Original Paper

Application of response surface methodologies in capillary electrophoresis

Wen Jun Gong¹, Yu Ping Zhang^{1,*}, Seong Ho Choi², Yi Jun Zhang¹, and Kwang-Pill Lee³

¹ Henan Institute of Science and Technology, Xinxiang 453003, P.R. China

² Department of Chemistry, Hannam University, Daejon, South Korea

³ Department of Chemistry, Graduate School, Kyungpook National University, South Korea

Received December 9, 2005; accepted March 22, 2006; published online June 29, 2006 \oslash Springer-Verlag 2006

Abstract. A fast method of determining ascorbic acid and isoascorbic acid by capillary zone electrophoresis with a photodiode array detector was developed. Response surface methodologies based on three-level, three-variable designs, such as the Box-Behnken design, central composite face-centered and full fractional design, were used comparatively for optimization of buffer pH, buffer concentration and operation voltage. Statistical interpretation of the variables concerning different responses, such as resolution and migration time of the last migrated analyte, were performed. The optimum conditions of these variables were predicted using a second-order polynomial model fitted to the results obtained by applying three designs. The response surface plots using three experimental designs revealed a separation optimum with Tris–HCl buffer of pH 8.5, a concentration of 50 mM, and an operation voltage of 30 kV. The significance of the statistical designs were confirmed by the generally good agreement obtained between predicted responses and actual experimental data. We concluded that experimental designs offer a rapid means of optimizing several variables and provide an efficient test for the robustness of the analytical method.

Key words: Response surface methodology; capillary electrophoresis; ascorbic acid; isoascorbic acid; multi-criteria decision.

Many methods have been developed in order to optimize the parameters of interest in capillary electrophoresis and related techniques [1, 2]. In chemometric approaches experimental measurements are performed in such a way that all variables vary at the same time. An objective function is utilized in which the analyst introduces the desired criteria (selectivity, resolution, time of analysis). The advantage of chemometrics tools is that no explicit models are required, and when models are available, optimization is easier to perform by regression methods. Optimization of a CE separation condition is a critical step, since the wide array of variables, such as applied voltage, buffer composition, ionic strength, temperature, capillary length, and injection time, can influence the separation efficiency, migration time, resolution etc. and a complete and quite general physicochemical model in CE is still missing [3–5]. One approach to achieving optimum separation is to vary the experimental parameter steps while keeping other parameters constant. Yet searching for the optimum separation condition using this approach requires too much experimental work and is tedious and time-consuming. Furthermore, when interaction appears, univariate optimization is not suitable for finding Author for correspondence. E-mail: yupzhang@hotmail.com the best experimental conditions since the influence of

any given variable depends on the magnitude of the other variables.

A suitable alternative to overcoming the aforementioned shortcomings lies in experimental design techniques. Moreover, the number of experiments to be carried out can be reduced drastically when following these chemometric strategies [6–9]. Chemometrics are involved in the preliminary stages of establishing a CE method and the analysis of CE data to extract the maximum amount of significant information. It allows a large number of parameters to be screened simultaneously and to achieve this in a small number of mathematical runs, it is the most important aspect of mathematical design and provides a mathematical framework. Experimental designs such as the Plackeet-Burman design (PBD), Box-Behnken design (BBD), central composite face-centered design (CCF), central composite circumscribed design (CCC), fullfactorial design (FFD) etc. have been used for CE separation studies [10–16]. Several studies have been reported on the use of multivariate statistical analysis to optimize CE methods. Depending on the design, the response model can show the relationship between each parameter. BBD, CCF are the response surface methods used to optimize CE separation [17, 18], which is an efficient statistical tool for optimization of multiple variables to predict the best performance conditions using a minimum number of experiments.

L-ascorbic acid (L-AA), also known as vitamin C, is a natural antioxidant in food and biological systems with important nutritional benefits for human health, e.g. its effect against scurvy, cardiovascular disease and cancer. D-isoascorbic acid (D-IAA), also known as erythorbic acid, the C_5 epimer of L-AA, displaying about 10% of the bioactivity of L-AA is usually added to foods for nonvitamin purposes [19]. However, the stereochemical structures and properties of L-AA and D-AA are so close that quantitative analysis of L-AA in food systems with D-IAA additives is difficult. Several methods used to determine ascorbic acid have been published, including high-performance liquid chromatography (HPLC), gas chromatography (GC), capillary electrophoresis (CE) [20]. Capillary zone electrophoresis (CZE) possesses the advantages of short analysis time, low sample consumption, high separation efficiency, simple experiment operation and so on. In the present work, BBD, CCD and FFD have been compared to determine the optimum separation conditions by CZE. Furthermore, in order to find the best compromise between several responses, a multicriteria

decision-making approach was used, where the resolution response and migration time response can be simultaneously optimized. Baseline separation of the analytes was obtained in 5 min with a resolution larger than 1.5.

Experimental

Chemicals

L-AA, HCl, oxalic acid, EDTA.2Na, tris (hydroxymethyl) aminomethane were purchased from Beijing Chemical Reagent Company, D-Isoascorbic acid was obtained from Beijing Bailingwei Chemical Reagent Company (www.jkchemical.com), doubly distilled water was obtained from a super-purification system (Danyangmen Corporation, Jiangshou, China). The mixed standard of L-AA and D-LAA with a concentration of $10 \mu g/mL$, respectively, were dissolved in a solution containing 0.1% oxalic acid and 1 mM EDTA for stabilization. All electrolyte solutions and samples were filtered through a $0.45 \mu m$ pore size filter and degassed in an ultrasonic bath for 15 min before use.

Apparatus

CE separation was performed with an HP3D CE system with a photodiode array detector, an autosampler, and a power supply able to deliver up to 30 kV (Agilent Technologies, Inc., Walbronn, Germany). CE ChemStation (Hewlett-Packard) software was used for instrument control, data acquisition, and data handling. All experiments were carried out the cationic mode (anode at the inlet and cathode at the outlet). An uncoated fused silica capillary (Yongnian Optical Conductive Fiber Plant, Hebei, China, www.rui-feng.com) of $75 \mu m$ I.D. and 48.5 cm (40 cm effective length) was employed throughout all experiments. The capillary was conditioned before each analysis by flushing successively with H2O and buffer for 5 min, respectively. Samples were injected with 50 mbar pressure for 5 s and separated at the operating voltage.

Software

Polynomial equations and the statistical analysis of the response variables were supported by Microsoft Excel 2000 software (version 5.0; Microsoft Corp., Redmond, WA). Origin 6.0 (Origin Corp., MA, USA) was used for making the response surface diagrams and contour plots.

Results and discussion

Statistical experimental design

The pH of the carrier electrolyte has a dramatic influence on the separation of L-AA and D-AA. The electroosmotic flow is a significant factor when L-AA is analyzed by CE, because it has a highly anionic character in neutral conditions, and can thus migrate quickly on CE using a fused silica capillary. With pKa values of 4.17 and 11.57 for L-AA, the buffer pH must

be above 4.2 to ionize the L-AA. L-AA migrates faster at a higher pH (pH \geq 7.0), whereas the peak of L-AA does not appear at the lower pH ($pH \leq 4.0$) because the electroosmotic flow is very slow under these conditions. However, at and above pH 8.5, imperfect baseline fluctuations occur, and the peak of L-AA is interfered with by the solvent peak. Therefore, the optimum pH of the buffer was about pH 8.0. Increasing the voltage results in shorter migration time and improved separation efficiency. However, as the voltage is further increased, excessive Joule heating results in band broadening. Hence, maximum resolution is obtained by maintaining the voltage below the level at which Joule heating becomes a limiting factor, so an organic buffer was used in our studies. The separation temperature affects the mobility of the species and of the EOF by changing the viscosity of the electrolyte. In our study, it was kept constant at 20° C.

Initial method screening to determine the most significant variables for the analytes did not require an experimental design approach due to previous CE studies [21–25]. As shown in Table 1, three important variables were chosen for the optimization designs, namely the pH, the concentration of the buffer and the applied voltage. In order to calculate quadratic regression model coefficients, each design variable has to be studied at three distinct levels at least.

The BBD, CCFD and FFD models were used comparatively for the multivariable approach. The CCD is based on a full factorial design (Fig. 1a), which is augmented by centre points, axial or start points, which are described as a star (Fig. 1b). So there are 8 cube points (for a full factorial) with levels of -1 and +1, 6 axial or start points with levels of $-\alpha$ and $+\alpha$ and 6 replicates of the centre point. Depending on the α value, three types of CCD are distinguished: central composite circumscribed if $\alpha > 1$, central composite inscribed if $\alpha < 1$, and central composite face centre if $\alpha = 1$. The first two types are spherical

Table 1. Coded and true values of BBD, CCF and FFD model variables

Variables	Code		Level		
	Coded value	True value	-1	0	
pH Buffer concentration	x_1 χ_2	X_1 X_2	7.5 30	8.0 40	8.5 50
(mM) Voltage (kV)	x_3	X_3	20	25	30

Fig. 1. Representation of FFD, CCFD and BBD models

designs, while the latter (CCFD) is a cubic design. Here, for three variables and three levels, a total of 20 experiments were considered.

As with the CCD model, BBD is a response surface method used to examine the relationship between one or more response variables and a set of quantitative experimental parameters. The BBD is not directly based on a full factorial design, as it uses middle points instead of corner points (Fig. 1c). The experimental plan for a three-parameter design is laid out according to the following pattern: two variables have a combination of their extreme levels, while the other is set to its mean value. For a three-parameter design, all experimental points are located on the edges of a cube around the centre points. BBD requires fewer experiments than CCD but covers a slightly smaller experimental region. It is also a spherical design.

The levels of three variables for three designs are shown in Table 1. The maximum and minimum concentration for buffer pH (x_1) were fixed at 7.5 and 8.5, respectively. Likewise, Tris–HCl buffer concentration $(x₂)$ was used in the range of 30–50 mM with an operation voltage (x_3) of 20–30 kV. The BBD model with a total of fifteen experiments (the twelve middle points of the edges on a cube and 3 centre points), the CCFD model with a total of twenty experiments, and FFD with twenty seven experiments are depicted in Tables 2, 3 and 4, respectively. All other experimental plans and the runs were randomized to exclude any bias. The resolution response (R_2) and the migration time (t_2) were monitored during processing. The respective calculated responses for the analytes are also included in Tables 2, 3 and 4.

Table 2. Experimental design and response results using the BBD model

Run	x_1	x_2	x_3	$R_{\rm s}$		t_2	
				Exp.	Pred.	Exp.	Pred.
1	-1	-1	θ	1.20	1.24	4.40	4.46
2	1	-1	θ	0.94	0.98	4.24	4.33
3	$^{-1}$	1	$\overline{0}$	2.22	2.18	6.11	6.02
4	1	1	$\overline{0}$	1.57	1.53	5.28	5.22
5	-1	0	-1	1.75	1.74	7.17	7.02
6	1	0	-1	1.28	1.26	6.43	6.26
7	-1	$\overline{0}$	1	1.69	1.70	4.26	4.43
8	1	0	1	1.26	1.27	4.11	4.26
9	$\overline{0}$	-1	$^{-1}$	0.53	0.50	4.70	4.78
10	0	1	-1	1.23	1.28	6.09	6.33
11	0	-1	1	0.59	0.53	3.05	2.81
12	0	1	1	1.20	1.23	3.80	3.72
13	Ω	$\overline{0}$	θ	0.97	0.97	4.37	4.37
14	0	0	$\overline{0}$	0.97	0.97	4.37	4.37
15	0	$\overline{0}$	$\overline{0}$	0.97	0.97	4.37	4.37

Table 3. Experimental design and response results using the CCFD model

Table 4. Experimental design and response results using the FFD

In order to define the relationship between the responses and the variables, a quadratic regression model should be applied on the basis of a multiple linear regression (MLR). The selected model included 10 coefficients (the constant term, B_0 , three main effects, three quadratic terms and three interaction terms, as indicated in the equation [26–28]:

27 1 1 1 1.23 1.31 4.22 4.22

$$
y = B_0 + \sum_{i=1}^{n} B_i X_i + \sum_{i=j=1}^{n} B_{ij} X_i X_j \tag{1}
$$

In our studies, it can be changed into the following equation according to the n value and the coded values of three variables.

$$
y = B_0 + B_1x_1 + B_2x_2 + B_3x_3 + B_{12}x_1x_2 + B_{13}x_1x_3
$$

+ B₂₃x₂x₃ + B₁₁x₁² + B₂₂x₂² + B₃₃x₃² (2)

where y is the response to be modeled $(R_s \text{ and } t_2)$, B_i are the coefficients of the models by MLR, x_1 is the buffer pH (in coded variable), x_2 is the buffer concentration (in coded variable) and x_3 is the operation voltage (in coded variable).

Fig. 2. (a) Relationship between the resolution response and the coefficients of three models with corresponding SD; (b). The relationship between retention time of the second peak and the coefficients of three models with corresponding SD. Error bars represent \pm SD, \diamond -significant at level 0.05

Coefficients using three mathematical models of different responses could be calculated with the help of the Microsoft Excel MLR kit. They were calculated and plotted from Fig. 2a and b. The comparative effects for three models have been clearly shown. When the coefficient is not included in the 95% confidence interval, this means that it is statistically different from 0 and that therefore the variable associated with this has a significant influence on the response. If it is positive, it favorably influences the response, no matter whether it represents a main, a quadratic effect or a first-order interaction.

As shown in Fig. 2a, buffer concentration, pH and its quadratic term have strong, significant influences on the resolution in the three models. As for retention the time, Fig. 2b shows that pH, buffer concentration, voltage and pH quadratic terms have significant influences on the retention time in the three models.

The significance of the parameters estimated by the least-squares can be assessed using classical statistical tools such as ANOVA. When different single-responses or multi-responses are chosen as the objective function, the most significant variables for different responses will be different. The three models were found to describe the experimental data adequately, with a high confidence level ($p < 0.05$), and led to a coefficient of determination (r^2) of 0.9940, 0.9840 and 0.9646 for R_s and 0.9855, 0.9871 and 0.9567 for t_2 , respectively, indicating the suitability of these models, and allowing the establishment of response surfaces and contour plots and the prediction of any responses within the experimental domain.

On knowing the coefficients, it was possible to obtain the function of the experimental responses related to three variables. Based on the mathematical model, the response surface can be explored graphically. In this case, one can plot the response surfaces and their two-dimensional contour plots against two of the variables, while the third is held constant at a specified level, usually the center value. Figures 3–8 show the

Fig. 3. Response surface plot of R_s using BBD model

 $8₄$

 $\overline{7}$

6

5

 $\overline{4}$

 $3\frac{1}{20}$

 22

 t_2 (min)

 8.4

 8.2 8.0

 $\hat{\gamma}$

7.8

7.6

30

26

28

24

 $V(kV)$

response surfaces and contour plots for the epimeric separation optimization, obtained plotting pH versus buffer concentration, and pH versus voltage. It can be observed that the responses (R_s) show the same behavior among the three models studied. An increase in buffer concentration (C) will clearly improve the resolution, while the pH has a negative effect and voltage a small effect on it. For the migration time (t_2) , voltage has an apparent negative effect, while buffer concentration and pH have little effect on it.

Derringer's desirability function

Generally, responses were transformed into an appropriate desirability scale to balance between different responses. Frequently, different weight factors should be assigned to each response, with larger weights corresponding to more important responses and smaller weights to less important responses [17, 29]. After the individual desirabilities were calculated for each response, they were combined to provide a measure of the composite desirability of the multi-response

system. This measure is the weighted geometric average of the individual desirabilities or the responses. Sometimes, it is very difficult to choose different weights according to the importance of different variables. The most popular methodology applied to multiple response optimization is the desirability function approach [1].

The measured properties of each response Y_i , $i = 1$, 2, ... m, are transformed into a dimensionless desirability scale (d_i) , defined as partial desirability function. This allows combining the results obtained for properties measured on different scales. The scale of the desirability function ranges between $d = 0$, for a completely undesirable response, and $d = 1$, if the response is at the target value. Once the function d_i is defined for each of the m responses of interest, an overall objective function (D) , representing the global desirability function, is calculated by determining the geometric mean of the individual desirabilities. Therefore, the function D over the experimental domain is calculated as follows: $D = (\Pi d_i)^{1/m}$. Taking into account all requirements for m responses, we can

choose the conditions of the design variables that maximize D. In our study, only the resolution response (R_s) and the migration time of the second enantiomer (t_2) were considered. In order to define D quality response, t_2 , and R_s were normalized. The shortest t_2 and the highest R value (the more desirable situation) in all experiments were given the value 1 (maximum), while the longest t_2 and the lowest R_2 value (the unwanted one) were given the value 0 (minimum). Linear interpolation allowed us to calculate the normalized values for the remaining t_2 and R, normalized values were called d_1 , and d_2 , which could be calculated according to the following equations: $d_1 = (R_i - R_{min})/$ $(R_{\text{max}} - R_{\text{min}}); d_2 = (t_{\text{max}} - t_i)/(t_{\text{max}} - t_{\text{min}}).$

A value of D different from zero implies that all responses are in a desirable range simultaneously and consequently. For a value of D close to 1, the combination of the different criteria is globally optimal, so that the response values are near target values (Fig. 9).

After calculation by the SOLVER program in Microsoft Excel, the parameters obtained for the com-

bination responses of the studied analytes using the three experimental designs were as follows:

$$
D_{\rm BBD} = 0.0000 - 0.1401x_1 + 0.2549x_2 + 0.0331x_3
$$

+ 0.0747x₁₂ - 0.0291x₁₃ + 0.005779x₂₃
+ 0.4538x₁² + 0.06186x₂² + 0.1559x₃² (3)

$$
D_{\text{CCD}} = 0.1560 - 0.1524x_1 + 0.1983x_2 + 0.03787x_3
$$

+ 0.01916x₁₂ - 0.04602x₁₃ + 0.08129x₂₃
+ 0.3275x₁² + 0.004924x₂² - 0.1092x₃² (4)

$$
D_{\text{FFD}} = 0.2777 - 0.1469x_1 + 0.2235x_2 + 0.03576x_3
$$

- 0.03767x₁₂ - 0.04038x₁₃ + 0.05613x₂₃
+ 0.2872x₁² - 0.08167x₂² - 0.05698x₃² (5)

One optimal condition $(-1, 1, 1)$ for three experimental designs was obtained with a global degree of satisfaction of D for the combination responses. The coded variable values (x, x, x_3) corresponded to maximum D_{BBD} (1.0), D_{CCD} (0.88) and D_{FFD} (0.97), respectively. An optimal experiment that tested the predictability of

Fig. 8. Response surface plot of t_2 using FFD model

Fig. 9. Shape of the d_i function associated with the response Y_i , resolution (R_s) ; (b) migration time of the last peak

the three models was performed using the optimal condition with buffer of pH 7.5, 50 mM Tris–HCl, and 30 kV. Figure 10 shows the electrophoretogram of the

Fig. 10. Electropherogram of the studied analysis using the optimal conditions: Tris–HCl = 50 mM , pH = 7.5 , 30 kV , wavelength 254 nm. The first peak is ascorbic acid, the second peak is isoascorbic acid

Table 5. Observed and predicted response for testing the predictability of the models

	Response Experimental BBD average value pred. value pred. value pred. value $(n=3)$		CCF	FFD
R_{s}	1.86	2.10	1.89	1.95
t ₂	4.83	5.12	5.01	4.89

studied analytes achieved using the optimized experimental conditions. The results of the experiment were compared to the predicted values from the three models (see Table 5). Compared with the BBD method, closer agreement using the CCF and FFD models could be found in most cases between the observed and predicted responses.

Conclusions

The simultaneous evaluation of the experimental variables involved in the CE optimization using the typical analytes was carried out by means of the BBD, CCFD and FFD models with efficient estimation of the first and second-order coefficients. BBD has fewer design points, which is less expensive to run than the CCFD design with the same number of variables. Moreover, CCFD usually has axial points on the surface of the "cube". FFD is the combination of designed points using BBD and CCFD. Therefore, for the three designs, all design points are sure to fall within a safe operating zone, with similar results for optimization and prediction in our studies. An appropriate use of experimental designs ensures that experimental data contain a maximum of information and provide answers to real chemical problems, and confirm the necessity of applying chemometric techniques in analytical chemistry and show their successful implementation. Compared to empirical methods, chemometrics can greatly simplify the optimization procedure of finding the appropriate experimental conditions.

Acknowledgements. This study was supported by the Program for New Century Talents of the University in Henan Province, Program for Backbone Teachers in Henan Province and Beijing Kangbeide Pharmaceutical Development Company Limited.

References

- [1] Siouffi A M, Phan-Tan-Luu R (2000) J Chromatograph A 892: 75
- [2] Gabrielsson J, Lindberg N O, Lundstedt T (2002) J Chemometrics 16: 141
- [3] Zhang Y P, Lee K P, Kim S H, Gopalan A, Yuan Z B (2004) Electrophoresis 25: 2711
- [4] Suresh Bahu C V, Chung B C, Yoo Y S (2004) Anal Lett 37: 2485
- [5] Jimidar M, Aguiar P F, Pintelon S, Massart D L (1997) J Pharm Biomed Anal 15: 709
- [6] Sentellas S, Saurina J (2003) J Sep Sci 26: 875
- [7] Sentellas S, Saurina J (2003) J Sep Sci 26: 1395
- [8] Gabrielsson J, Lindberg N O, Lundstedt T (2002) J Chemometrics 16: 141
- [9] Zhang Y X, Li H, Havel J (2005) Talanta 65: 853
- [10] Rudaz S, Cherkaoui S, Gauvrit J Y, Lanteri P, Veuthey J L (2001) Electrophoresis 22: 3316
- [11] Ficarra R, Cutroneo P, Aturki Z, Tommasini S, Ficarra P (2002) J Pharm Biomed Anal 29: 987
- [12] Hillaert S, Heyden Y V, Bossche W V (2002) J Chromatogra A 978: 231
- [13] Smadja C, Le Potier I, Chaminade P, Jacquot C, Taverna M (2003) Chromatographia 58: 79
- [14] Servais A C, Fillet M, Chiap P, Abushoffa A M, Hubert P, Crommen J (2002) J Sep Sci 25: 1087
- [15] Avois L M, Mangin P, Saugy M (2003) J Chromatogr B 791: 203
- [16] Loukas Y L, Sabbah S, Scriba G K E (2001) J Chromatogra A 931: 141
- [17] Ragonese R, Macka M, Hughes J, Petocz P (2002) J Pharm Biomed Anal 27: 995
- [18] Hows M E P, Perrett D, Kay J (1997) J Chromatograph A 768: 97
- [19] Liao T, Wu S B, Wu W C, Chang H M (2000) J Agric Food Chem 48: 37
- [20] Tang Y J, Wu M J (2005) Talanta 65: 794
- [21] Herrero-Martinez J M, Simo-Alfonso E, Deltoro V I, Calatayud A, Ramis-Ramos G (1998) Anal Biochem 265: 275
- [22] Versari A, Mattioli A, Parpinello G P, Galassi S (2004) Food Control 15: 355
- [23] Fukushi K, Takeda S, Wakida S I, Yamane M, Higashi K, Hiiro K (1997) J Chromatogr A 772: 313
- [24] Choi O K, Jo J S (1997) J Chromatogr A 781: 435
- [25] Davey M W, Bauw G, Montagu M V (1996) Anal Biochem 239: 8
- [26] Harang V, Tysk M, Westerlund D, Isaksson D, Johansson G (2002) Electrophoresis 23: 2306
- [27] Ivanovic D, Medenica M, Jancic B, Maenovic A, Markovic S (2004) Chromatographia 60: 87
- [28] Jimidar M I, Vennekens T, Ael W V, Redlich D, Smet M D (2004) Electrophoresis 25: 2876
- [29] Alesolo U, Gonzalez L, Jimenez R M, Alonso R M (2003) J Chromatogr A 990: 271