.115	5646.459
8	70	0.8	50	5.459	1.311	0.000	3486.538
9	65	0.6	30	9.019	1.308	0.000	4389.452
10	65	1.0	30	5.514	1.158	0.000	3803.609
11	65	0.6	50	8.048	1.507	1.158	4675.234
12	65	1.0	50	4.784	1.410	0.000	3975.597
13	65	0.8	40	6.468	1.413	0.000	4581.871
14	65	0.8	40	6.472	1.417	0.000	4527.589
15	65	0.8	40	6.440	1.395	0.000	4408.380
16	65	0.8	40	6.452	1.381	0.000	4485.048

Note. met: methanol percentage (%); flow: flowrate (mL/min); temp: column temperature (°C); ret: retention time (min); res: resolution; tai: tailing factor; n: theoretical plates number

following formula:

```
#setting up coded levels for genistein model
> gen.rsm <- coded.data(gen, x1 ~ (methanol - 65)/5,
                        x2 ~ (flowrate - 0.8)/0.2,
                        x3 ~ (temp - 40)/10)
#setting up coded levels for daidzein model
> dai.rsm <- coded.data(dai, x1 ~ (methanol - 65)/5,
                        x2 ~ (flowrate - 0.8)/0.2,
                        x3 ~ (temp - 40)/10)
#setting up coded levels for glycitein model
> gly.rsm <- coded.data(gly, x1 ~ (methanol - 65)/5,
                        x2 ~ (flowrate - 0.8)/0.2,
                        x3 ~ (temp - 40)/10)
```

The coded data sets facilitated to ensure the RSM analysis can be employed properly. It was necessary to understand that the medium level for methanol percentage was set to 65% with the interval of ± 5%, the medium level for flow rate was set to 0.8 mL/min with the interval of ± 0.2 mL/min, and the medium level for column temperature was set to 40 °C with the interval of ± 10 °C for genistein, daidzein, and glycitein.

After successfully coded all data sets, RSM analysis was performed to assess the optimized condition for separation of genistein, daidzein, and glycitein. The empiric second-order polynomial models and RSM analysis of each response for genistein, daidzein, and glycitein were presented in Table 3. The RSM perspective plots of all responses for genistein (Fig. 3), daidzein (Fig. 4), and glycitein (Fig. 5) were also generated using R software.

Independent variables have significantly affected the responses only if multiple $R^2 \geq 0.8$ and adjusted $R^2 > 0.8$; in addition, the difference between multiple R^2 with the adjusted R^2 must be less than 0.2 [37]. These results indicated that the experimental model was a good fit with polynomial equations [18]. Except for the tailing factor for genistein, all equation models generated were resulted acceptable both multiple R^2 and adjusted R^2. All responses resulted desirable p-value with $p < 0.05$. These data implied that all the generated models were significant.

Table 3. The empiric second-order polynomial model and RSM analysis of retention time, resolution, tailing factor, and theoretical plate number for genistein, daidzein, and glycitein

Responses	Model equations	Multiple R^2	Adjusted R^2	p-value	Stationary points and eigenvalues		
					Methanol (%)	Flow rate (mL/min)	Column temp. (°C)
Genistein							
Retention	$Y1=7.467-1.816X_1-1.992X_2-0.832X_3+0.412X_1X_2+0.432X_1X_3+0.101X_2X_3+0.655X_1^2+0.361X_2^2+0.126X_3^2$	0.9990	0.9976	2.439e-08	63.09 (0.827)	1.31 (0.267)	69.18** (0.047)
Resolution	$Y2=2.685-1.145X_1+-0.067X_2-0.722X_3+0.046X_1X_2+0.337X_1X_3-0.074X_2X_3+0.249X_1^2-0.009X_2^2-0.058X_3^2$	0.9912	0.9779	1.808e-05	74.79 (0.324)	0.39** (0.005)	47.87 (-0.146)
Tailing factor	$Y3=1.232+0.007X_1-0.019X_2-0.007X_3+0.002X_1X_2-0.029X_1X_3+0.012X_2X_3+0.018X_1^2+0.006X_2^2-0.004X_3^2$	0.8870	0.7175*	2.844e-02	66.57 (0.011)	1.07 (0.002)	39.88 (-0.020)
N	$Y4=6636.241-1412.779X_1-331.238X_2+86.715X_3+188.381X_1X_2-66.206X_1X_3-41.046X_2X_3+440.932X_1^2-254.798X_2^2-438.799X_3^2$	0.9966	0.9914	1.064e-06	73.06 (454.870)	0.80 (-265.862)	39.80 (-441.673)
Daidzein							
Retention	$Y1=5.953-1.076X_1-1.579X_2-0.452X_3+0.249X_1X_2+0.212X_1X_3+0.058X_2X_3+0.396X_1^2+0.306X_2^2+0.055X_3^2$	0.9995	0.9986	4.270e-09	63.34 (0.507)	1.28 (0.225)	75.07** (0.024)
Resolution	$Y2=5.745-2.835X_1+0.257X_2+0.080X_3+2.020X_1X_2-0.478X_1X_3+1.267X_2X_3+1.780X_1^2+0.631X_2^2-0.016X_3^2$	0.9221	0.8053	1.028e-02	67.82 (2.373)	0.82 (0.679)	27.78 (-0.657)
Tailing factor	$Y3=1.312-0.257X_1+0.143X_2+0.481X_3+0.304X_1X_2+0.389X_1X_3-0.002X_2X_3-0.051X_1^2-0.363X_2^2-0.325X_3^2$	0.9275	0.8187	8.418e-03	63.26 (0.090)	0.81 (-0.346)	45.31 (-0.483)
N	$Y4=5037.560-431.136X_1-216.560X_2-1.048X_3+110.144X_1X_2+384.965X_1X_3+42.914X_2X_3+465.304X_1^2+95.278X_2^2-333.937X_3^2$	0.9203	0.8008	1.095e-02	66.32 (517.345)	0.99 (87.337)	42.11 (-378.037)
Glycitein							
Retention	$Y1=6.458-1.240X_1-1.716X_2-0.423X_3+0.279X_1X_2+0.182X_1X_3+0.060X_2X_3+0.472X_1^2+0.318X_2^2+0.065X_3^2$	0.9992	0.9980	1.306e-08	66.30 (0.571)	1.28 (0.238)	57.61** (0.046)
Resolution	$Y2=1.402-0.300X_1-0.040X_2+0.171X_3+0.012X_1X_2-0.092X_1X_3+0.013X_2X_3+0.152X_1^2-0.018X_2^2-0.038X_3^2$	0.9603	0.9006	1.514e-03	71.11 (0.162)	0.71 (-0.016)	47.09 (-0.050)
Tailing factor	$Y3=2.7756e-17-0.584X_1-0.149X_2+0.138X_3+0.008X_1X_2+0.013X_1X_3-0.290X_2X_3+0.439X_1^2+0.158X_2^2+0.132X_3^2$	0.9265	0.8161	8.753e-03	67.63 (0.439)	1.97** (0.290)	98.72** (-0.001)
N	$Y4=4500.722-1181.620X_1-299.906X_2+78.735X_3+179.021X_1X_2-48.063X_1X_3-28.449X_2X_3+400.746X_1^2+88.247X_2^2-377.996X_3^2$	0.9845	0.9613	9.520e-05	72.07 (425.467)	0.85 (64.552)	40.04 (-379.024)

Note.

* indicated unacceptable R^2 ($R^2 < 0.8$)

** indicated inapplicable suggestion (too slow or too fast flowrate and too high column temperature)

R software also generated stationary points for each independent variable. However, the stationary points were nowhere near the experiment as the extrapolation. In several cases, stationary points in the original unit as an estimation of optimum condition resulted in a negative number or inapplicable estimation for the specific experimental process [29]. Hence, it is recommended to observe each estimation to collect confirmatory data of stationary points prediction. In this RSM analysis, a model for a retention time of genistein, daidzein, and

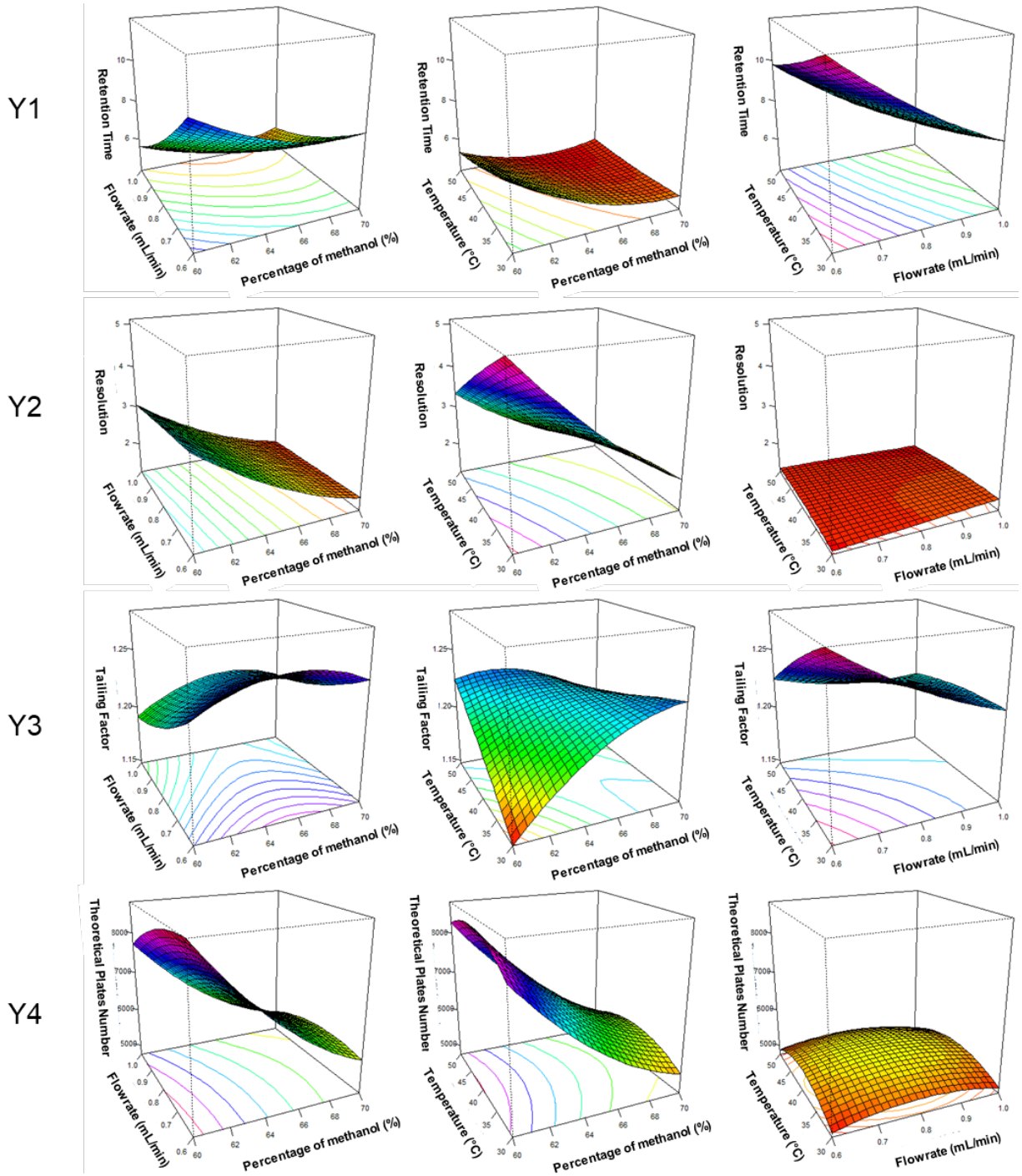

Fig 3. The perspective plot of retention time (Y1), resolution (Y2), tailing factor (Y3), and theoretical plates number (Y4) for genistein

glycitein, and tailing factor for glycitein resulted from high column temperature (more than 50 °C) and did not recommend to be applied in the experiment in order to avoid band broadening and loss of separation efficiency. Unsuitable flow rate estimations were found in the model of genistein resolution and glycitein tailing factor which were too slow and too fast, respectively. If the flow rate was set too slow, the analytical time will be too long and lead to the band broadening. On the other hand, the faster flow rate may increase the column pressure and

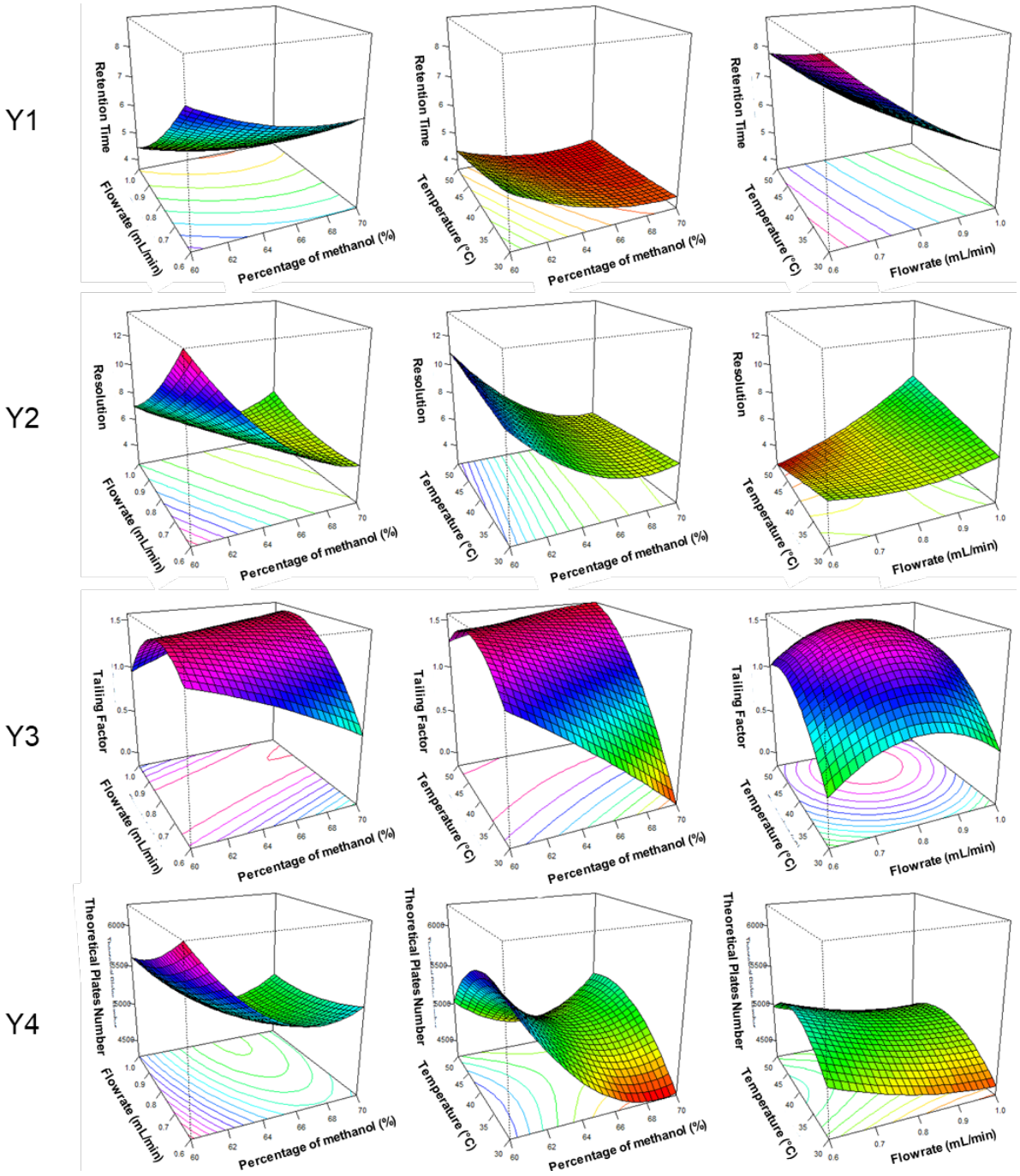

Fig 4. The perspective plot of retention time (Y1), resolution (Y2), tailing factor (Y3), and theoretical plates number (Y4) for daidzein

affecting column lifetime. The eigenvalues indicated the shape of the response surface. The negative eigenvalues direct a downward curvature, whereas positive eigenvalues direct an upward curvature. Hence, all negative eigenvalues indicated the stable point is a maximum, all positive eigenvalues indicated the stable point is a minimum, and the mixed sign of eigenvalues indicated the stable point is a saddle.

The optimized condition has successfully resulted in the good separation of daidzein, glycitein, and genistein for both standard and sample solutions (Fig. 6). The three compounds were completely separated from

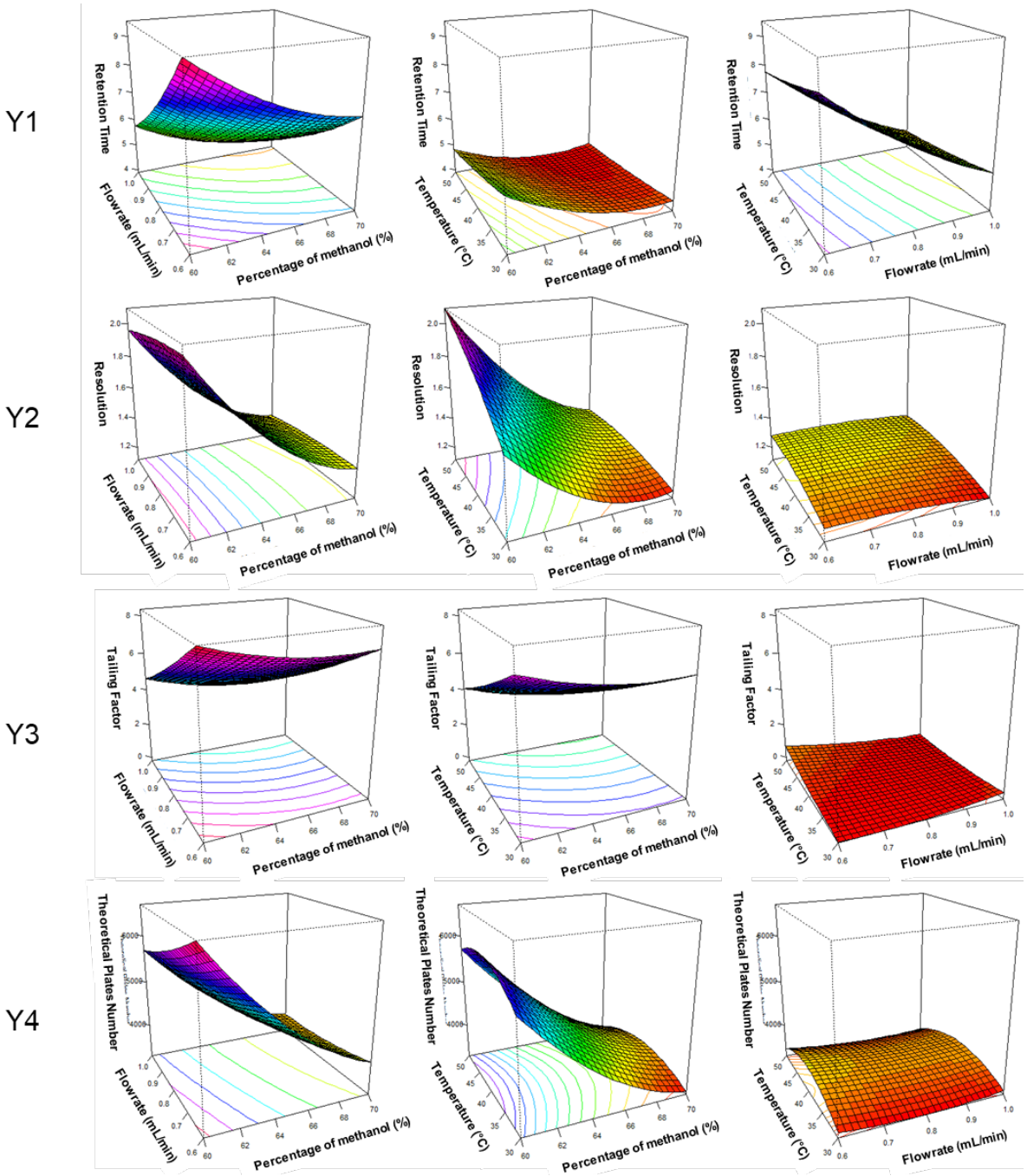

Fig 5. The perspective plot of retention time (Y1), resolution (Y2), tailing factor (Y3), and theoretical plates number (Y4) for glycitein

the soy milk matrix with a short analytical time (less than 10 min). Besides the statistical approach, the equation model of the daidzein tailing factor was applicable to estimate the optimized condition of HPLC separation since the peak of glycitein was detected near the daidzein peak. If the tailing factor of the daidzein peak was too high, it has potentially resulted in a peak broadening, which can be affected by the separation between daidzein and glycitein.

Genistein, daidzein, and glycitein can be separated by reversed-phase HPLC using the C_{18} column [11-12,38-39] due to the possibility of interaction between isoflavone aglycones as organic compounds toward the

silica-based column [40]. Methanol was chosen as one of the mobile phase compositions since it was reported that aglycone isoflavone such as genistein was played an important role as either a hydrogen bond donor or acceptor to interact with methanol and resulted in extra stability for the whole system [41]. The mobile phase composition of methanol-water resulted in stronger interaction with the residual surface silanols compared to acetonitrile-water composition in the same organic mol fraction [34]. However, methanol was chosen in this study since it was less toxic and significantly lower in the cost compared to acetonitrile [42]. The optimized method resulted in a more effective HPLC method compared to the previous study since it can separate not only genistein and daidzein but also glycitein [39]. From the point of view of retention time, the optimized method resulted in the lower retention time of genistein and daidzein compared to other previous studies [11], indicating the time efficiency of HPLC separation.

System Suitability Test

The system suitability test ensures that the method can work properly to be applied in a further analytical stage. HPLC separation properties, including retention time (RT), area, capacity factor (k'), resolution (Rs), tailing factor (TF), and theoretical plate number, were observed. The result of the system suitability test was exhibited in Table 4.

From the system suitability test results, a relative standard deviation of retention time and area were met the acceptance criteria (RSD < 2.0). Capacity factor, tailing factor, and the number of the theoretical plate were also met the acceptance criteria of k' > 2.0, TF < 2.0, and N > 2000, respectively [13]. The general recommendation for peak resolution was Rs > 2.0 as the resolution of genistein and daidzein met this criterion. However, a resolution of 1.5 between two adjacent peaks was quite sufficient for accurate peak integration, if the peak areas are not too much different [43].

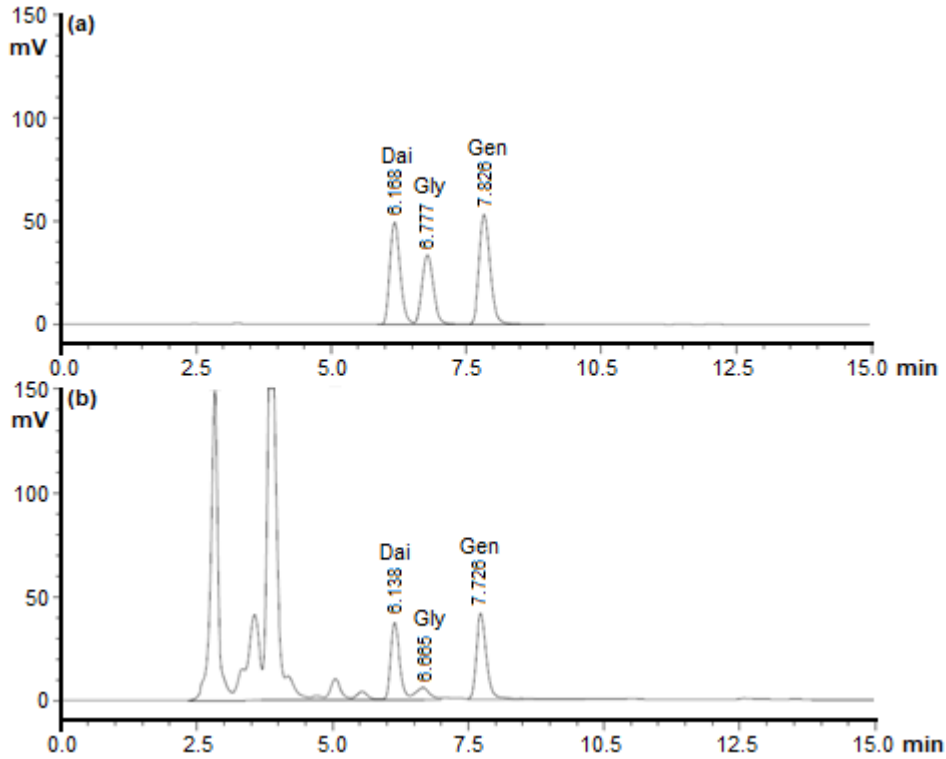

Fig 6. HPLC chromatograms of mixture standard solution (A) and soy milk sample solution (B) containing genistein (Gen), daidzein (Dai), and glycitein (Gly). Mobile phase: methanol-water (63.26:36.74). Flowrate: 0.81 mL/min. Column: C_{18} column of Hibar® 250-4,6 Purospher® STAR RP-18 endcapped (5 µm). Column temperature: 45.31 °C. Wavelength detection at 260 nm. Volume injection: 10 µL

Table 4. Results of system suitability test (n = 5)

Analytes	RT		Area		k'	Rs	TF	N
	Mean	RSD (%)	Mean	RSD (%)				
Genistein	7.611	0.554	804077.6	0.295	8.888	2.043	1.208	5817.173
Daidzein	6.060	0.418	714179.0	0.398	6.875	2.479	1.198	3870.025
Glycitein	6.769	0.655	521229.0	0.342	7.771	1.742	1.111	4070.633

Table 5. Quantitative estimation of genistein, daidzein, and glycitein

Compounds	Standard solution			Sample solution		
	Retention time (min)	Peak Area	Concentration (µg/mL)	Retention time (min)	Peak Area	Concentration (µg/mL)
Genistein	7.826	347095	4.002	7.726	552680	6.372
Daidzein	6.168	545258	7.961	6.138	429648	6.273
Glycitein	6.777	122714	3.204	6.665	109279	2.853

Quantitative Estimation

The research of RSM development for HPLC separation, followed by quantitative estimation, has been performed in the previous studies [27]. A single point calibration method was applied for estimating the content of genistein, daidzein, and glycitein using a single-point calibration method. Quantitative estimation of genistein, daidzein, and glycitein was presented in Table 5. Estimated concentration of genistein, daidzein, and glycitein in soy milk sample were 6.372, 6.273, and 2.853 µg/mL, respectively.

■ CONCLUSION

The employment of rsm package of R software for optimizing of genistein, daidzein, and glycitein separation has successfully performed and useful for soy milk analysis by HPLC method. Independent variables, including methanol percentage in mobile phase composition, flow rate, and column temperature, were optimized using Box-Behnken Design involving response surface methodology. Methanol percentage of 63.26%, the flow rate of 0.81 mL/min, and column temperature of 45.31 °C were applied in the HPLC system and resulted in quantitative estimation of genistein, daidzein, and glycitein in soy milk sample were 6.372, 6.273, and 2.853 µg/mL, respectively. In the future, it is recommended to perform analytical method validation after employing an experimental design in the optimization stage. Employing RSM in HPLC optimization can be applied in other analytical samples such as natural products, pharmaceutical products, and biological samples. However, the economical and effective optimization HPLC method can be carried out with RSM employed with R statistical software. Using the rsm package, it is also possible to generate a sophisticated visualization of the response surface model for supporting data interpretation.

■ SUPPLEMENTARY

S1. R formula for analyzing genistein data
S2. R formula for analyzing daidzein data
S3. R formula for analyzing glycitein data

■ ACKNOWLEDGMENTS

The authors are grateful to the Faculty of Pharmacy, Gadjah Mada University, Yogyakarta, Indonesia, for the provided facilities. This research was financially funded by The Thesis Recognition Program 2019, Directorate of Research, Universitas Gadjah Mada (No.: 3034/UNI/DITLIT/DIT-LIT/LT/2019).

■ REFERENCES

[1] Yimit, D., Hoxur, P., Amat, N., Uchikawa, K., and Yamaguchi, N., 2012, Effects of soybean peptide on immune function, brain function, and neurochemistry in healthy volunteers, *Nutrition*, 28 (2), 154–159.

[2] Messina, M., 2016, Soy and health update: Evaluation of the clinical and epidemiologic literature, *Nutrients*, 8 (12), E754.

[3] Aspiyanto, Susilowati, A., Lotulung, P.D., Melanie, H., and Maryati, Y., 2019, Potency of stirred microfiltration cell in the separation of fermented beans as protein isolate for a natural source of folic acid, *Indones. J. Chem.*, 19 (1), 9–18.

[4] Valsecchi, A.E., Franchi, S., Panerai, A.E., Rossi, A., Sacerdote, P., and Colleoni, M., 2011, The soy isoflavone genistein reverses oxidative and inflammatory state, neuropathic pain, neurotrophic and vasculature deficits in a diabetes mouse model, *Eur. J. Pharmacol.*, 650 (2-3), 694–702.

[5] Lissin, L.W., and Cooke, J.P., 2000, Phytoestrogens and cardiovascular health, *J. Am. Coll. Cardiol.*, 35 (6), 1403–1410.

[6] Yuliani, S.H., Istyastono, E.P., and Riswanto, F.D.O., 2016, The cytotoxic activity on T47D breast cancer cell of genistein-standardized ethanolic extract of tempeh - A fermented product of soybean (*Glycine max*), *Orient. J. Chem.*, 32 (3), 1619–1624.

[7] Jiang, S., Cai, W., and Xu, B., 2013, Food quality improvement of soy milk made from short-time germinated soybeans, *Foods*, 2 (2), 198–212.

[8] Huang, H., Liang, H., and Kwok, K.C., 2006, Effect of thermal processing on genistein, daidzein and glycitein content in soymilk, *J. Sci. Food Agric.*, 86 (7), 1110–1114.

[9] Messina, M., Nagata, C., and Wu, A.H., 2006, Estimated Asian adult soy protein and isoflavone intakes, *Nutr. Cancer*, 55 (1), 1–12.

[10] Abdella, A., El Baz, A.F., Mahrous, E.E., El Maksoud, A.A.A., Ibrahim, I.A., Abdel-Monem, A.R., and Yang, S.T., 2018, Response surface methodology for optimization of genistein content in soy flour and its effect on the antioxidant activity, *Iran. J. Pharm. Res.*, 17 (3), 1026–1035.

[11] Yuliani, S.H., Gani, M.R., Istyastono, E.P., and Riswanto, F.D.O., 2018, Optimization of genistein and daidzein extraction from tempeh- a fermented product of soybean, *J. Pharm. Pharmacogn. Res.*, 6 (4), 231–241.

[12] Rostagno, M.A., Palma, M., and Barroso, C.G., 2007, Ultrasound-assisted extraction of isoflavones from soy beverages blended with fruit juices, *Anal. Chim. Acta*, 597 (2), 265–272.

[13] Snyder, L.R., Kirkland, J.J., and Dolan, J.W., 2010, *Introduction to Modern Liquid Chromatography*, John Wiley & Sons, New Jersey, United States.

[14] Zečevič, M., Živanovič, L., Agatonovic-Kustrin, S., and Minic, D., 2001, The use of a response surface methodology on HPLC analysis of methyldopa, amiloride and hydrochlorothiazide in tablets, *J. Pharm. Biomed. Anal.*, 24 (5-6), 1019–1025.

[15] Riswanto, F.D.O., Rohman, A., Pramono, S., and Martono, S., 2019, Application of response surface methodology as mathematical and statistical tools in natural product research, *J. Appl. Pharm. Sci.*, 9 (10), 125–133.

[16] Yusof, N.A.A., Hadzir, N.M., Ashari, S.E., Hanapi, N.S.M., and Hamid, R.D., 2019, Optimization of enzymatic synthesis of betulinic acid amide in organic solvent by response surface methodology (RSM), *Indones. J. Chem.*, 19 (4), 849–857.

[17] Gunawan, E.R., and Suhendra, D., 2008, Four-factor response surface optimization of the enzymatic synthesis of wax ester from palm kernel oil, *Indones. J. Chem.*, 8 (1), 83–90.

[18] Kalariya, P.D., Namdev, D., Srinivas, R., and Gananadhamu, S., 2017, Application of experimental design and response surface technique for selecting the optimum RP-HPLC conditions for the determination of moxifloxacin HCl and ketorolac tromethamine in eye drops, *J. Saudi Chem. Soc.*, 21 (Suppl. 1), S373–S382.

[19] Ganesamoorthi, B., Kalaivanan, S., Dinesh, R., Kumar, T.N., and Anand, K., 2015, Optimization technique using response surface method for USMW process, *Procedia Soc. Behav. Sci.*, 189, 169–174.

[20] Behera, S.K., Meena, H., Chakraborty, S., and Meikap, B.C., 2018, Application of response surface methodology (RSM) for optimization of leaching parameters for ash reduction from low-grade coal, *Int. J. Min. Sci. Technol.*, 28 (4), 621–629.

[21] Bhavsar, S., Dudhagara, P., and Tank, S., 2018, R software package based statistical optimization of

process components to simultaneously enhance the bacterial growth, laccase production, and textile dye decolorization with cytotoxicity study, *PLoS One*, 13 (5), e0195795.

[22] R Core Team, 2013, R Development Core Team, *R A Lang. Environ. Stat. Comput.*, 55, 275–286.

[23] Beringhs, A.O., Dalmina, M., Creczynski-Pasa, T.B., and Sonaglio, D., 2015, Response surface methodology IV-optimal design applied to the performance improvement of an RP-HPLC-UV method for the quantification of phenolic acids in *Cecropia glaziovii* products, *Rev. Bras. J. Farmacogn.*, 25 (5), 513–521.

[24] Sahoo, D.K., and Sahu, P.K., 2015, Chemometric approach for RP-HPLC determination of azithromycin, secnidazole, and fluconazole using response surface methodology, *J. Liq. Chromatogr. Relat. Technol.*, 38 (6), 750–758.

[25] Kayesh, R., Jahan, M.S., and Sultan, M.Z., 2017, Development using response surface methodology and validation of a stability-indicating RP-HPLC method for simultaneous estimation of azilsartan medoxomil and chlorthalidone in solid dosage form, *Chromatographia*, 80, 593–603.

[26] Bae, H., Jung, H.S., and Jung, J.Y., 2018, Optimization of HPLC-tandem mass spectrometry for chlortetracycline using response surface analysis, *Environ. Eng. Res.*, 23 (3), 309–315.

[27] Munawar, T.M., Prakash, D.V.S., and Vangalapati, M., 2019, Development of response surface methodology for optimization of parameters and quantitative analysis of chebulinic acid from the composition of medicinal herbs by HPLC, *Saudi J. Biol. Sci*., 26 (7), 1809-1814.

[28] Chaudhari, S.R., and Shirkhedkar, A.A., 2019, Design of experiment avenue for development and validation of RP-HPLC-PDA method for determination of apremilast in bulk and in *in-house* tablet formulation, *J. Anal. Sci. Technol.*, 10, 10.

[29] Lenth, R.V., 2009, Response-surface methods in R, using rsm, *J. Stat. Software*, 32 (7), 1–17.

[30] Box, G.E.P., and Behnken, D.W., 1960, Some new three level designs for the study of quantitative variables some new three level designs for the study of quantitative variables, *Technometrics*, 2 (4), 455–475.

[31] Myers, R.H., Montgomery, D.C., and Anderson-Cook, C.M., 2009, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 3rd Ed., John Wiley & Sons, New Jersey, United States.

[32] Chan, C.C., Lee, Y.C., Lam, H., and Zhang, X.M., 2004, *Analytical Method Validation and Instrument Performance Verification*, John Wiley & Sons, New Jersey, United States.

[33] Bezerra, M.A., Santelli, R.E., Oliveira, E.P., Villar, L.S., and Escaleira, L.A., 2008, Response surface methodology (RSM) as a tool for optimization in analytical chemistry, *Talanta*, 76 (5), 965–977.

[34] Rafferty, J.L., Siepmann, J.I., and Schure, M.R., 2011, Mobile phase effects in reversed-phase liquid chromatography: A comparison of acetonitrile/water and methanol/water solvents as studied by molecular simulation, *J. Chromatogr. A*, 1218 (16), 2203–2213.

[35] Meurs, J., 2016, Flow rates in liquid chromatography, gas chromatography, and supercritical fluid chromatography: A tool for optimization, *J. Open Res. Software*, 4 (32), e32.

[36] Åsberg, D., Enmark, M., Samuelsson, J., and Fornstedt, T., 2014, Evaluation of co-solvent fraction, pressure and temperature effects in analytical and preparative supercritical fluid chromatography, *J. Chromatogr. A*, 1374, 254–260.

[37] Purba, N.B.R., Rohman, A., and Martono, S., 2019, The optimization of HPLC for quantitative analysis of acid orange 7 and Sudan II in cosmetic products using box behnken design, *Int. J. Appl. Pharm.*, 11 (2), 130–137.

[38] Chaiyasut, C., Kumar, T., Tipduangta, P., and Rungseevijitprapa, W., 2010, Isoflavone content and antioxidant activity of Thai fermented soybean and its capsule formulation, *Afr. J. Biotechnol.*, 9 (26), 4120–4126.

[39] Sulistyowati, E., Martono, S., Riyanto, S., and Lukitaningsih, E., 2019, Development and

validation for free aglycones daidzein and genistein in soybeans (*Glycine max* (L.) Merr.) using RP HPLC method, *Int. J. Appl. Pharm.*, 11 (2), 138–142.

[40] Ruiz-Angel, M.J., Pous-Torres, S., Carda-Broch, S., and García-Alvarez-Coque, M.C., 2014, Performance of different C18 columns in reversed-phase liquid chromatography with hydro-organic and micellar-organic mobile phases, *J. Chromatogr. A*, 1344, 76–82.

[41] Zhao, H., Song, X., Zhang, Y., and Sheng, X., 2019, Molecular interaction between MeOH and genistein during soy extraction, *RSC Adv.*, 9 (67), 39170–39179.

[42] Anonymous, 2019, The role of methanol and acetonitrile as organic modifiers in reversed-phase liquid chromatography, *Chromatogr. Today*, 12 (1), 24–26.

[43] Ermer, J., and Miller, J.H.M., 2005, *Method Validation in Pharmaceutical Analysis: A Guide to Best Practice*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.