

Valérie Harang<sup>1,2</sup>  
Sven P. Jacobsson<sup>1</sup>  
Douglas Westerlund<sup>2</sup>

<sup>1</sup>Analytical Development,  
Pharmaceutical and  
Analytical R&D, AstraZeneca,  
Södertälje, Sweden

<sup>2</sup>Department of Analytical  
Pharmaceutical Chemistry,  
Biomedical Centre,  
University of Uppsala,  
Uppsala, Sweden

## Microemulsion electrokinetic chromatography of drugs varying in charge and hydrophobicity Part II: Strategies for optimization of separation

The separation of anionic, cationic, and neutral drugs in microemulsion electrokinetic chromatography (MEEKC) was studied. The concentration of sodium dodecyl sulfate (SDS; surfactant) and 2-propanol (organic solvent) was varied in a three-level full factorial design. 29 different model substances were chosen with different hydrophobicities and charges (neutral, positive, and negative). The models were calculated by means of multiple linear regression (MLR). The compounds were divided into five different subgroups, and different strategies for optimization of the separation within each group were investigated. The optimization was done by maximizing the selectivity using response surface plots in MODDE, by calculation of different chromatographic functions, and by using the software DryLab™. For all the different groups, MODDE, almost all chromatographic functions and DryLab™ gave approximately the same settings of the factors for optimum separation. Attempts were made to fit descriptors of the compounds to the retention data from the three-level full factorial design by means of partial least squares projection to latent structures (PLS). Between 86 and 89% of all predictions of migration times were acceptable (80–120% of the observed value).

**Keywords:** Chromatographic functions / Microemulsion electrokinetic chromatography / Molecular modelling / Statistical experimental design / Three-level full factorial design

DOI 10.1002/elps.200305812

### 1 Introduction

Microemulsion electrokinetic chromatography (MEEKC) is a technique similar to micellar electrokinetic chromatography (MEKC), with the main difference that the microemulsion has a core of tiny droplets of oil inside the micelles. By using a surfactant and a cosurfactant, the oil

droplets are stabilized and the surface tension between the oil and the water phase is reduced. A typical microemulsion consists of 0.8% w/w octane (oil), 3.3% w/w sodium dodecyl sulfate (surfactant), 6.6% w/w 1-butanol (cosurfactant) and 89.3% w/w 10 mM tetraborate buffer (pH 9.2) [1, 2]. Phosphate or borate buffers at a high pH are often used, which will generate a high electroosmotic flow (EOF) when applying the voltage. The EOF is towards the cathode (detector side) and the negatively charged oil droplets migrate in the opposite direction (towards the anode), although the EOF is strong enough to sweep the oil droplets to the cathode. Neutral compounds have a retention time between  $t_0$  (time for an unretained substance, EOF) and  $t_m$  (time for the microemulsion droplets to reach the detector). More hydrophobic compounds have the longest retention times since they are more heavily distributed to the oil droplets. Methanol is often used to measure  $t_0$ , and dodecyl benzene is used as an oil droplet marker ( $t_m$ ). Ionized solutes can also be separated by MEEKC. Electrostatic interactions (repulsion for negative ions and attraction for positive ions) will, in addition to other interactions, play a role. Since ionized solutes have their own electrophoretic mobility, the migration time observed will be the result of both partitioning between the oil droplets and the water phase and their electrophoretic behavior [3].

**Correspondence:** Valérie Harang, Product Analysis, Pharmaceutical and Analytical R&D, AstraZeneca, SE-151 85 Södertälje, Sweden

**E-mail:** valerie.harang@astrazeneca.com

**Fax:** +46-8-55328886

**Abbreviations:** **A36**, AR-P016336; **A37**, AR-P016337; **A51**, AR-P017151; **ATR**, arcs tangens resolution; **BUP**, bupivacaine; **CEF**, chromatographic exponential function; **COF**, chromatographic optimization function; **CRF**, chromatographic response function; **CRS**, chromatographic resolution statistic; **DIS**, disopyramide; **DOD**, dodecyl benzene; **EPH**, ephedrine; **ESO**, estrone; **EST**, estradiol; **F97**, FLA797; **F08**, FLA708; **F40**, FLA740; **GUA**, guaifenesin; **IPA**, 2-propanol; **KET**, ketoprofen; **LID**, lidocaine; **log P**, octanol-water partition coefficient; **MEEKC**, microemulsion electrokinetic chromatography; **MEO**, metoprolol; **MEP**, mepivacaine; **MLR**, multiple linear regression; **NAP**, naproxen; **NOR**, norethisterone; **pK<sub>a</sub>**, acid dissociation constant; **PLS**, partial least squares projection to latent structures; **PRI**, prilocaine; **PRO**, propranolol; **r**, relative resolution product; **REM**, remoxipride; **Rp**, resolution product; **SAL**, salicylic acid; **SOB**, sodium benzoate; **TER**, terbutaline; **TRI**, trimethoprim

A large number of parameters in MEEKC can be manipulated during method development. The type and concentration of the oil, buffer, surfactant, cosurfactant, organic solvent, counter-ion and the pH will affect the separation performance. Furthermore, changing instrument parameters such as the temperature and the voltage can have an effect on the separation. In a previous study [4], a number of parameters (concentration of SDS, Brij 35, 1-butanol, 2-propanol, buffer, and temperature) were screened for their contribution in the separation performance of some analytes with differing hydrophobicity and charge, by means of a fractional factorial design. It was concluded that SDS and 2-propanol had the largest effects on the migration and on selectivity changes. Statistical experimental designs have been used for the development or optimization of MEKC methods [5–8]. Klampfl *et al.* [9] optimized the separation of nine peaks (UV filters in suntan lotions) with a MEEKC method by means of two factors (the ratio SDS/Brij 35 and 2-propanol), using artificial neural networks.

The scope of our work was to study the two factors (SDS, 2-propanol) more thoroughly in a three-level full factorial design using 29 different compounds. The modelling was done with multiple linear regression (MLR). Furthermore, different strategies for optimization of the separation were investigated. The compounds were divided into five subgroups and each group was optimized separately by maximizing the selectivity of the critical peak pair using response surface plots in MODDE, by calculating the chromatographic response function (CRF) [10–13], chromatographic resolution statistic (CRS) [10, 14], chromatographic exponential function (CEF) [10, 15–16], chromatographic optimization function (COF) [10, 17], arcs tangens resolution (ATR) [18], resolution product (Rp) [19] and relative resolution product (*r*) [20], and by using the software DryLab™ [21–27]. CRF, CEF, COF, ATR, and *r* have all been used for optimization of separation in electrodriven techniques. To our knowledge, no articles have been published where chromatographic functions or DryLab™ have been used for optimization of separation in MEEKC. Furthermore, descriptors from in-house software, SELMA [28–32], were calculated for all the compounds investigated and fitted by means of partial least squares projections to latent structures (PLS) to the retention data from the three-level full factorial design [33–38]. The data were divided into a training set and a test set, and the training set was used for generating a model, which subsequently was used for the prediction of the retention data of the compounds in the test set for the different microemulsions used in the three-level full factorial design.

## 2 Materials and methods

### 2.1 Materials

Brij 35 (polyoxyethylene (23) lauryl ether), 2-propanol, *n*-octane, dodecyl benzene, and 1-butanol were purchased from Fluka (Buchs, Switzerland). Boric acid and SDS were obtained from Sigma (St. Louis, MO, USA). Sodium hydroxide, methanol, and acetonitrile came from Merck (Darmstadt, Germany). The water used was of Millipore quality (Watford, Herts, UK). Salicylic acid was purchased from Fluka (Buchs, Switzerland). Disopyramide and propranolol were bought from Sigma. Trimethoprim, ephedrine, sodium benzoate, naproxen, terbutaline, guaifenesin, metoprolol, ketoprofen, lidocaine, mepivacaine, bupivacaine, prilocaine, AR-P016336, AR-P016337, AR-P017151, norethisterone acetate, estrone, estradiol, remoxipride, FLA797, FLA708, and FLA740 were all obtained from AstraZeneca. Four confidential substances from AstraZeneca (Södertälje, Sweden) other than those mentioned above were also included. These substances are designated M, P, R, and G. All chemicals were of analytical grade.

#### 2.1.1 Preparation of microemulsions

Borate buffer was prepared from boric acid, titrated to pH 9.2 with sodium hydroxide. The microemulsion was prepared by weighing together in a flask SDS, octane, 1-butanol, Brij 35, 2-propanol, and borate buffer. More details of the different microemulsions used in the experiments are given in tables or figures. The solutions were placed in an ultrasonic bath for 30 min to obtain a clear solution, and filtered through a 0.45 µm membrane filter (GHP, Bulk Acrodisc) before use.

#### 2.1.2 Sample preparation

A stock solution of each compound was prepared by dissolving it in methanol. Stock solutions containing several analytes were also made. Test solutions were then prepared by diluting the stock solutions with the actual filtered microemulsion to give a concentration in the range between 20–1000 µg/mL.

### 2.2 Apparatus, computational and analytical methods

Experiments were performed on a Hewlett Packard 3D Capillary Electrophoresis system (Agilent Technologies, Waldbronn, Germany) using Chemstation (Version A.06.01) for system control, data collection, and data analysis. UV detection was carried out at 210 nm and 235 nm with a

bandwidth of 4 nm. The sample solutions were hydrodynamically injected at the anode using a pressure of 7 mbar during 5 s (1 nL). The electroosmotic flow (EOF) was measured by injecting methanol, and dodecyl benzene was used as the oil droplet marker. The pH was measured using a Metrohm 632 pH meter (Calomel reference combined pH meter electrode, Radiometer Analytical SA, France). The chemometric design and evaluation of the models were carried out using the software package MODDE 6.0 (Umetrics, Umeå, Sweden). The program SIMCA-P+ 10.0.2. (Umetrics, Umeå, Sweden) was used for modelling of the molecular descriptors and retention data with PLS. ACDLabs [39] was used for prediction of log *P* and p*K*<sub>a</sub> of the analytes, and SELMA [28] was used for calculation of the molecular descriptors of each compound. Separation was performed on 33 cm × 50 μm internal diameter fused-silica capillaries (Skandinaviska GeneTec, Sweden), detection window at 24.5 cm. Detection windows were prepared using an electrical burner device from Capital HPLC (Broxburn, UK). New capillaries were flushed with 1 M sodium hydroxide for 30 min and with water for 30 min. The capillary was rinsed between injections with 0.1 M sodium hydroxide for 1 min followed by the microemulsion solution for 1 min. The applied voltage was set to 10 kV.

## 2.3 Calculations

### 2.3.1 Equations

The resolution was calculated using Eq. (1), where *t*<sub>R</sub> is the migration time and *w*<sub>0.5</sub> is the peak width at 50% of the peak height:

$$R_s = \frac{1.18(t_{R,2} - t_{R,1})}{(w_{0.5,1} + w_{0.5,2})} \quad (1)$$

The migration factor defined according to Eq. (2) was used for calculation of the selectivity (3):

$$k_M = \left( \frac{t_R - t_0}{t_0} \right) \quad (2)$$

The selectivity was calculated according to Eq. (3):

$$\alpha = \frac{k_{M,2}}{k_{M,1}} \quad (3)$$

### 2.3.2 Modelling

The data from the statistical experimental design can be fitted by means of MLR and the responses can be described by a polynomial function [40, 41]:

$$y = \text{constant} + b_1k_1 + b_2k_2 + b_{12}k_1k_2 + b_{11}k_1^2 + b_{22}k_2^2 + \text{error} \quad (4)$$

where *k*<sub>1</sub> and *k*<sub>2</sub> are the factors included in the model and *b*<sub>1..n</sub> are the regression coefficients that are estimated by the MLR model. The interaction between the factors is described by *b*<sub>12</sub>*k*<sub>1</sub>*k*<sub>2</sub>, and curvature by *b*<sub>11</sub>*k*<sub>1</sub><sup>2</sup> + *b*<sub>22</sub>*k*<sub>2</sub><sup>2</sup>. For each new MLR model, all linear, interaction, and quadratic terms were included. The fraction of variation of the response that can be explained by the model, *R*<sup>2</sup> = (total sum of squares – sum of squares for residuals)/total sum of squares, and the fraction of variation of the response that can be predicted by the model, *Q*<sup>2</sup> = 1 – (prediction residual sum of squares/total sum of squares), was then examined. For a good model, *R*<sup>2</sup> and *Q*<sup>2</sup> should be as close to 1 as possible. The model estimated the coefficients (*b*<sub>*n*</sub>), which represent half the effect of a factor. Some of the coefficients (interaction or quadratic terms) that did not have a significant effect were then removed from the model and a new model was constructed. If *R*<sup>2</sup> and *Q*<sup>2</sup> decreased following the removal of an insignificant coefficient from the model, the coefficient was added to the model again. To identify outliers, a normal probability plot of the residuals was examined. The observed response vs. predicted plot was also examined to evaluate the predictability of each model and the observed response vs. run order plot was examined to make sure that there was no systematic error. A logarithmic transformation of the responses improved some of the models, and was therefore used. From the in-house SELMA software [28], 93 different molecular descriptors could be calculated for each compound. A PLS model [42] was fitted between the molecular descriptors and the migration times of the different compounds from the experimental design. See Section 3.3 for more details about refining the model and the evaluation.

## 3 Results and discussion

### 3.1 Statistical experimental design

In a previous investigation [4] six factors – SDS (% w/w), Brij 35 (% w/w), 1-butanol (% w/w), 2-propanol (IPA; % w/w), concentration of buffer (mM) and temperature (°C) – were studied with a fractional factorial design (2<sup>6-2</sup> + 6 center points). Eight different analytes were used: disopyramide (DIS), estradiol (EST), lidocaine (LID), naproxen (NAP), norethisterone (NOR), propranolol (PRO), salicylic acid (SAL), and trimethoprim (TRI). It was concluded that the factors SDS and IPA had the largest effects on the migration times and on the selectivity.

SDS and IPA were therefore chosen as factors in a three-level full factorial design. The concentration of SDS was set to 2, 3.5, and 5% w/w, and the concentration of IPA was varied between 2 and 10% w/w. In all experiments

the following factors were constant: Brij 35: 1% w/w, 1-butanol: 7% w/w, octane: 0.8% w/w, concentration of borate buffer: 20 mM with pH 9.2, temperature: 40°C, and voltage: 10 kV. The experimental domain and the design are shown in Table 1.

A larger number of analytes were investigated in the factorial design compared to the previous study. In Fig. 1 the molecular structures of 25 compounds are shown. In addition, four confidential substances from AstraZeneca were also included. These substances are designated M, P, R, and G. Methanol was used as the electroosmotic flow (EOF) marker, and dodecyl benzene (DOD) was the microemulsion droplet marker.

Table 2 lists predicted log *P* and p*K*<sub>a</sub> from ACDLabs software [39] and charge (or partial charge) of molecule at pH 9.2 for all compounds. The hydrophobicities (log *P*) of the analytes range from -2.31 (M) to 4.13 (EST). At pH 9.2 PRO, DIS, ephedrine (EPH), M, remoxipride (REM),

**Table 1.** Experimental design according to a three-level full factorial design

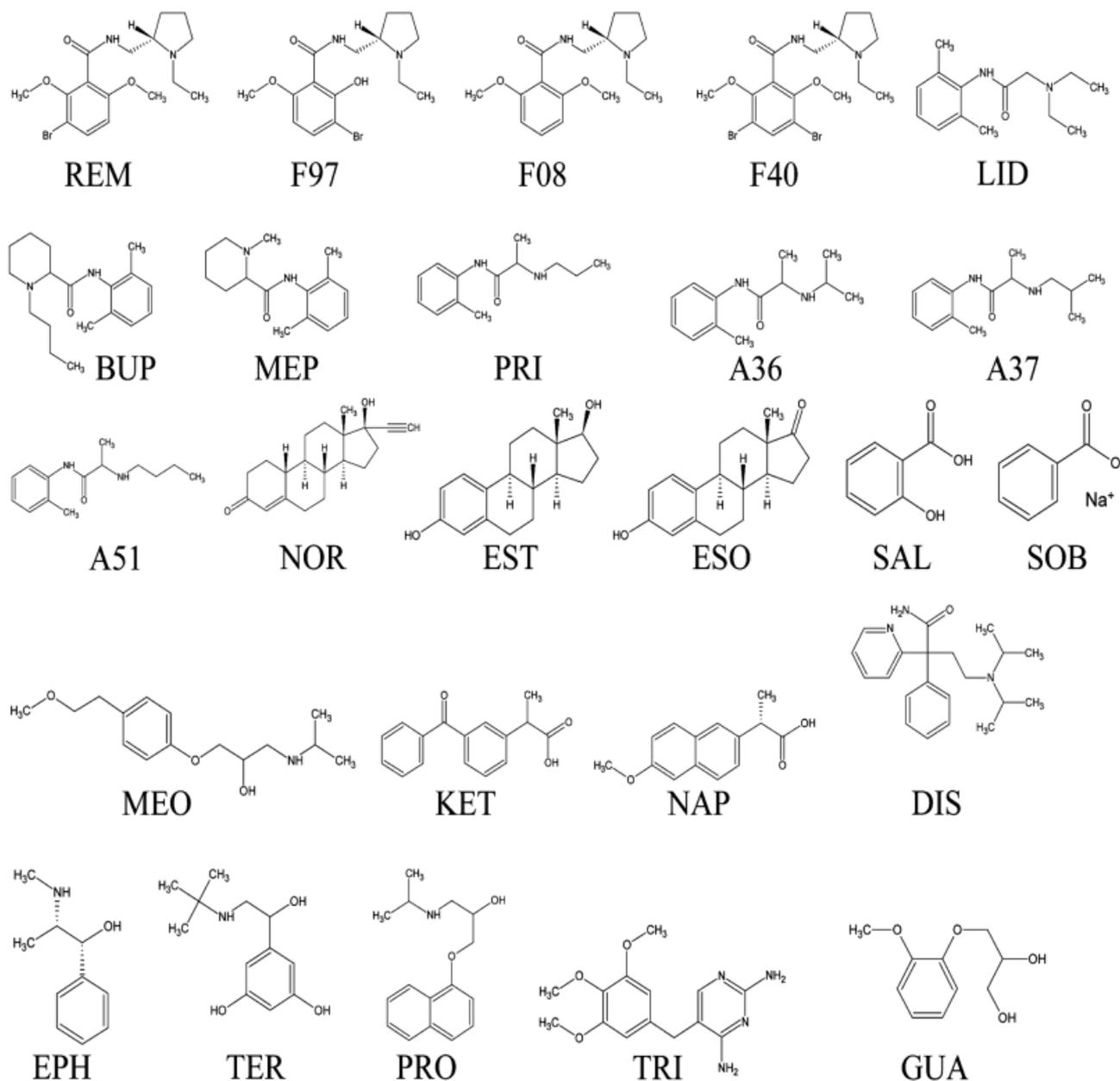
Exp. name	Run order	SDS (% w/w)	IPA (% w/w)	Current (μA)
N1	1	2.0	2.0	30
N2	9	3.5	2.0	50
N3	10	5.0	2.0	72
N4	6	2.0	6.0	27
N5	4	3.5	6.0	45
N6	5	5.0	6.0	65
N7	3	2.0	10.0	24
N8	7	3.5	10.0	42
N9	2	5.0	10.0	60
N10	8	3.5	6.0	48
N11	11	3.5	6.0	46

Constant factors: Brij 35 1% w/w, 1-butanol 7% w/w, buffer concentration 20 mM, temperature 40°C, voltage 10 kV. The current from each experiment is shown in the last column.

**Table 2.** Names of compounds, abbreviations, predicted log *P* and p*K*<sub>a</sub> from the software ACDLabs and charge of compound at pH 9.2

Abbreviation	Name of compound	Predicted log <i>P</i> [35]	Predicted p <i>K</i> <sub>a</sub> [35]	p <i>K</i> <sub>a</sub> from literature [44, 45]	Charge <sup>h)</sup> at pH 9.2
REM <sup>a)</sup>	Remoxipride	2.20	8.97		0.4+
F97 <sup>a)</sup>	FLA797	3.31	6.31, 8.97		0.6-
F08 <sup>a)</sup>	FLA708	1.14	8.97		0.4+
F40 <sup>a)</sup>	FLA740	3.21	8.97		0.4+
LID <sup>b)</sup>	Lidocaine	2.36	8.53	7.9 <sup>f) g)</sup>	0.2+
BUP <sup>b)</sup>	Bupivacaine	3.64	8.17	8.1 <sup>g)</sup>	0
MEP <sup>b)</sup>	Mepivacaine	2.04	8.09	7.7 <sup>g)</sup>	0
PRI <sup>b)</sup>	Prilocaine	1.74	7.95	7.9 <sup>f) g)</sup>	0
A36 <sup>b)</sup>	AR-P016336AA	1.55	7.95		0
A37 <sup>b)</sup>	AR-P016337AA	2.09	7.95		0
A51 <sup>b)</sup>	AR-P017151AA	2.27	8.02		0
P <sup>c)</sup>	P	3.16	5.56, 6.35		1-
R <sup>c)</sup>	R	2.56	6.34, 6.48		1-
G <sup>c)</sup>	G	3.93	5.36		1-
M <sup>c)</sup>	M	-2.31	2.95		1+
NOR <sup>d)</sup>	Norethisterone	3.38	13.10		0
EST <sup>d)</sup>	Estradiol	4.13	10.27		0
ESO <sup>d)</sup>	Estrone	3.69	10.25		0
TER <sup>e)</sup>	Terbutaline	0.48	9.12, 9.33	8.7 <sup>f) g)</sup> , 10.0 <sup>f) g)</sup>	0
GUA <sup>e)</sup>	Guaifenesin	0.57	13.45		0
SOB <sup>e)</sup>	Sodium benzoate	1.89	4.20	4.2 <sup>f) g)</sup>	1-
SAL <sup>e)</sup>	Salicylic acid	2.06	3.01	3.0 <sup>f) g)</sup>	1-
KET <sup>e)</sup>	Ketoprofen	2.81	4.23		1-
NAP <sup>e)</sup>	Naproxen	3.00	4.40	4.2 <sup>f) g)</sup>	1-
TRI <sup>e)</sup>	Trimethoprim	0.79	7.34	7.2 <sup>g)</sup>	0
MEO <sup>e)</sup>	Metoprolol	1.39	8.70	9.7 <sup>f) g)</sup>	0.2+
PRO <sup>e)</sup>	Propranolol	3.10	9.15	9.5 <sup>f) g)</sup>	0.5+
DIS <sup>e)</sup>	Disopyramide	2.86	3.72, 10.10	8.4 <sup>g)</sup>	0.9+
EPH <sup>e)</sup>	Ephedrine	1.05	9.38	9.6 <sup>f)</sup>	0.6+

a) Group 1; b) Group 2; c) Group 3; d) Group 4; e) Group 5; f) [44]; g) [45]; h) Charge or partial charge of the compound in aqueous solution



**Figure 1.** Molecular structures of the compounds investigated. Structures not shown for R, M, G, and P. See Table 2 for explanation of abbreviations.

FLA708 (F08), FLA740 (F40), LID, and metoprolol (MEO) have a positive net charge, and NAP, ketoprofen (KET), SAL, sodium benzoate (SOB), FLA797 (F97), P, R and G are negatively charged. The charges of the analytes can change when an organic modifier is added to the background electrolyte (BGE) due to a shift in the dissociation constant. Furthermore, the apparent pH of the buffer can also change when adding an organic modifier. The remaining compounds are neutral. The analytes were classified into five

different groups based on their structures. Group 1 consisted of REM, F40, F97, and F08. LID, bupivacaine (BUP), mepivacaine (MEP), prilocaine (PRI), AR-P016336 (A36), AR-P016337 (A37), and AR-P017151 (A51) were part of group 2. Group 3 contained P, R, G and M. Group 4 consisted of NOR, EST, and estrone (ESO). The remaining analytes varying largely in hydrophobicity and charge were placed in group 5 (terbutaline (TER), guaifenesin (GUA), SOB, SAL, NAP, KET, TRI, MEO, PRO, DIS, and EPH).

The migration times for the analytes from each experiment in the design are listed in Table 3. For experiment N4 (low level of SDS, medium level of IPA) and N7 (low level of SDS, high level of IPA), almost all the peaks of the analytes were split into two. This was not observed at medium (3.5% w/w) or high (5% w/w) levels of SDS, or if both factors were at a low level (2% SDS, 2% IPA). SDS is added to the

microemulsion to introduce charge, to stabilize the oil droplets, and to reduce the surface tension between the oil and the water phase. IPA is added to the microemulsion to increase the solubility of the analytes in the aqueous phase. A fraction of IPA is distributed to the oil droplets, which will lead to a decrease in the charge density of the oil droplets [43]. A low amount of SDS with a combination of

**Table 3.** Migration times (min) from the experimental design for all analytes

Exp.	EOF	DOD	REM <sup>a)</sup>	F97 <sup>a)</sup>	F08 <sup>a)</sup>	F40 <sup>a)</sup>	LID <sup>b)</sup>	BUP <sup>b)</sup>	MEP <sup>b)</sup>	PRI <sup>b)</sup>	A36 <sup>b)</sup>
N1	3.20	6.51	5.25	5.10	4.18	5.91	5.65	6.31	5.05	5.25	5.01
N2	3.62	12.22	9.09	8.65	6.20	11.07	9.89	11.46	7.98	8.76	8.19
N3	4.10	22.06	14.20	13.82	8.46	18.36	16.15	19.44	11.95	13.53	12.51
N4	3.97	7.87	6.10	6.25	4.84	7.00	6.70	7.32	5.88	6.21	5.96
N5	4.71	13.81	9.67	9.88	6.83	11.87	11.00	12.50	9.16	9.84	9.34
N6	5.27	30.77	16.11	16.54	9.30	22.16	19.74	25.18	14.55	16.46	15.47
N7	4.93	9.17	7.02	7.42	5.79	8.00	7.72	8.41	6.49	7.24	7.03
N8	6.09	18.24	10.42	11.41	8.05	12.84	12.88	14.31	10.62	11.74	10.51
N9	6.94	41.91	17.76	20.01	10.99	25.76	23.41	29.44	17.39	19.44	18.20
N10	4.68	13.56	9.11	9.38	6.43	11.01	10.56	11.93	8.55	9.65	8.57
N11	4.82	15.31	10.05	10.22	7.00	12.39	11.77	13.22	9.37	10.21	9.70
Exp.	A37 <sup>b)</sup>	A51 <sup>b)</sup>	P <sup>c)</sup>	R <sup>c)</sup>	G <sup>c)</sup>	M <sup>c)</sup>	NOR <sup>d)</sup>	EST <sup>d)</sup>	ESO <sup>d)</sup>	TER <sup>e)</sup>	GUA <sup>e)</sup>
N1	5.65	5.67	5.97	5.84	5.83	2.49	6.30	5.80	6.18	3.67	4.12
N2	10.12	10.15	10.77	10.28	10.38	3.06	10.84	11.07	11.11	4.65	6.68
N3	16.42	16.58	18.35	17.11	17.71	3.56	20.49	18.79	18.80	5.81	7.68
N4	6.78	6.73	7.44	7.38	7.21	2.79	7.41	7.23	7.17	4.43	4.94
N5	11.23	11.29	12.43	12.11	12.13	3.38	12.70	12.19	11.97	5.70	6.85
N6	20.92	20.90	24.19	23.03	22.97	3.84	26.60	24.28	23.12	6.99	9.26
N7	7.81	7.79	8.83	8.86	8.68	3.17	8.35	8.17	8.04	5.46	5.96
N8	12.62	12.44	16.28	15.98	14.81	3.60	14.60	13.97	14.44	7.25	8.39
N9	24.19	24.62	32.94	30.86	29.69	4.28	29.63	27.17	27.11	9.12	11.47
N10	10.66	10.50	12.41	12.17	11.17	3.20	12.10	11.59	11.94	5.69	6.71
N11	11.74	11.71	13.51	12.88	12.69	3.41	13.95	13.29	12.93	5.72	7.00
Exp.	SOB <sup>e)</sup>	SAL <sup>e)</sup>	KET <sup>e)</sup>	NAP <sup>e)</sup>	TRI <sup>e)</sup>	MEO <sup>e)</sup>	PRO <sup>e)</sup>	DIS <sup>e)</sup>	EPH <sup>e)</sup>		
N1	7.68	7.83	5.72	5.69	4.10	5.11	6.14	5.12	4.30		
N2	10.60	13.27	8.46	8.48	5.71	8.33	11.40	8.49	6.52		
N3	15.45	13.73	12.82	13.00	7.85	12.45	19.19	12.61	9.17		
N4	10.99	11.52	7.25	7.57	4.88	5.70	7.13	5.66	4.72		
N5	18.58	21.30	10.37	11.83	6.72	8.61	11.93	8.57	6.78		
N6	34.77	50.05	19.39	19.77	9.02	13.36	23.42	13.32	9.48		
N7	16.32	17.15	9.28	9.68	5.81	6.47	7.98	6.46	5.41		
N8	41.73	46.77	14.65	16.28	7.90	10.03	13.53	9.56	7.26		
N9	n.a.	40.98	29.18	31.52	11.02	14.83	25.15	14.47	10.50		
N10	18.55	19.81	10.72	11.32	6.47	8.63	11.41	8.25	6.56		
N11	20.42	23.63	12.00	12.67	7.03	8.95	13.03	9.05	6.94		

See Table 1 for the experimental design and Table 2 for explanation of abbreviations. a) Group 1; b) Group 2; c) Group 3; d) Group 4; e) Group 5; n.a., value is missing for SOB, since the analyte did not reach the detection window during the run (80 min). Most of the peaks from experiment N4 (2% w/w SDS, 6% w/w IPA) and N7 (2% w/w SDS, 10% w/w IPA) were split into two. The migration time of the first peak is then shown in the table above.

medium to high amounts of IPA might lead to a micro-emulsion that is not stable enough, which could be a possible explanation of the observed double peaks of the analytes. If a peak was split into two, the migration time of the first peak was used.

One MLR model (with linear, quadratic, and interaction terms) was calculated for all the compounds simultaneously using the logarithm of the migration time as response. Results from the oil droplet marker (dodecyl benzene) and the EOF marker (methanol) were also included in the calculation. Table 4 shows the coefficients (half the effect) of all terms. Significant coefficients (95% of confidence) are marked in bold. In addition,  $R^2$  and  $Q^2$  are also listed in the same table.  $Q^2$  for SAL was very low (0.025), but when a new model was calculated with only the main

terms included,  $Q^2$  increased to 0.622. All  $R^2$  and  $Q^2$  are reasonably high (0.983–0.999, 0.853–0.993, respectively), except for SAL ( $R^2$ : 0.841,  $Q^2$ : 0.622). In part I of this work [4] it was observed that the migration time of SAL seemed to be sensitive to the condition of the capillary surface. New capillaries contain higher amounts of metal impurities. Formation of metal complex with the negatively charged SAL could be a possible explanation of the variation of the migration time, something that also impeded the quality of the model.

As can be seen from the coefficients (half the effect) in Table 4, increasing the concentration of SDS and IPA in the microemulsion will increase the migration times of all analytes, methanol, and dodecyl benzene. Increasing the concentration of SDS will increase the charge density of

**Table 4.** Scaled and centered coefficients ( $b_{xx}$ ) from MLR models of log (migration times) from the experimental design

Compound	$R^2, Q^2$	$b_{\text{SDS}}$	$b_{\text{IPA}}$	$b_{\text{SDS}*\text{SDS}}$	$b_{\text{IPA}*IPA}$	$b_{\text{SDS}*IPA}$
EOF	0.998, 0.983	<b>0.063</b>	<b>0.107</b>	-0.012	< -0.001	<b>0.010</b>
DOD	0.996, 0.980	<b>0.297</b>	<b>0.100</b>	0.025	0.007	<b>0.032</b>
REM <sup>a)</sup>	0.993, 0.957	<b>0.210</b>	<b>0.047</b>	0.009	0.001	-0.007
F97 <sup>a)</sup>	0.996, 0.974	<b>0.214</b>	<b>0.074</b>	0.011	0.001	-0.001
F08 <sup>a)</sup>	0.994, 0.983	<b>0.145</b>	<b>0.061</b>	-0.007	0.015	-0.007
F40 <sup>a)</sup>	0.992, 0.942	<b>0.250</b>	<b>0.057</b>	0.018	-0.001	0.004
LID <sup>b)</sup>	0.995, 0.981	<b>0.235</b>	<b>0.069</b>	0.008	<0.001	0.006
BUP <sup>b)</sup>	0.994, 0.963	<b>0.262</b>	<b>0.067</b>	0.023	-0.002	0.014
MEP <sup>b)</sup>	0.996, 0.987	<b>0.199</b>	<b>0.066</b>	0.002	<0.001	0.013
PRI <sup>b)</sup>	0.998, 0.989	<b>0.211</b>	<b>0.071</b>	0.001	0.002	0.004
A36 <sup>b)</sup>	0.992, 0.960	<b>0.204</b>	<b>0.070</b>	0.014	-0.001	0.004
A37 <sup>b)</sup>	0.994, 0.958	<b>0.241</b>	<b>0.067</b>	0.017	-0.006	0.007
A51 <sup>b)</sup>	0.993, 0.951	<b>0.243</b>	<b>0.066</b>	0.020	-0.004	0.008
P <sup>c)</sup>	0.997, 0.984	<b>0.262</b>	<b>0.101</b>	0.012	0.006	0.021
R <sup>c)</sup>	0.998, 0.987	<b>0.251</b>	<b>0.105</b>	0.011	0.004	<b>0.019</b>
G <sup>c)</sup>	0.994, 0.974	<b>0.253</b>	<b>0.092</b>	0.022	0.005	0.013
M <sup>c)</sup>	0.985, 0.934	<b>0.071</b>	<b>0.043</b>	-0.004	0.002	-0.006
NOR <sup>d)</sup>	0.995, 0.986	<b>0.270</b>	<b>0.069</b>	0.031	-0.016	0.009
EST <sup>d)</sup>	0.992, 0.960	<b>0.260</b>	<b>0.068</b>	0.015	-0.012	0.003
ESO <sup>d)</sup>	0.997, 0.987	<b>0.253</b>	<b>0.065</b>	0.009	0.002	0.011
TER <sup>e)</sup>	0.999, 0.992	<b>0.103</b>	<b>0.094</b>	-0.009	<b>0.009</b>	<b>0.006</b>
GUA <sup>e)</sup>	0.983, 0.853	<b>0.138</b>	<b>0.072</b>	-0.024	0.020	0.003
SOB <sup>e)</sup>	0.997, 0.974	<b>0.278</b>	<b>0.302</b>	0.014	<b>0.046</b>	<b>0.141</b>
SAL <sup>e)</sup>	0.841, 0.622	<b>0.210</b>	<b>0.227</b>	*	*	*
KET <sup>e)</sup>	0.993, 0.971	<b>0.213</b>	<b>0.134</b>	0.030	0.003	<b>0.037</b>
NAP <sup>e)</sup>	0.996, 0.986	<b>0.215</b>	<b>0.150</b>	0.015	-0.003	<b>0.038</b>
TRI <sup>e)</sup>	0.995, 0.984	<b>0.138</b>	<b>0.073</b>	-0.002	0.003	-0.001
MEO <sup>e)</sup>	0.998, 0.993	<b>0.186</b>	<b>0.043</b>	-0.008	0.012	-0.007
PRO <sup>e)</sup>	0.993, 0.965	<b>0.252</b>	<b>0.051</b>	0.013	-0.004	0.001
DIS <sup>e)</sup>	0.994, 0.976	<b>0.186</b>	<b>0.035</b>	-0.003	0.013	-0.010
EPH <sup>e)</sup>	0.995, 0.970	<b>0.153</b>	<b>0.034</b>	-0.002	0.010	-0.010

Significant coefficients (95% of confidence) are in bold. See Table 1 for the experimental design and Table 2 for explanation of abbreviations. a) Group 1; b) Group 2; c) Group 3; d) Group 4; e) Group 5; \* terms not included in the model

the oil droplets and since the oil droplets are negatively charged, they will move faster to the anode (against the EOF). Longer retention times will therefore be observed for neutral analytes which partition into the oil droplets. The migration times observed for charged compounds are a combination of electrostatic/hydrophobic interaction to the oil droplets and the analytes' own mobility in the aqueous phase. Additionally, an increase of the SDS concentration will increase the time of the EOF marker (decrease the EOF), which will increase the migration times of all analytes. An increase of the concentration of IPA will increase the viscosity of the BGE and decrease the EOF (increase the time of the EOF marker), leading to longer migration times of all analytes. Furthermore, increasing the concentration of IPA in the microemulsion will increase the solubility of hydrophobic analytes in the aqueous phase. A fraction of IPA can be distributed to the oil droplet, leading to an increase of size and decrease of charge density. A significant interaction between SDS and IPA was found for the negatively charged compounds R, SOB, KET, and NAP. One quadratic term (IPA\*IPA) was found significant for SOB and all terms were found significant for TER.

In summary, the effects on the migration times observed when changing the concentration of SDS and IPA are a combination of changes in the EOF (decreases with increasing SDS and IPA), changes in the size and charge density of the oil droplet (increases with increasing SDS, decreases with increasing IPA), and changes in partition of the analytes into the oil droplet. For charged analytes, changes in electrostatic interactions with the oil droplet will also occur when changing the concentration of SDS and IPA. In addition, charged analytes have their own mobility in the aqueous phase in the background electrolyte, which also contributes to the total migration time.

### 3.2 Optimization of separation

Three different optimization strategies were evaluated, (i) by optimizing the selectivity ( $\alpha$ ) of the critical pair of peaks, (ii) by calculating chromatographic functions (CRF, CRS, CEF, COF, ATR, Rp, r), and (iii) by using Drylab™ [25].

(i) The program MODDE was used for the setup of the experimental design and for evaluation of the MLR models for the migration times of the different compounds. In addition, the selectivity ( $\alpha$ ) calculated according to Eq. (3) for the pair of peaks that were closest to each other (within each group) was also used as a response in MODDE. Response surface plots were generated and the setting of the factors giving the highest selectivity was chosen.

(ii) The chromatographic functions CRF [10–13], CRS [10, 14], CEF [10, 15, 16], COF [10, 17], ATR [18], Rp [19], and r [20] were calculated according to Eq. (5)–(10) within each group and used as a response in MODDE. The separation within each group was optimized by generating response surface plots and choosing the setting of the factors that gave the highest (for CRF, COF, ATR, Rp, and r) or the lowest (for CRS and CEF) response.

$$\text{CRF}_1 = \sum_{i=1}^L R_i + L^{w_1} - w_2 |T_A - T_L| - w_3 (T_1 - T_0) \quad (5a)$$

where  $R_i$  is the resolution of the  $i^{\text{th}}$  peak,  $L$  is the number of peak pairs,  $T_A$  is the maximum acceptable time (40 or 50 min),  $T_L$  is the migration time of the last peak,  $T_1$  is the migration time of the first peak,  $T_0$  is the minimum retention time of the first peak (2 min), and  $w_1 - w_3$  are weighting factors selected by the operator (here 1 has been selected to give equal weights). The sum of all resolutions is calculated in the first term. Unresolved peaks have little influence on the function since a high resolution between other peaks contributes to a high value of CRF. A modified equation of CRF was therefore tested, where all resolution values exceeding 3 were not included in the sum (first term).

Chromatographic resolution functions according to Eqs. 5b [12] and 5c [13] were also evaluated:

$$\text{CRF}_2 = a \sum \ln \frac{R_{so}}{R_s} + b \sum \ln \frac{R_s}{R_{so}} + c \ln \frac{T_0}{T} \quad (5b)$$

$$\text{CRF}_3 = a \sum \ln \frac{R_{so}}{R_s} + b \sum \ln \frac{R_s}{R_{so}} \quad (5c)$$

where  $R_{so}$  is the optimum resolution (1.5),  $R_s$  is the resolution between two neighboring peaks,  $T$  is the total time, and  $T_0$  the optimum total time (40 or 50 min). The weighting factors,  $a$ ,  $b$ , and  $c$ , were selected according to [12, 13], where  $a$  (excess resolution factor) and  $c$  (time factor) were set to 5 and  $b$  (overlap degradation factor) to 50. Only peak pairs with  $R_s > 2$  were included in the sum of the first term. Furthermore, only  $R_s < 1.5$  was included in the sum of the second term. More weight was given to the separation (second term) by giving  $b$  the value of 50.

$$\text{CRS} = \left\{ \sum_{i=1}^{n-1} \left[ \frac{(R_i - R_{opt})^2}{R_i(R_i - R_{min})^2} \right] + \sum_{i=1}^{n-1} \frac{(R_i)^2}{aR^2} \right\} \frac{T_f}{n} \quad (6)$$

where  $R_i$  is the resolution of the  $i^{\text{th}}$  peak,  $R$  is the average resolution,  $R_{opt}$  is the desired optimum resolution (1.5),  $R_{min}$  is the minimum acceptable resolution (1.0),  $T_f$  is the

migration time of the final peak,  $a$  is the number of resolution elements, and  $n$  is the number of peaks. Very high values of CRS were obtained when the resolution between two peaks was close to the minimum acceptable resolution. For optimum resolution, the value of CRS must be minimized. A modified equation of CRS, where  $T_f/n$  was set equal to 1, was also used and compared to the original equation.

$$\text{CEF} = \left[ \left( \sum_{i=1}^{n-1} \left( 1 - e^{a(R_{\text{opt}} - R_i)} \right)^2 \right) + 1 \right] \left[ 1 + \frac{t_f}{t_{\text{max}}} \right] \quad (7)$$

where  $R_i$  is the resolution of the  $i^{\text{th}}$  peak,  $R_{\text{opt}}$  is the desired optimum resolution (1.5),  $t_f$  is the migration time of the final peak,  $t_{\text{max}}$  is the maximum acceptable time (40 or 50 min),  $a$  is the slope adjustment factor (here set to 1) and  $n$  is the number of expected peaks. The slope adjustment factor ( $a$ ) was set to 1 so that the significance of the resolution term was not increased compared to the time term (no weighing). Optimum resolution is obtained if CEF is minimized. A modified equation of CEF, where the last term ( $1 + t_f/t_{\text{max}}$ ) was excluded, was also used and compared to the original equation.

$$\text{COF} = \sum_{i=1}^n A_i \ln(R_i/R_{\text{id}}) + B(t_m - t_n) \quad (8)$$

where  $R_i$  is the resolution of the  $i^{\text{th}}$  pair and  $R_{\text{id}}$  is the desired resolution ( $= 1.5$ ),  $t_m$  is the desired maximum analysis time (40 or 50 min) and  $t_n$  the time of the last eluting peak.  $A$  and  $B$  are weights chosen by the operator [17]. Different weights were tested on  $A$  ( $= 2$ ) and  $B$  ( $= 0.1$ ) to give more weight to the separation and less to the time, although the last term (time) was still too dominant. A modification of the equation where only the separation term was included worked out better for the optimization and was therefore used.

$$F(R_i) = \frac{\arctan[a(R_i - b)] + \pi/2}{\pi} \quad (9a)$$

$$\text{ATR} = \sum_{i=1}^{n-1} F(R_i) \quad (9b)$$

where  $R_i$  is the resolution of the  $i^{\text{th}}$  pair, the values of  $a$  and  $b$  being chosen by the researcher. ATR is the sum of the function  $F(R_i)$  [18]. The constant  $b$  can be chosen as the minimum acceptable resolution (e.g., 1.0). Two different values of ATR were calculated, the first with  $a = 1$  and  $b = 0$ , and the second with  $a = 1$  and  $b = 1$ . The same conclusion was made regarding the setting of the factors for optimum separation, so that the first one was chosen.

$$R_p = \prod_{i=1}^{n-1} R_{S_{i,i+1}} \quad (10a)$$

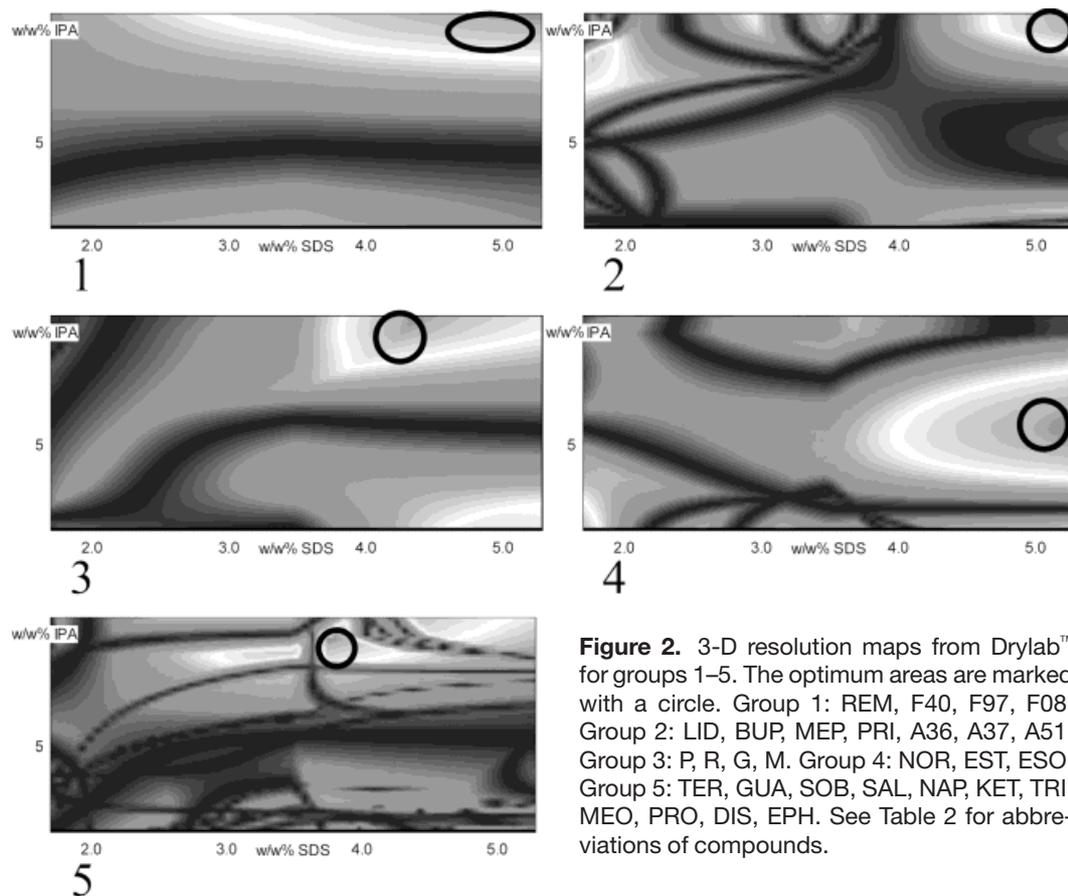
$$r = \frac{\prod_{i=1}^{n-1} R_{S_{i,i+1}}}{\left[ \frac{\sum_{i=1}^{n-1} R_{S_{i+1,i}}}{n-1} \right]^{n-1}} \quad (10b)$$

The resolution product (Eq. 10a) is calculated by multiplying all the resolutions [19]. The relative resolution product (Eq. 10b) is similar to the resolution product, except that it also takes into account how the peaks are spread in the electropherogram [20].

(iii) The migration times were entered into the Drylab™ program, and choosing the area with the highest resolution in the calculated 3-D resolution map optimized the separation [21–24]. The retention times or retention factors are adapted by a cubic fit (polynom containing linear and quadratic terms) to predict the 3-D resolution map [25–27]. Figure 2 shows 3-D resolution maps for five different groups of compounds.

Table 5 shows a compilation of the different optimization strategies and the optimum setting of the factors (SDS, IPA) for five different groups of compounds. For almost all groups, CRS (group 1–4) and CEF (1, 3, 5) gave different optimum settings of the factors compared to the other equations (CRF<sub>1-3</sub>, CRF<sub>1</sub> modified, CRS modified, CEF modified, COF modified, ATR, Rp and r), DryLab™ and for optimization of the selectivity in MODDE. This is due to the fact that the equations (6 and 7) take account of both resolution and total analysis time. Since increased amounts of SDS and IPA will increase the analysis time, the settings of the factors will differ compared to the other equations. For the modified equations of CRS, CEF, and COF (no time term included), the optimum setting of the factors was the same as for equations focused only on the resolution (ATR, Rp). Furthermore, CRF<sub>1</sub>, CRF<sub>1</sub> modified and CRF<sub>2</sub> have all a time term added (not multiplied) to the equation, but nevertheless gave the same optimum setting of the factors as for the equations only containing resolution terms. Since the equations CRS and CEF suggested areas in the experimental domain that were not most favorable for the separation, more account was taken of the other functions.

For group 1 (REM, F40, F97, F08)  $\alpha$ , Drylab™, CRF<sub>1-3</sub>, CRF<sub>1</sub> modified, CRS modified, CEF modified, COF modified, ATR and Rp were optimized and gave approximately the same optimum settings (4.7–5% SDS, 7.8–10% IPA). The responses CRS (2% SDS, 7.7% IPA), CEF (2.6% SDS, 2% IPA) and r (2% SDS, 10% IPA) had other optimum settings. Furthermore, two optima were found when CEF modified was used as a response. REM and F97 were the analytes that came closest together, but



**Figure 2.** 3-D resolution maps from Drylab™ for groups 1–5. The optimum areas are marked with a circle. Group 1: REM, F40, F97, F08. Group 2: LID, BUP, MEP, PRI, A36, A37, A51. Group 3: P, R, G, M. Group 4: NOR, EST, ESO. Group 5: TER, GUA, SOB, SAL, NAP, KET, TRI, MEO, PRO, DIS, EPH. See Table 2 for abbreviations of compounds.

the separation was easy to optimize. Two additional experiments were carried out: one with 5% w/w SDS and 10% w/w IPA (Fig. 3B) and the other with 5% w/w SDS and 0% w/w IPA (Fig. 3A). The other variables were set to 1% w/w Brij 35, 7% w/w 1-butanol, 0.8% w/w octane, 20 mM borate buffer pH 9.2, 10 kV, and 40°C. As can be seen from the figure, all peaks are well separated in both cases. F97 migrates before the REM peak when no IPA is present in the microemulsion and after the REM peak when 10% w/w IPA is added. This is due to the fact that F97 is partially negatively charged and the other compounds are partially positively charged. Furthermore, the peak shape of F97 was different compared to the other peaks for the same reason. All peaks migrate more slowly when IPA is added to the microemulsion, due to the increase in viscosity and decrease in EOF.

In group 2 (LID, BUP, MEP, PRI, A36, A37, A51) all the responses gave approximately the same optimum settings of the factors, 3.4–5% SDS and 10% IPA, except when CRS was used (see explanation above). A37 and A51 were the most difficult peaks to separate. Even at the highest amounts of SDS and IPA, the resolution was

not sufficient. The optimum settings of the factors should be outside the experimental domain investigated. One additional experiment with 6% w/w SDS and 12% w/w IPA was performed. Figure 3C shows that a good separation between the peaks was observed, but that baseline separation between A37 and A51 was still not achieved.

In group 3 (P, R, G, M) two critical pair of peaks were found: G and R and R and P. Different settings of the variables were obtained when optimizing  $\alpha$ . For the separation between G and R, SDS and IPA should be 2.8 and 10% w/w, respectively. Other settings of the factors were found for best separation between R and P (5% w/w SDS, 2% w/w IPA). Interestingly, the responses CRF<sub>2-3</sub>, CRS modified, CEF modified, COF modified, ATR and Rp had two optima in the response surface plot, where the factors should be either 3.5–4.5% w/w SDS and 2% w/w IPA, or 3.5–5% w/w SDS and 10% w/w IPA. The relative resolution product ( $r$ ) gave two optimum settings as well, but with a medium amount of IPA (5.5–6.5% IPA). This should reflect the problem that two different settings of the factors were best for the separation between

**Table 5.** Optimum settings of variables for best separation of different groups of the analytes

	Group 1		Group 2		Group 3		Group 4		Group 5	
	(%) SDS	(%) IPA	(%) SDS	(%) IPA	(%) SDS	(%) IPA	(%) SDS	(%) IPA	(%) SDS	(%) IPA
$\alpha$ of critical peak pair	5.0	10.0	5.0	10.0	5.0 <sup>c)</sup> 2.8 <sup>d)</sup>	2.0 <sup>c)</sup> 10.0 <sup>d)</sup>	5.0	6.0	3.5	10.0
3-D Rs maps in DryLab™	5.0	10.0	5.0	10.0	4.4	10.0	5.0	6.0	3.8	9.8
CRF <sub>1</sub>	5.0	7.8	5.0	10.0	5.0	10.0	5.0	6.0	5.0	10.0
CRF <sub>1</sub> modified <sup>a)</sup>	5.0	10.0	5.0	10.0	5.0	10.0	5.0	7.3	5.0	10.0
CRF <sub>2</sub>	5.0	10.0	4.3 <sup>e)</sup> 4.3 <sup>f)</sup>	2.0 <sup>e)</sup> 10.0 <sup>f)</sup>	3.8 <sup>e)</sup> 4.5 <sup>f)</sup>	2.0 <sup>e)</sup> 10.0 <sup>f)</sup>	3.5	6.5	5.0	10.0
CRF <sub>3</sub>	5.0	10.0	4.4 <sup>e)</sup> 4.4 <sup>f)</sup>	2.0 <sup>e)</sup> 10.0 <sup>f)</sup>	3.8 <sup>e)</sup> 4.5 <sup>f)</sup>	2.0 <sup>e)</sup> 10.0 <sup>f)</sup>	3.5	6.5	5.0	10.0
CRS	2.0	7.7	2.6	10.0	2.0	2.0	2.0	5.6	5.0	10.0
CRS modified <sup>b)</sup>	5.0	10.0	4.5 <sup>e)</sup> 3.8 <sup>f)</sup>	2.0 <sup>e)</sup> 10.0 <sup>f)</sup>	4.5 <sup>e)</sup> 5.0 <sup>f)</sup>	2.0 <sup>e)</sup> 10.0 <sup>f)</sup>	4.5	7.6	5.0	10.0
CEF	2.6	2.0	4.1	10.0	3.6	2.0	3.2	6.1	5.0	2.0
CEF modified <sup>b)</sup>	2.9 <sup>e)</sup> 4.7 <sup>f)</sup>	2.0 <sup>e)</sup> 10.0 <sup>f)</sup>	4.3	10.0	3.7 <sup>e)</sup> 4.3 <sup>f)</sup>	2.0 <sup>e)</sup> 10.0 <sup>f)</sup>	3.8	7.4	5.0	10.0
COF modified <sup>b)</sup>	5.0	10.0	5.0	10.0	3.8 <sup>e)</sup> 4.5 <sup>f)</sup>	2.0 <sup>e)</sup> 10.0 <sup>f)</sup>	4.4	6.7	5.0	10.0
ATR	5.0	10.0	5.0	10.0	4.2 <sup>e)</sup> 4.2 <sup>f)</sup>	2.0 <sup>e)</sup> 10.0 <sup>f)</sup>	4.4	6.4	4.9	10.0
Rp	5.0	10.0	5.0	10.0	3.5 <sup>e)</sup> 3.5 <sup>f)</sup>	2.0 <sup>e)</sup> 10.0 <sup>f)</sup>	4.1	5.9	5.0	10.0
r	2.0	10.0	3.4	10.0	5.0 <sup>e)</sup> 2.0 <sup>f)</sup>	5.5 <sup>e)</sup> 6.5 <sup>f)</sup>	2.8	7.2	5.0	10.0

Comparison of different optimization strategies. For optimization of separation:  $\alpha$ , 3-D resolution map (DryLab™), CRF<sub>1-3</sub>, CRF<sub>1</sub> modified, COF, ATR, Rp and r should be maximized and CRS and CEF should be minimized. Group 1: REM, F40, F97, F08. Group 2: LID, BUP, MEP, PRI, A36, A37, A51. Group 3: P, R, G, M. Group 4: NOR, EST, ESO. Group 5: TER, GUA, SOB, SAL, NAP, KET, TRI, MEO, PRO, DIS, EPH. See Table 2 for abbreviations of compounds.

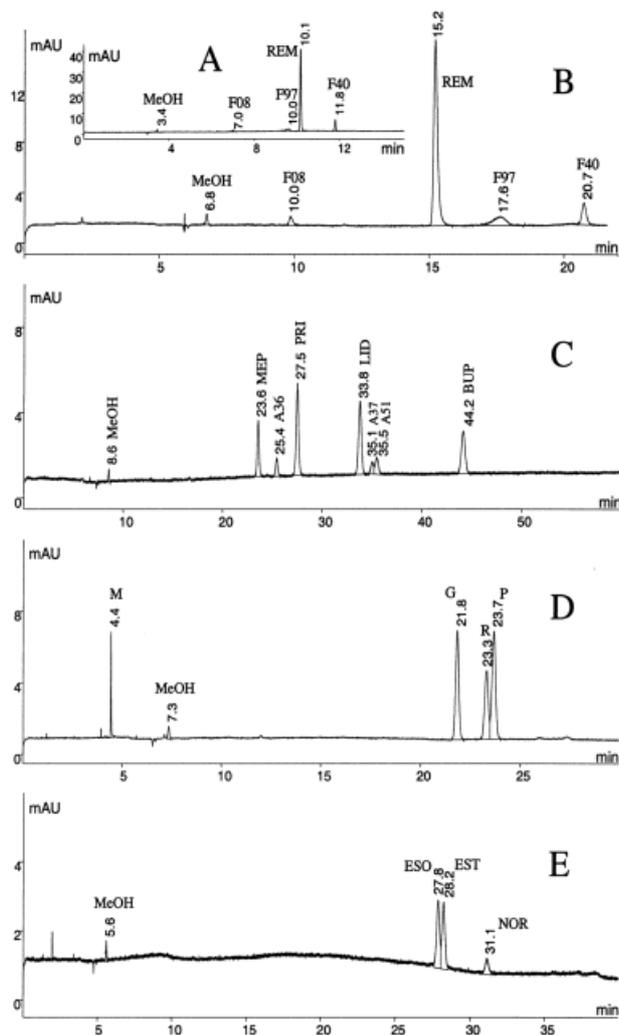
$\alpha = k_{M,2}/k_{M,1}$ , where  $k_M$  is the migration factor.

- a) Resolution values > 3 were not included in the sum;  
 b) the time term was excluded from the equation;  
 c) maximum selectivity between R and G;  
 d) maximum selectivity between P and R;  
 e), f) two maxima were found in the response surface plot.

G/R and R/P. The responses CRF<sub>1-3</sub>, CRF<sub>1</sub> modified and Drylab™ gave approximately the same settings of the factors, 4.4–5% SDS and 10% IPA. For the additional experiment, an SDS concentration of 4.5% w/w was chosen. Peak movements were studied by using MODDE for prediction of migration times at different levels of IPA. Low or medium concentrations of IPA gave poor separation between G and R, but at higher concentrations better separations could be achieved between G, R and P. The concentration of IPA (12% w/w) was therefore set higher in the additional ex-

periment compared to the highest concentration in the experimental domain. Figure 3D shows that G and R are well separated, but that R and P are not baseline-separated.

Approximately the same setting of the factors for optimum separation of the compounds in group 4 (NOR, EST, ESO) was found for all the different optimization strategies, a medium to high level of SDS and a medium level of IPA. The separation between ESO and EST was difficult, so the additional experiment was done



**Figure 3.** Optimization of the separation of the different groups of analytes. (A) (B) REM, F08, F97, and F40 (group 1); (C) MEP, PRI, LID, BUP, A36, A37, and A51 (group 2); (D) M, G, R, and P (group 3); (E) ESO, EST, and NOR (group 4). Concentration of SDS and IPA (w/w): (A) 5% SDS, 0% IPA; (B) 5% SDS, 10% IPA; (C) 6% SDS, 12% IPA; (D) 4.5% SDS, 12% IPA; (E) 6% SDS, 6% IPA. Settings of other factors: 1% w/w Brij 35, 7% w/w 1-butanol, 0.8% w/w octane, 20 mM borate buffer, pH 9.2, 10 kV, 40°C. MeOH, methanol; EOF marker.

with 6% w/w SDS and 6% w/w IPA (Fig. 3E). Even with an increased amount of SDS, a complete separation between ESO and EST was not achieved.

Group 5 had several pairs of peaks (DIS/MEO, NAP/KET, EPH/GUA, TER/EPH) that could be optimized in separation. When  $\alpha$  was maximized in the “Optimizer” (Simplex optimization) in MODDE, a compromise was chosen since differ-

ent settings gave the best separation for different peak pairs. For MEO/DIS, NAP/KET, and GUA/EPH, the concentration of SDS should be at a medium and the amount of IPA should be at a high level for maximum resolution, while a high amount of SDS and a low level of IPA are best for the separation of EPH/TER. A compromising setting of the factors was therefore 3.5% SDS and 9.7% IPA. This was close to the optimum setting from Drylab™ (3.8% SDS, 9.8% IPA). All the other responses (except CEF) gave 4.9–5% w/w SDS and 10% w/w IPA as optimum settings of the factors. Group 5 consists of the remaining analytes that did not belong to groups 1–4. Different settings of the factors can be chosen, depending on which compounds are to be separated; hence, no additional experiment was carried out for optimization of the separation of all compounds in group 5.

Generally, separation between peaks is achieved at long migration times, but at the same time the efficiency is reduced for late migrating peaks and the total analysis time will be too long. Logically, when using the different optimization methods, the optimum settings of SDS and IPA were found in one corner, namely at the highest level for both factors. However, the optimum condition of a factor could also be found at a medium level (e.g., 6% w/w IPA for group 4). If the resulting separation within the experimental domain is not adequate, a small extrapolation outside the area is one possible way to go to achieve sufficient separation. In some cases (A37/A51 in group 2, R/P in group 3 and ESO/EST in group 4), the extrapolation of one or both factors tested was not enough.

Optimizing the separation of critical peak pairs by modeling in MODDE the selectivity ( $\alpha$ ) between them worked well. The risk is that while optimizing the separation for one peak pair, the resolution between others can deteriorate. It is also possible to optimize several critical peak pairs in the same model. MODDE can be used for the prediction of peak movements. One factor is then set at a fixed level and the migration time can be predicted at several levels of the other factor. A combination of optimizing  $\alpha$  and at the same time studying peak movements in MODDE is one possible strategy for optimizing the separation.

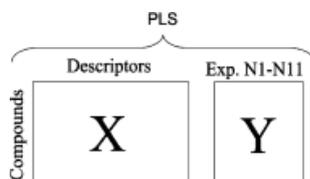
Drylab™ enables one to predict the resolution between all peaks within the experimental domain. The 3-D resolution plot presented is for the most critical peak pair. Different pairs of peaks can be critical in different parts of the experimental domain. The same optimum settings of the factors were found in both DryLab™ and MODDE, except for group 3. Most of the chromatographic functions worked well except for CRS and CEF. This was probably due to the fact that a time term was multiplied by the resolution term, giving too much weight to the total time compared to the resolution.

### 3.3 Molecular modelling

#### 3.3.1 Fitting a PLS model between descriptors of the compounds and migration times from the experimental design

SELMA [28–32] is a program used for describing chemical information of different compounds. The program calculated 93 descriptors for each analyte used in the three-level full factorial design (29 compounds, see Fig. 1). Example of descriptors are: number of bonds, number of atoms, nitrogen counts, highest positive atomic charge, topological dipole moment, polarizability, molecular weight, number of H-bond donors, and log *P*

(Table 6). The descriptors of each compound were then fitted by PLS with four components [42] to the migration times from experiments N1–N11 in the three-level full factorial design (Fig. 4). The PLS model was refined by



**Figure 4.** Fitting molecular descriptors of the compounds to the retention data (log(migration times)) from the three-level full factorial design (experiments N1–N11) with PLS.

**Table 6.** Name and explanation of descriptors from the in-house software SELMA [28] used in the refined PLS model

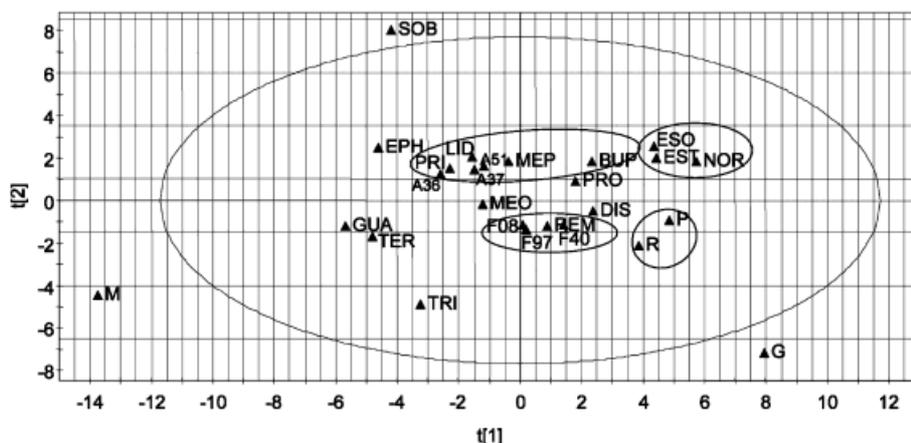
No.	Name of descriptor	Explanation
1	Numb. of bonds	Number of bonds
2	Numb. of rings	Number of rings
3	MaxRing1	Size of the largest ring
4	MaxRing3	Size of the third largest ring
5	Numb. of rig. bond	Number of rigid bonds
6	Min eV #3	Third smallest eigenvalue of the graph adjacency matrix
7	Graph radius	Smallest of the largest values found in the graph distance matrix
8	Graph diameter	Length of the longest chain in molecular graph
9	Wiener index	Half of the sum of all distance matrix elements
10	Balaban index	$J = \frac{2}{m - n + 1} \sum_{\text{bond } i,j} \frac{1}{\sqrt{(v_i * v_j)}}; v_i = \sum_{j \neq i} d_{ij}$ <p><i>m</i> = number of bonds; <i>n</i> = number of atoms; <i>d</i> = distance matrix element [46]</p>
11	Motoc index	$M = \sqrt{\frac{\sum_{d_1}^{d_i=d_{\max}} g_i d_i^2}{\sum_{d_1}^{d_i=d_{\max}} g_i}}$ <p><i>g<sub>i</sub></i> is the number of atomic pairs separated by the distance <i>d<sub>i</sub></i>, where <i>d<sub>i</sub></i> varies from 1 to <i>d<sub>max</sub></i>.</p>
12	Inform content	<p>Information content: <math>I = - \sum_{\text{sym.cl.}} p_i * \log_2(p_i)</math></p> <p><i>p<sub>i</sub></i> is the number of atoms within each topological symmetry class divided by total number of atoms.</p>
13	K&H Kappa1	$\text{Kappa1} = {}^1\kappa = \frac{n(n-1)^2}{({}^1P)^2}$ <p><i>n</i> is the number of atoms and <i><sup>1</sup>P</i> is the number of topological paths of length 1 [47, 48].</p>

**Table 6.** Continued

No.	Name of descriptor	Explanation
14	Kier Chi0	Connectivity index
15	Kier Chi2	Connectivity index
16	Kier Chi3p	Connectivity index
17	Kier Chi4p	Connectivity index
18	Kier Chi6p	Connectivity index
19	Carbon count	Count of element C
20	Nitrogen count	Count of element N
21	Max. pos. charge G	Highest positive atomic charge
22	Charge range Gast.	Difference between the highest and the lowest atomic charges
23	Aver. neg. charge 1	An average negative atomic charge
24	Dipole moment Gast	Topological dipole moment
25	HMO reson. Energy	Resonance energy of $\pi$ -electron system in $\beta$ -units
26	Aver. pos. charge 2	An average positive atomic charge
27	Aver. neg. charge 2	An average negative atomic charge
28	Dip. mom. 2D (G+H)	Topological dipole moment
29	Polar count	Number of polar atoms
30	Nonpolar count	Number of nonpolar atoms
31	Polar count / MW	Number of polar atoms divided by molecular weight
32	Nonpolar count / MW	Number of nonpolar atoms divided by molecular weight
33	Mol. volume 2D	Molecular volume
34	PSA	Polar surface area
35	NPSA	Nonpolar surface area
36	TSA	Total surface area
37	Polarizability	Molecular polarizability
38	HB-donors	Number of H-bond donors
39	HB-acceptors	Number of H-bond acceptors
40	Neg. ioniz.	Number of negatively ionizable groups
41	logP	Octanol/water partition coefficient (fragment based method by Suzuki [49])
42	clogP	Octanol/water partition coefficient (fragment based method by Leo and Weininger)
43	CMR	Molar refractivity [50]

excluding descriptors with values near 0, *i.e.*, descriptors with a low influence on the responses (see Table 6 for an explanation of the remaining descriptors). Furthermore, the model had difficulties with the prediction of migration times for compounds SAL, NAP, and KET. The model improved dramatically when these compounds were excluded. SAL, NAP, and KET are all negatively charged, but there are still five negatively charged compounds included in the model (F97, P, R, G, and SOB). The refined model had a moderate to high  $R^2$  (0.901) and  $Q^2$  (0.709).

Figure 5 shows a score plot of the compounds. Naturally, analytes with similar structures will group together in the plot. The most hydrophobic solutes are situated to the right of the figure (G, NOR, EST, ESO), while the most hydrophilic compounds are to the left (M, TRI, GUA, TER, EPH). The loadings for X (descriptors) and Y (migration times) from the three-level full factorial design are placed in the same plot (Fig. 6). Descriptors that are positively correlated with the migration times (according to experiments N1–N11) are found in the upper right square. These descriptors explain the hydrophobicity of the compounds



**Figure 5.** Score plot of 26 compounds from the PLS model. SAL, NAP, and KET are excluded from the model. See Table 2 for abbreviations of compounds.

like octanol/water partition coefficient ( $\log P$ ,  $\text{clog } P$ ) and number of nonpolar atoms (nonpolar count, nonpolar count/MW), as well as the charge of the analytes (negatively ionizable groups, average negative charge, average positive charge, dipole moment). Higher values of these descriptors will lead to longer migration times. Descriptors such as the size of the largest ring (max ring1), size of the third largest ring (max ring3), graph radius and graph diameter can also be found in the same corner. In the lower left corner descriptors are found that are negatively correlated with the responses (migration times according to experiments N1–N11). These descriptors give details of the number of polar atoms (polar count, polar count/MW), number of H-bond donors and acceptors (HB donors, HB acceptors), number of nitrogen atoms (nitrogen count) and polar surface area (PSA). Higher values of these descriptors will lead to shorter migration times. A representative plot of predicted migration times plotted against observed migration times for experiment N10 are shown in Fig. 7. As can be seen from the plot, a close correlation exists between predicted and observed migration times.

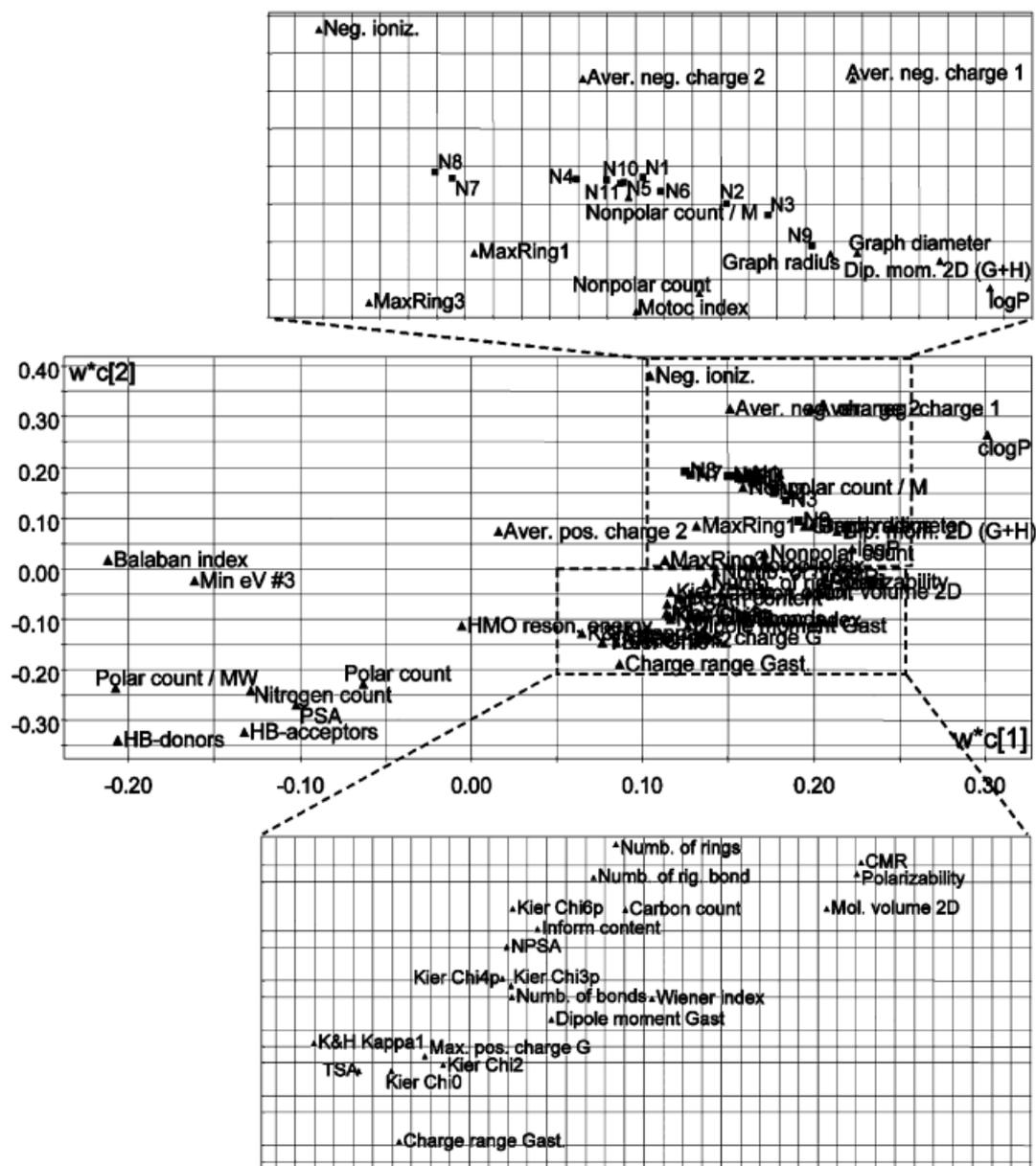
### 3.3.2 Validity of the PLS model

A model like the one described above can be used for the prediction of migration times with different settings of SDS and IPA for new compounds. Descriptors can be calculated for new compounds and the model will predict the migration times, giving the analyst an idea of the migration times of the new compound when using the different microemulsions before any laboratory work begins. Good predictions can be obtained when cross-validation ( $Q^2$ ) is used, as in the model above. The question is whether it is possible to predict the migration times of new molecules that are not involved in the model for different settings of SDS and IPA according to experi-

ments N1–N11. This was tested by randomly selecting seven compounds (NOR, GUA, TRI, G, REM, BUP, P) and using them as a test set. Two of the seven compounds in the test set were negatively charged (P, G), one was partly positive (REM) and the others were neutral. The remaining substances are then used as a training set, and a new PLS model is calculated. Consequently, the training set is used for prediction of migration times for the compounds in the test set. The  $R^2$  and  $Q^2$  of the training set were 0.882 and 0.624, respectively. The quotient between the predicted and the observed value was multiplied by 100 (a value close to 100% signifying a good prediction). If the prediction is acceptable in a range of 80–120%, the migration times of the compounds of the test set were within that range for 66 out of 77 predictions (85.7%). The compound G could not be predicted correctly for 6 out of 11 cases. Furthermore, TRI and P could not be predicted correctly for 2 out of 11 predictions, and the prediction of the retention times of GUA failed once. The predictions that were outside the acceptable range were 0.6–0.7 times lower (TRI, GUA) or 1.2–1.6 times higher (G, P) than the experimental values.

A second test set was created in the same way as the first. This time REM, A51, BUP, R, MEO, and EPH were randomly selected. The test set contained one negatively charged (R), two positively charged (EPH, REM) and three neutral compounds. The  $R^2$  and  $Q^2$  of the training set were 0.907 and 0.707, respectively. Fifty-nine out of 66 (89.4%) predictions were within acceptable range (80–120%). EPH could not be predicted properly for 7 out of 11 cases, the predicted migration times were 1.3–1.5 higher than the experimental values.

In conclusion, 86–89% of all predictions of new compounds not included in the model were acceptable (80–120% of the observed value). The models had difficulties with the prediction of migration times for a hydrophobic negatively charged compound (G) in the first test



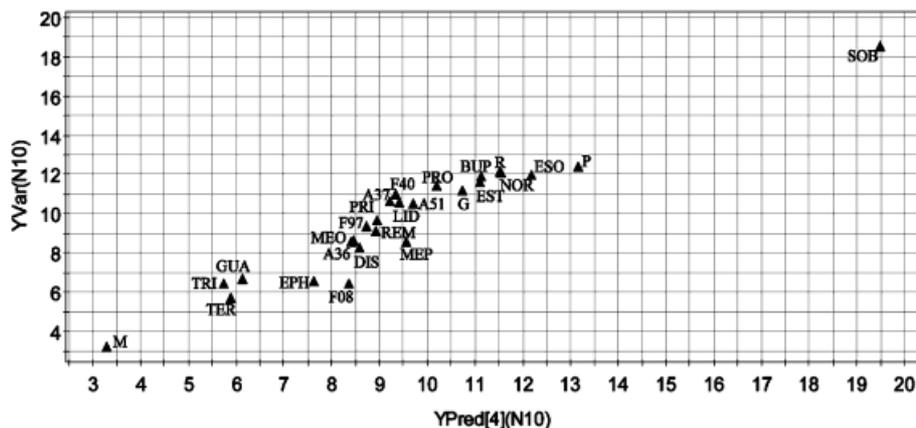
**Figure 6.** Loading plot of X (descriptors) and Y (log (migration times)) from the three-level full factorial design (experiments N1–N11).

set, and hydrophilic positively charged in the second test set (EPH). The models are still useful for prediction of retention times in different microemulsions of new compounds.

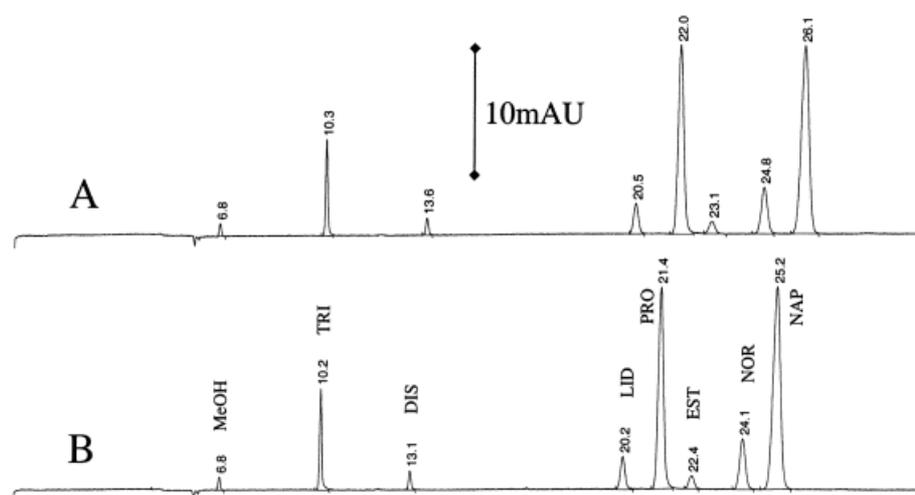
### 3.4 Stability of the microemulsion

A microemulsion consisting of 5% w/w SDS, 1% Brij 35, 7% w/w 1-butanol, 10% w/w IPA, 0.8% w/w octane, and 76.2% w/w 20 mM borate buffer, pH 9.2, was refrigerated

for 10 months. A new microemulsion with the same proportions of ingredients was freshly made. Both microemulsions were used for the analysis of a sample containing TRI, DIS, LID, PRO, EST, NOR and NAP. Figure 8 shows that if the right proportions in a microemulsion are used, the solution can remain stable for many months. The same separation is obtained for both microemulsions. A small but acceptable shift in migration times can be observed (26.1 compared to 25.2 min in migration times for the last peak (NAP)).



**Figure 7.** Predicted vs. observed migration times of 26 compounds. Settings of the factors are according to experiment N10 (3.5% w/w SDS, 6% w/w IPA).



**Figure 8.** Stability of microemulsion N9. The microemulsion consists of 5% w/w SDS, 1% w/w Brij 35, 7% w/w 1-butanol, 10% w/w IPA, 0.8% w/w octane, 20 mM borate buffer, pH 9.2, 10 kV, 40°C. (A) The microemulsion was stored in the refrigerator for 10 months; (B) the microemulsion was freshly made. MeOH, methanol, EOF marker. See Table 2 for abbreviations of compounds.

#### 4 Concluding remarks

The factors SDS (% w/w) and IPA (%w/w) in microemulsion electrokinetic chromatography were investigated in a three-level full factorial design. The effects of these factors on the migration times of 29 different compounds were explored. Generally, an increased amount of SDS and IPA will increase the migration times of all analytes. Interaction between the factors SDS and IPA were found significant (95% of confidence) for R, SOB, KET, and NAP. The analytes were divided into five subgroups, and three different strategies for optimizing the separation of each group were investigated. The separation was optimized by (i) maximizing the selectivity of critical peak pairs in the software MODDE, (ii) by using chromatographic functions, and (iii) by using the software DryLab™. For all groups, MODDE, the chromatographic functions and DryLab™ gave approximately the same settings of the factors for optimum resolution. Two chromatographic

functions (CRS, CEF) did not work here. This was probably due to the fact that a time term is multiplied by the resolution term, giving too much weight to the total time compared to the resolution. A modified version of the two equations, where the time term was excluded, gave better predictions of the optimum setting of the factors.

In-house software (SELMA) was used for calculation of descriptors of the same compounds used in the three-level full factorial design. The compounds were divided into a training set and a test set and the training set could be used to predict the migration times of the compounds in the test set for different microemulsions (according to the different experiments in the three-level full factorial design). Using two different training sets, 86–89% of the observed migration times in the two test sets were predicted correctly (in the range of 80–120% of the experimental values). Highly stable microemulsions can be made if the right proportions of the ingredients are used.

A microemulsion (5% w/w SDS, 1% w/w Brij 35, 7% w/w 1-butanol, 10% w/w IPA, 0.8% w/w octane, and 76.2% w/w 20 mM borate buffer, pH 9.2) stored for 10 months in a refrigerator gave the same separation performance of a test mixture as a freshly made microemulsion with the same proportion of ingredients.

We would like to thank Ms. Jessica Eriksson for skilful experimental assistance. We would also like to thank Johan Gottfries and Ulf Norinder for helping us with the calculation of descriptors with the SELMA software.

Received September 28, 2003

## 5 References

- [1] Watarai, H., *Chem. Lett.* 1991, 391–394.
- [2] Altria, K. D., Mahuzier, P.-E., Clark, B. J., *Electrophoresis* 2003, 24, 315–324.
- [3] Strasters, J. K., Khaledi, M. G., *Anal. Chem.* 1991, 63, 2503–2508.
- [4] Harang, V., Ericsson, J., Sanger-van de Griend, C., Jacobsson, S. P., Westerlund, D., *Electrophoresis* 2004, 25, 80–93.
- [5] Thorsteinsdottir, M., Ringbom, C., Westerlund, D., Andersson, G., Kaufmann, P., *J. Chromatogr. A* 1999, 831, 293–309.
- [6] Persson-Stubberud, K., astrom, O., *J. Chromatogr. A* 1998, 798, 307–314.
- [7] Nielsen, M. S., Nielsen, P. V., Frisvad, J. C., *J. Chromatogr. A* 1996, 721, 337–344.
- [8] He, Y., Lee, H. K., *J. Chromatogr. A* 1998, 793, 331–340.
- [9] Klampfl, C. W., Leitner, T., Hilder, E. F., *Electrophoresis* 2002, 23, 2424–2429.
- [10] Siouffi, A. M., Phan-Tan-Luu, R., *J. Chromatogr. A* 2000, 892, 75–106.
- [11] Berridge, J. C., *J. Chromatogr.* 1982, 244, 1–14.
- [12] Stebbins, M. A., Davis, J. B., Clark, B. K., Sepaniak, M. J., *J. Microcol. Sep.* 1996, 8, 485–494.
- [13] Schaepper, J. P., Fox, S. B., Sepaniak, M. J., *J. Chromatogr. Sci.* 2001, 39, 411–419.
- [14] Schlabach, T. D., Excoffier, J. L., *J. Chromatogr.* 1988, 439, 173–184.
- [15] Morris, V. M., Hughes, J. G., Marriott, P. J., *J. Chromatogr. A* 1996, 755, 235–243.
- [16] Morris, V. M., Hargreaves, C., Overall, K., Marriott, P. J., Hughes, J. G., *J. Chromatogr. A* 1997, 766, 245–254.
- [17] Miyawa, J. H., Alasandro, M. S., Riley, C. M., *J. Chromatogr. A* 1997, 769, 145–153.
- [18] Mikaeli, S., Thorsen, G., Karlberg, B., *J. Chromatogr. A* 2001, 907, 267–277.
- [19] Klein, E. J., Rivera, S. L., *J. Liq. Chromatogr. Rel. Technol.* 2000, 23, 2097–2121.
- [20] Billiet, H. A. H., Andersson, P. E., Haddad, P. R., *Electrophoresis* 1996, 17, 1367–1372.
- [21] Haber, P., Baczek, T., Kaliszczan, R., Snyder, L. R., Dolan, J. W., Wehr, C. T., *J. Chromatogr. Sci.* 2000, 38, 386–392.
- [22] Molnar, I., *J. Chromatogr. A* 2002, 965, 175–194.
- [23] Jupille, T. H., Dolan, J. W., Snyder, L. R., Molnar, I., *J. Chromatogr. A* 2002, 948, 35–41.
- [24] Hoang, T. H., Cuerrier, D., McClintock, S., Di Maso, M., *J. Chromatogr. A* 2003, 991, 281–287.
- [25] Drylab™ 2000, Version 3.0.09, LC Resources Inc., Walnut Creek, CA 94596, USA.
- [26] *User manual DryLab™ 2000.*
- [27] *Reference guide DryLab™ 2000.*
- [28] SELMA, *Synthesis and Structure Administration (SaSA)*, Olsson, T., Sherbukhin, V., AstraZeneca R&D Molndal, Molndal, Sweden.
- [29] Oprea, T. I., Gottfries, J., Sherbukhin, V., Svensson, P., Kuhler, T. C., *J. Mol. Graphics Modelling* 2000, 18, 512–524.
- [30] Linusson, A., Gottfries, J., Lindgren, F., Wold, S., *J. Med. Chem.* 2000, 43, 1320–1328.
- [31] Oprea T. I., Gottfries, J., *J. Comb. Chem.* 2001, 3, 157–166.
- [32] Bergstrom C., Norinder, U., Luthman, K., Artursson, P., *J. Chem. Inf. Comput. Sci.* 2004, in press.
- [33] Baczek, T., Kaliszczan, R., *J. Chromatogr. A* 2003, 987, 29–37.
- [34] Nord, L. I., Fransson, D., Jacobsson, S. P., *Chemometrics Int. Lab. Systems* 1998, 44, 257–269.
- [35] Lee, S. K., Park, Y. H., Yoon, J. Y., Lee, D. W., *J. Microcol. Sep.* 1998, 10, 133–139.
- [36] Liang, H.-R., Vuorela, H., Vuorela, P., Hiltunen, R., Riekkola, M.-L., *J. Liq. Chromatogr. Rel. Technol.* 1998, 21, 625–643.
- [37] Schaeper, J. P., Fox, S. B., Sepaniak, M. J., *J. Chromatogr. Sci.* 2001, 39, 411–419.
- [38] Kilar, F., Visegrady, B., *Electrophoresis* 2002, 23, 964–971.
- [39] ACD/Log P DB, Version 5.15, Advanced Chemistry Development Inc., Toronto, Canada.
- [40] Box, G. E. P., Hunter, W. G., Hunter, J. S., *Statistics for Experimenters*, John Wiley & Sons, New York 1978.
- [41] *User manual of program MODDE 6.0.*
- [42] Hoskuldsson, A., *J. Chemom.* 1988, 2, 211–228.
- [43] Klampfl, C. W., *Electrophoresis* 2003, 24, 1537–1543.
- [44] Albert A., Serjeant, E. P., *The Determination of Ionization Constants – a Laboratory Manual*, third edition, Chapman and Hall, London, New York 1984.
- [45] Moffat, A. C., *Clarke's Isolation and Identification of Drugs in Pharmaceuticals, Body Fluids, and Post-mortum Material*, second edition, The Pharmaceutical Press, London 1986.
- [46] Balaban, A., *Pure Appl. Chem.* 1983, 55, 199–206.
- [47] Kier, L. B., *Quant Struct.-Act. Relat.* 1985, 4, 109–116.
- [48] Kier, L. B., *Quant. Struct.-Act. Relat.* 1986, 5, 1–7.
- [49] Suzuki, T. et al., *J. Chem. Eng. Japan* 1993, 26, 114–115.
- [50] <http://www.daylight.com/dayhtml/doc/cmr/>