Carmen García-Ruiz
Olga Jiménez
Maria L. Marina

Departamento de
Química Analítica,
Facultad de Química,
Universidad de Alcalá,
Alcalá de Henares (Madrid), Spain

Retention modeling and resolution optimization for a group of *N*-phenylpyrazole derivatives in micellar electrokinetic chromatography using empirical and physicochemical models

The optimization of the separation resolution for a group of *N*-phenylpyrazole derivatives in micellar electrokinetic chromatography (MEKC) as a function of the separation buffer composition (surfactant and organic modifier concentration) has been performed. In order to achieve our purpose, the first step has been the prediction of the migration times of the electroosmotic flow (t_0) and micelles (t_m), and the retention factors of solutes (k), as a function of surfactant (sodium dodecyl sulfate) and alcohol (*n*-propanol or *n*-butanol) concentrations, by means of empirical equations. Also, some physicochemical models have been applied to relate the retention factors to the surfactant and the organic modifier concentrations in order to optimize the separation resolution and to increase our knowledge of the separation process. Finally, a comparison of the resolution optimization through the use of the physicochemical and empirical models selected has been made in order to obtain the optimum separation buffer composition for the separation of a group of 17 *N*-phenylpyrazole derivatives as test solutes.

CE and CEC

1 Introduction

The optimization of the separations performed by micellar electrokinetic chromatography (MEKC) is a complex task due to the great number of variables affecting the process: pH, type and concentration of buffer, surfactant, and organic modifiers. In recent years, in order to predict the optimal separation conditions with the minimum number of experiments, several strategies in MEKC have been reported [1, 2]. As an example, it can be cited the overlapping resolution mapping (ORM) [3–9], iterative regression strategies [10], physicochemical approaches [11–14], empirical equations [15–18], and artificial neural networks (ANNs) [19]. In order to carry out these studies, several designs can be used, *e.g.*, the Plackett-Burman design [20, 21] and the orthogonal array design (OAD) [22] which are factorial designs suitable for screening the influence of many parameters and to monitor possible interactions among a large number of factors, or the central composite design [23, 24] that can provide a response surface for the prediction of areas of optimum performance. From the different strategies cited, the ORM has been one of the most used [3–9] because it allows the deduction of the optimal separation conditions from an overlay of all the graphs obtained plotting the resolution *versus* different separation conditions. However, migration of solutes is not followed. Also, a great number of experiments is required to carry out the optimization process.

Physicochemical models describing the migration behavior of individual solutes have been reported. Thus, Khaledi *et al.* [11–14] introduced physicochemical models describing the migration behavior of both acidic and basic solutes as a function of the separation buffer composition and physicochemical constants. First, a description of migration in terms of pK_a, micelle-water binding constant, and mobility of the anionic solutes in the absence of micelles has been performed [11]. A description of the solute mobility in terms of physical and chemical constants of each solute, the pH of the buffer, and the micelle concentration in the buffer has also been reported [12]. Finally, a model describing the migration of ionizable (acidic and basic) solutes as a function of the simultaneous variation of the pH and micelles concentration has

Correspondence: Dr. Maria Luisa Marina, Departamento de Química Analítica, Facultad de Química, Universidad de Alcalá, E-28871 Alcalá de Henares (Madrid), Spain
E-mail: mluisa.marina@uah.es
Fax: 34-91-8854971

Abbreviations: MPRE, mean prediction relative errors; **PC**, principal component

been proposed [13]. Recently, our research team has developed a physicochemical model to predict the retention of neutral compounds as a function of micellized surfactant and the organic modifier concentration [14]. Although these models usually require the use of non-linear fitting software during the process of optimization, physicochemical migration characteristics of compounds are obtained. Insight is gained in the mechanism of migration in MEKC.

On the other hand, empirical equations have been considered very valuable tools to predict the retention behavior of solutes as a function of different variables in MEKC. Pyell and Bütehorn [15] have proposed linear first-degree models for the migration time of the electroosmotic flow marker, the logarithm of the migration times of the micelles and the logarithm of the retention factors of the solutes. They used these models to predict the resolution between peaks, which were calculated assuming a constant plate number. In practice, the predicted optimum diverges from the real optimum, indicating that the underlying linear function is not able to describe the retention behavior accurately. Later, Bütehorn and Pyell [16] introduced the interaction between the variables considered (surfactant and modifier concentrations) to model the migration time of the electroosmotic flow marker, the logarithm of the migration times of the micelles and the retention factors of the solutes, however, the number of parameters in their empirical equations was equal to the number of experiments achieved, so their models appeared to have better descriptive than predictive quality. Recently, Zomeren *et al.* [18] studied which response should be modeled preferently to enable resolution optimization in MEKC (resolution, separation factor, apparent and effective mobility). Although their results are very promising too much data were required (14 data to obtain the model parameters).

ANNs can be considered as soft models because they do not need mathematical equations [19, 25] and have been usually applied to classification, modeling, association, and mapping [19, 26]. Havel *et al.* [19] have examined the modeling capabilities of the ANN approach in MEKC, with comparison to hard models and the use of ANNs in combination with suitable designs to facilitate the optimization and/or prediction of electrophoretic mobilities in MEKC. Jalali-Heravi and Garkani-Nejad [27] have compared the prediction power of ANNs and multiple linear regression in capillary electrophoresis, pointing out that their results were superior with the ANN. Farková *et al.* [28] have shown that ANNs can be used to estimate peak parameters and which experimental design can be applied for efficient prediction of optimal separation conditions. Srecnik *et al.* [29] have developed an ANN model,

which can be generalized and used in a variety of applications for retention modeling in ion chromatography. Agatonovic-Krustin *et al.* [30] have compared the usefulness of ANNs for response surface modeling in HPLC optimization with multiple regression methods. Gao *et al.* [25] have carried out the optimization of gas chromatographic experimental parameters and compared their results with that obtained by the orthogonal method. Loukas [31] has examined the behavior of a series of training algorithms in the behavior of ANNs and the results were compared from the partial least square (PLS) method. Zhao *et al.* [32] have applied an ANN to model the retention behavior of several solutes in ternary systems in HPLC and to predict two groups of different liver and bile diseases. Madden *et al.* [33] have used ANNs to predict the retention times of anions when eluted with linear hydroxide gradients of varying slope. Jiménez *et al.* [34] have examined the usefulness of ANNs to model the retention behavior of organic solutes in micellar liquid chromatography.

From the results comparison with other empirical and theoretical models it seems that ANNs are the best choice but some drawbacks must be clear. The use of ANNs is not a simple task. Previously, the ANN architecture must be optimized and it means to optimize the number of hidden layers, the number of neurons in the hidden layer(s), the normalization or transformation of the experimental data, the data used to train the network (not only the representativity but also the number of them), the *momentum* and the learning rate, the training algorithm, the transfer functions for the hidden layer(s) and the output layer, *etc.* There are no rules of thumb and "trial and error" must be adopted. Moreover, although the network structure could be carefully optimized some other problems must be taken in mind, that is, the effect of overfitting and, what it is more important, there is no guarantee to find the global minimum. So, recently, Siouffi and Phan-Tan-Lun [2] in a very interesting review about optimization methods in chromatography and capillary electrophoresis have pointed out that the trend is towards the use of ANNs but they are still in infancy when applied in separations. Moreover, if a hard model (empirical or theoretical model) is available and it can predict accurately the magnitude we are interested in, the use of ANNs is not justified. So, our aim in this work has been to perform a global resolution optimization by using empirical and physicochemical models to predict the retention factors of 17 *N*-phenylpyrazole derivatives as a function of the surfactant and modifier concentrations in the buffer composition. Also, it has been considered of great importance to predict the migration time of the electroosmotic flow and micelles due to the scarce data found in the literature.

## 2 Materials and methods

2.1 Chromatographic data

Retention data for the group of 17 *N*-phenylpyrazole derivatives considered as model solutes (Table 1 shows their structures and names) when separation buffer contains 0.08 M of 2-(*N*-cyclohexylamino)ethanesulfonic acid (CHES) in alkaline medium (pH 10) with SDS as surfactant and *n*-propanol or n-butanol as the organic modifier have been used in this work [35]. Although the experimental conditions employed to perform the separation of these compounds are widely explained in [35], they are summarized below: an Applied Bio-systems capillary electrophoresis instrument 279A-HT model (Norwalk, CT, USA), with UV detection at 238 mn and temperature controlled at 30°C was used to obtain the electropherograms of the compounds under study. The dimensions of the fused-silica capillary (Polymicro Technologies, Phoenix, AZ, USA) used were 75 cm of total length and 50 cm of effective length with 25 μm of inner diameter and 375 mm of outer diameter. The applied voltage was 15 kV.

Table 1. Identification numbers, names and structures of *N*-phenylpyrazole derivatives studied

No.	Name	R_3	R_4	R_5	R_2'
1	DNPP	H	H	H	NO_2
2	3-Methyl DNPP	Me	H	H	NO_2
3	4-Methyl DNPP	H	Me	H	NO_2
4	4,5-Dimethyl DNPP	H	Me	Me	NO_2
5	3-Ethyl DNPP	Et	H	H	NO_2
6	4,5-Dimethyl *p*NPP	H	Me	Me	H
7	3,4,5-Trimethyl DNPP	Me	Me	Me	NO_2
8	4-Methyl *p*NPP	H	Me	H	H
9	3-Methyl-4-nitro-5-cloro DNPP	Me	NO_2	Cl	NO_2
10	3,5-Dimethyl *p*NPP	Me	H	Me	H
11	4-Bromo *p*NPP	H	Br	H	H
12	3-Bromo-4-methyl DNPP	Br	Me	H	NO_2
13	3,5-Dimethyl-4-bromo DNPP	Me	Br	Me	NO_2
14	5-Methyl-4-bromo *p*NPP	H	Br	Me	H
15	3-*tert*-Butyl *p*NPP	*t*-Bu	H	H	H
16	3,4-Dibromo DNPP	Br	Br	H	NO_2
17	3-Ethyl-4-bromo DNPP	Et	Br	H	NO_2

Structure
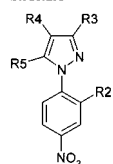
DNPP, dinitrophenylpyrazole; *p*NPP, paranitrophenylpyrazole

2.2 Data treatment

Data have been divided into two sets, the model and the prediction data sets following the factorial design plotted in Fig. 1. Model data set (crosses) permits the calculation of the parameters of the different equations and the test data set (circles) have been used to evaluate the prediction capability of the empirical and physicochemical equations used in this work. It should be mentioned that when the buffer consisted of 0.03 M SDS and 0.03 as the volume fraction of *n*-butanol, results obtained have proved to be outliers and were not used to build any of
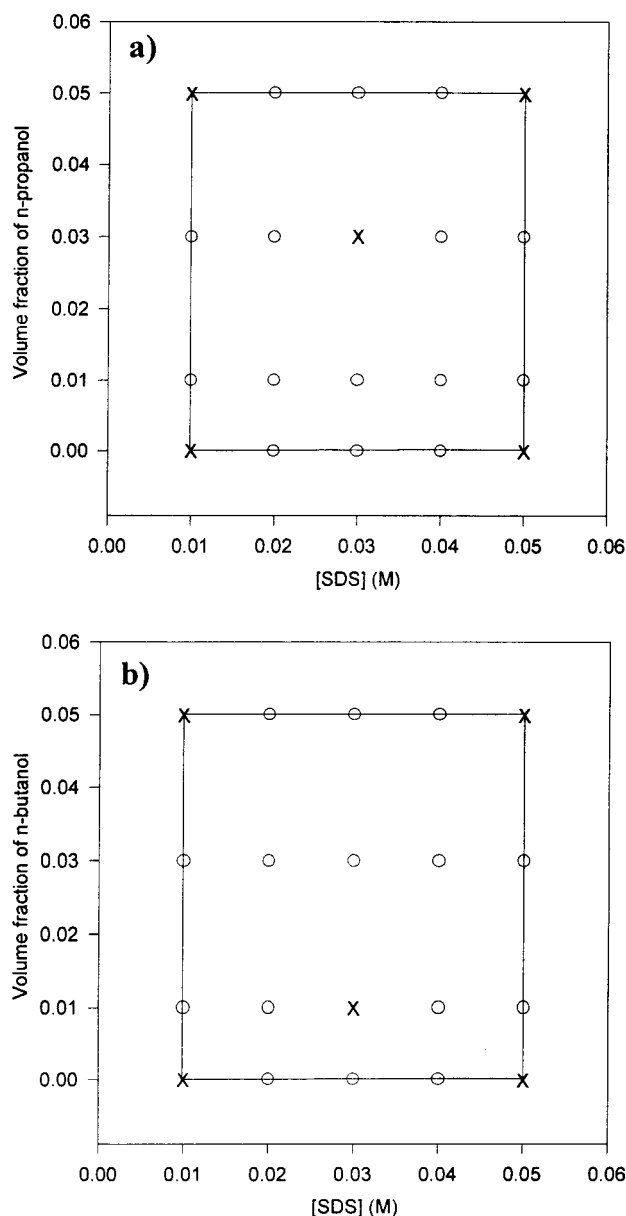

Figure 1. Composition of the MEKC systems using a 0.08 M CHES (pH 10) buffer. Circle data means the prediction data set and crosses the model data set.

the models. The evaluation of empirical equations has been made using box plots performed by the Sigma Plot software [36]. Data treatment was made using Microsoft Excel [37]. Multiple regression analysis and tests performed in this work to compare the error averages obtained in different experimental conditions were carried out using the Statgraphics Plus software [38]. Nonlinear regressions were performed by using the Sigma Plot software [36]. The performance of equations considered in this work has been evaluated comparing the relative errors defined as follows:

$$PRE(\%) = \frac{|k_{cal} - k_{exp}|}{k_{exp}} 100 \tag{1}$$

$$MPRE(\%) = \sum_{i=1}^{i=N} \frac{PRE_i}{N} \tag{2}$$

where k_{cal} is the calculated retention factor by the equation that is being testing, k_{exp} is the experimental retention factor, N is the number of different separation buffers (with different concentrations of surfactant and alcohol).

3 Results and discussion

As mentioned, the optimization of separations performed by MEKC is complex and difficult due to the high number of parameters affecting the separation process. Thus, resolution of two closely adjacent peaks (R_s), in MEKC, is dependent on the selectivity factor, the mean retention factor and the ratio of the migration time of the electroosmotic flow to the migration time of the micelles according to the following equation proposed by Terabe and Cheng [39]:

$$R_S = \frac{\sqrt{N}}{4} \frac{\alpha - 1}{\alpha} \frac{k}{k+1} \frac{1 - \frac{t_0}{t_m}}{1 + \frac{t_0}{t_m} k} \tag{3}$$

where N is the theoretical plate number; k is the mean retention factor of the two peaks considered; α is the selectivity factor (k_j/k_i being $k_j > k_i$), and t_0 and t_m are the migration times of the electroosmotic flow and micelles, respectively. According to Eq. (3), resolution optimization requires the knowledge of t_0, t_m, k, and N values. In the following calculations, the efficiency of the chromatographic system (N) has been considered independent of the electrolyte composition [15, 16]. However, the values of t_0, t_m, k, and α are dependent on the electrolyte composition. In this study, the prediction of these parameters has been performed by means of empirical and physicochemical equations in order to optimize the resolution.

3.1 Use of empirical equations for the prediction of electroosmotic flow and micelles migration times

Pyell and Bütehorn [15] first stated the importance of predicting the migration times of the electroosmotic flow (t_0) and micelles (t_m) as a function of the buffer composition. However, although they have proposed the use of different equations [15, 16] no validity studies have been performed. So, in this work a systematic study on equations relating migration times of the electroosmotic flow and micelles with the surfactant concentration and the volume fraction of the organic modifier (μ and ϕ, respectively) has been performed (Table 2). It must be mentioned that two of them have been used previously by Pyell and Bütehorn [15], Eq. (22), and by Bütehorn and Pyell [16], Eq. (4). In order to evaluate the empirical equations (Eqs. 4–33) that best predict the migration times of the electroosmotic flow and micelles, multiple regression analysis has been performed by using the model data set shown in Fig. 1. Once the parameters have been calculated, the prediction has been performed in the conditions shown for the test data set (Fig. 1).

Table 2. Empirical equations used for the prediction of electroosmotic flow (t_0) and micelles (t_m) migration times

Basic equation	z value	Equation No.	z value	Equation No.
$z = A + B\mu + C\phi$ $+ D\mu\,\phi$	t_0	4	t_m	5
	$1/t_0$	6	$1/t_m$	7
	$\lg t_0$	8	$\lg t_m$	9
$z = A + B\mu\,\phi$	t_0	10	t_m	11
	$1/t_0$	12	$1/t_m$	13
	$\lg t_0$	14	$\lg t_m$	15
$z = A + B\phi + C\mu\,\phi$	t_0	16	t_m	17
	$1/t_0$	18	$1/t_m$	19
	$\lg t_0$	20	$\lg t_m$	21
$z = A + B\mu + C\phi$	t_0	22	t_m	23
	$1/t_0$	24	$1/t_m$	25
	$\lg t_0$	26	$\lg t_m$	27
$z = A + B\mu + C\mu\,\phi$	t_0	28	t_m	29
	$1/t_0$	30	$1/t_m$	31
	$\lg t_0$	32	$\lg t_m$	33

In order to establish the validity of the tested equations, different criteria have been taken into account. That is, the most simple equation for which the lowest mean prediction relative errors are obtained is our objective. Nevertheless, we must take in mind that all the terms must be statistically significant. In this respect, for the prediction of t_0 when *n*-propanol is used, the lowest mean prediction

relative errors are obtained for Eqs. 4, 6, 8, 16, 22, 24, and 26 (no statistically significant differences, $P > 0.05$) as can be observed in Fig. 2a. From these equations, we have chosen Eq. (4) because all the terms are statistically significant (indicating that it is not possible to simplify it to
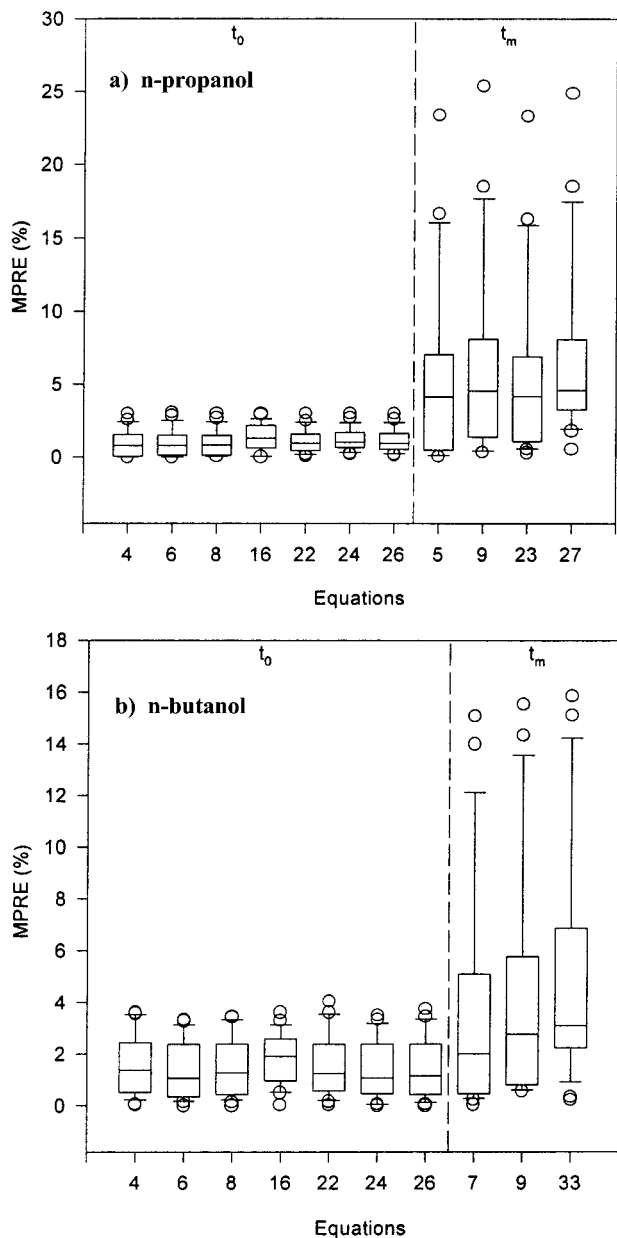

Figure 2. MPRE for electroosmotic flow (t_0) and micelle (t_m) migration times by using empirical equation in a MEKC system with (a) *n*-propanol or (b) *n*-butanol as the organic modifier. These plots are defined in terms of percentiles and take a quick look at the median and spread of the data, as well as the mean, minimun, and the maximum values for the variable studied. Outliers of percentiles are also represented by circles. The lengths of the upper and lower lines associated to each box show how stretches the tails of the distribution are.

obtain Eqs. 16, 22, 24, or 26) and due to it is the simplest one compared with Eqs. (6) and (8). Following the same reasoning exposed previously, we have chosen Eq. (4) to predict t_0 (Fig. 2b) when *n*-butanol is considered as the organic modifier and Eqs. (23) and (7) (Figs. 2a and b) to predict t_m when *n*-propanol and *n*-butanol are considered as the organic modifier, respectively. It can be noted that the equation that best predicts the migration time of the separation buffer is the same when the organic modifier used is *n*-propanol and *n*-butanol (Eq. 4), but that the equations differ when the prediction of the migration time of the micelle is achieved. This fact is in agreement with Van Hove *et al.* [40] that considers these two alcohols belonging to different categories.

3.2 Use of empirical equations for retention prediction

In order to predict the retention of the compounds under study, four empirical equations (Eqs. 34–37) proposed by Jiménez *et al.* [17] and another one empirical equation (Eq. 38) proposed by Pyell and Bütehorn [15] relating retention factors with the total surfactant concentration and the volume fraction of the alcohol (μ and ϕ, respectively) have been employed:

$$\frac{1}{k} = A + \frac{B}{\mu} + C\phi \tag{34}$$

$$\frac{1}{k} = A + \frac{B}{\mu} + C\frac{\phi}{\mu} \tag{35}$$

$$\frac{1}{k} = A + \frac{B}{\mu} + C\frac{\phi^2}{\mu} \tag{36}$$

$$\frac{1}{k} = A + \frac{B}{\mu} + C\frac{\phi}{\mu} + D\phi^2 \tag{37}$$

$$\ln k = A + B\ln\mu + C\phi \tag{38}$$

Multiple regression analysis has been performed to obtain the equation parameters (Eqs. 34–38) by using the model data set shown in Fig. 1. Then, retention factors for the prediction data set (Fig. 1) were calculated.

From the comparison of the mean prediction relative errors it can be observed (Figs. 3a and b) that when *n*-propanol was used, the lowest errors were obtained by means of Eqs. (35–38) (no statistically significant differences observed, $P > 0.05$). When *n*-butanol is considered, the lowest errors were obtained by means of Eq. (37) and (38) (no statistically significant differences observed, $P > 0.05$). However, the multiple regression analysis shows that Eq. (37) contains a nonsignificant term, reason for which Eq. (38) was selected as the most appro-
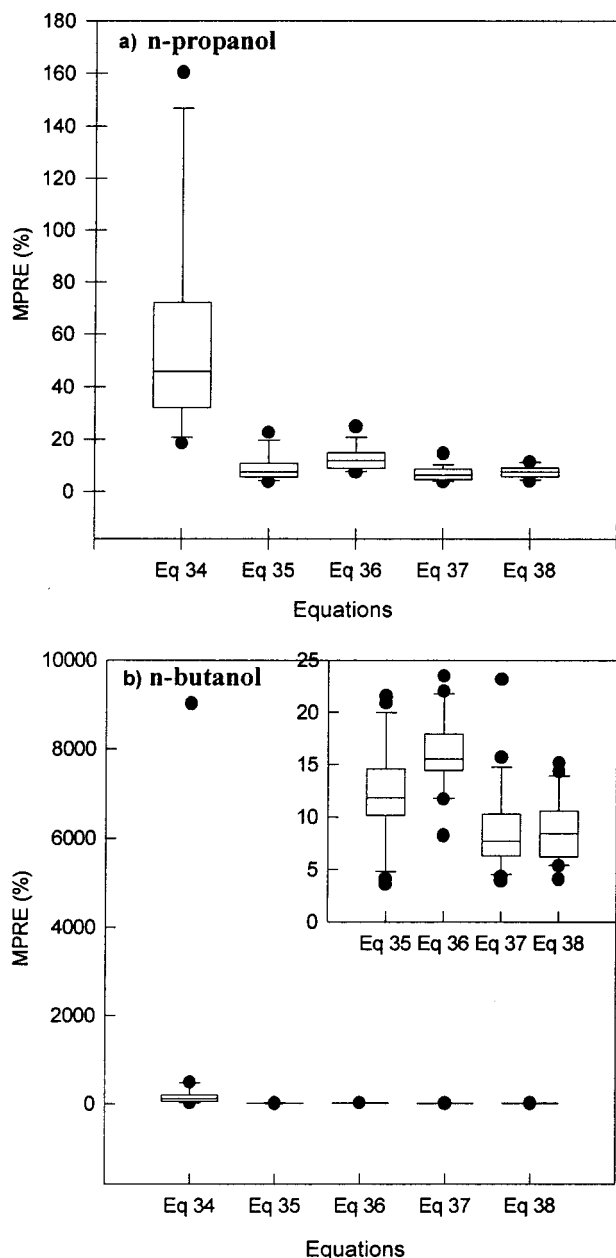
Figure 3. MPRE for the retention factors by using empirical equations in a MEKC system with (a) *n*-propanol or (b) *n*-butanol as the organic modifier.

priate to predict the retention of the solutes considered (prediction errors 3.9–11.5% for *n*-propanol and 5.4–15.1% for *n*-butanol).

Nevertheless, Jiménez *et al.* [17] found that Eq. (35) was the most appropriate to predict the retention behavior of a group of dihydropyridines in similar MEKC systems using the same five empirical equations employed in this work. These results could be explained by the different characteristics of each group of compounds. As an example, the
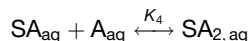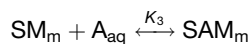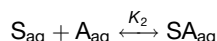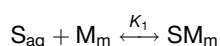
literature shows that the group of the *N*-phenylpyrazole derivatives studied in this work are less hydrophobic (with values of the logarithm of the octanol-water distribution coefficient, log P_{ow}, ranging from 2.08 to 2.60 [41]) than the group of dihydropyridines (with values of the log P_{ow} ranging from 2.43 to 4.31 [42]).

3.3 Use of physicochemical models for retention prediction

The retention prediction through the use of a physicochemical model relating the retention factor with the micellized surfactant (SDS) and the organic modifier (*n*-propanol and *n*-butanol) concentrations has been performed. The physicochemical model [14] considers different interactions among the species present in the system according to several equilibria. Relating these equilibria with the solute retention obtained by MEKC the following equation can be obtained:

$$k = \frac{fK_1\left(1 + K_3\left[A_{aq}\right]\right)\left[M_m\right]}{1 + K_2\left[A_{aq}\right] + K_2K_4\left[A_{aq}\right]^2} \tag{39}$$

where, k is the retention factor, f is the phase ratio, $[A_{aq}]$ and $[M_m]$ are the alcohol and the micellized surfactant concentrations (molar concentrations), respectively, and K_1, K_2, K_3, and K_4 are equilibrium constants. These constants take into account the association of the solute with the micelle to form a complex in the micellar pseudophase (K_1), the enhancement of solubility of solute in the separation buffer modified by alcohols (K_2), the formation of complexes among the solute, the alcohol and the micelle in the micellar pseudophase (K_3), and the complex formation between the solute-alcohol complex and other molecules of alcohol (K_4), as it is shown in the following equilibria:

$$S_{aq} + M_m \xrightarrow{K_1} SM_m$$

$$S_{aq} + A_{aq} \xrightarrow{K_2} SA_{aq}$$

$$SM_m + A_{aq} \xrightarrow{K_3} SAM_m$$

$$SA_{aq} + A_{aq} \xrightarrow{K_4} SA_{2,aq}$$

Nevertheless, some simplifications to this equation can be done, according to the following approximations [14]:

(i) If $1 \gg K_3\,[A_{aq}]$ the simplified expression is:

$$k = \frac{fK_1[M_m]}{1 + K_2\left[A_{aq}\right] + K_2K_4\left[A_{aq}\right]^2} \tag{40}$$

(ii) If $1 \ll K_3\,[A_{aq}]$ the simplified expression is:

$$k = \frac{fK_1K_3\left[A_{aq}\right][M_m]}{1 + K_2\left[A_{aq}\right] + K_2K_4\left[A_{aq}\right]^2} \tag{41}$$

(iii) If $1 \gg K_4 [A_{aq}]$ the simplified expression is:

$$k = \frac{fK_1 \left(1 + K_3 [A_{aq}]\right) [M_m]}{1 + K_2 [A_{aq}]} \qquad (42)$$

In order to check the validity of this model, the retention data for a group of test solutes (17 *N*-phenylpyrazole derivatives) have been used. First of all, nonlinear regressions were achieved to obtain the equilibrium constant values from Eqs. (39)–(42) in order to predict the retention factors. These values were used to calculate the mean prediction relative errors according to the four physicochemical equations described. Mean relative errors for the four equations considered in MEKC systems with *n*-propanol (5.7, 20.2, 31.4, 5.8 for Eqs. 39, 40, 41, and 42, respectively) and *n*-butanol (4.8, 5.4, 30.1, and 6.9 for Eqs. 39, 40, 41, and 42, respectively) have been compared with a multiple comparison procedure [38] to establish the equation that best model the experimental retention behavior. The results from the comparison indicate that Eqs. (39) and (42) are not statistically different when *n*-propanol is used as the organic modifier and Eqs. (39), (40) and (42) are not statistically different when *n*-butanol is employed as the organic modifier. Although for the two systems considered (*n*-propanol and n-butanol), Eqs. (39) and (42) do not differ significantly, the mean error values are lower when Eq. (39) is used, so it can be considered that Eq. (39) is the best equation, although the use of nonlinear regression is necessary. From the results exposed in the two preceding sections we can conclude that Eqs. (38) and (39) are, respectively, the best equations to predict retention factors by means of empirical and physicochemical models. Moreover, from the comparison of mean prediction errors obtained, the statistical test indicates that the physicochemical model is the best to explain the retention behavior of the solutes under study (confidence level of 95%).

3.4 Resolution optimization through the use of physicochemical and empirical models

The evaluation of all these equations has permitted to choose the best equations for the prediction of k, t_0, and t_m with the aim of optimizing the resolution (R_s) between two consecutive peaks using Eq. (3). In this equation it has been considered an N value constant and equal to 150 000 [43]. The mean retention factor of the two peaks considered has been calculated according to the empirical Eq. (38) or the physicochemical Eq. (39); and t_0 and t_m have been obtained, respectively, from empirical Eqs. (4) and (23) for *n*-propanol, and from Eqs. (4) and (7) for *n*-butanol. The resolution has been optimized according to the criterion of global resolution given in [44]. This opti-

mizing criterion is based on the normalized product, r, of different properties, $X_{i,i+1}$, associated to pairs of consecutive peaks.

$$r = \prod_{i=1}^{n-1} \frac{X_{i,i+1}}{\left(\sum_{i=1}^{n-1} X_{i,i+1}(n-1)\right)^{n-1}} \qquad (43)$$

In this work, X is the resolution predicted with the above-mentioned equations and n is the number of solutes.

Figure 4 shows the response surfaces obtained for systems containing SDS as the surfactant and *n*-propanol or *n*-butanol as the organic modifier. The equations used for the prediction of retention factors are Eqs. (38) (Figs. 4a and b) and (39) (Figs. 4c and d), respectively. In both cases, the prediction of t_0 and t_m was achieved by empirical Eqs. (4) and (23) when *n*-propanol was used and Eqs. (4) and (7) when *n*-butanol was employed. It can be observed that the maximum resolution to separate the group of *N*-phenylpyrazole derivatives studied using empirical equations corresponds to a 0.08 M CHES buffer (pH 10) with 0.02 M in SDS and 0.03 volume fraction of *n*-propanol (Fig. 4a). On the other hand, the maximum resolution obtained for the model solutes studied using the empirical equations selected correspond to 0.08 M CHES buffer (pH 10) with 0.05 M SDS and 0.03 volume fraction of *n*-butanol (1), or with 0.01 M SDS without *n*-butanol (2) (Fig. 4b). When the physicochemical model (Eq. 39) has been used, the response surface for *n*-propanol shows two maxima which correspond to 0.08 M CHES buffer (pH 10) with 0.05 M SDS and a 0.05 volume fraction of *n*-propanol (1) or with 0.02 M SDS and a 0.01 volume fraction of *n*-propanol (2) (Fig. 4c). The best resolution is obtained for maximum 2. On the other hand, when *n*-butanol is used as organic modifier only a maximum is obtained, corresponding to 0.08 M CHES buffer (pH 10) with 0.02 M SDS and a 0.03 volume fraction of *n*-butanol (Fig. 4d).

Figure 5 shows the experimental electropherograms for a selected mixture of solutes and for the best conditions chosen from the response surfaces shown in Fig. 4. From the experimental electropherograms shown in Fig. 5 some conclusions can be drawn: (i) As expected, Eq. (39) leads to the best optimal conditions for the separation of the solutes considered in this study if we compare with the results obtained by means of Eq. (38). For example, if *n*-propanol is considered (Figs. 5a and c) solutes 5 and 6 are resolved (although not completely) but when Eq. (38) is used in the retention prediction these peaks coelute. Moreover, when butanol is used as the organic modifier (Figs. 5b and d), and Eq. (39) is used, solutes 15 and 16 are resolved (although not completely) but when Eq. (38) is used these two solutes coelute. (ii) It is interesting to
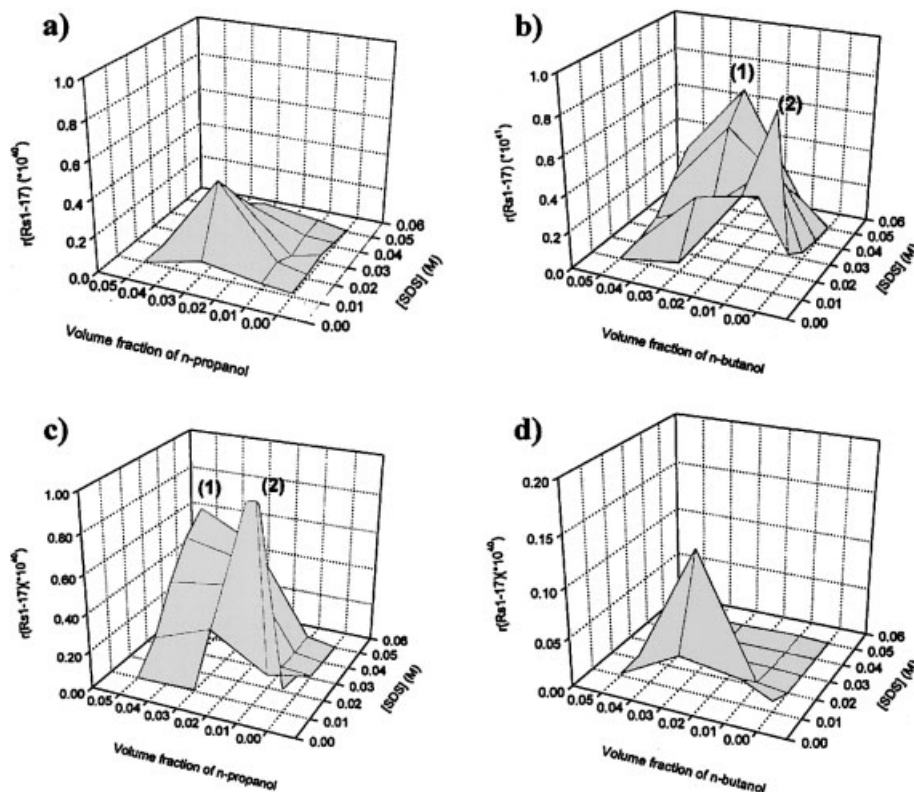
Figure 4. Response surfaces obtained for the group of *N*-phenylpyrazole derivatives studied in the MEKC system with SDS as surfactant and *n*-propanol or *n*-butanol as the organic modifier using the predicted retention factors with (a, b) the empirical Eq. (38) or (c, d) the physicochemical Eq. (39). Conditions of maximum resolution: (a) 0.08 M CHES buffer (pH 10) with 0.02 M SDS and 3% *n*-propanol; (b) 0.08 M CHES buffer (pH 10) with 0.05 M SDS and 3% *n*-butanol (1) or with 0.01 M SDS without *n*-butanol (2); (c) 0.08 M CHES buffer (pH 10) with 0.05 M SDS and 5% *n*-propanol (1) or with 0.02 M SDS and 1% *n*-propanol (2); (d) 0.08 M CHES buffer (pH 10) with 0.02 M SDS and 3% *n*-butanol.
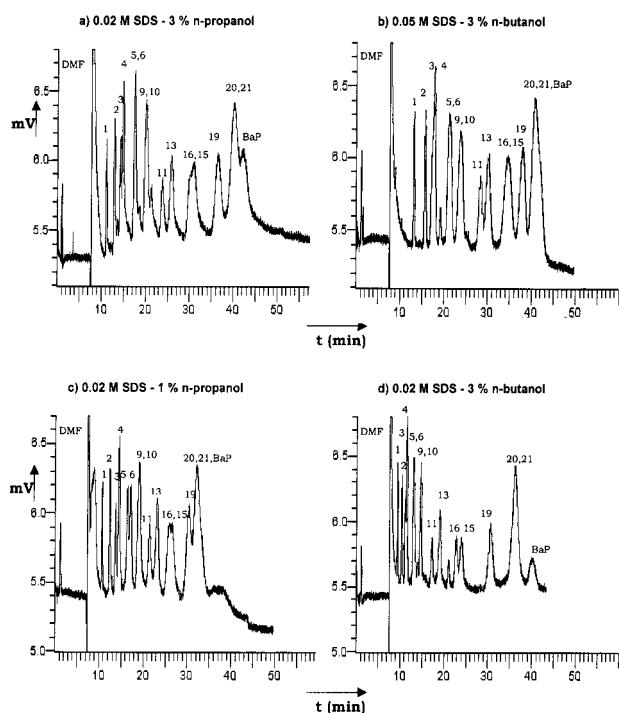

Figure 5. Electropherograms corresponding to the injection of a mixture of 15 *N*-phenylpyrazole derivatives in the optimal resolution conditions obtained in Fig. 4. DMF: dimethylformamide; BaP, benzo[*a*]pyrene.

note that *n*-propanol (Fig. 5c) better separates the first-migrating compounds in the electropherograms while *n*-butanol better separates the last-migrating compounds as has been reported previously in [35]. So, these two alcohols work complementarily and the most adequate one would depend on the solutes we are interested in.

3.5 Mechanistical approaches through the physicochemical model

The values of the physicochemical constants obtained from Eq. (39) are shown in Table 3. These results clearly show that the constant values depend not only on the solute nature but also on the alcohol nature. From the fK_1 values obtained, we can conclude that although the hydrophobicity is the main driving force affecting the chromatograhic behavior, it is not the only one. Thus, a plot of the fK_1 values *versus* log P_{ow} shows a moderately strong relationship between the variables (correlation coefficient, r, 0.8297). A better correlation is obtained when the logarithm values of fK_1 are plotted *versus* the log P_{ow} ($r = 0.9564$), but solutes with the same hydrophobicity differ in fK_1 values, probably due to electrostatic interactions (*i.e.*, solutes 2 and 4, both with log P_{ow} 2.16, solutes 11 and 16, both with log P_{ow} 2.43, and solutes 8 and 12, both with log P_{ow} 2.41).

Table 3. Equilibrium constants obtained from physicochemical model considered (Eq. 39)

Solute No.	*n*-Propanol				*n*-Butanol			
	fK_1	K_2	K_3	K_4	fK_1	K_2	K_3	K_4
1	38	2.52	0.89		38	3.06	0.27	
2	67	2.67	0.87		67	3.43	0.19	
3	96	5.15	2.11		67	3.06	0.03	
4	109	3.46	0.94		109	4.18		
5	162	2.16	0.36		161	3.43		0.44
6	188	6.11	1.98		188	4.04	2.83	2.49
7	196	4.75	1.71		196	4.29	0.05	
8	233	4.35	1.70		232	3.81	0.09	
9	236	17.50	8.28		236	3.53		
10	268	4.85	1.46		267	5.63	0.06	
11	363	2.13	0.48		357	2.75		0.70
12	470	1.61	0.45		403	1.78		1.22
13	493	4.19	1.48		757	4.41		
14	757	6.15	2.69		753	3.86		0.41
15	795	1.12	0.38		784	7.95	7.71	3.91
16	740	1.59	0.99	1.18	736	2.28		24.70
17	1550	5.01	5.50		1550	0.44		18.70

The values of K_2 and K_3 for the solute number nine are significantly higher than those for the other solutes when *n*-propanol is considered, indicating that favorable interactions between the alcohol and the nitrogen atom in the pyrazole ring of this compound are possible, both in the micellar phase and in the electrophoretic buffer. This fact leads to contrary effects to the retention, first the formation of complexes in the micellar pseudophase increases the retention in this phase and, second, the formation of complexes in the electrophoretic buffer drifts the retention towards the hydro-organic phase. When *n*-butanol is considered, the complex formation between the alcohol, the solute and the micelle is negligible. This alcohol is more hydrophobic than *n*-propanol, so the aqueous phase polarity and the micelle surface charge density of the micelle diminish [45]. As a consequence, the electrostatic interactions are diminished, fact that can be viewed by the lower value of K_2 and the negligible value of K_3 constants. Generally, K_3 values are lower for *n*-butanol than for *n*-propanol with two exceptions (solutes 6 and 15). These solutes have a very voluminous substituent in the R_3 position (see Table 1), so perhaps steric impediments can be expected. *n*-Butanol could enhance the retention of this compound (K_3 value higher in *n*-butanol than in *n*-propanol) due to the expansion of the micelle or to the better solubilization power of this solvent inside the core of the micelle. The highest values of K_4 (solute 16 for systems containing *n*-propanol and solutes 16 and 17 for systems containing *n*-butanol) can be attributed to interactions between the alcohol head group and the positive charge density of the two nitrogen atoms (in the pyra-

zole ring). For the other compounds, the K_4 values are low or negligible, so in these cases the retention could be explained by means of a more simplistic model (Eq. 42).

In order to clarify the relationship among the retention behavior and the structural properties of the solutes, principal component (PC) and cluster analysis have been achieved. Massart *et al.* [46] highly recommended that a clustering method is combined with a PC output. So, in this work the four constant values have been reduced to a lower number of variables by means of the PC analysis (two and three components have been extracted when *n*-propanol and *n*-butanol are considered, respectively).

The PCs extracted were: (i) for *n*-propanol:
$PC1 = 0.999998 fK_1 - 0.000781284 K_2 + 0.00159507 K_3 + 0.000170362 K_4$;
$PC2 = -0.0000369037 fK_1 + 0.888103 K_2 + 0.459474 K_3 - 0.0125016 K_4$.
(ii) For *n*-butanol:
$PC1 = 0.999924 fK_1 - 0.000893064 K_2 + 0.000838577 K_3 + 0.0122405 K_4$;
$PC2 = -0.0121667 fK_1 - 0.127968 K_2 - 0.0691081 K_3 + 0.989293 K_4$;
$PC3 = -0.00177059 fK_1 + 0.563511 K_2 + 0.815836 K_3 + 0.129861 K_4$.

Then, the cluster analysis has been applied to the components extracted in every case. In this way, we can classify the compounds with the same retention behavior. In Fig. 6, the cluster scatterplots for systems containing *n*-propanol and *n*-butanol are shown (*PC* 2 vs. *PC* 1).
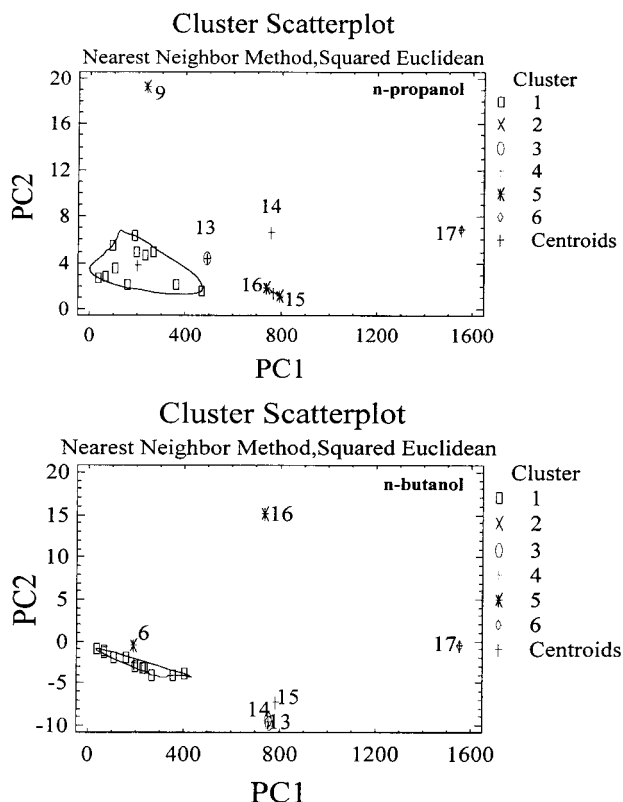
Cluster Scatterplot

Nearest Neighbor Method, Squared Euclidean

Cluster Scatterplot

Nearest Neighbor Method, Squared Euclidean

Figure 6. Cluster scatterplots of the *PC*2 *versus* the *PC*1 for MEKC systems containing *n*-propanol and *n*-butanol, respectively.

The cluster analysis indicated that compounds 9, 13, 14, 15, 16, and 17 for *n*-propanol and compounds 6, 13, 14, 15, 16, and 17 for *n*-butanol are out of the homogeneous group that form the other compounds. Considering this, several observations can be made. Thus, $fK1$ values, that are representative of solute-micelle association constants, are not statistically different for a 95% of significance level to *n*-propanol and *n*-butanol. This observation is in agreement with results shown in [24] except for the buffer containing 5% *n*-butanol. In this case, if the phase ratio is considered f(butanol)$>f$(propanol) results would show that association constants for *n*-butanol are lesser than for *n*-propanol. In addition, $fK1$ values increase with the solute number according to their hydrophobicity as they have been ordered. On the other hand, the K_2 and K_3 values, that represent the equilibrium constants of solute-alcohol complex and of solute-micelle-alcohol complex, respectively, have little values in all cases considered. Finally, K_4, that represents the equilibrium constants between the solute-alcohol complex and other alcohol molecules is negligible for all solutes considered (except for 16) for *n*-propanol while it has bigger values for *n*-butanol. These results indicate that the enhancement of solubility for *n*-butanol is bigger than for

n-propanol, probably because this alcohol can react with the solute-micelle complex more easily than *n*-propanol, which is in agreement with the structure of *n*-butanol (which has a hydrocarbon chain longer than *n*-propanol and so it can form a more stable complex with the micelle). The information obtained by means of the physicochemical model used can help us to classify the solutes with respect to the retention behavior, so we will be able to apply different equations to model it.

4 Concluding remarks

From the different empirical models used to predict the retention of the group of *N*-phenylpyrazole derivatives studied in this work as a function of the surfactant and alcohol concentration in a MEKC system with SDS as surfactant and *n*-propanol or *n*-butanol as organic modifier, Eq. (38) leads to the lowest mean relative prediction errors. In addition, for these compounds, the empirical equations which enable the best prediction of the migration times of the electroosmotic flow and micelles are Eqs. (4) and (23) for *n*-propanol and Eqs. (4) and (7) for *n*-butanol. A physicochemical model has also been employed to predict the retention of the group of *N*-phenylpyrazole derivatives observing that the best results correspond to the unsimplified model (Eq. 39). This model has also enabled to obtain interesting physicochemical information since physicochemical constants values depend on the solute and alcohol nature.

A comparison of the resolution optimization through the use of the equations selected for physicochemical and empirical models has been made for the solutes studied (*N*-phenylpyrazole derivatives) under the conditions used in this work (SDS as surfactant and *n*-propanol or *n*-butanol as organic modifier), obtaining as the optimum separation buffer compositions: 0.02 M SDS with 0.01 or 0.03 volume fractions of *n*-propanol or 0.03 volume fraction of *n*-butanol as well as 0.05 M SDS with a 0.03 volume fraction of *n*-butanol. The experimental electropherograms obtained under these optimal conditions enabled to observe that (i) the use of the physicochemical equation selected leads to the best optimal conditions and (ii) the use of *n*-propanol in the separation buffer enabled a better separation of the first-migration compounds while *n*-butanol enabled a better separation of the last-migrating compounds as it has been reported previously [35].

5 References

[1] Corstjens, H. H., Billiet, H. A., Frank, J., Luyben, K. C. A. M., *J. Chromatogr. A* 1995, *715*, 1–11.

[2] Siouffi, A. M., Phan-Tan-Lun, R., *J. Chromatogr. A* 2000, *892*, 75–106.

[3] Yeo, S. K., Ong, C. P., Li, S. F. Y., *Anal. Chem.* 1991, *63*, 2222–2225.

[4] Ng, C. L., Ong, C. P., Lee, H. K., Li, S. F. Y., *Chromatographia* 1992, *34*, 166–172.

[5] Yao, Y. J., Lee, H. K., Li, S. F. Y., *J. Chromatogr.* 1993, *637*, 195–203.

[6] Wu, J., Wong, M. K., Li, S. F. Y., Ong, C. N., *J. Chromatogr. A* 1995, *709*, 351–359.

[7] Sun, S. W., Chen, L. Y., *J. Chromatogr. A* 1997, *766*, 215–224.

[8] Sun, S. W., Wu, A. C., *J. Chromatogr. A* 1998, *814*, 223–231.

[9] Sun, S. W., Chen, L. Y., *J. Liq. Chromatogr. Rel. Technol.* 1998, *21*, 1613–1627.

[10] Bütehorn, U., Pyell, U., *J. Chromatogr. A* 1997, *792*, 157–163.

[11] Khaledi, M. G., Smith, S. C., Strasters, J. K., *Anal. Chem.* 1991, *63*, 1820–1830.

[12] Smith, S. C., Khaledi, M. G., *J. Chromatogr.* 1993, *632*, 177–184.

[13] Quang, C., Strasters, J. K., Khaledi, M. G., *Anal. Chem.* 1994, *66*, 1646–1653.

[14] Jiménez, O., Benito, I., Marina, M. L., *6th SAS, Symposium on Analytical Sciences*, Deauville Conference, Valencia, Spain 1998, p. 40.

[15] Pyell, U., Bütehorn, U., *J. Chromatogr. A* 1995, *716*, 81–95.

[16] Bütehorn, U., Pyell, U., *J. Chromatogr. A* 1997, *772*, 27–38.

[17] Jiménez, O., Benito, I., García, M. A., Marina, M. L., *8th Symposium on Handling of Environmental and Biological Samples in Chromatography*, Almería, Spain 1997, p. 245.

[18] Van Zomeren, P. V., Hilhorst, M. J., Coenegracht, P. M. J., de Jong, G. J., *J. Chromatogr. A* 2000, *867*, 247–259.

[19] Havel, J., Breadmore, M., Macka, M., Haddad, P. R., *J. Chromatogr. A* 1999, *850*, 345–353.

[20] Vindevogel, J., Sandra, P., *Anal. Chem.* 1991, *63*, 1530–1536.

[21] Rogan, M. M., Altria, K. D., Goodall, D. M., *Chromatographia* 1994, *38*, 723–730.

[22] He, Y., Lee, H. K., *J. Chromatogr. A* 1998, *793*, 331–340.

[23] Jumppanen, J. H., Wiedmer, S. K., Sirén, H., Riekkola, M. L., Haario, H., *Electrophoresis* 1994, *15*, 1267–1272.

[24] Altria, K. D., Howells, J. S., *J. Chromatogr. A* 1995, *696*, 341–348.

[25] Guo, W., Zhu, P., Brodowsky, H., *Talanta* 1997, *44*, 1995–2001.

[26] Zupan, J., Gasteiger, J., *Neural Networks for Chemists. An Introduction*, VCH, Weinheim 1993.

[27] Jalali-Heravi, M., Garkani-Nejad, Z., *J. Chromatogr. A* 2002, *971*, 207–215.

[28] Farková, M., Peña-Méndez, E. M., Havel J., *J. Chromatogr. A* 1999, *848*, 365–374.

[29] Srecnik, G., Debeljak, Z., Cerjan-Stefanovic, S., Novic, M., Bolanca, T., *J. Chromatogr. A* 2002, *973*, 47–59.

[30] Agatonovic-Kustrin, S., Zecevic, M., Zivanovic, Lj., Tucker, I. G., *Anal. Chim. Acta* 1998, *364*, 265–273.

[31] Loukas, Y. L., *J. Chromatogr. A* 2000, *904*, 119–129.

[32] Zhao, R. H., Yue, B. F., Ni, J. Y., Zhou, H. F., Zhang, Y. K., *Chem. Intell. Lab. Systems* 1999, *45*, 163–170.

[33] Madden, J. E., Nebojsa, A., Haddad, P. R., Havel, J., *J. Chromatogr. A* 2001, *910*, 173–179.

[34] Jiménez, O., Benito, I., Marina, M. L., *Anal. Chim. Acta* 1997, *353*, 367–379.

[35] García-Ruiz, C., García, M. A., Marina, M. L., *Electrophoresis* 2000, *21*, 2424–2431.

[36] Sigma Plot for Windows, Version 1.02, Jandel Corporation, 1986–1994.

[37] Microsoft Excel 97, Microsoft Corporation, 1985–1997.

[38] Statgraphics Plus for Windows, Version 2.1, Microsoft Corporation, 1994–1996.

[39] Cheng, N., Terabe, S., *Electrophoresis* 1995, *16*, 2100–2103.

[40] Van Hove E., Szücs R., Sandra P., *J. High Resolut. Chromatogr.* 1996, *19*, 674–678.

[41] Gago, F., Pastor, M., Pérez-Butragueño, J., López, R., Alvarez-Builla, J., Elguero, J., *Quant. Struct.-Act. Relat.* 1994, *13*, 165–171.

[42] Benito, I., García, M. A., Díez-Masa, J. C., Marina, M. L., *J. Microcol. Sep.* 1997, *9*, 87–95.

[43] Bütehorn, U., Pyell, U., *J. Chromatogr. A* 1997, *792*, 157–163.

[44] Drouen, A. C. J. H., Billiet, H. A. H., Schoenmakers, P. J., De Galan, L., *Chromatographia* 1982, *16*, 48–52.

[45] Farrau, A., *Anal. Chim. Acta* 1995, *317*, 181–188.

[46] Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y., Kaufman, L., in: Vandeginste, B. G. M., Michotte, Y. (Eds.), *Data Handling in Science and Technology 2*, Elsevier, Amsterdam 1988.