Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Pharmacy 300

Aspects of Optimisation of Separation of Drugs by Chemometrics

BY

VALÉRIE HARANG

ACTA UNIVERSITATIS UPSALIENSIS
UPPSALA 2003

Dissertation presented at Uppsala University to be publicly examined in B41, Uppsala Biomedical Centre, Uppsala, Friday, November 28, 2003 at 13:15 for the degree of Doctor of Philosophy (Faculty of Pharmacy). The examination will be conducted in Swedish.

Abstract

Harang, V. 2003. Aspects of Optimisation of Separation of Drugs by Chemometrics. Acta Universitatis Upsaliensis. *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy* 300. 75 pp. Uppsala. ISBN 91-554-5778-9

Statistical experimental designs have been used for method development and optimisation of separation. Two reversed phase HPLC methods were optimised. Parameters such as the pH, the amount of tetrabutylammonium (TBA; co-ion) and the gradient slope (acetonitrile) were investigated and optimised for separation of erythromycin A and eight related compounds. In the second method, a statistical experimental design was used, where the amounts of acetonitrile and octane sulphonate (OSA; counter ion) and the buffer concentration were studied, and generation of an α-plot with chromatogram simulations optimised the separation of six analytes.

The partial filling technique was used in capillary electrophoresis to introduce the chiral selector Cel7A. The effect of the pH, the ionic strength and the amount of acetonitrile on the separation and the peak shape of R- and S-propranolol were investigated.

Microemulsion electrokinetic chromatography (MEEKC) is a technique similar to micellar electrokinetic chromatography (MEKC), except that the microemulsion has a core of tiny droplets of oil inside the micelles. A large number of factors can be varied when using this technique. A screening design using the amounts of sodium dodecyl sulphate (SDS), Brij 35, 1-butanol and 2-propanol, the buffer concentration and the temperature as factors revealed that the amounts of SDS and 2-propanol were the most important factors for migration time and selectivity manipulation of eight different compounds varying in charge and hydrophobicity. SDS and 2-propanol in the MEEKC method were further investigated in a three-level full factorial design analysing 29 different compounds sorted into five different groups. Different optimisation strategies were evaluated such as generating response surface plots of the selectivity/resolution of the most critical pair of peaks, employing chromatographic functions, simplex optimisation in MODDE and 3D resolution maps in DryLab™.

Molecular descriptors were fitted in a PLS model to retention data from the three-level full factorial design of the MEEKC system. Two different test sets were used to study the predictive ability of the training set. It was concluded that 86 – 89% of the retention data could be predicted correctly for new molecules (80 – 120% of the experimental values) with different settings of SDS and 2-propanol.

Statistical experimental designs and chemometrics are valuable tools for the development and optimisation of analytical methods. The same chemometric strategies can be employed for all types of separation techniques.

Keywords: optimisation, separation, chemometrics, high performance liquid chromatography, electrodriven techniques, microemulsion electrokinetic chromatography, statistical experimental design, molecular modelling, chiral separation, cellobiohydrolase

Valérie Harang, Department of Medicinal Chemistry, Analytical Pharmaceutical Chemistry, Box 574, Uppsala University, SE-75123 Uppsala, Sweden

ISSN 0282-7484
ISBN 91-554-5778-9
urn:nbn:se:uu:diva-3738 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-3738)

I like it
I do it
That's my code

Alain Delon
(French actor)

Papers Discussed

This thesis is based on the following papers, which are referred to by their Roman numerals in the text.

I. V. Harang, D. Westerlund
Optimization of an HPLC method for the separation of erythromycin and related compounds using factorial design.
Chromatographia, 1999, 50, 525 – 531

II. V. Harang, A. Karlsson, M. Josefson
Liquid chromatography method development and optimization by experimental design and chromatogram simulations.
Chromatographia, 2001, 54, 703 – 709

III. V. Harang, M. Tysk, D. Westerlund, R. Isaksson, G. Johansson
A statistical experimental design to study factors affecting enantioseparation of propranolol by capillary electrophoresis with cellobiohydrolase (Cel7A) as chiral selector.
Electrophoresis, 2002, 23, 2306 – 2319

IV. Valérie Harang, Jessica Eriksson, Cari Sänger-van de Griend, Sven P. Jacobsson, Douglas Westerlund
Microemulsion electrokinetic chromatography of drugs varying in charge and hydrophobicity. Part I: Impact of parameters on separation performance evaluated by multiple linear regression models.
Electrophoresis, accepted August 2003

V. Valérie Harang, Sven P. Jacobsson, Douglas Westerlund
Microemulsion electrokinetic chromatography of drugs varying in charge and hydrophobicity. Part II: Strategies for optimisation of separation.
In manuscript

Reprints were made by kind permission of the journals.

Contents

1 Introduction....1

2 Aim of study....5

3 Separation methods....6
3.1 Principles of liquid chromatography....6
3.2 Principles of electrodriven techniques....8
3.3 Equations....11
3.3.1 Efficiency, plate height and symmetry....11
3.3.2 Mobility....12
3.3.3 Retention factor and migration factor....13
3.3.4 Selectivity, apparent selectivity, relative mobility difference and resolution....14

4 Chemometrics....16
4.1 Experimental design....16
4.1.1 Screening design....16
4.1.2 Optimisation design, response surface methodology designs (RSM)....17
4.2 Modelling....17
4.3 Choice of factors....22

5 Effect of factors from the experimental designs....25
5.1 Papers I and II (LC)....25
5.2 Paper III (CE with chiral selector)....27
5.3 Papers IV and V (MEEKC)....31

6 Responses for optimisation....38
6.1 Number of peaks, number of resolved peaks....38
6.2 Noise, efficiency, symmetry....38
6.3 Retention/migration time, retention factor and mobility....39
6.4 Resolution, selectivity, relative mobility difference....39
6.5 Chromatographic functions....39
6.6 Response surface or contour plots....45

7 Optimisation of separation using optimiser in MODDE (simplex), α-plot or DryLab™....47

7.1 Simplex 47
7.2 Alpha-plot 48
7.3 DryLab™ 49
7.4 Discussions of the different optimisation tools 50

8 Molecular modelling 54
8.1 Fitting a PLS model between descriptors of the compounds and migration times from the experimental design in MEEKC (paper V) 54
8.2 Validity of the PLS model 57

9 Conclusions 59

10 Acknowledgements 67

11 References 69

12 Appendix: Structure of compounds used in papers I – V 73

Abbreviations

ACN	acetonitrile
AEA	anhydroerythromycin A
α	selectivity
A36	AR-P016336
A37	AR-P016337
A51	AR-P017151
BGE	background electrolyte
Brij	Brij 35
Buf	buffer concentration
BUP	bupivacaine
But	1-butanol
CCC	central composite circumscribed design
CCD	central composite design
CCF	central composite face-centred design
CEC	capillary electrochromatography
CZE	capillary zone electrophoresis
DIS	disopyramide
DKP	diketopiperazine
DOD	dodecyl benzene
EA	erythromycin A
EANO	erythromycin A N-oxide
EB	erythromycin B
EC	erythromycin C
EE	erythromycin E
EEA	erythromycin A enol ether
EOF	electroosmotic flow
EPH	ephedrine
ESO	estrone
EST	estradiol
F08	FLA708
F97	FLA797
F40	FLA740
Gr	gradient slope

GUA	guaifenesin
HPLC	high-performance liquid chromatography
Ion	buffer concentration (paper II) or ionic strength (paper III)
IPA	2-propanol
k	retention factor
KET	ketoprofen
kI	an unknown peak in paper II
k21	H 314/21
k26	H 314/26
k27	H 314/27
k99	H 299/87
LID	lidocaine
Log P	octanol-water partition coefficient
MEA	N-demethylerythromycin A
MEEKC	microemulsion electrokinetic chromatography
MEKC	micellar electrokinetic chromatography
MEO	metoprolol
MEP	mepivacaine
MLR	multiple linear regression
μ	mobility
μ_{eff}	effective mobility
μ_{app}	apparent mobility
N	efficiency
NAP	naproxen
NOR	norethisterone
OSA	octane sulphonate
pK_a	acid ionisation constant
PLS	partial least squares projections to latent structures
PRI	prilocaine
PRO	propranolol
q	ratio of stationary phase to mobile phase volume in the column: W_s/V_m
q	quotient between the effective mobility of the analyte in the microemulsion and the effective mobility of the analyte in the corresponding buffer system (paper IV)
REM	remoxipride
RMD	relative mobility difference
R_S	resolution
RSM	response surface methodology design
SAL	salicylic acid
SDS	sodium dodecyl sulphate

SOB	sodium benzoate
Sym	symmetry
TBA	tetrabutylammonium hydrogen sulphate
T	temperature
TER	terbutaline
TRI	trimethoprim

1 Introduction

An essential part of the pharmaceutical industry is the department of Analytical Development, whose goals include "contributing to the development of new active substances and pharmaceutical formulations by providing information based on pharmaceutical analytical chemistry, by developing analytical methods and specifications used in quality control of material for toxicological and clinical trials and by subsequent transfer of these methods and specifications to operations".

The development of robust and efficient methods is a crucial process in order to provide other departments with chemical information. New projects are started continuously, and this, combined with a shortage of personnel, means that fast and efficient method development is needed.

High-performance liquid chromatography (HPLC) is one of the most frequently used techniques in analytical chemistry. The availability of a large amount of different column-packing materials, combined with the ability to manipulate several parameters (the pH, the type and concentration of the buffer, the type and concentration of organic modifier (acetonitrile, tetrahydrofurane, methanol), the mixture of organic modifiers, the temperature, the flow, gradient and additives (ion-pair or competing reagents)), makes the technique useful for method development and optimisation of separation.

In the past decade electrodriven techniques, such as capillary zone electrophoresis (CZE), capillary electrochromatography (CEC), micellar electrokinetic chromatography (MEKC) and microemulsion electrokinetic chromatography (MEEKC), have been used for method development. In these techniques a large number of parameters can also be manipulated (type and concentration of the buffer, surfactant, organic solvent, ion-pair reagent, pH, temperature, voltage, and type and dimension of the capillary) for optimisation of separation.

Traditionally, method development and optimisation have been performed by varying one factor at a time, while keeping all the other factors constant. When the first factor had been optimised, the second one was varied, and so on. Unfortunately, by using this strategy many experiments are carried out without necessarily finding the optimum. When statistical experimental designs are used, several parameters are varied simultaneously in a limited number of experiments. The calculated model based on the

design and the results will provide information about the importance of each factor and whether there are interactions between the factors. Still if the optimum is not found within the experimental domain, an indication of better conditions will be given.

Depending on the compounds to be separated and on the chosen separation technique (HPLC or electrodriven techniques), different parameters are of importance. A knowledge of which factors to choose and the range of each factor can be gained by a search in the literature or by pre-experiments. Furthermore, several factors can be screened with a fractional factorial design. After the screening phase, if necessary, a central composite design can be performed with two or three of the most important factors for further optimisation of the separation.

Knowledge and consideration of the effects and processes of each parameter should be helpful and valuable for the choice of parameters in the screening and optimisation phase. However, interpreting of the effect of one factor in HPLC or electrodriven techniques can be difficult, since a change in one factor can affect others. For instance, changing the pH in LC may alter the charge of an analyte, which in turn affects the degree of ion pairing with counter ion added to the mobile phase. Adding an organic modifier to the background electrolyte (BGE) may affect the viscosity and the pH of the BGE. A change of viscosity and pH may in turn change the magnitude of the electroosmotic flow (EOF). This will also occur when changing one factor at a time. In order to recognise such interactions the parameters should be varied in a multivariate way.

J. P. Bounine et al. [1] presented a plot with connections between interacting variables, parameters and separation quality criteria (Figure 1). According to J. C. Berridge [2], the most important factors for maximising selectivity in HPLC are those involved in the mobile phase composition.

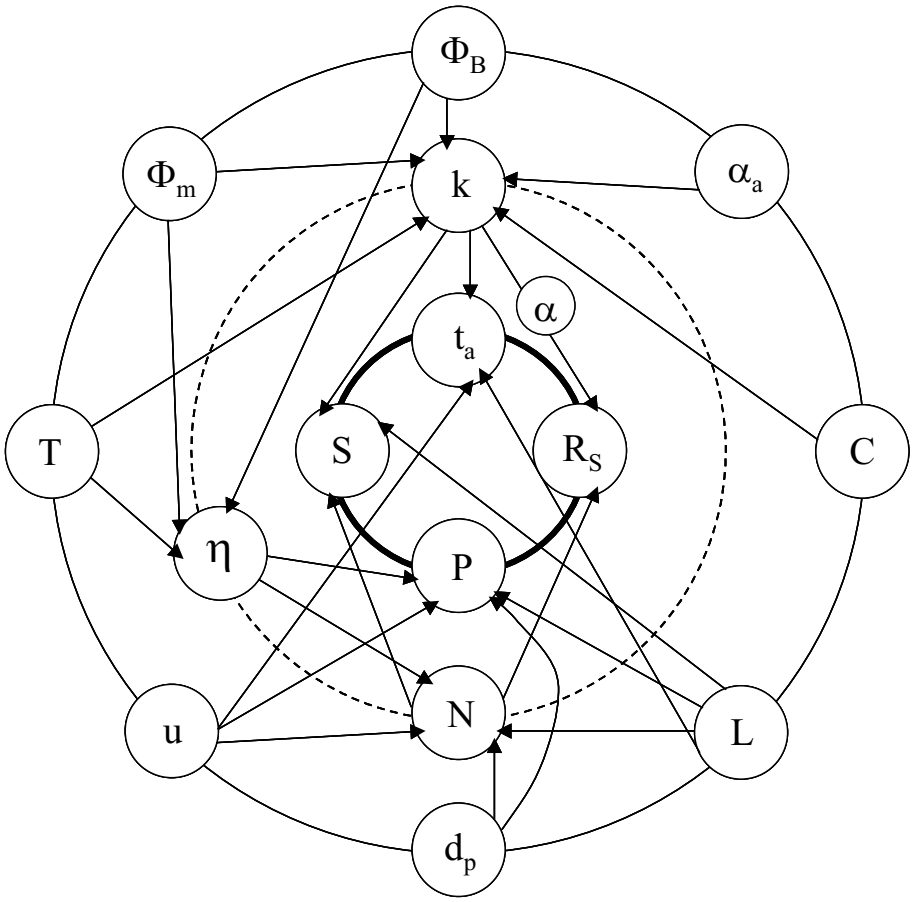


Figure 1: Interaction between variables (outer circle), intermediate parameters (dotted circle) and separation quality criteria (inner circle). Variables: Φ_B = volume fraction of strongest solvent, Φ_m = volume fraction of modifying solvent, T = temperature, u = flow (linear velocity), d_p = particle diameter, L = column length, α_a = adsorbent activity, C = carbon content of alkyl-bonded stationary phase (or counter-ion concentration). Intermediate parameters: α = selectivity, k = retention factor, N = column efficiency, η = solvent viscosity. Quality criteria: t_a = analysis time, R_S = resolution, P = pressure drop, S = peak height (sensitivity).

An overview of the available parameters and their effect on the selectivity in MEEKC is illustrated in Figure 2. There are a choice of numerous factors for method development and optimisation. A screening design revealed that the amounts of SDS and 2-propanol (paper IV) were the factors that had the largest impact on the migration time and the selectivity of eight different test compounds.

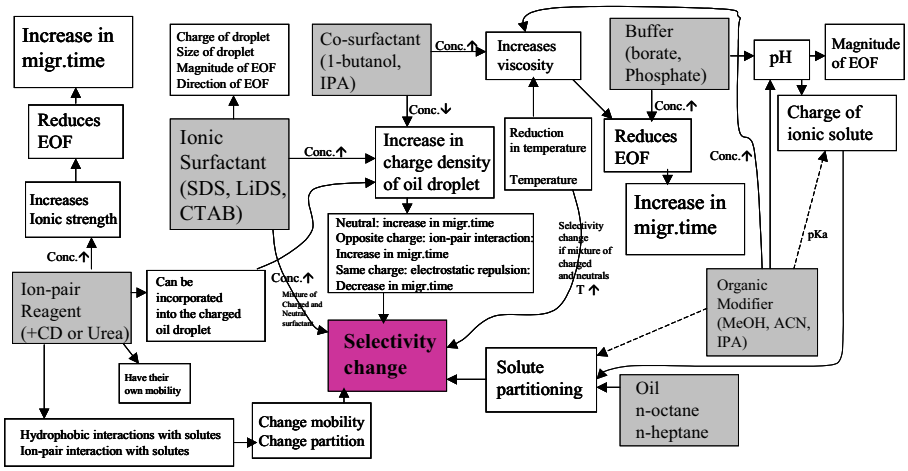


Figure 2: Overview of available factors and their effects of them on selectivity in MEEKC. The factors that can be manipulated in MEEKC are marked grey, and the effect of the manipulation is marked dark grey.

For complicated methods where several compounds are to be separated from each other, performing experiments according to a central composite design is often necessary for optimisation of separation. Usually two or three of the most important factors are used in the design. A difficult step in the optimisation is the selection of adequate response(s). Maximising the resolution or the selectivity of critical peak pairs can be done in response surface plots; however this method has its drawbacks since optimising one pair of peaks may impair the separation of another pair of peaks. Furthermore, response surface plots can only be plotted for two factors at a time, which means that if several factors are investigated, the setting of the third factor can be difficult. Chromatographic functions take the whole chromatogram/electropherogram into account. Advantages of using chromatographic functions are that only one value represents each chromatogram/electropherogram and no peak tracking is necessary. However, the drawback is that sometimes separation of peaks in one part is more important compared to another part of the chromatogram/ electropherogram. Simplex optimisation is another way of optimising a response. The response is optimised within the experimental domain for all the factors simultaneously. Furthermore, 3D resolution plots can be generated in DryLab™, giving a map of critical resolution between peaks. Unfortunately, the plots can only be generated for two factors at a time, and more experiments are needed compared to the use of MODDE. Depending on the separation problem, different approaches to optimisation is needed.

2 Aim of study

This thesis deals with optimisation of separation by means of experimental design with HPLC and electrodriven techniques. Considerations of the choice of factors, designs and responses are made, and the effects of factors on retention, selectivity and peak performance are studied. In addition, strategies for efficient method development and optimisation are outlined. Finally, an evaluation of the ability to predict the retention of new molecules when using a model of molecular descriptors and retention data in MEEKC is performed.

3 Separation methods

3.1 Principles of liquid chromatography

Adsorption chromatography (liquid-solid chromatography) was invented by the Russian botanist M. S. Tswett in 1903, who used the technique to separate plant pigments [3-6]. The technique was rediscovered in the 1930's and has since then developed rapidly to one of the most frequently used techniques in modern analytical chemistry. High-performance liquid chromatography (HPLC) developed in the 1970's have extensively been described in the literature in early textbooks [7-9], monograph series [10] and reviews [11]. Modern HPLC equipment consists of a pump, a detector (often a diode-array UV detector), an autoinjector and an oven usually containing a column with reversed-phase packing material (silica with hydrophobic alkyl groups attached to the surface). A mobile phase is pumped through the column and the analytes in the injected sample are separated depending on their degree of interaction to the packing material.

A relationship between the retention time of the solute and the equilibrium conditions is described by:

$$t_R = t_0(1+k) \qquad (1)$$

where t_R is the retention time of the solute and t_0 is the time of an unretained solute, and k is the retention factor, in adsorption chromatography defined according to:

$$k = \frac{C_s W_s}{C_m V_m} \qquad (2)$$

where C_sW_s is the total amount of the solute in the stationary phase and C_mV_m is the total amount of the solute in the mobile phase. C_s/C_m is the distribution ratio of the solute between the stationary phase and the mobile

phase and W_s/V_m (short form: q) is the ratio of stationary phase weight to mobile phase volume in the column.

The stationary phase (adsorbent) contains binding groups (active sites, adsorption sites). The active sites can vary, depending on the surface of the stationary phase. The retention of a solute in a chromatographic system is due to competition for the adsorption sites by the solute and the eluent.

The retention of an uncharged solute (A) will depend on its ability to compete for the active sites (binding sites) on the adsorbent (stationary phase) with the solvent molecules (S). The retention factor of an uncharged solute can be described by [8]:

$$k_A = \frac{qK^0 K_A}{1 + K_S [S]_m} \tag{3}$$

where K^0 is the capacity of the adsorbent. K_A is the equilibrium constant for the distribution of the solute to the stationary phase, K_S is the equilibrium constant for the distribution of the solvent to the stationary phase and $[S]_m$ is the concentration of eluent component S in the mobile phase.

For charged compounds (HA^+), the retention factor will depend on ion-pair interaction between the analyte (HA^+) and the ions of opposite charge (X^-) in the eluent. The ion pair will in turn compete with other ion pairs formed by components of the mobile phase for the adsorption sites on the stationary phase. The competition may be summarised in an equation for the retention factor [8]:

$$k_A = \frac{qK^0 K_{HAX} [X^-]}{1 + K_{QX} [Q^+][X^-]} \tag{4}$$

where K_{HAX} and K_{QX} are the solid-phase extraction constants of the analyte (HAX) and the buffer co-ion (Q^+) and counter ion (X^-).

The resolution between two solutes depends on three parameters (N, k_2 and α) and is described by the following equation:

$$R_S = \frac{\sqrt{N}}{4}\left(\frac{k_2}{1+k_2}\right)\left(\frac{\alpha-1}{\alpha}\right) \qquad (5)$$

where N is the column efficiency, k_2 is the retention factor of the last solute and α is the selectivity ($=k_2/k_1$) between the two solutes.

A gradient elution might be necessary if the retention differences between the eluting compounds are too large. Two different mobile phases are used during a gradient run. The elution power increases with time by progressively taking a larger portion of the mobile phase containing more organic modifier in reversed phase chromatography.

3.2 Principles of electrodriven techniques

Hjertén first described open tubular electrophoresis in 1967 [12]. The technique was further developed by Virtanen [13] and Mikkers [14] using 200-µm internal diameter capillaries, and in 1981 by Jorgenson and Lukacs using 75-µm internal diameter fused silica capillaries [15, 16].

A capillary electrophoresis instrument consists of a high-voltage power supply and a capillary filled with a buffer and with the ends placed in vials containing the same buffer (background electrolyte). The two vials also contain one electrode each to make an electrical contact between the high-voltage power supply and the capillary. Usually the sample is loaded onto one end of the capillary by replacing one of the vials with a vial containing the sample and by applying pressure to the vial. At the opposite end of the capillary, detection can take place directly through the capillary wall. UV detectors are often used.

The separation of solutes by electrophoresis is based on their differences in velocity in an electric field [17]. The velocity depends on the charge and size of the solute and can be calculated according to (6):

$$v = \mu E \qquad (6)$$

where v is the velocity (cm/s) of the ion, µ (cm^2/Vs) is the electrophoretic mobility and E is the applied electric field strength (V/cm). The measured mobility (apparent mobility) can be calculated by:

$$\mu_{app} = \frac{L_{tot} L_{eff}}{t_R U} \quad (7)$$

where L_{eff} is the effective capillary length to the detector, L_{tot} is the total capillary length, t_R is the migration time and U is the applied voltage.

The silanol groups of the fused silica capillary start to dissociate at pH>2, and the capillary wall is increasingly negatively charged at higher pH. The cations in the background electrolyte (BGE) will form a double-layer close to the wall. When the voltage is applied, the cations forming the double-layer are attracted toward the cathode, and an electroosmotic flow (EOF) is formed.

The mobility of the EOF is described by:

$$\mu_{EOF} = \frac{\varepsilon \zeta}{\eta} \quad (8)$$

where ε is the dielectric constant, ζ is the zeta potential (potential difference of the double layer) and η is the solution viscosity.

All solutes will move in the same direction regardless of charge if the EOF is strong enough. Normally the flow is from the anode to the cathode, which means that the cations will be recorded by the detector first, followed by neutral solutes (with the same velocity as the EOF) and lastly the anions.

A solutes effective mobility can be calculated by:

$$\mu_{eff} = \mu_{app} - \mu_{EOF} \quad (9)$$

Capillary zone electrophoresis (CZE) can be used for ionic species, but not for neutral analytes since they will have the same velocity as the EOF. Terabe introduced micellar electrokinetic chromatography (MEKC) in 1984 [18, 19] for the separation of both neutral and charged compounds. Surfactants in a concentration above the critical micelle concentration are used in MEKC. Micelles are formed in the BGE with the hydrophobic tails inside the micelle and with the hydrophilic heads oriented towards the buffer solution. SDS has been widely used as the surfactant. Neutral species are separated by the difference in partitioning in and out of the micelles. Increasing partition into the micelles gives longer migration time. The

elution range (time window) for neutral compounds is between the EOF (no partitioning into the micelles) and the time for the micelles to reach the detector (fully incorporated into the micelles). Charged compounds can have a retention time outside the time window.

The retention factor (the ratio of the total moles of solute in the micelle (=pseudostationary phase) divided by the total moles of solutes in the BGE) is given by:

$$k = \frac{(t_R - t_0)}{t_0(1 - \frac{t_R}{t_m})} \qquad (10)$$

where t_R is the retention time of the solute, t_0 is the retention time of unretained solute moving at the EOF rate and t_m is the micelle retention time.

Microemulsion electrokinetic chromatography (MEEKC) [20, 21] is a technique similar to MEKC, except that the microemulsion has a core of tiny droplets of oil inside the micelles. By using a surfactant and a co-surfactant, the oil droplets are stabilised 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 SDS (surfactant), 6.6% w/w 1-butanol (co-surfactant) and 89.3% w/w 10 mM tetraborate buffer (pH 9.2).

CE can also be used for separation of enantiomers if a chiral selector is added to the BGE. The most frequently used group of selectors is the cyclodextrins [22]. Examples of other types of chiral selectors are macrocyclic antibiotics [23], crown ether [24] and proteins [25, 26]. Cellobiohydrolase Cel7A (previously denoted CBH 1) from the cellulose-degrading fungus *Trichoderma reesei* is a protein that is used as a chiral selector in CE. Cel7A absorbs ultraviolet light and causes disturbances in the UV detection of the analytes. This problem is avoided by using the partial filling technique [26]. A protein plug is injected first followed by the injection of the analyte. When the electric field is applied, the protein and the analyte migrate in opposite directions, which allows them to interact. As a consequence, the protein will not reach the detector at the cathodic end.

3.3 Equations

The following equations were used in papers I–V:

3.3.1 Efficiency, plate height and symmetry

The equation in USP24 [27] (half-height equations) was used to compute the number of theoretical plates (papers III–V):

$$N = 5.54\left(\frac{t_R}{w_{0.5}}\right)^2 \qquad (11)$$

where t_R is the migration time and $w_{0.5}$ is the peak width at half the peak height.

The plate height was calculated by (paper IV):

$$H = \frac{L_{eff}}{N} \qquad (12)$$

where L_{eff} is the capillary length to the detector and N is the efficiency.

The symmetry is a factor describing the shape of a peak and was calculated as a pseudomoment by the integrator using a number of moment equations [28] (paper III):

$$m_1 = a_1\left(t_2 + \frac{a_1}{1.5H_f}\right) \qquad (13)$$

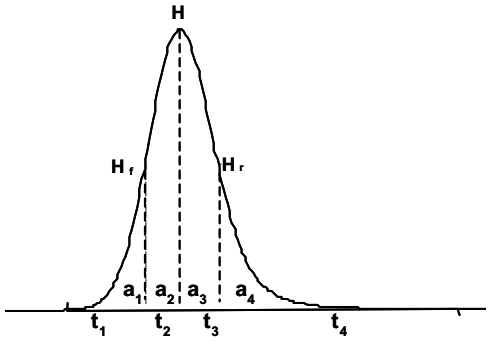
$$m_2 = \left(\frac{{a_2}^2}{0.5H_f + 1.5H}\right) \qquad (14)$$

$$m_3 = \left(\frac{{a_3}^2}{0.5H_r + 1.5H}\right) \qquad (15)$$

$$m_4 = a_4\left(t_3 + \frac{a_4}{1.5H_r}\right) \qquad (16)$$

$$Symmetry = \sqrt{\frac{m_1 + m_2}{m_3 + m_4}} \qquad (17)$$

where H_f = height of the front inflection point, H_r = height of the rear inflection point, H = height at the apex, a_i = area of slice and t_i = time of slice.



If no inflection points were found, the peak symmetry was calculated as follows;

$$Symmetry = \frac{a_1 + a_2}{a_3 + a_4} \qquad (18)$$

where a_1 and a_2 denote the areas of the slices situated at the left of the apex, while a_3 and a_4 are the areas of the slices situated at the right of the apex. When the peak is Gaussian, the symmetry factor becomes 1. A symmetry factor <1 describes a tailing triangular peak, while a symmetry factor >1 describes a leading triangular peak. This applies for both equations (17, 18).

3.3.2 Mobility

Apparent mobilities were calculated according to (papers III and IV):

$$\mu_{app} = \frac{L_{tot} L_{eff}}{t_R U} \quad (19)$$

where L_{tot} is the total length of the capillary, L_{eff} is the effective length of the capillary, t_R is the migration time of the analyte and U is the voltage.

Effective mobilities were calculated according to (papers III and IV):

$$\mu_{eff} = \mu_{app} - \mu_{EOF} \quad (20)$$

where μ_{EOF} is the electroosmotic mobility.

The quotient between the effective mobility of the analyte in the microemulsion and the effective mobility of the analyte in the corresponding buffer system (without the microemulsion) was used as a new response, since the retention factor according to equation (24) could not be used for anions in the MEEKC system (paper IV):

$$q = \frac{\mu_{eff}}{\mu_{eff,aq}} \quad (21)$$

3.3.3 Retention factor and migration factor

Retention factor for LC was calculated according to (papers I and II):

$$k = \frac{t_R - t_0}{t_0} \quad (22)$$

where t_R is the retention time and t_0 is the time for an unretained peak in the LC system.

Two different equations for the retention factor, k (k_T, k_C), were used in MEEKC, employing the same theory as for MEKC. Equation (23) is the retention factor (k_T) for neutral analytes in MEKC derived by Terabe [18], where t_0 is the time for an unretained substance (EOF) and t_m the time for the micelles to reach the detector. Equation (24) can be used for calculation of the retention factor (k_C) for charged analytes in MEKC [29-31], where μ_{eff} is the effective mobility of the charged compound in the MEKC system, $\mu_{eff,aq}$

is the effective mobility of the charged compounds in the pure buffer solution without any micelles and μ_m is the mobility of the micelles. Using equation (24) for cations, it is advisable to employ the effective mobility ($\mu_{eff,\ aq}$) of the cations in a buffer solution containing surfactant monomers below the critical micelle concentration (CMC) to compensate for eventual ion-pair interaction [32, 33]. Equations (23) and (24) also apply to MEEKC, t_m and μ_m then being the migration time and the mobility of the oil droplets, respectively:

$$k_T = \frac{t_R - t_0}{t_0(1 - t_R / t_m)} \tag{23}$$

$$k_C = \frac{\mu_{eff} - \mu_{eff,aq}}{\mu_m - \mu_{eff}} \tag{24}$$

The migration factor defined according to equation (25) was used in paper V:

$$k_M = \left(\frac{t_R - t_0}{t_0} \right) \tag{25}$$

3.3.4 Selectivity, apparent selectivity, relative mobility difference and resolution

Equation (26) was used to calculate the selectivity (α) in LC (papers I and II):

$$\alpha = \frac{k_2}{k_1} \tag{26}$$

where k is the retention factor according to equation (22).

The selectivity in MEEKC was calculated according to (paper V):

$$\alpha = \frac{k_{M,2}}{k_{M,1}} \tag{27}$$

where equation (25) is used for calculation of k_M.

The equation below was used for selectivity between enantiomers (α) in CE (paper III):

$$\alpha = \frac{\mu_{eff\,1}}{\mu_{eff\,2}} \qquad (28)$$

where μ_{eff1} is the effective mobility of the first enantiomer and μ_{eff2} is the effective mobility of the second enantiomer.

If no EOF was measured, the apparent selectivity in CE (paper III) was calculated instead:

$$\alpha^* = \frac{\mu_{app1}}{\mu_{app2}} \qquad (29)$$

where μ_{app1} is the apparent mobility of the first enantiomer and μ_{app2} is the apparent mobility of the second enantiomer.

Another way of defining the separation between the enantiomers is the relative mobility difference (paper III):

$$\mathrm{RMD} = \Delta\mu / \overline{\mu} \qquad (30)$$

where the difference in mobility between the enantiomers is divided by the average mobility.

The resolution was calculated using equation (31) (paper V):

$$R_S = \frac{1.18(t_{R,2} - t_{R,1})}{(w_{0.5,1} + w_{0.5,2})} \qquad (31)$$

where t_R is the migration time and $w_{0.5}$ is the peak width at 50% of the peak height.

4 Chemometrics

Chemometrics, coined by Svante Wold in 1972, is a mathematical/statistical tool for extracting information from large tables of measured chemical data or for designing chemical experiments to gain maximum information out of as few experiments as possible [34-40].

4.1 Experimental design

When using statistical experimental designs (examples below), several parameters (factors) are investigated simultaneously in a predefined scheme. The number of experiments carried out is dependent on how many factors are of interest. Two or three levels of each factor are chosen and a so-called work sheet can be created. The experimental domain is found within the boundaries of all the factors used. Three or more experiments are often made in the centre of the design to obtain information about the repeatability and the occurrence of curvature in the responses (measurements). All the experiments are done in a random order to avoid systematic errors. Different types of designed experiments are used, depending on the chemical problem to be solved. Screening and optimisation designs are described below.

4.1.1 Screening design

In a full factorial design [41-47], experiments using all possible combinations of all the factors at the predefined levels are performed. The number of experiments needed is 2^k, where k is the number of factors chosen (4, 8, 16, 32 experiments for 2, 3, 4, 5 factors, respectively). A three-level full factorial design with two factors was used in paper V. If several factors are to be investigated, a full factorial design is perhaps not an option; fractional factorial designs [48-54] can then be used instead. Here, the number of experiments is reduced, which means that not all possible combinations of the factors are investigated. Depending on the reduction, the design will have different resolutions (III or IV). A full factorial design has

the resolution V. If the design has resolution III, the main effect of the factors will be confounded with the interaction terms. Resolution IV is a better design since the main effects are not confounded, although the interaction terms are confounded with each other. Designs with resolution III can be used for robustness testing of methods, but are not suitable for method development or optimisation. Designs with resolution IV are suitable for the screening of several factors. Plackett Burman designs are similar to strongly reduced fractional factorial designs. A maximum of 7 factors can be investigated in 8 experiments, or 15 factors in 16 experiments. Ref [55] investigated 4 factors with 8 experiments when optimising the BGE. Furthermore, eight factors were explored with 15 experiments by Mikaeli et al. [45]. A D-optimal design is like a fractional factorial design, but with additional experiments to resolve some of the interaction terms. This type of design is used when several factors are investigated so that it is not possible to use a full factorial design, but at the same time some of the interaction terms are of interest. Thorsteinsdottir et al. investigated 8 factors and 6 interaction terms by a D-optimal design [54]. A fractional factorial design with $2^{6-2}+6=22$ experiments (resolution IV) was used in paper IV.

4.1.2 Optimisation design, response surface methodology designs (RSM)

After the screening phase it might be necessary to continue with a central composite design [43-45, 48, 50, 53, 56-67]. In such a design (CCF: central composite face-centred design or CCC: central composite circumscribed design) two or three factors are investigated at several levels. These designs make it possible to investigate whether any curvature exists in the response. In papers I–III, a central composite design (CCF) with three factors was used.

4.2 Modelling

The data from the statistical experimental design can be fitted by means of MLR (multiple linear regression) and the responses can be described by a polynomial function [37]. For a screening design, the polynomial can consist of the following terms:

$$Y = \beta_0 + \sum \beta_i x_i + \sum \beta_{ij} x_i x_j + \varepsilon \qquad (32)$$

For a central composite design, the polynomial includes the following terms:

$$Y = \beta_0 + \sum_1^k \beta_i x_i + \sum_1^k \beta_{ii} x_i^2 + \sum_{i<j} \sum \beta_{ij} x_i x_j + \varepsilon \quad (33)$$

where $x_i \ldots x_k$ are the factors included in the model and $\beta_i \ldots \beta_k$ the regression coefficients that are estimated by the MLR model. The interaction between the factors is described by $\beta_{ij} x_i x_j$, and curvature by $\beta_{ii} x_i^2$.

For each new model, all terms were included. The fraction of variation of the response that can be explained by the model (R^2=(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^2=(1− prediction residual sum of squares/total sum of squares)) were then examined. For a good model, R^2 and Q^2 should be as close to 1 as possible. The model estimated the coefficients ($\beta_i \ldots \beta_k$), which represent half the effect of a factor. The confidence interval of each coefficient (95% of confidence) was studied to see if the factors had any effect on the responses. Some of the coefficients (interaction terms and quadratic terms) that did not have a significant effect were then removed from the model, and a new model was made. If R^2 and Q^2 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. Logarithmic transformation of the responses can sometimes improve the models.

The data from the statistical experimental design can also be fitted by means of PLS (partial least squares projections to latent structures) [68, 69] and the responses can be described by the same polynomial function as when MLR is used (equations 32 and 33). The difference between MLR and PLS is that PLS fits a model simultaneously of all the responses, while MLR fits all the responses separately. PLS accepts a small amount of missing data (<10%), which is not the case for MLR. Furthermore, PLS is the method of choice for fitting if covariance exists between the responses. PLS finds the relationship between a matrix Y (the responses) and a matrix X (factors) expressed as:

$$Y = XB + E \quad (34)$$

The PLS regression coefficients (B) are identical to the coefficients calculated by multiple regression if there is only one response. PLS creates new variables called X-scores according to $t_a=Xw_a$, where w_a are called the weights. The X-scores are then used to model the responses (Y). The Y-variables are combined to few Y-scores (u_a) using weights c_a, $u_a=Yc_a$. The PLS estimations is done so that it maximises the correlation in each model dimension between t_a and u_a. For each PLS component (number a), X-scores (t_a), Y-scores (u_a), X-weights (w_a) and Y-weights (c_a) are obtained. The PLS model can be described mathematically by equations (35-37) and geometrically by Figures 3 and 4.

$$X = 1 * \overline{X} + TP' + E \quad (35)$$

$$Y = 1 * \overline{Y} + UC' + F \quad (36)$$

$$B = W(P'W)^{-1}C' \quad (37)$$

where T and U are the matrix of scores in X- and Y-space, respectively, P' is the matrix of loadings showing the influence of the X-variables, C' is the matrix of weights expressing the correlation between Y and T (X). W is the matrix of weights expressing the correlation between X and U (Y), E and F are the matrix of the residuals, and B is the matrix of the coefficients.

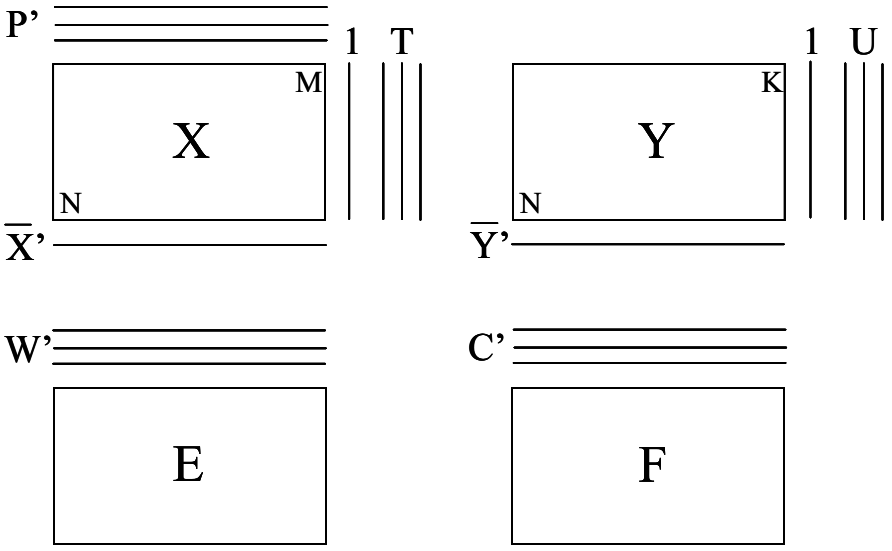


Figure 3: Matrix relationship of PLS. T and U are the matrix of scores in X- and Y-space, respectively, P' is the matrix of loadings showing the influence of the X-variables, C' is the matrix of weights expressing the correlation between Y and T (X), W is the matrix of weights expressing the correlation between X and U (Y), and E and F are the matrix of the residuals.

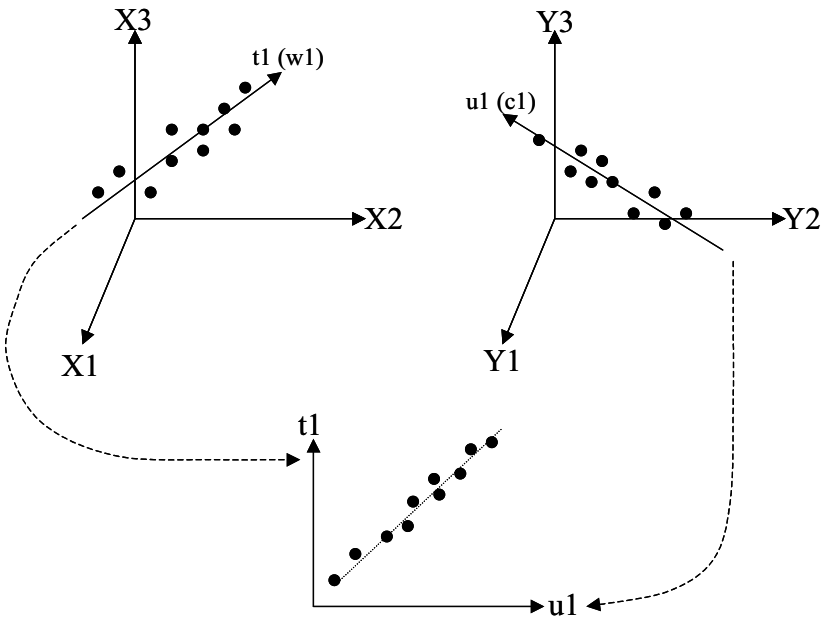


Figure 4: Geometrical interpretation of the PLS procedure.

The scores for the first component in X (t_1) plotted against the scores for the first component in Y (u_1) shows the correlation between X and Y (Figure 4). Plotting the scores from the two first components from the X space (t_1, t_2) displays how the observations are situated when projected onto a plane. This was made in paper V when molecular descriptors of different compounds were fitted to retention data from MEEKC with PLS. The score plot showed how the compounds with similar structure grouped together. Plotting the PLS weights (w and c), also called loadings, in the same graph will give information about the relationship between X and Y variables. X variables close to zero have a small impact on the responses, while X-variables far from zero influences the responses. A positive correlation between X and Y exist if they are situated on the same side in the plot, and a negative correlation if an X-variable is on the opposite side of the responses in the plot. An increase of the X-variable will increase the response if there is a positive correlation, while an increase of an X-variable will decrease the response if the correlation is negative. A loading plot (wc-plot) from paper V is shown in Figure 21.

MLR was used in papers I, III, IV and V and PLS in paper II for modelling of retention and peak shape data. By using PLS instead of MLR (paper II), a loading plot could be generated. An unknown degradation product (kI) came close to the main peak (k27) in the loading plot (Figure 5), and combined with the coefficient plots of both k27 and the unknown, it could be concluded that the unknown compound had a similar structure compared to k27 and that both had the same charge (1+). PLS was used in paper V to fit molecular descriptors of different compounds with retention data from MEEKC.

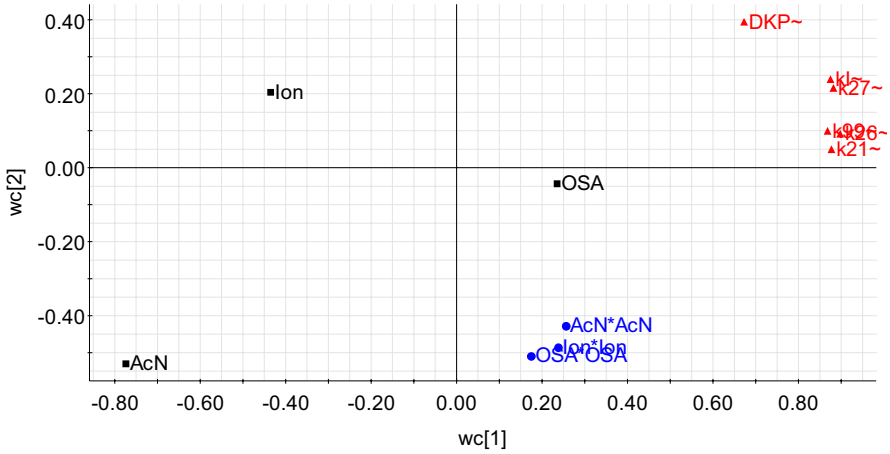


Figure 5: Loading plot from a PLS model (paper II). The amount of acetonitrile (AcN), the buffer concentration (Ion) and the amount of octane sulphonate (OSA) were varied in a central composite design. The responses used in the modelling were the logarithm of the retention factors of six different compounds (see Appendix II for molecular structures).

4.3 Choice of factors

Factors that can be used for method development in LC are, for instance, the pH, the type and concentration of buffer, the type and concentration of organic modifier (acetonitrile, tetrahydrofurane, methanol), the mixture of organic modifiers, the temperature, the flow, gradient and additives (counter ion or competing reagents) [41, 48, 49, 56-62, 70-72]. The pH, the amount of the counter-ion tetrabutylammonium and the steepness of the gradient were investigated in a central composite design in paper I for optimisation of the separation of erythromycin and related compounds. The factors used for method development and optimisation in paper II were the buffer concentration and the amounts of acetonitrile and octane sulphonate.

Factors that often are used for method development and optimisation in CE, CEC, MEKC and MEEKC are the pH, the concentration of the buffer component, the concentration of SDS, the field strength and temperature. Other relevant factors may be injection time, sample stacking, ionic strength, concentrations of acetonitrile, methanol, 2-propanol, cyclodextrin, sodium deoxycholate, Brij 35, urea, sodium heptyl sulphate and imidazole and type of cyclodextrin [42-47, 50-55, 63-67]. In paper III, the factors pH, ionic strength and the amount of acetonitrile were chosen for optimisation of chiral separation of propranolol with Cel7A as chiral selector. The amounts

of SDS, Brij 35, 1-butanol and 2-propanol, the buffer concentration and the temperature were the factors used in a screening design for eight different analytes in MEEKC (paper IV). The effect of concentration variations of SDS and 2-propanol in the MEEKC method were further investigated in a three-level full factorial design analysing 29 different compounds (paper V).

It is important to determine which factors are the most significant in order for the method to be optimised before performing a central composite design. Such a design should be done with a minimum of factors (2–3), since more factors will generate designs with a large number of experiments. Information can be gathered by literature search, by performing pre-experiments (one variable at a time) or by carrying out a screening design with several factors. The LC method in paper I was further developed and optimised compared to a method found in the literature. Several pre-experiments were done in paper III, where the pH, the type and concentration of different buffers and the type and concentration of different organic modifiers were investigated for the separation of R- and S-propanolol in CE with Cel7A as chiral selector. The choice of the factors (pH, concentration of bis-tris acetate buffer and of acetonitrile) in the central composite design was based on these pre-experiments. A large number of parameters in MEEKC can be manipulated, such as the type and concentration of the oil, buffer, surfactant, co-surfactant, organic solvent and ion-pair reagent and the pH. Furthermore, changing instrument parameters such as the temperature and the voltage can have an effect on the separation. A screening design with six parameters was performed to evaluate which factors had the largest effect on the migration times of eight test substances in MEEKC (paper IV). As a result two factors (SDS, 2-propanol) were then chosen for the three-level full factorial design (paper V).

For totally unknown samples, the method development is initially performed continuously. This is due to the fact that more information is gained with time about the degradation path of a substance. Furthermore, changes in the manufacture of the substance or the formulation can give new peaks that could disturb the analysis. By having information on how and by how much the different factors affect the separation or the selectivity, further method development is easier to perform. Knowledge about the robustness of the method is also gained from the experimental design.

Interpretation of the effect of one factor in LC or CE can be difficult, since a change in one factor may affect other factors. For instance, changing the pH in LC may alter the charge of an analyte, which in turn affects the degree of ion pairing with a counter ion added to the mobile phase. Adding an organic modifier to the BGE in CE may change the viscosity and the pH of the BGE. A change of viscosity and pH may in turn affect the magnitude of the EOF. This also happens when changing one factor at a time instead of

varying them in a multivariate way. When an effect from each factor used in an experimental design is evaluated, the analyst should bear in mind that a part of the effect is a consequence of secondary effects. If possible, the experiments should be planned in such a way that "secondary" effects are minimised. In paper III, Cel7A was used as chiral selector to separate R- and S-propranolol. The partial filling technique was used since the selector absorbs ultraviolet light, which causes disturbances in the UV detection of the analytes. A long plug with a low concentration of the selector was used to minimise the change in conductivity between the two zones. Furthermore, the application time was adjusted depending on the viscosity of the BGE (longer application times for higher viscosity) so that the same amount of selector with the same plug length was obtained in each experiment.

5 Effect of factors from the experimental designs

Method development and optimisation may be performed using experimental designs without reflecting on the underlying processes of the factors. However, knowledge and consideration of the processes are essential for the relevant choice of factors in the screening phase, and also later in the optimisation phase. A summary of the effect of the factors used in papers I through V is given below.

5.1 Papers I and II (LC)

The pH (6.5 – 7.5), the concentration of TBA (0.002 – 0.022 M) and the steepness of a gradient with acetonitrile (1.33 – 1.68) were investigated in paper I. The modelled responses were the retention factor (eq. 22) for erythromycin A (EA) and related compounds EC, MEA, AEA, EB, UK, EEA and diethylphthalate (plo=placebo) from the formulation. The retention factor increased with the pH for all substances except diethylphthalate, a neutral compound. Since the pK_a value of EA, and probably also its related substances, is approximately 8.6, the fraction of positively charged molecules will decrease with increasing pH. Thus, the molecules will interact more with the stationary phase at increasing pH, resulting in higher k-values (Figure 6). TBA will compete with the analytes for sites on the stationary phase according to equation (4). An increase in the concentration of TBA will result in a decrease in k for all peaks related to erythromycin. The k for diethylphthalate was affected negatively with an increasing concentration of TBA, but not as much as for EA and related compounds, indicating a difference in retention behaviour between the cationic erythromycins and the neutral diethylphthalate. The interaction between the pH and the concentration of TBA was not significant for all the responses. The gradient had a negative effect on all responses. A decrease in k was observed when changing the gradient from 1.33 to 1.68 (from a flatter to a

steeper slope). However, the effect varies between the different compounds, showing that the gradient also has an impact on selectivity.

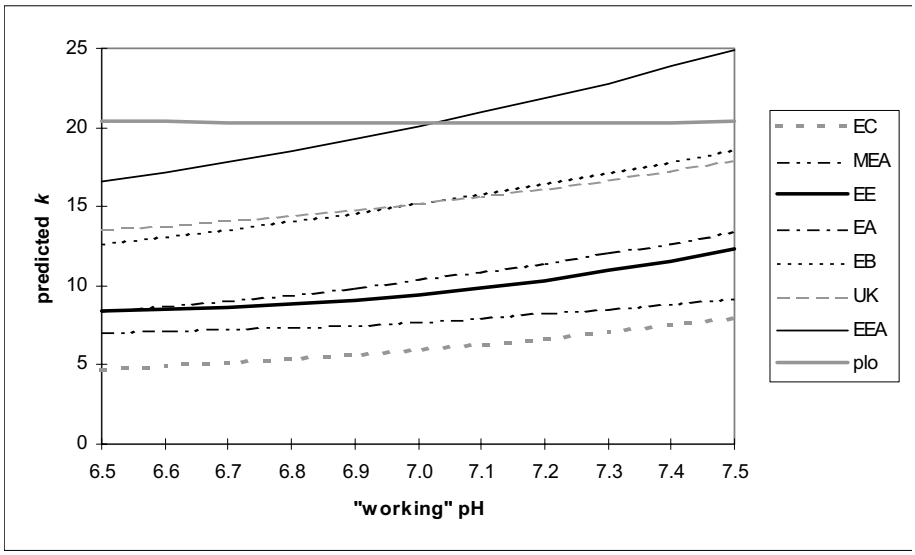


Figure 6: Model predictions of the retention factors as a function of "working" pH in LC (paper I). The other factors were set at constant levels, 0.012 M TBA and gradient 1.33. EC: erythromycin C, MEA: N-demethylerythromycin A, EE: erythromycin E, EA: erythromycin A, EB: erythromycin B, UK: unknown, EEA: erythromycin A enol ether, *plo=placebo*: diethylphthalate.

In paper II, the amount of acetonitrile (20 – 30%), the buffer concentration (0.01 – 0.10 M) and the amount of OSA (1 – 5 mM) were investigated in a central composite design. The retention factors (eq. 22) for H314/26 (k26), H314/27 (k27), H314/21 (k21), H299/87 (k99), diketopiperazine (DKP) and an unknown peak (kI) were calculated and used as responses in the model. The only factor that had any effect on DKP was the amount of acetonitrile (negative effect). This was expected, as the buffer concentration and the concentration of OSA did not have any effect on DKP, which is uncharged at pH 2.6. Both factors (the amount of acetonitrile and buffer concentration) had a negative effect, and the concentration of OSA had a positive effect on the retention factors. At higher buffer concentrations the molecules have less access to the stationary phase, possibly due to increasing competition from the buffer co-ion for the adsorption sites, and will therefore elute faster. When the mobile phase contains more OSA, the ion-pair distribution will increase, resulting in higher retention of the analytes. The effect of OSA was larger for compounds k21 and k99 compared to the other compounds (k26

and k27), since they were the most charged analytes (2+). It was assumed that the unknown analyte (kI) had the same charge as k26 and k27 (1+), given that the effect of the factors was of the same magnitude for all of them (Figure 7).

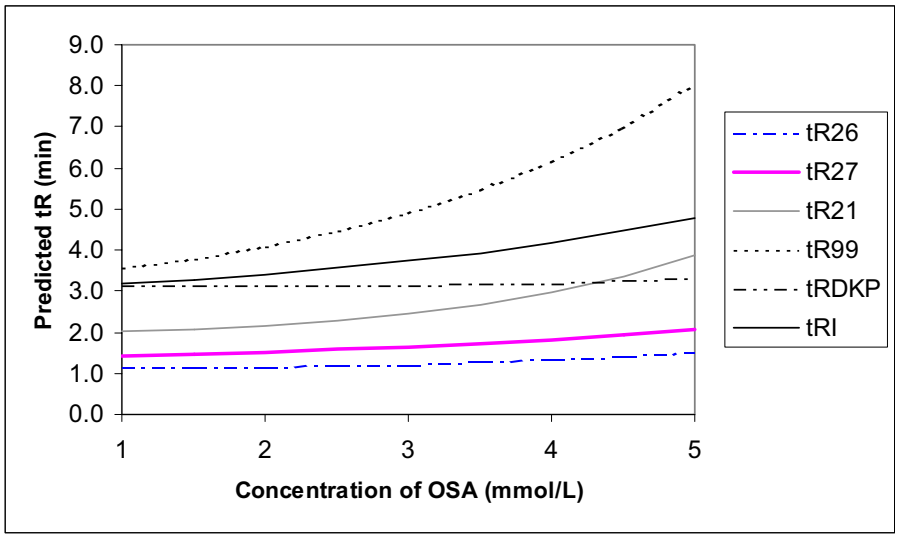


Figure 7: Predicted retention times of all analytes when varying the concentration of octane sulphonate (OSA) in LC (paper II). The concentration of acetonitrile (AcN) was set constant to 20%, and the buffer concentration (Ion) was 0.055 M.

5.2 Paper III (CE with chiral selector)

Cellobiohydrolase Cel7A (previously denoted CBH 1) from the cellulose-degrading fungus *Trichoderma reesei* has successfully been used as a chiral selector in liquid chromatography for separation of enantiomers [73-75]. In 1993 Valcheva et al. [26] introduced both Cel7A as the chiral selector and the partial filling technique in capillary electrophoresis. Cel7A has a molecular mass of 64000 and an isoelectric point (pI) of 3.9 [76]. The three-dimensional crystal structure of the Cel7A catalytic domain [77-79] reveals a 50-Å long cellulose-binding tunnel [77, 78]. Three carboxylate residues (E212, D214 and E217) in the tunnel are crucial for both the catalysis and the chiral recognition mechanism [79]. A recently reported X-ray structure of a complex between (S)-propranolol and Cel7A [80] proves that the product-binding site and the enantioselective binding site for the chiral compound overlap.

The pH (5.0 – 7.0), the ionic strength (0.01 – 0.02) and the amount of acetonitrile in the BGE (1 – 19 %) were investigated in paper III (CE). Several responses were modelled: the effective mobility, the efficiency and the symmetry of the R- and S-propranolol peaks. The same responses were used for rac-propranolol, when no selector was present in the system. The selectivity, the relative mobility difference, the mobility difference and the resolution between the enantiomers were also studied.

The effects of the different factors on the effective mobility are shown in Figure 8. Increasing the pH decreased the effective mobilities not only due to a generally stronger binding to the selector protein migrating in the opposite direction, but also because the selector migrates faster in that direction as the pH increases. A small decrease in mobility for rac-propranolol without selector was observed. Furthermore, a quadratic effect of the pH (pH*pH) was significant only for rac-propranolol. The ionic strength had a large negative effect on rac-propranolol mobility in the absence of the selector, but small positive effects for both (R)- and (S)-propranolol when the selector was present. In the latter case the weaker binding to the selector more than compensates for the typical decrease in free mobility as the ionic strength is increased. Increasing the content of acetonitrile in the BGE decreases the mobilities. The largest effect could be found for rac-propranolol in the absence of the selector and for (R)-propranolol when the selector was present, while the effective mobility of (S)-propranolol was not affected significantly. This is probably due to a trade-off between influence on free mobility and on the interaction also in this case.

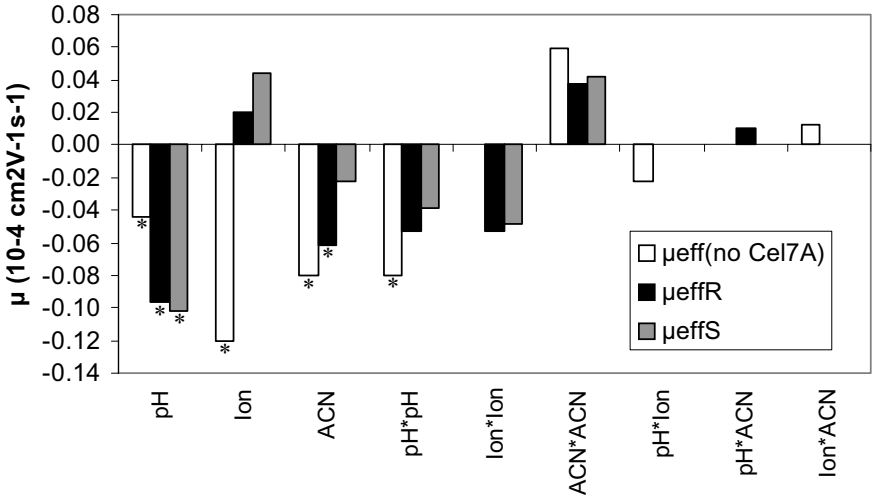


Figure 8: Effect plot showing the change of the effective mobility of (R)- and (S)-propranolol and *rac*-propranolol in the absence of Cel7A when each factor was increased from a low to a high level. Significant effects are marked with an asterisk.

Four different responses for the separation between the peaks were evaluated (resolution, selectivity, relative mobility difference (RMD) and mobility difference between enantiomers (delta μ)). For all models, the same tendency was seen for the effect of the factors. The pH was not significant, but the quadratic term pH*pH was significant and negative, indicating that the response is not linear and has an optimum between pH 5 and 7. Regardless of the response used, both ionic strength and acetonitrile decrease the separation between the enantiomers, demonstrating that both these factors contribute to the interaction with the selector (Figure 9).

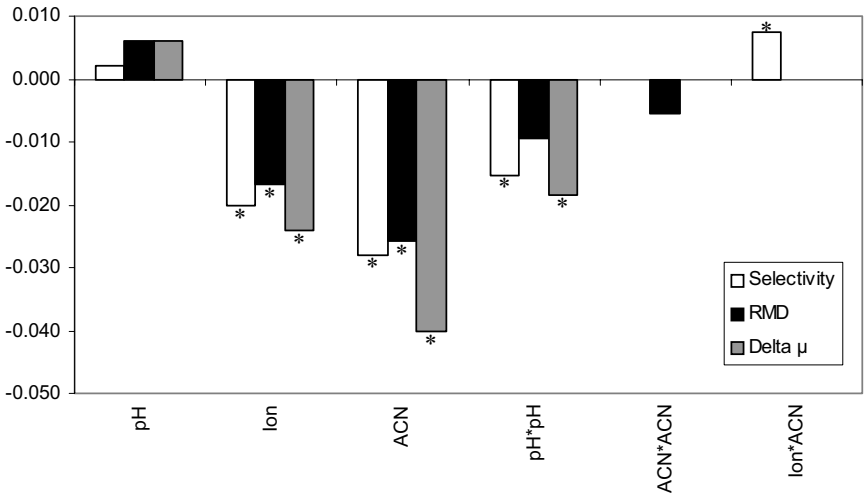


Figure 9: Effect plot showing the change of the selectivity, the relative mobility difference (RMD) and the effective mobility difference between (R)- and (S)-propranolol when each factor was increased from a low to a high level. Significant effects are marked with an asterisk.

The efficiency increased by increasing concentration of acetonitrile being the most important factor for (R)- and (S)-propranolol, while increasing the pH decreased the efficiencies. Both factors affected the (S)-propranolol peak more than the (R)-propranolol peak. Similar effects of pH and concentration of acetonitrile were also found for rac-propranolol without the selector.

Generally, the peak symmetries of (R)- and (S)- propranolol and of rac-propranolol were negatively affected by increasing pH and positively affected by an increase in the ionic strength and the concentration of acetonitrile. The fact that the factors have similar effects on rac-propanolol in the absence of selector as on the enantiomers in its presence may indicate that the effects originate from the buffer properties. However, the changing conditions in the selector zone may still be dominant for the enantiomers but work in the same direction as the buffer effects. The conditions are very complex but some conclusions may be drawn from the results obtained. In chiral separations the most retained peak is most often more asymmetric (tailing) than the less retained. The reason has been interpreted as being a consequence of slow kinetics; the enantiomer with highest association constant has the slowest off rate constant. The ionic strength improved the symmetry for all peaks. In this system an increasing buffer concentration will increase the mobility of the enantiomers due to decreasing binding to the selector, which probably will decrease the tailing, since a smaller fraction of the enantiomer is bound to the selector. Another factor contributing to tailing may be due to mobility mismatch between the buffer ions and the

analyte. An increasing concentration of the buffer will decrease the difference in conductivity between the two zones. The concentration of acetonitrile improved the symmetry of the (S)-propranolol peak more compared to the other peaks. (S)-propranolol has the strongest interaction with the selector. When using the partial filling technique, the analyte must pass different zones: first the selector zone (90% of effective capillary length) and then the remaining zone to the detection window containing BGE without the selector. The analyte will have different mobilities in the different zones, slower in the selector zone due to interaction with the selector, and faster in the BGE zone without any selector. When, for instance, the sample plug with (S)-propranolol starts leaving the selector zone, the molecules in the first part of the plug will start to move with a higher mobility compared to those still interacting with the selector in the selector zone. This might explain why the peak symmetry of (S)-propranolol is poorer compared to the other peaks. Adding acetonitrile to the BGE will decrease the interaction between (S)-propranolol and the selector, and might therefore be the reason for the large positive effect on the peak shape of (S)-propranolol.

5.3 Papers IV and V (MEEKC)

Six factors (SDS, Brij 35, 1-butanol, 2-propanol, buffer concentration and temperature) in MEEKC were investigated in a fractional factorial design (paper IV). The responses investigated were the separation window (t_m–t_0), t_0 (time for EOF marker), t_m (time for oil droplet marker), the retention factor (k_T, k_C (eq. 23, 24)) and the plate height (H).

An MLR model was calculated for each response t_m–t_0, t_0 and t_m, respectively (Figure 10). R^2 varied between 0.949 and 0.993 and the range of Q^2 was 0.820–0.968. The time (t_0) for methanol (EOF-marker) increased with increasing concentration of SDS, Brij 35,1-butanol, 2-propanol and buffer, and by reducing the temperature. 2-Propanol had the largest effect on t_0. The reduction of EOF (increased t_0) by increased concentration of 1-butanol, 2-propanol, and by decreased temperature is expected due to an increased viscosity. A reduced EOF was observed when the buffer concentration was increased due to a decrease of the zeta potential. Brij 35 could reduce the wall charge if adsorbed to the capillary surface leading to a possible explanation of reduced EOF when increasing the concentration of Brij 35. The reduction of EOF (increased t_0) by increasing SDS is unexpected, since this would rather be expected to increase the EOF due to increasing capillary surface charges caused by adsorption of additional SDS. An explanation might be that the zeta potential will decrease due to the

increased ionic strength, and/or that with an increasing SDS concentration most of the SDS will prefer to partition into the microemulsion rather than adsorb to the capillary surface also indicated by the increasing separation window. The amounts of SDS and 2-propanol were the most important factors for the separation window and for t_m. Increasing the concentration of SDS increases those two responses. The effect of an increasing SDS concentration is obvious, since this would increase the charge density of the microemulsion, resulting in increasing mobility of the oil droplets towards the anode, and this effect is larger than the small positive effect on t_0. An increasing amount of 2-propanol gave positive effects on both t_m and t_0, but with a larger effect on t_m resulting in a larger separation window. A similar effect on the separation window was observed by increasing the buffer concentration, which is due to a decrease of the mobilities by shielding of charges at the capillary surface and of the oil droplets. An increase in the concentration of Brij 35 and 1-butanol gives the opposite effect, a reduction of the separation window (t_m-t_0). In contrast to 2-propanol and buffer concentration, those additives reduces t_m, which can be explained by a decrease of the charge density of the oil droplets combined with a reduction of the EOF. An increase of the temperature reduced the separation window due to decreasing viscosity affecting both t_m and t_0 in a negative direction.

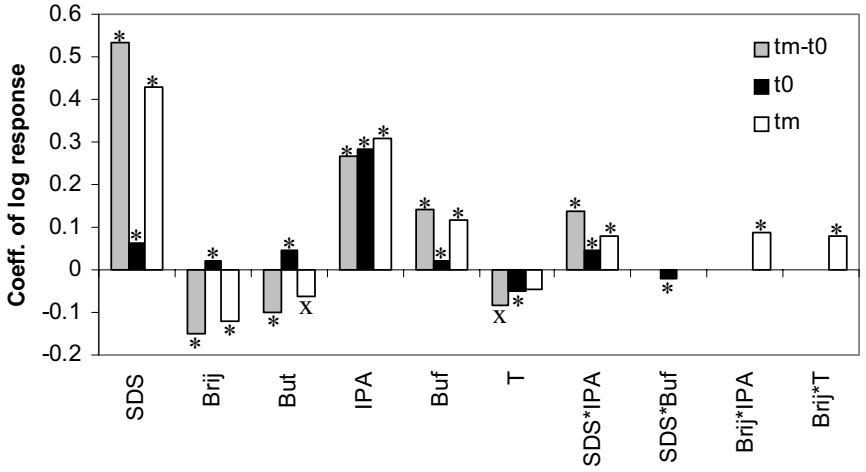


Figure 10: Scaled and centred coefficients from MLR model for responses log (t_m–t_0), log (t_0) (EOF) and log (t_m) (oil droplet marker) (95% of confidence). Significant coefficients are marked with an asterisk and coefficients that are on the brink of being significant are marked with x.

The model using plate heights as responses had medium values of R^2 (0.455–0.900) and Q^2 (0.144–0.568), which means that only a few

conclusions can be drawn. The three most important factors affecting the plate heights were the amounts of SDS and 2-propanol and the temperature (Figure 11). Increasing the amount of SDS increased the plate height for TRI, LID, NOR, NAP and PRO. Furthermore, an increase in 2-propanol will increase the plate height for TRI, LID, SAL and NAP. An elevated temperature will increase the plate height for TRI, LID, NOR and PRO. No factors had any significant effect on the plate height for EST and DIS. An increase in buffer concentration increased the plate height for NAP.

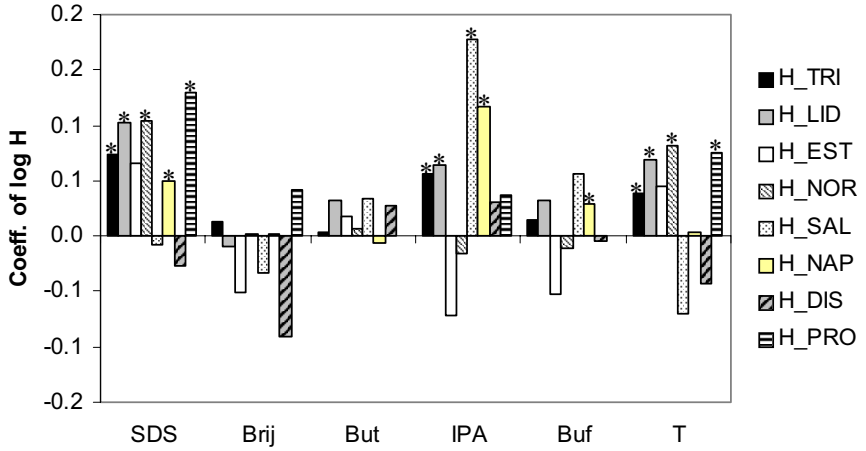


Figure 11: Scaled and centred coefficients from the PLS model for responses log (H) (plate height) (95% of confidence). Significant coefficients are marked with an asterisk. TRI = trimethoprim, LID = lidocaine, EST = estradiol, NOR = norethisterone, SAL = salicylic acid, NAP = naproxen, DIS = disopyramide, PRO = propranolol.

According to the references [81, 82], the total band broadening in the capillary can be described as the sum of the plate heights caused by five factors. This equation applies to both MEKC and MEEKC:

$$H_{tot} = H_l + H_m + H_{aq} + H_T + H_{ep} \qquad (38)$$

where H_{tot} is the overall plate height and H_l, H_m, H_{aq}, H_T and H_{ep} are the plate heights generated by longitudinal diffusion, sorption-desorption kinetics in micellar (microemulsion) solubilisation, intermicelle mass transfer in the aqueous phase, radial temperature gradient effects on the electrophoretic velocity of the micelles (Joule heat effect), and dispersion of electrophoretic mobilities of the micelles, respectively. The longitudinal diffusion decreases

with increasing migration velocity, whereas H_m and H_{aq} will increase with increasing migration velocity, while H_T and H_{ep} should be independent of migration velocity.

To obtain lower plate heights, the concentrations of SDS and 2-propanol should be low, which are the settings of the factors that will lead to higher mobilities of the analytes. Furthermore, a lower temperature would also lead to reduced plate heights. The combination of lower plate heights observed with increasing velocities and with decreasing temperature suggests that the longitudinal diffusion (H_l) is the predominant factor in H_{tot}.

Adding 2-propanol (IPA) to the BGE had the largest effect on the retention factors (k_T, k_C according to eq. 23 and 24 for neutral and positively charged compounds) (Figure 12). Changing the amount of IPA from 0 to 20% w/w decreased the retention factors for all analytes. IPA will mainly be in the aqueous phase of the BGE, although a fraction is distributed to the oil droplets. This will increase the size and thus decrease the charge density of the oil droplet. Cations will therefore have a reduced electrostatic interaction with the surface of the oil droplet. The solubility of hydrophobic analytes in the aqueous phase will increase, resulting in a decreased partitioning to the oil droplet.

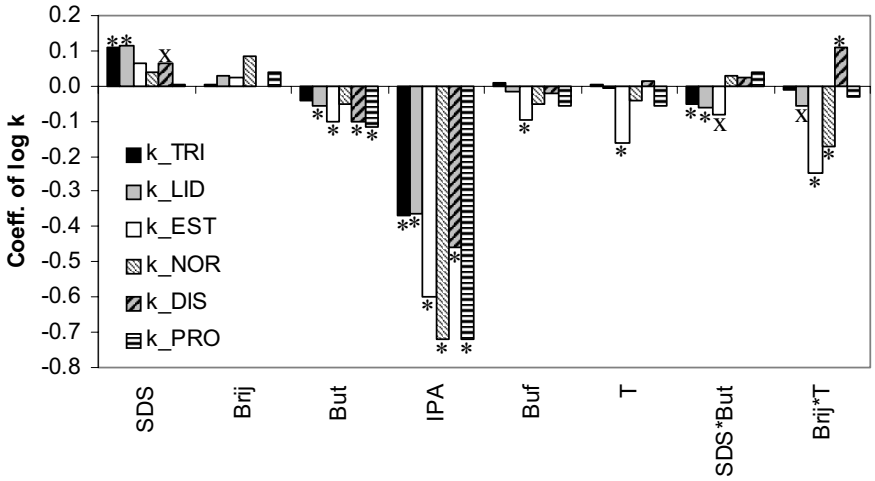


Figure 12: Scaled and centred coefficients from MLR model for responses log (retention factors, k_T, k_C) (95% of confidence). Calculated according to eq. (23) for TRI, LID, EST and NOR. Calculated according to eq. (24) for DIS and PRO. Significant coefficients are marked with an asterisk and coefficients that are on the brink of being significant are marked with x. TRI = trimethoprim, LID = lidocaine, EST = estradiol, NOR = norethisterone, DIS = disopyramide, PRO = propranolol.

The amount of SDS had a significant effect only for TRI and LID. The retention factors increased with the concentration of the surfactant. The factor verged on being significant for DIS. By adding more SDS to the microemulsion, the charge density of the oil droplet will increase, providing not only a stronger electrostatic negative surface area of the oil droplet, but also an increased hydrophobic character of the internal part of the oil due to the nonpolar character of the SDS tails. This is reflected in the tendency for a stronger distribution of the analytes to the microemulsion, although the effect is relatively weak.

Changing 1-butanol from 5 to 9% w/w will significantly decrease the retention factors of the cations (DIS, PRO) and of two of the neutral compounds (LID, EST). Increasing the amount of 1-butanol will increase the size, decrease the charge density and affect the hydrophobicity of the oil droplet. Fewer electrostatic interactions can be observed for the positively charged analytes, and the partitioning of LID and EST will decrease due to changes in the size and character of the oil droplets.

Only the retention factor for EST was reduced significantly with increasing temperature (25 to 40 °C). Higher temperatures may influence the distribution to the microemulsion through thermodynamic effects.

An increased buffer concentration (10 to 50 mM) had only a small negative effect on the retention factor for EST. No other analytes are affected significantly by the buffer concentration. An increase in the buffer concentration will increase the ionic strength, which might influence the distribution equilibrium to the oil phase.

Adding Brij 35 to the microemulsion did not have any significant effect on the retention factor for any of the analytes. A significant effect of Brij 35 might be observed if the factor had been varied over a wider range. The tendency for Brij 35 is to have a positive effect on the retention factors of all analytes except DIS. Adding Brij 35 to the microemulsion will in this case slightly decrease the negative charge density of the oil droplets, so the deviating behaviour of the most positively charged analyte, DIS, is natural.

The interaction term SDS*But was significant for TRI and LID and almost significant for EST. This interaction is confounded with Brij*But, so the magnitude of the term is a combination of the two terms. The situation is similar for the interaction term Brij*T, which is combined with But*IPA. Significant effects of this term were found for EST, NOR and DIS.

Using equation (24, k_C) to calculate the retention factors of anions (SAL, NAP) resulted in negative values for some of the experiments in the factorial design. Furthermore, for SAL and NAP in experiment N10 and for NAP in experiment N16, the retention factor could not be calculated, since the analytes and dodecylbenzene (oil droplet marker) had the same effective mobilities. It is obvious that the retention factor is an inadequate response for

negatively charged compounds. The reason is probably the repelling electrostatic forces between the negatively charged oil droplet and the negatively charged compounds. An alternative response for the anions was used instead. The quotient between the effective mobility of the anions in the microemulsion and the effective mobility in the corresponding buffer (or buffer containing IPA) was calculated and used as a response in the factorial design. A quotient close to 1 would imply a mobility of the analytes in the microemulsion that is similar to the mobility in the corresponding buffer system, and would indicate a negligible distribution to the microemulsion. The three factors that dilute the charge of the oil droplets and change its chemical character, i.e. Brij 35 (only for NAP), IPA and But, decrease the ratios for both compounds, indicating increasing distribution to the micelles (Figure 13). Interestingly, however, the effect of SDS is the opposite for the two compounds: increasing the ratio for NAP (less distribution to the emulsion) and decreasing it for SAL (higher distribution). This reflects the different chemical characters of the two compounds as discussed later in this section. Increasing the temperature increased the response for both SAL and NAP, but an increased buffer concentration increased the ratio only for NAP. Three interaction terms were included in the model for SAL and NAP. The largest effect from an interaction between two factors was found for SDS*IPA for both SAL and NAP, but the interaction term SDS*IPA is confounded with Buf*T.

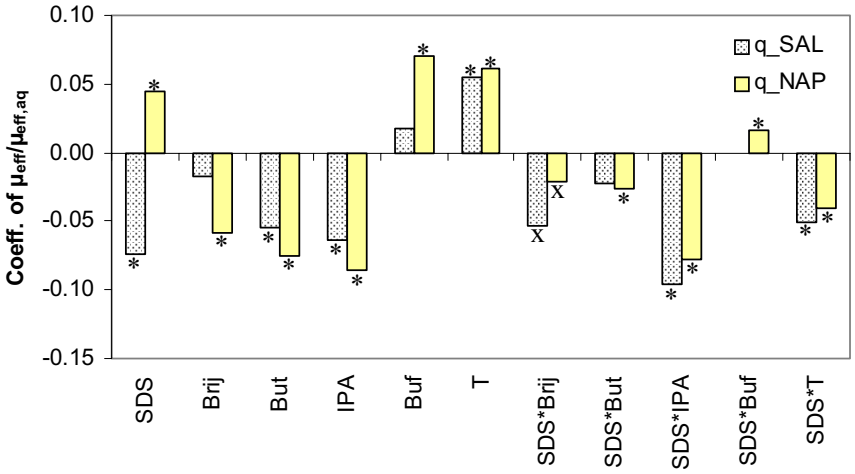


Figure 13: Scaled and centred coefficients from the MLR model for responses $\mu_{eff}/\mu_{eff,aq}$ for SAL and NAP (95% of confidence). Significant coefficients are marked with an asterisk and coefficients that are on the brink of being significant are marked with x. SAL = salicylic acid, NAP = naproxen.

For NAP, a high amount of SDS and buffer concentration and a low level of Brij 35, 1-butanol and IPA gave a higher value of the quotient (μeff/μeff,aq). This means that when the oil droplet has the highest charge density, it will have the maximum electrostatic repulsion to the negatively charged NAP. NAP stays in the aqueous phase and will have a mobility close to the mobility in the corresponding buffer system. The impact of charge for SAL, and consequently the electrostatic repulsion, may be diminished by the possibility of an internal H-bonding, which may give the compound a more neutral character compared to NAP. The negative effect of SDS on the ratio is a further indication in this direction.

The amounts of SDS and 2-propanol were studied in a three-level full factorial design in paper V. One MLR model (with linear, quadratic and interaction terms) was calculated for all the compounds (29) simultaneously, using the logarithm of the migration time as response. Results from the oil droplet marker (dodecylbenzene) and the EOF marker (methanol) were also included in the calculation. The migration times were used as the response since the focus of paper V was more on using different strategies for optimisation of separation than studying the analytes' distribution in the pseudostationary phase (oil droplets in the microemulsion). An increase in the amount of SDS or 2-propanol will increase the migration times of all analytes (hydrophilic, hydrophobic, neutral, negative or positive charge) due to 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 BGE, which also contributes to the total migration time.

6 Responses for optimisation

Several responses can be used for optimisation, depending on the analytical problem. The choice depends on the actual separation problem, and may be the resolution, efficiency, retention factor, migration time etc.

6.1 Number of peaks, number of resolved peaks

When numerous peaks are to be separated, a suitable response may be to count the number of peaks or the number of resolved peaks. Identification of each peak is then not necessary [42, 44, 50]. Alfazema et al. [42] separated 70 peaks of UV-absorbing constituents in urine by MEKC in less than 12 minutes by using the number of peaks as a response in the chemometric optimisation. If fewer peaks are to be analysed, advantages may exist in identifying each peak so that changes in elution order can be observed.

6.2 Noise, efficiency, symmetry

Other responses that can be studied to gain optimal separation methods are the noise [55], the efficiency [54, 55, 63, 64] and the symmetry [43]. By modelling these parameters information about which factors cause a high noise level or poor peak shapes can help the analyst to choose which areas in the experimental domain to avoid. In paper III the optimisation of the separation of propranolol enantiomers with Cel7A as chiral selector was done with a central composite design. The efficiency and the symmetry were optimised together with the resolution and selectivity between the enantiomers, but the factors had opposite effects. High ionic strength, a high amount of acetonitrile and a low pH gave high efficiencies and a good peak shape but affected the resolution and the selectivity. A compromise for the setting of the factors had to be made.

6.3 Retention/migration time, retention factor and mobility

Migration times of first and last peaks or the migration time window were used to evaluate the span of the separation window [paper IV, 43, 51, 55, 63, 65]. Retention factors or retention/migration times were used as responses to study the influence of different factors [papers I, II, IV and V, 54, 63, 64]. Furthermore, in papers I, II and IV predicted retention factors/times from the MLR/PLS model were calculated to study peak movements when changing a factor. The mobility of (R)- and (S)-propranolol was calculated and compared to the mobility of rac-propranolol when no chiral selector was added to the BGE in paper III. In paper IV the retention factor (k_C) could not be calculated for anions, so a new response had to be evaluated, i.e. the quotient between the effective mobility of the anion in the microemulsion and the effective mobility in the corresponding buffer ($\mu_{eff}/\mu_{eff,aq}$).

6.4 Resolution, selectivity, relative mobility difference

The most efficient way of optimising the separation between peaks is to optimise the resolution or the selectivity between them. This has been done by several authors [43, 51-55, 63, 64, 67]. Varesio et al. [43] optimised the relative total resolution of their electrophoretic system by adding the resolution between the two first peaks to the resolution of the two last peaks and dividing it by n−1, where n is the number of analytes separated. The selectivity or the resolution between critical pair of peaks was optimised in papers I, III and V. In addition, the relative mobility difference between R- and S-propranolol (RMD) was used as a response (paper III).

6.5 Chromatographic functions

If several peaks are to be separated, a chromatographic function can be used involving that one value is calculated for each chromatogram/ electropherogram. Some functions take into account both analysis time and resolution (CRF [83-86], CRS [83, 87], CEF [53, 83, 88], COF [67, 83]), while other functions are only focused on the resolution (ATR [45], Rp [89], r [90]). Some of the functions have constants that are set by the researcher giving weights to different parts of the equation. Depending on the importance, weights can be set for the resolution or the analysis time. The

advantage of chromatographic functions is that the whole chromatogram/ electropherogram is taken into account and no peak tracking is needed. The disadvantage is that sometimes the resolution of peaks is important in one part of the electropherogram compared to another, where it may be acceptable with co-eluting peaks.

The chromatographic functions CRF, CRS, CEF, COF, ATR, Rp and r according to equations (39-48) were used for optimisation of separation within five different groups of compounds (paper V). A three-level full factorial design with SDS and 2-propanol as factors in MEEKC was carried out. Within each group, a response was calculated for each experiment in the design using the chromatographic functions. The separation was then optimised by generating response surface plots in MODDE 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.

$$CRF_1 = \sum_{i=1}^{L} R_i + L^{w_1} - w_2 \left| T_A - T_L \right| - w_3 \left(T_1 - T_0 \right) \quad (39)$$

where [84] R_i is the resolution of the ith 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_1. A modified equation of CRF_1 was therefore tested, where all resolution values exceeding 3 were not included in the sum (first term).

$$CRF_2 = a \sum \ln \frac{R_{so}}{R_s} + b \sum \ln \frac{R_s}{R_{so}} + c \ln \frac{T_0}{T} \quad (40)$$

$$CRF_3 = a \sum \ln \frac{R_{so}}{R_s} + b \sum \ln \frac{R_s}{R_{so}} \quad (41)$$

where [85, 86] R_{so} is the optimum resolution (1.5), R_s is the resolution between two neighbouring 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 [85, 86], 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 was 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.

$$CRS = \left\{ \sum_{i=1}^{n-1} \left[\frac{(R_i - R_{opt})^2}{R_i (R_i - R_{\min})^2} \right] + \sum_{i=1}^{n-1} \frac{(R_i)^2}{a\overline{R}^2} \right\} \frac{T_f}{n} \quad (42)$$

where [87] R_i is the resolution of the ith peak, $\overline{R}$ 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 minimised. A modified equation of CRS, where T_f/n was set equal to 1, was also used and compared to the original equation.

$$CEF = \left[\left(\sum_{i=1}^{n-1} \left(1 - e^{a(R_{opt} - R_i)} \right)^2 \right) + 1 \right] \left[1 + \frac{t_f}{t_{\max}} \right] \quad (43)$$

where [53, 88] R_i is the resolution of the ith peak, R_{opt} is the desired optimum resolution (1.5), t_f is the migration time of the final peak, t_{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 minimised. A modified equation of CEF, where the last term $(1+t_f/t_{max})$ was excluded, was also used and compared to the original equation.

$$COF = \sum_{i=1}^{n} A_i \ln(R_i / R_{id}) + B(t_m - t_n) \quad (44)$$

where [67] R_i is the resolution of the ith pair and R_{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 optimisation and was therefore used.

$$F(R_i) = \frac{\arctan\left[a(R_i - b)\right] + \pi/2}{\pi} \quad (45)$$

$$ATR = \sum_{i=1}^{n-1} F(R_i) \quad (46)$$

where R_i is the resolution of the ith pair, the values of a and b being chosen by the researcher. ATR is the sum of the function $F(R_i)$ [45]. 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.

$$R_p = \prod_{i=1}^{n-1} Rs_{i,i+1} \quad (47)$$

$$r = \frac{\prod_{i=1}^{n-1} Rs_{i,i+1}}{\left[\frac{\sum_{i=1}^{n-1} Rs_{i+1,i}}{n-1}\right]^{n-1}} \quad (48)$$

The resolution product (eq. 47) is calculated by multiplying all the resolutions [89]. The relative resolution product (eq. 48) is similar to the resolution product, except that it also takes into account how the peaks are spread in the electropherogram [90]

CRF, CEF, COF, ATR and r have all been used for optimisation of separation in electrodriven techniques. To our knowledge, no articles have been published where chromatographic functions have been used for optimisation of separation in MEEKC.

Table 1 shows a compilation of the different optimisation strategies and the optimum setting of the factors (SDS, 2-propanol) 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_{1-3}, CRF_1 modified, CRS modified, CEF modified, COF modified, ATR, Rp and r), DryLab™ and for optimisation of the selectivity in MODDE. This is due to the fact that the equations (42 and 43) take account of both resolution and total analysis time. Since increased amounts of SDS and 2-propanol 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, r). Furthermore, CRF_1, CRF_1 modified and CRF_2 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 favourable for the separation, more account was taken of the other functions.

Table 1: Optimum settings of variables (SDS, IPA) for best separation of different groups of the analytes in MEEKC. Comparison of different optimisation strategies.

	Group 1		Group 2		Group 3		Group 4		Group 5	
	% SDS	% IPA	% SDS	% IPA	% SDS	% IPA	% SDS	% IPA	% SDS	% IPA
α of critical peak pair	5.0	10.0	5.0	10.0	5.0[c] 2.8[d]	2.0[c] 10.0[d]	5.0	6.0	3.5	10.0
3D 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_1	5.0	7.8	5.0	10.0	5.0	10.0	5.0	6.0	5.0	10.0
CRF_1 modified[a]	5.0	10.0	5.0	10.0	5.0	10.0	5.0	7.3	5.0	10.0
CRF_2	5.0	10.0	4.3[e] 4.3[f]	2.0[e] 10.0[f]	3.8[e] 4.5[f]	2.0[e] 10.0[f]	3.5	6.5	5.0	10.0
CRF_3	5.0	10.0	4.4[e] 4.4[f]	2.0[e] 10.0[f]	3.8[e] 4.5[f]	2.0[e] 10.0[f]	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[b]	5.0	10.0	4.5[e] 3.8[f]	2.0[e] 10.0[f]	4.5[e] 5.0[f]	2.0[e] 10.0[f]	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[b]	2.9[e] 4.7[f]	2.0[e] 10.0[f]	4.3	10.0	3.7[e] 4.3[f]	2.0[e] 10.0[f]	3.8	7.4	5.0	10.0
COF modified[b]	5.0	10.0	5.0	10.0	3.8[e] 4.5[f]	2.0[e] 10.0[f]	4.4	6.7	5.0	10.0
ATR	5.0	10.0	5.0	10.0	4.2[e] 4.2[f]	2.0[e] 10.0[f]	4.4	6.4	4.9	10.0
Rp	5.0	10.0	5.0	10.0	3.5[e] 3.5[f]	2.0[e] 10.0[f]	4.1	5.9	5.0	10.0
r	2.0	10.0	3.4	10.0	5.0[e] 2.0[f]	5.5[e] 6.5[f]	2.8	7.2	5.0	10.0

For optimisation of separation: α, 3D resolution map (DryLab™), CRF_{1-3}, CRF_1 modified, COF, ATR, Rp and r should be maximised and CRS and CEF should be minimised. 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. $\alpha=k_{M,2}/k_{M,1}$, where k_M is the migration factor. IPA: 2-propanol. [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. See Appendix III for molecular structures.

6.6 Response surface or contour plots

Starting with a full factorial design (a model containing linear and interaction terms) or with a central composite design (a model with linear, interaction and quadratic terms), a response surface plot can be constructed with two factors at a time. The x- and y-axes represent the factors, while the response is plotted on the z-axis. Suitable responses include the resolution or the selectivity for a pair of peaks, the retention/migration time of the last peak, the efficiency or the noise. The optimum condition can then be selected in the plot [41, 48, 56-58, 62]. Examples based on this optimisation method from the literature are: retention time window or retention factor of a peptide as a function of acetonitrile and temperature [63], difference in relative migration times plotted against molarity of the buffer and the SDS concentration [65], and the logarithm of the chromatographic function CEF plotted as a function of voltage and pH [53].

Response surface plots were used in paper III, where the symmetry of the (R)- or (S)-propranolol peak was plotted as a function of ionic strength and the amount of acetonitrile. The optimum setting of the factors resulting in a good peak shape was different for the two enantiomers (Figure 14).

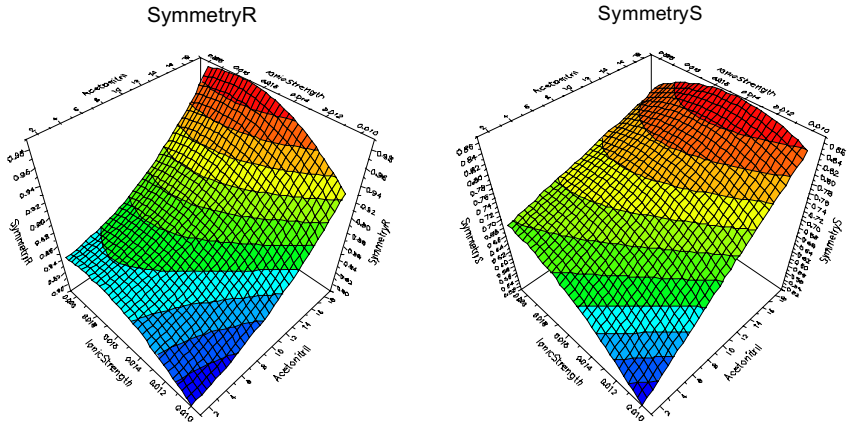


Figure 14: Response surface plot of peak symmetry for (R)- and (S)-propranolol (paper III). Ionic strength vs. concentration of acetonitrile. The pH was set constant at 6.0.

The selectivity between (R)- and (S)-propranolol (paper III) was plotted as a function of ionic strength and pH. A medium level of the pH combined with a low ionic strength gave the highest selectivity (Figure 15).

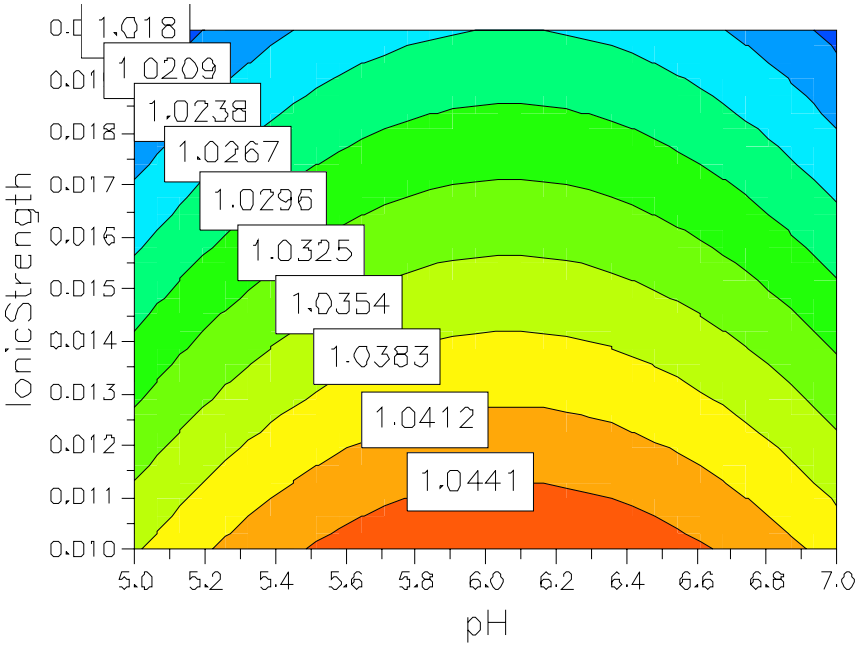


Figure 15: Contour plot of the selectivity between (R)- and (S)-propranolol. Ionic strength vs. the pH. Acetonitrile was set constant at 10% v/v.

Response surface plots were also used to optimise MEEKC separation of five different groups of compounds in paper V. The selectivity or the calculated chromatographic functions were plotted as a function of SDS and 2-propanol.

7 Optimisation of separation using optimiser in MODDE (simplex), α-plot or DryLab™

7.1 Simplex

The simplex optimisation is a sequential search for the optimum. Experiments according to a central composite design or a Box Behnken design are first performed. The predicted response within the experimental domain is then optimised by a simplex search. The coordinates of a triangle (for 2 factors) or a tetrahedron (for 3 factors) are chosen. The point with the lowest response value is reflected to the opposite side to give a new coordinate. The point with the lowest response is again reflected to the opposite side. This process is continued until the optimum setting is found [70-72]. For Alfazema et al. [42] the optimum conditions for an MEKC method were set as a starting point for the simplex calculation, and the search was performed in a narrow area around the optimum settings. Better conditions compared to the starting point were not found in the simplex search. However, Pretswell et al. [66] found a better condition within the experimental domain when performing a simplex search for determination of cations by CZE.

A sequential optimisation with simplex can also be performed without carrying out initial experiments according to a design [91]. If two factors are used, three experiments are first performed. New settings of the factors for the next experiment are found by reflecting the point with the lowest response (worst separation) to the other side. This optimisation strategy has many drawbacks. The process is “blind”; consequently, the analyst does not get the whole picture of the experimental domain. Furthermore, the simplex is often trapped on local optima instead of the global optimum, and a large number of experiments must be carried out.

The software MODDE is used for the set-up of the design and for calculation of MLR and PLS models; however MODDE also has an “optimiser” included in the software, which is based on a Nelder Mead simplex method (no fixed step size) [69, 92]. The responses can be

maximised or minimised within the experimental domain investigated. Eight starting points within the experimental domain are used for the simplex calculation to avoid being trapped at a local minimum/maximum. This optimisation method has an advantage over the response surface plot, since the response is optimised using all factors at the same time. The response surfaces are plotted for two factors at a time and the third factor must be set at one level. Several response surface plots can be made setting the third factor at different levels, but it can be difficult to get an overview and find the optimum if the model contains interaction and quadratic terms. Using the "optimiser" or response surface plots for optimisation of separation in paper V gave the same result since only two factors were investigated (SDS and 2-propanol). One drawback of using the "optimiser" in MODDE is that when optimising the separation for one peak pair, the resolution between others can deteriorate. Optimising several critical peak pairs in the same model can solve this problem.

7.2 Alpha-plot

The separation of six peaks (HPLC) was optimised by using an α-plot (paper II). Three factors (ionic strength, amount of octane sulphonate and of acetonitrile) were first investigated in a central composite design. A PLS model was then calculated using the retention factor of all analytes as responses. Unscaled and uncentred beta coefficients for all solutes generated in the MODDE software were then transferred to an in-house simulation program developed in Matlab™. Two factors were chosen for the α-plot, while the third factor was set to a constant level. For each of the two factors, an interval and ten segments were selected. Two factors with ten segments each will form a grid. At every grid point the retention factors for all solutes were predicted. Subsequently, α ($=k_2/k_1$) was calculated for all possible combinations of retention factor pairs. If the quotient was less than 1, the inverted value was used. A surface plot (α-plot) was then produced by selecting the lowest α at each grid point. This resulted in a non-continuous surface with minimum α, showing the best overall chromatographic separation at the surface peaks and ridges. The α-plot was presented both as a 3D plot (Figure 16) and as a contour plot. By choosing the ridges in the α-plot, the simulation program will produce a chromatogram which shows the best overall chromatographic separation. The setting of the factors at the ridge that gave the shortest run time was chosen as the optimised method. The advantage of optimisation using α-plots is that new α-plots can be generated with a new combination of the factors from the same design. One

drawback is that only two factors can be plotted at a time; however, the α-plot is superior compared to a response surface plot since the selectivity is based on calculation from a combination of all peaks, while only one response can be chosen in a response surface plot. Another drawback is that new Matlab™ codes must be written for each new MLR/PLS model calculated, depending on which terms are included in the model (linear, interaction and quadratic terms).

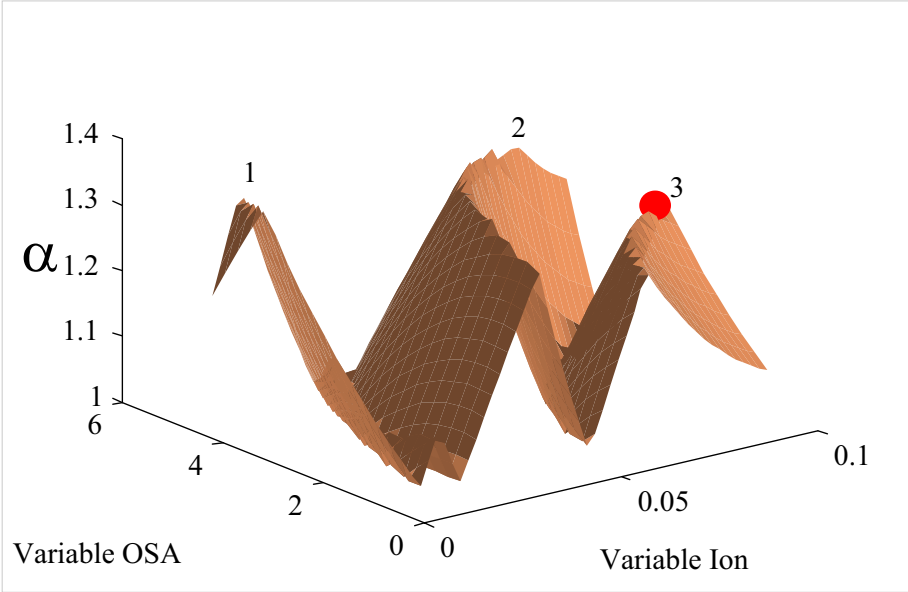


Figure 16: The calculation of α is repeated for a grid of mobile-phase compositions. The α values are then plotted as a function of the two varied mobile-phase constituents: octane sulphonate (OSA) and buffer concentration (Ion). The amount of acetonitrile (AcN) is set constant at 20%. The resulting surface has 3 ridges, numbered 1–3. Choosing the third ridge will give a chromatogram with all peaks separated within approx. 12 minutes.

7.3 DryLab™

The software DryLab™ [93-102] has been used for development of LC methods for a long time. In the first version one predefined factor could be chosen at a time (e.g. % of organic modifier, gradient time, flow rate), two experiments were done and linear models of log (k=retention factor) were used to predict a resolution map for critical resolution ($R_s=1/4N^{1/2}(\alpha-1)(k/[1+k])$) of the peaks involved. In a later version two predefined factors could be chosen (e.g. % of organic modifier and temperature). At least four experiments were needed and linear models of log(k) were used to predict

three-dimensional resolution maps (x- and y-axes for the factors and z-axis for the resolution). When the pH was included as a factor in the software, at least three experiments were necessary and log(k) was adapted with a standard cubic spline fit (polynomial with linear and quadratic terms) [94, 95]. In the latest version of DryLab™ additional factors can be defined and combined for two factors at a time. If nine experiments are done, log k will be adapted by a cubic fit to predict the 3D resolution map. The models used in DryLab™ are linear or cubic, so that no interaction between the two investigated factors can be included. In paper V two factors (concentration of SDS and IPA) were investigated at three levels (9 experiments + 2 extra in the centre point). Using the 3D maps from the software optimised the separation of five groups of different compounds.

DryLab™ suffers from the disadvantage that only two factors can be chosen at a time. At least nine experiments must be performed to obtain a cubic fit of the responses. Fewer experiments are needed when MODDE is used for optimisation, if two factors are chosen; a minimum of 7 experiments is carried out (4 for each corner and 3 in the centre). Furthermore, the terms that give the best fit are included in the model (linear, interaction and/or quadratic). The quality of the model can be examined by studying different plots (e.g. summary of fit, residuals). No measure of the model quality can be obtained in DryLab™. The only way to test model quality is to perform additional experiments and compare the predicted with the experimental response.

7.4 Discussions of the different optimisation tools

Plotting a map of the lowest resolution (3D resolution from DryLab™) compared to the lowest α values (in-house program in Matlab™) is better, since the efficiency of the peaks is also taken into account, along with the resolution. Only the selectivity between peaks is used in the α-plot. However, it is also possible to include the peak width in the same MLR/PLS model to account for peaks of different sizes and widths in the Matlab™ simulation. The software can also plot the simulated widths in the predicted chromatogram, although this was not done in paper II.

In one situation fewer experiments were needed in DryLab™ to make a 3D resolution map. Usually, with two factors, 4, 6 or 9 experiments are required. If curvature is to be expected, nine experiments should be performed. Results from the central composite design (CCF) with three factors from paper II were used in DryLab™. One side of the “cube” was chosen (five experiments), setting the amount of acetonitrile at a low level (20%). Results from at least six experiments were needed from the design;

consequently, one of the experiments from the central composite design was used twice (see Figure 17).

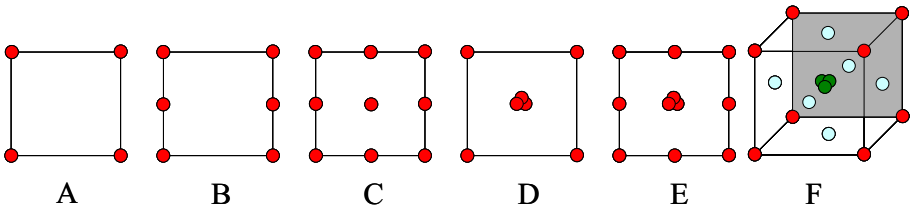


Figure 17: Example of different experimental designs used in DryLab™ (A – C) and MODDE (D – F).

Since the fitted model in MODDE only contained linear and quadratic terms, the α-plot from Matlab™ (based on the model from MODDE) is similar to the 3D resolution plot from DryLab™ (Figure 18). This would not have been the case if the model in MODDE also had interaction terms included, since only linear and quadratic terms are used in DryLab™. A buffer concentration of 100 mM and 2.9 mM OSA (grey dot) were first chosen as an optimum setting of the factors. The amount of OSA was then changed to 2.2 mM (black dot) resulting in shorter retention time for the last peak and a small decrease in selectivity between the two last peaks. A minor and acceptable change of selectivity of the other peaks was observed by the change in the OSA concentration. Close agreement between predicted retention times from MODDE, Matlab™ and DryLab™ and the experimental values was found (Table 2).

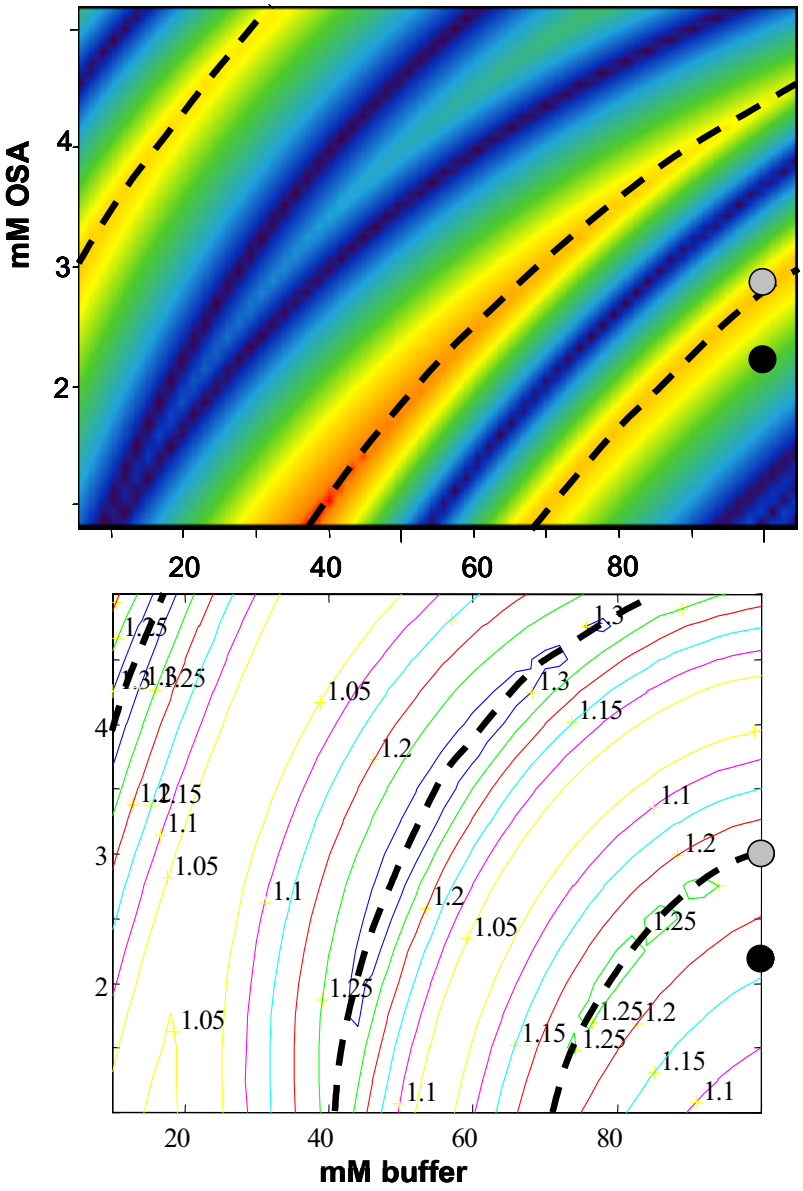


Figure 18: Comparison of 3D resolution map (DryLab™) and α-plot (MODDE and Matlab™). The resolution or the selectivity of critical pair of peaks is plotted as a function of buffer concentration and amount of octane sulphonate (OSA). The ridges, where the best separation is obtained, are marked with a dashed line. A buffer concentration of 100 mM and 2.9 mM OSA (grey dot) were first chosen as an optimum setting of the factors. The amount of OSA was then changed to 2.2 mM (black dot) resulting in shorter retention time for the last peak and a small decrease in selectivity between the two last peaks.

Table 2: Predicted retention times from [1]MODDE, [2]Matlab™ and [3]DryLab™ compared to the actual values from the experiment (paper II).

Peak name	Predicted[1] t_R (min)	Predicted[2] t_R (min)	Predicted[3] t_R (min)	Actual t_R (min)
H 314/26 (k26)	1.81	1.81	2.07	1.95
H 314/27 (k27)	2.88	2.89	3.03	2.89
H 314/21 (k21)	4.35	4.35	4.39	4.44
DKP	6.11	6.11	6.84	6.16
Unknown (kI)	7.43	7.43	8.47	7.49
H 299/87 (k99)	8.66	8.65	10.08	9.42

8 Molecular modelling

8.1 Fitting a PLS model between descriptors of the compounds and migration times from the experimental design in MEEKC (paper V)

SELMA [103-107] is a program used for describing chemical information about different compounds. The program calculated 93 descriptors for each analyte used in the three-level full factorial design (paper V, 29 compounds, see Appendix III). Example of descriptors are: number of bonds, number of atoms, nitrogen counts, highest positive atomic charge, topological dipole moment, polarisability, molecular weight, number of H-bond donors and log P.

The descriptors of each compound were then fitted by PLS with 4 components [68] to the migration times from experiments N1–N11 in the three-level full factorial design (Figure 19). The PLS model was refined by excluding descriptors with values near 0, i.e. descriptors with a small influence on the responses. 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).

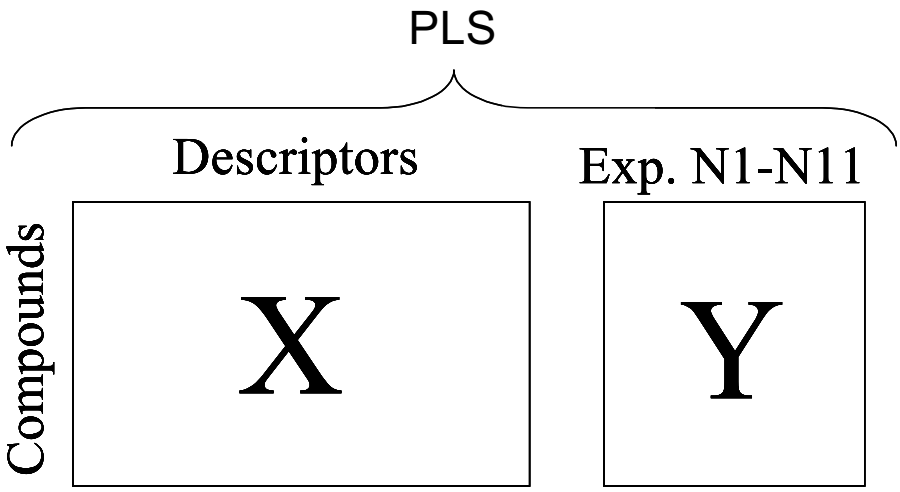


Figure 19: Fitting molecular descriptors of the compounds to the log (migration times) from the three-level full factorial design (experiments N1-N11) with PLS.

Figure 20 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 (log (migration times)) from the three-level full factorial design are placed in the same plot (Figure 21). Descriptors that are positively correlated with the migration times are found in the upper right square. These descriptors explain the hydrophobicity of the compounds like octanol/water partition coefficient (log P, clog P) and number of nonpolar atoms (nonpolar count, nonpolar count/MW), as well as the charge of the analytes (negatively ionisable groups, average negative charge, average positive charge, dipole moment). 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 are found descriptors that are negatively correlated with the responses (log (migration times)). 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).

A representative plot of predicted migration times plotted against observed migration times for experiment N10 (centre point in the three-level full factorial design) is shown in Figure 22. As can be seen from the plot, a close correlation exists between predicted and observed migration times.

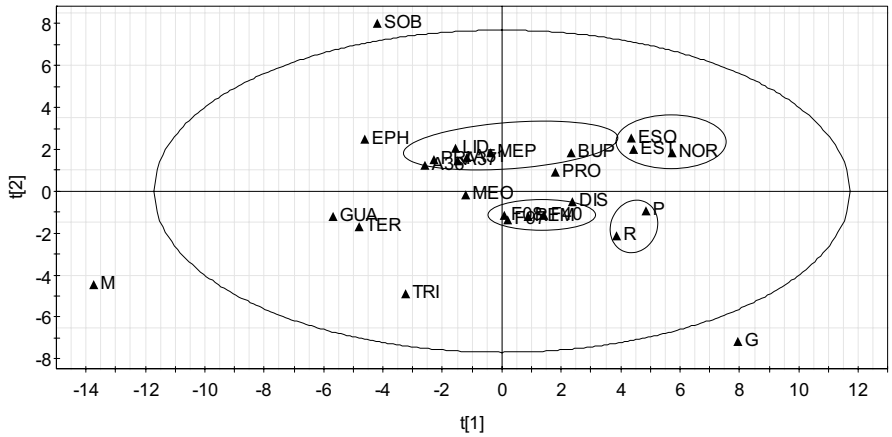


Figure 20: Score plot of 26 compounds from the PLS model. SAL, NAP and KET are excluded from the model.

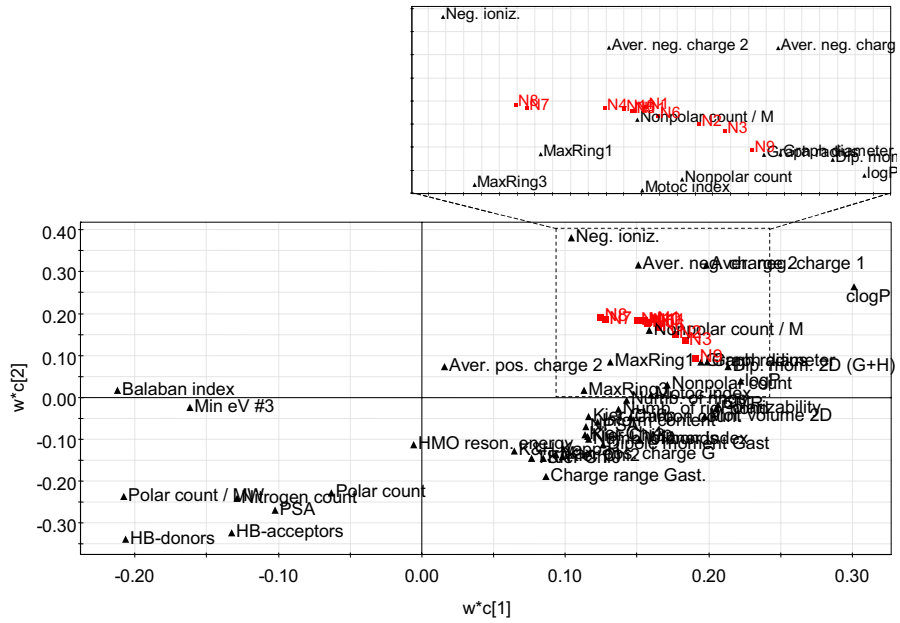


Figure 21: Loading plot of X (descriptors) and Y (log (migration times)) from the three-level full factorial design (experiments N1 – N11).

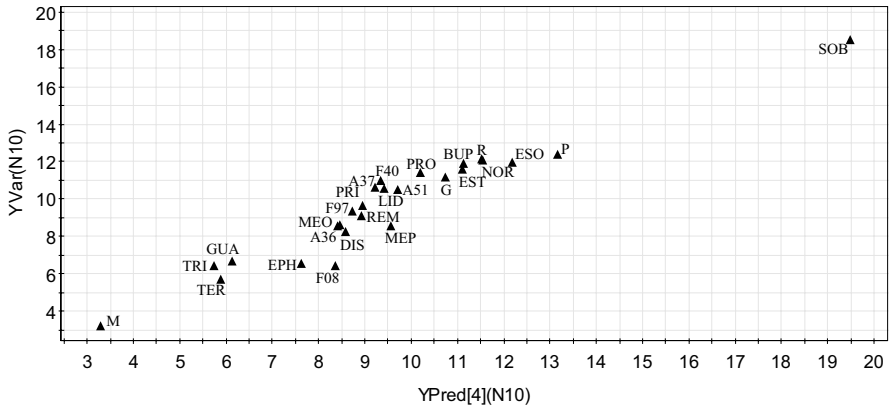


Figure 22: Predicted vs. Observed migration times of 26 compounds. The settings of the factors are according to experiment N10 (3.5% w/w SDS, 6% w/w IPA).

A literature search shows that molecular descriptors are often used in the area of drug discovery, where structure-property relationships are modelled [104, 105]. Furthermore, quantitative structure-retention relation (QSRR) has been modelled for retention data from GC [108, 109] and LC [110-112]. Kaliszan et al. [113] were able to predict the retention times of new molecules (3 descriptors) on three different HPLC columns and two different gradient elutions. Molecular modelling has been used in MEKC by Liang et al. [114], where migration parameters of flavonoids were studied with reference to structural descriptors. In addition, mechanistic and molecular modelling studies were carried out by Copper et al. [115] for separation of derivatised amino acid enantiomers with cyclodextrin-modified capillary electrophoresis. Molecular modelling fitted with retention data from MEEKC was not found in the literature.

8.2 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 experiments N1–N11. This was tested by randomly selecting 7 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 in 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. 59 out of 66 (89.4%) predictions were within an acceptable range (80–120%). EPH could not be predicted properly in 7 out of 11 cases, the predicted migration times being 1.3 – 1.5 times higher than the experimental values.

In conclusion: the models are useful for prediction of migration times in different microemulsions for a majority of new compounds; 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 set, and hydrophilic positively charged in the second test set (EPH).

9 Conclusions

Two different reversed-phase HPLC methods have been optimised. In paper I, a gradient method with ammonium phosphate pH 7, acetonitrile and a co-ion TBA added to the mobile phase, separated several related compounds from erythromycin (Figure 23).

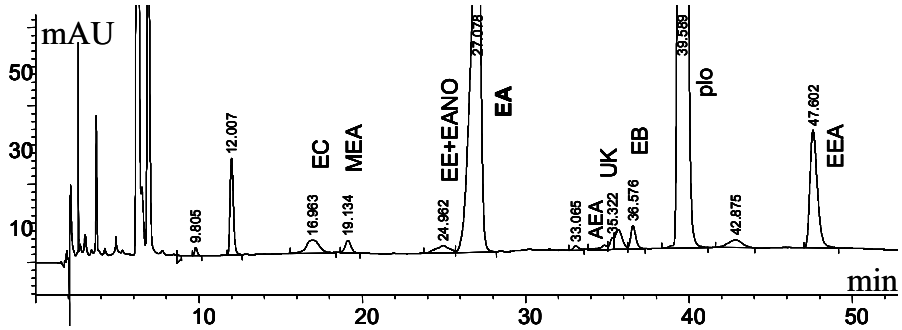


Figure 23: A chromatogram of a spiked sample solution obtained with the optimised method. The factors were set to 0.012 M TBA, "working" pH 7.4, gradient 1.33 and temperature 40 °C. EC: erythromycin C, MEA: N-demethylerythromycin A, EE: erythromycin E, EANO: erythromycin A N-oxide, EA: erythromycin A, AEA: anhydroerythromycin A, UK: unknown, EB: erythromycin B, *plo=placebo*: diethylphthalate, EEA: erythromycin A enol ether.

An isocratic HPLC method with a mobile phase containing phosphate buffer pH 2.6, acetonitrile and a counter ion (OSA) separated H314/27 from five related compounds (Figure 24).

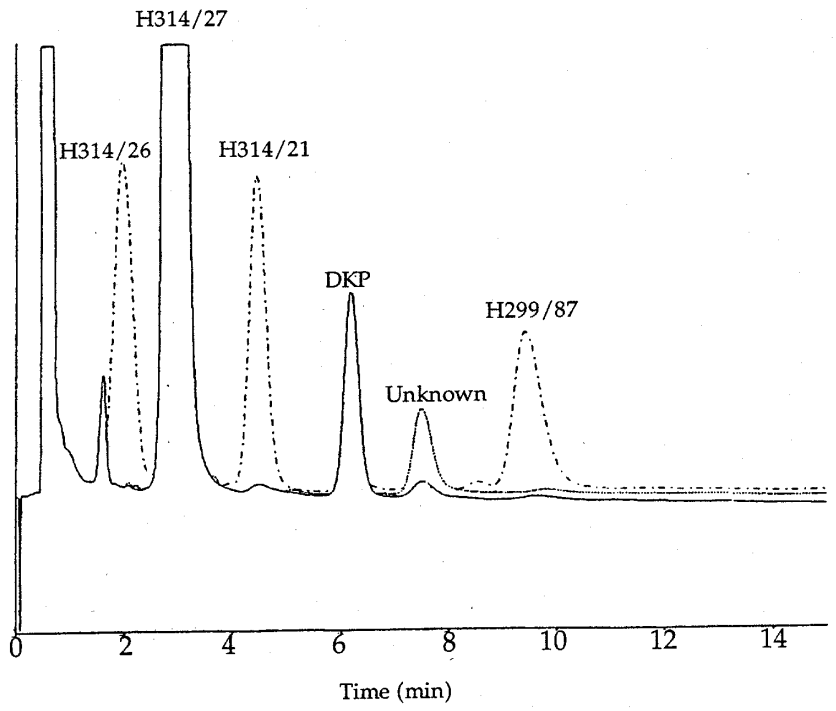


Figure 24: The optimised mobile phase (AcN: 20%, OSA: 2.2 mM, buffer concentration (Ion): 0.1 M) from the simulation was tested in the laboratory. The peak eluting before H 314/26 is a system peak. Superimposing three chromatograms on each other has produced this Figure. The reason is that two analytes (DKP and I = Unknown) are not available in pure form.

A chiral separation of propranolol was achieved by capillary electrophoresis with cellobiohydrolase (Cel7A) as chiral selector (paper III). The optimisation was not only focused on the separation, but on the efficiencies and symmetries of the peaks as well (Figure 25).

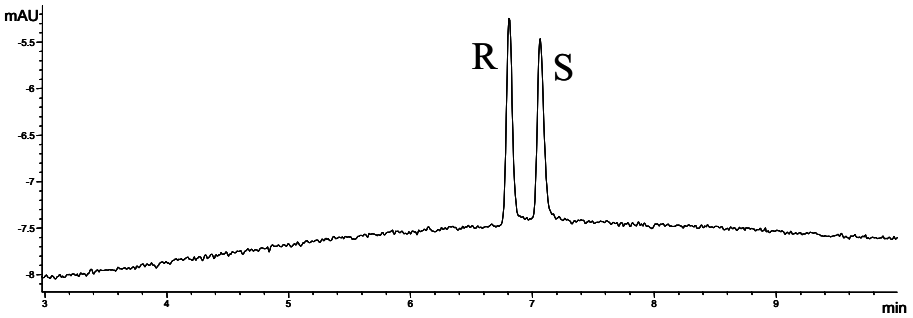


Figure 25: Electropherogram from the optimisation of the enantioseparation of (R)- and (S)-propranolol with Cel7A as selector. BGE: 0.015 M bistris-acetate pH 6.5, 17% v/v acetonitrile. The effective length of the capillary was filled with 90% of 31.2 µM Cel7A, 2% 15 µM *rac*-propranolol and 0.8% of BGE before the voltage (constant current of 7 µA) was applied. The temperature was set at 22°C

Eight compounds (lidocaine, trimethoprim, propranolol, naproxen, estradiol, norethisterone, disopyramide and salicylic acid) varying in charge and hydrophobicity were first chosen to study six factors (SDS, Brij 35, 1-butanol, 2-propanol, buffer concentration and temperature) in a screening design in MEEKC (paper IV). SDS and 2-propanol had the largest effect on migration time and were further studied in a three-level full factorial design (paper V). Twenty-nine different compounds were divided into five groups and the separation was optimised within each group using different optimisation tools (see Table 1 sections 6.4-6.5 and Table 3 below). The optimised separation of groups 1 – 4 is given in Figure 26.

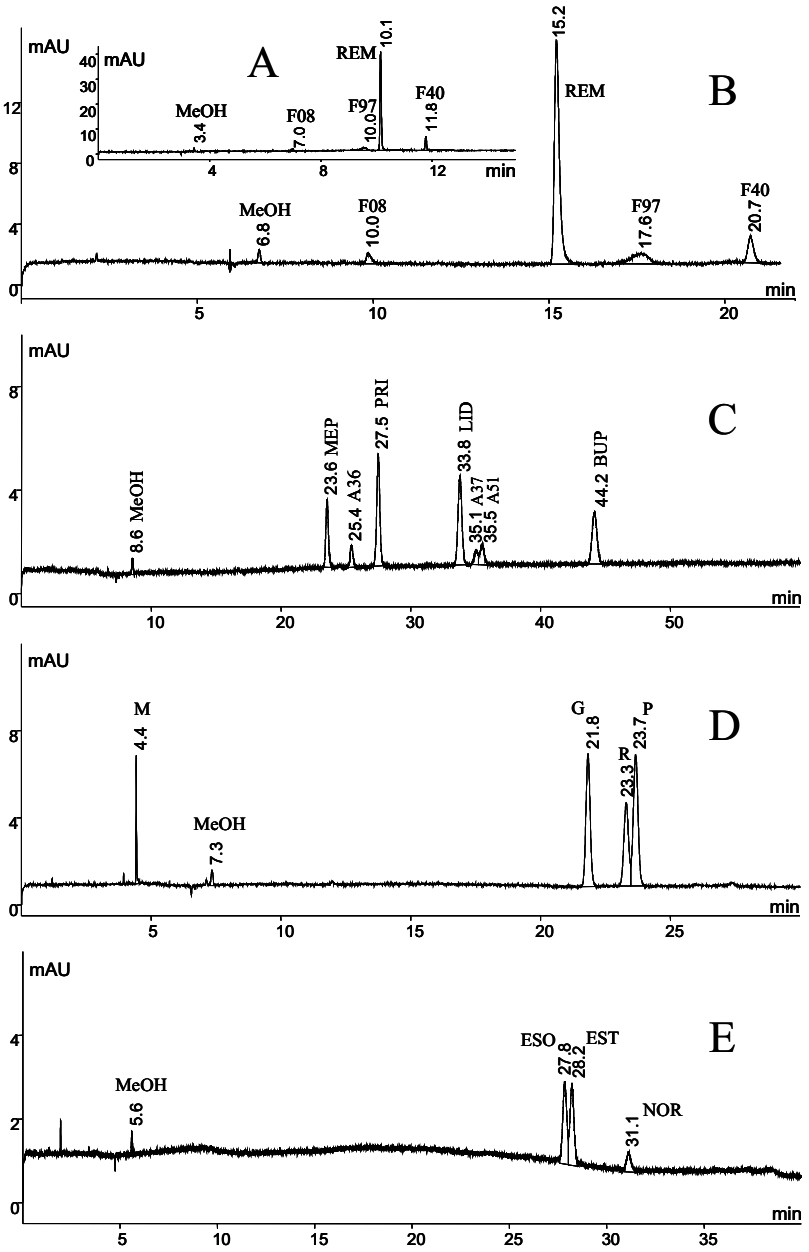


Figure 26: Optimisation of the separation of the different groups of analyte. A, B: REM, F08, F97 and F40 (group 1), C: MEP, PRI, LID, BUP, A36, A37 and A51 (group 2), D: M, G, R and P (group 3) and E: ESO, EST and NOR (group 4). Concentration of SDS and IPA (w/w): A: 5% SDS, 0% IPA, B: 5% SDS, 10% IPA, C: 6% SDS, 12% IPA, D: 4.5% SDS, 12% IPA and 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, 10kV, 40 °C. MeOH: methanol, EOF marker.

Statistical experimental design is a powerful tool for the development and optimisation of separation methods. For development of a method separating many compounds, a screening of factors should first be performed (literature search, pre-experiments or a screening design). For optimisation of the separation, a central composite design (2-3 factors) or a three-level full factorial design (2 factors) should be carried out (see Figure 27).

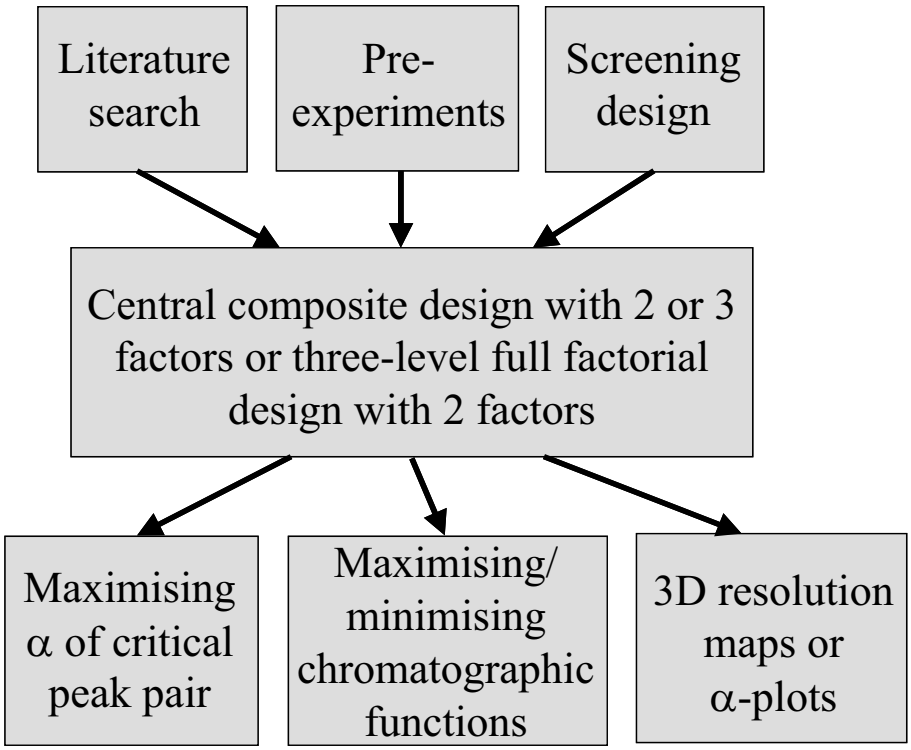


Figure 27: Method development and optimisation strategies.

With a central composite design as a base, several optimisation strategies can be used, depending on the complexity of the separation. A compilation of advantages and drawbacks of the different optimisation methods evaluated is given in Table 3.

Table 3: Advantages and drawbacks of different optimisation methods.

Optimisation strategy	Advantages	Drawbacks
Response surface plots (in MODDE) of selectivity/resolution of critical peak pair or chromatographic functions.	Useful plots. Easy to get an overview and to choose the maximum/minimum value. In combination with a response surface plot of the last eluting peak, the analyst gets a picture of the separation in combination with the total analysis time.	Response surface plots can only be plotted with two factors at a time. If three factors are investigated, the third factor must be set at a certain level. Several response surface plots can be made setting the third factor at different levels. The choice of factor levels will be even more complicated if four factors are used.
Simplex ("Optimiser" in MODDE). Responses are selectivity/resolution of critical peak pair(s) or chromatographic functions.	Finds the optimum/minimum within the experimental design for all the factors in the design simultaneously. This is an advantage when more than two factors are investigated. Several responses can be optimised simultaneously. Several responses can be optimised with different criteria, e.g. maximising the resolution and minimising the total analysis time.	A simplex can be trapped in local instead of global optima, but the risk is minimised in MODDE since 8 simplexes start from 8 different runs.
α-plot in Matlab™	A surface map of the lowest selectivity of all combination of peak pairs is calculated for two factors at a time. New plots can be made with different levels on the third factor. By pointing on the plot, a predicted chromatogram is shown. Interactive optimisation using the α-plot and predicted chromatograms to find factor settings	The analyst must have knowledge of Matlab™ programming. A new program must be written for each new MLR/PLS model.

	giving good separation and acceptable analysis time.	
3D resolution maps in Drylab™	3D resolution maps of peak pairs with lowest resolution for two factors are obtained. Predicted chromatograms are generated within the experimental domain.	Only two factors at a time can be used. More experiments are needed compared to MODDE (9 experiments for 2 factors). Quadratic terms are included in the model, but not interaction terms. No values for model quality are obtained, leading to potentially of large deviations between predicted and experimental results.

The chromatographic functions $CRF_{1\text{-}3}$, ATR, Rp and r worked well as responses for optimisation of separation (paper V). COF, CRS and CEF have a time term included in the equation, which gave too much weight to the time compared to the resolution and suggested areas in the experimental domain that were not optimal for the separation. These chromatographic functions (COF, CRS and CEF) worked out better when the time term was excluded from the equations and gave then the same optimum setting of the factors as for equations focused only on the resolution (ATR, Rp, r). $CRF_{1\text{-}2}$ have a time term added (not multiplied) to the equations; nevertheless, they gave the same optimum setting of the factors as for the equations only containing resolution terms.

A user-friendly commercial program is needed where 3D resolution maps can be obtained based on models with functions where linear, quadratic and interaction terms (if necessary) are included. Predicted chromatograms /electropherograms obtained when pointing at the 3D map is a useful tool and should also be integrated in the software. Furthermore, if more than two factors are used, different settings of the other factors should be possible, resulting in an automatic update of the 3D resolution plot.

Results from a central composite design or a three-level full factorial design in MODDE can be used in DryLab™ to generate 3D resolution maps if: a) two factors are used and investigated at three levels each (9 experiments + 2 extra in the centre of the design = 11 experiments), or b) a central composite design with three factors is carried out instead and the results from one side of the "cube" is used in DryLab™. The MLR model must then only contain linear and quadratic terms. If significant interaction

exists between the factors, the consequence could be a large discrepancy between predicted and observed retention/migration times.

Descriptors of different compounds varying in hydrophobicity and charge were fitted by PLS to the migration times of the same compounds from a three-level full factorial design (MEEKC). The compounds were divided into training and test sets. The training set was then used to predict the migration times of the compounds in the test set. Two different test sets were used. It was concluded that 86–89% of all predictions were acceptable (within 80–120% of observed value).

10 Acknowledgements

This work was carried out at the Analytical Development departments of Astra Läkemedel AB, Astra Arcus AB and AstraZeneca R&D Södertälje, Sweden.

Over the years many people have been involved, all of whom I am gratefully indebted.

In particular I, would like to acknowledge the following people:

Dr. Ove Åström and Professor Christina Graffner for giving me the opportunity to start this research.

My supervisors, Professor Douglas Westerlund and Adjunct Professor Sven P. Jacobsson, for valuable discussions, comments and support throughout the work.

Dr. Astrid Arbin for allowing me to complete my studies.

My co-authors, Dr. Anders Karlsson, Dr. Mats Josefson, Professor Roland Isaksson, Associate professor Gunnar Johansson and Dr. Cari Sänger-van de Griend, for significant contributions.

The stimulating and enjoyable task of being a supervisor to Lena Kullstam, Dr. Oliver Klett, Katarina Lindberg, Marcus Tysk and Jessica Eriksson during their undergraduate work.

All my present and former colleagues at the department of Analytical Pharmaceutical Chemistry, Uppsala University, for interesting seminars.

All my colleagues at AstraZeneca for creating a great ambience and for good support.

Eva Neidenström for wonderful support and encouragement.

All the former and present people of “Babette”. It’s great working with you.

The CE-group at AstraZeneca, especially Helena Brunnkvist, Karin Sander, Dr. Ninni Granelli, Dr. Karin Stubberud, Dr. Ludmila Westermark, Dr. Joanna Oreskär, Dr. Olle Stålberg, Dr. Cari Sänger-van de Griend, Margareta Andersson and Dr. Anna Maria Enlund.

Stefan Rännar at Umetrics for taking the time to discuss chemometric problems.

Mikael R. Ek at the library for swift literature searches.

All my friends in Norway and Sweden, Ragnhild, Kristin, Synnøve, Merete, Sussi, Anna B, Kicki, Camilla, Pernilla, Annette, Rana, Abed, Sofia, Mattias, Ulrika, Göran, Christina, Monja, Henrik S, Raffe, Karin L, Patrik W, Larisa H, Anna S, Henrik L, Katarina L, Maria R, Peter, Elisabeth, Jens and Stephen for believing in me, supporting me and making me think of other things than chemistry.

My family in France, in particular Pierrette and Pierre, always receiving me with open arms and serving fantastic French food.

My family in Sweden, Svenborg, Åge, Hanna, Jonas, Martin, Monika, Mikael and Johan for many enjoyable family gatherings.

My sister Ingvild and her family, Knut, Michèlle, Sigrid and Magnus, for support and understanding, and for great skiing in Tromsø and lovely times in Estaing.

My father, Professor Emeritus Ove Harang, a great and inspiring scientist. I understand the origin of my interests in art, books and films. Dr. Sylvaine Perraut for always welcoming me in Paris.

My dear husband, Anders, for all his love, support, advice and encouragement.

11 References

1. Bounine, J. P., Guiochon, G., Colin, C., *J. of Chrom.*, 1984, *298*, 1-20
2. Berridge, J. C., *Chemometrics and Intelligent Laboratory System*, 1988, *3*, 175-188
3. M.S.Tswett, Tr.Varshav. Obshch. Estestvoispyt., Otd. Biol. 1903, 14, 20. reprint in G. Hesse, H. Weil, *Michael Tswett's first paper on chromatography*, M. Woelm, Eschwege 1954
4. M. S. Tswett, Ber. Dtsch. Botan. Ges. 1906, 24, 316-323. An English translation of this article has recently been published in: V. G. Berezkin, Editor, *Chromatographic Adsorption Analysis: Selected Works of M. S. Tswett*, Ellis Hoorwood, New York, London, 1990, pp.21-26
5. Smolkova-Keulemansova E., *J. High Resol. Chromatogr.*, 2000, *23*, 497-501
6. Ettre, L. S., *LCGC North America*, 2003, *21*, 5, 458-467
7. Snyder, L. R., Kirkland J. J., *Introduction to modern liquid chromatography*, John Wiley & sons Inc., 1979
8. Schill, G., Ehrsson, H., Vessman, J., Westerlund, D., *Separation methods for drugs and related organic compounds*, second edition, Swedish Pharmaceutical Press, Stockholm 1983
9. Giddings, J. C., *Unified separation sciences*, John Wiley & sons, New York 1991
10. Brown, P. R., Grushka, E., *Advances in chromatography*, Marcel Dekker Inc., New York
11. Vailaya, A., Horvath, C., *J. of Chromatogr. A*, 1998, *829*, 1-27
12. Hjertén, *Chromatogr. Rev.*, 1967, *9*, 122-219
13. Virtanen, R., *Acta Polytech. Scand.*, 1974, *123*, 7-67
14. Mikkers, F. E. P., Eveaerts, F. M., Verheggen, T. P. E. M., *J. Chromatogr.* 1979, *169*, 11-20
15. Jorgensson, J. W., Lukacs, K. D., *Anal. Chem.*, 1981, *53*, 1298-1302
16. Jorgensson, J. W., Lukacs, K. D, *Science*, 1983, *222*, 266-272
17. Heiger, D. N., *High performance capillary electrophoresis – an introduction*, second edition, Hewlett-Packard Company, France 1992
18. Terabe, S., Otsuka, K., Ichikawa, K., Tsuchiya, A., Ando, T., *Anal. Chem.*, 1984, *56*, 111-113
19. Terabe, S., Otsuka, K., Ando, T., *Anal. Chem.*, 1985, *57*, 834-841
20. Watarai, H., *Chem.Lett.* 1991, 391-394
21. Altria, K.D., Mahuzier, P.-E., Clark, B. J., *Electrophoresis*, 2003, *24*, 315-324
22. Chankvetadze, B., *Capillary electrophoresis in chiral analysis*, John Wiley & sons, Chichester, UK 1997
23. Armstrong, D. W., Rundlett, K. L., Reid III, G. L., *Anal. Chem.*, 1994, *66*, 1690-1695

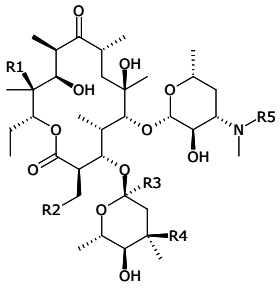
24. Fanali, S., Cristalli, M., Vespalec, R., Bocek, P., *Advances in electrophoresis*, 1994, *7*, 1-84
25. Barker, G. E., Russo, P., Hartwick, R. A., *Anal. Chem.*, 1992, *64*, 3024-3028
26. Valtcheva, L., Mohammad, J., Pettersson, G., Hjertén, S., *J. of Chromatogr.*, 1993, *638*, 263-267
27. US Pharmacopoeia 24 page 1923 National Publishing, Philadelphia PA, 1999
28. Hewlett Packard, *Understanding your Chemstation*, third edition, USA, 1994
29. Khaledi, M.G., Smith, S. C., Strasters, J. K., *Anal. Chem.*, 1991, *63*, 1820-1830
30. Smith, S., Khaledi, M. G., *J. of Chrom.*, 1993, *632*, 177-184
31. Muijselaar, P.G., Claessens, H. A., Cramers, C. A., *J. of Chrom. A*, 1997, *765*, 295-306
32. Strasters, J. K., Khaledi, M.G., *Anal. Chem.*, 1991, *63*, 2503-2508
33. Taillardat-Bertschinger, A., Carrupt, P.-A., Testa, B., *European J. of Pharm. Sci.*, 2002, *15*, 225-234
34. Wold, S., Technometrics, 1978, *20*, 397-405
35. Beebe, K. R., Kowalski, B. R., *Anal. Chem.*, 1987, *57*, 1007A-1017A
36. Wold, S., *J. Pharm. Biomed. Anal.*, 1991, *9*, 589-596
37. Box, G.E.P., Hunter, W.G., Hunter, J.S., *Statistics for experimenters*, John Wiley & Sons, Inc., New York 1978
38. Nortvedt, R., Brakstad, F., Kvalheim, O. M., Lundstedt, T., *Anvendelse av Kjemometri innen forskning og industri*, The Swedish Chemical society, 1996
39. Carlson, R., *Design and optimization in organic synthesis*, Elsevier, Amsterdam 1992
40. User guide and tutorial for MODDE 6.0
41. Schoenmakers, P. J., Mackie, N., Lopes Marques, R. M., *Chromatographia*, 1993, *35*, 18-32
42. Alfazema, L. N., Hows, M. E. P., Howells, S., Perrett, D., *Electrophoresis*, 1997, *18*, 1847-1856
43. Varesio, E., Gauvrit, J.-Y., Longeray, R., Lantéri, P., Veuthey, J.-L., *Chromatographia*, 1999, *50*, 195-201
44. Kaale, E., Van Schepdael, A., Roets, E., Hoogmartens, J., *J. of Chrom. A*, 2001, *924*, 451-458
45. Mikaeli, S., Thorsén, G., Karlberg, Bo, *J. of Chrom A*, 2001, *907*, 267-277
46. Havel, J., Breadmore, M., Macka, M., Haddad, P. R., *J. of Chrom. A*, 1999, *850*, 345-353
47. Klampfl, C. W., Leitner, T., Hilder, E. F., *Electrophoresis*, 2002, *23*, 2424-2429
48. Capote, R., Diana, J., Roets, E., Hoogmartens, J., *J. Sep. Sci*., 2002, *25*, 399-404
49. Marengo, E., Gennaro, M. C., Gianotti, V., Prenesti, E., *J. of Chrom. A*, 1999, *863*, 1-11
50. Nielsen, M. S., Nielsen, P. V., Frisvad, J. C., *J. of Chrom. A*, 1996, *721*, 337-344
51. Persson-Stubberud, K., Åström, O., *J. of Chrom. A*, 1998, *798*, 307-314
52. Fanali, S., Furlanetto, S., Aturki, Z., Pinzauti, S., *Chromatographia*, 1998, *48*, 395-401
53. Morris, V. M., Hargreaves, C., Overall, K., Marriott, P. J., Hughes, J. G., *J. of Chrom. A*, 1997, *766*, 245-254
54. Thorsteinsdottir, M., Westerlund, D., Andersson, G., Kaufmann, P., *Chromatographia*, 1997, *46*, 545-554
55. Vindvogel, J., Sandra, P., *Anal. Chem.*, 1991, *63*, 1530-1536
56. Marengo, E., Gennaro, M. C., Abrigo, C., *Anal. Chem.*, 1992, *64*, 1885-1893

57. Tucker, R. P., Fell, A. F., Berridge, J. C., Coleman, M. W., *Chirality*, 1992, *4*, 316-322
58. Rouberty, F., Fournier, J., *J. Liq. Chrom. & Rel. Technol.*, 1996, *19*, 1, 37-55
59. Karlsson, A., Hermansson, S., *Chromatographia*, 1997, *44*, 10-18
60. Coufal, P., Stulik, K., Claessens, H. A., Hardy, M. J., Webb, M., *J. of Chrom. A*, 1999, *732*, 437-444
61. Diana, J., Ping, G., Roets, E., Hoogmartens, J., *Chromatographia*, 2002, *56*, 313-318
62. Diana, J., Visky, D., G., Roets, E., Hoogmartens, J., *J. of Chrom. A*, 2003, *996*, 115-131
63. Thorsteinsdottir, M., Ringbom, C., Westerlund, D., Andersson, G., Kaufmann, P., *J. of Chrom. A*, 1999, *831*, 293-309
64. Wan, H., Öhman, M., Blomberg, L. G., *J. of Chrom. A*, 2001, *916*, 255-263
65. Hillaert, S., De Beer, T. R. M., De Beer, J. O., Van den Bossche, W., *J. of Chrom. A*, 2003, *984*, 135-146
66. Pretswell, E. L., Morrisson, A. R., *Analytical methods and instrumentation*, 1995, *2*, 87-91
67. Miyawa, J. H., Alasandro, M. S., Riley, C. M., *J. of Chrom. A*, 1997, *769*, 145-153
68. Höskuldsson, A., *J. Chemom.*, 1988, *2*, 211-228
69. Eriksson L., Johansson, E., Kettaneh-Wold, N., Wikström, C., Wold, S., *Design of experiments – Principles and applications*, Umetrics academy, Learnways, Stockholm Sweden 2000
70. Guillaume, Y., Guinchard, C., *J. of Liq. Chrom.*, 1993, *16(16)*, 3457-3470
71. Guillaume, Y., Guinchard, C., *J. of Liq. Chrom.*, 1994, *17(7)*, 1443-1459
72. Guillaume, Y., Cavalli, E. J., Peyrin, E., Guinchard, C., *J. Liq. Chrom. & Rel. Technol.*, 1997, *20(11)*, 1741-1756
73. Marle, I., Erlandsson, P., Hansson, L., Isaksson, R., Pettersson, C., Pettersson, G., *J. Chromatogr.,* 1991, *586(2)*, 233-248
74. Marle, I., Jönsson, S., Isaksson, R., Pettersson, C., *J. Chromatogr.,* 1993, *648(2),* 333-347
75. Hedeland, M., Isaksson, R., Pettersson, C., *J. Chromatogr. A,* 1998, *807(2)*, 297-305
76. Hedeland, M., Holmin, S., Nygard, M., Pettersson, C., *J. Chromatogr. A*, 1999, *864(1)*, 1-16
77. Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J.K.C., Teeri, T.T, Jones, T.A., *Science*, 1994, *265*, 524-528
78. Divne, C., Ståhlberg, J., Teeri, T.T., Jones, T.A., *J. Mol. Biol.,* 1998, *275*, 309-325
79. Ståhlberg, J., Divne, C., Koivula, A., Piens, K., Claeyssens, M., Teeri, T.T., Jones, T.A., *J. Mol. Biol.*, 1996, *264*, 337-348
80. Ståhlberg, J., Henriksson, H., Divne, C., Isaksson, R., Pettersson, G., Johansson, G., Jones, T.A., *J. Mol. Biol.*, 2001, *305*, 79-93
81. Terabe, S., Matsubara, N., *J. of Chrom.*, 1992, *608*, 23-29
82. Terabe, S., Otsuka, K., Ando, T., *Anal. Chem.*, 1989, *61*, 251-260
83. Siouffi, A. M., Phan-Tan-Luu, R., *J. of Chrom. A*, 2000, *892*, 75-106
84. Berridge, J.C., *J. of Chrom.,* 1982, *244*, 1-14
85. Stebbins, M. A., Davis, J. B., Clark, B. K., Sepaniak, M. J., *J. of Microcolumn sep.*, 1996, *8(7)*, 485-494

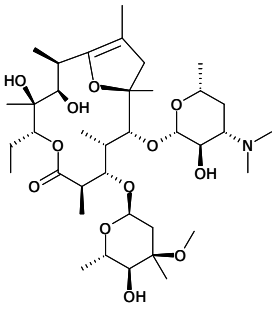
86. Schaepper, J. P., Fox, S. B., Sepaniak, M. J., *J. of Chrom. Science*, 2001, *39*, 411-419
87. Schlabach, T. D., Excoffier, J. L., *J. of Chrom.*, 1988, *439*, 173-184
88. Morris, V. M., Hughes, J. G., Marriott, P. J., *J. of Chrom. A*, 1996, *755*, 235-243
89. Klein, E. J., Rivera, S. L., *J. Liq. Chrom. & Rel. Technol.*, 2000, *23(14)*, 2097-2121
90. Billiet, H. A. H., Andersson, P. E., Haddad, P. R., *Electrophoresis*, 1996, *17*, 1367-1372
91. Berridge, J. C., *Chem. Int. Lab. Syst.*, 1989, *5*, 195-207
92. Nelder, J. A., Mead, R., *Computer Journal*, 1965, *7*, 308-313
93. Drylab™ 2000, version 3.0.09, LC Resources Inc., Walnut Creek, CA 94596, USA
94. User manual DryLab™2000
95. Reference guide DryLab™2000
96. Glajch, J. L., Kirkland, J. J., J. *of Chrom.*, 1989, *485*, 51-63
97. Snyder, L. R., *Analyst*, 1991, *116*, 1237-1244
98. Däppen, R., Molnar, I., *J. of Chrom.*, 1992, *592*, 133-141
99. Dolan, J. W., Snyder, L. R., Djordjevic, N. M., Hill, D. W., Saunders, D. L., Van Heukelem, L., Waeghe, T. J., *J. of Chrom. A*, 1998, *803*, 1-31
100. Haber, P., Baczek, T., Kaliszan, R., Snyder, L. R., Dolan, J. W., Wehr, C. T., *J of Chrom. Sci.,* 2000, *38*, 386-392
101. Molnar, I., *J. of Chrom. A*, 2002, *965*, 175-194
102. Jupille, T. H., Dolan, J. W., Snyder, L. R., Molnar, I., *J. of Chrom. A*, 2002, *948*, 35-41
103. SELMA, Synthesis and Structure Administration (SaSA), Olsson, T., Sherbukhin, V., AstraZeneca R&D Mölndal, Mölndal, Sweden
104. Oprea, T. I, Gottfries, J., Sherbukhin, V., Svensson, P., Kühler. T. C., *J. of Molecular Graphics and Modelling*, 2000, *18*, 512-524
105. Linusson, A., Gottfries, J., Lindgren, F., Wold, S., *J. of Med.Chem.*, 2000, *43*, 1320-1328
106. Oprea T. I., Gottfries, J., *J. of Comb. Chem.*, 2001, *3*, 157-166
107. Bergström C., Norinder, U., Luthman, K., Artursson, P., *Molecular descriptors influencing melting point and their role in classification of solid drugs, J. Chem. Inf. Comput. Sci.*, 2003, accepted for publication
108. Lin, Z., Liu, S:, Li, Z., *J. of Chrom. Sci.*, 2002, *40(1)*, 7-13
109. Dimov, N., Osman, A., Mekenyan, O., Papazova, D., *Analytical chimica Acta*, 1994, *298(3)*, 303-317
110. Baczek, T., Kaliszan, R., *J. of Chrom. A*, 2003, *987*, 29-37
111. Nord, L. I., Fransson, D., Jacobsson, S. P., *Chemometrics and Int. Lab. Systems*, 1998, *44*, 257-269
112. Lee, S. K., Park, Y. H., Yoon, J. Y., Lee, D. W., *J. of Microcolumn Sep.*, 1998, *10(1)*, 133-139
113. Kaliszan, R., Baczek, T., Bucinski, A., Buszewski, B., Sztupecka, M., *J. of Sep. Science*, 2003, *26*, 271-282
114. Liang, H.-R., Vuorela, H., Vuorela, P., Hiltunen, R., Riekkola, M.-L., J. *Liq. Chrom. and Rel. Technol.*, 1998, *21(5)*, 625-643
115. Copper, C. L., Davis, J. B., Cole, R. O., Sepaniak, M. J., *Electrophoresis*, 1994, *15*, 785-792

12 Appendix: Structure of compounds used in papers I – V

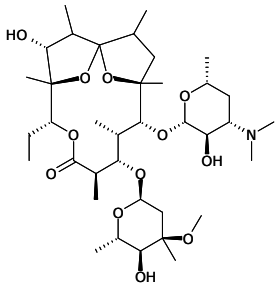
I. Structures of erythromycin and related compounds (paper I).



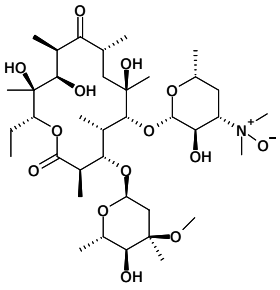
	R1	R2	R3	R4	R5
Erythromycin A (EA)	OH	H	H	OCH_3	CH_3
Erythromycin B (EB)	H	H	H	OCH_3	CH_3
Erythromycin C (EC)	OH	H	H	OH	CH_3
Erythromycin E (EE)	OH	-O-		OCH_3	CH_3
N-demethylerythromycin A (MEA)	OH	H	H	OCH_3	H



Erythromycin A enol ether
(EEA)

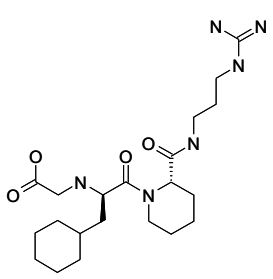


Anhydroerythromycin A
(AEA)

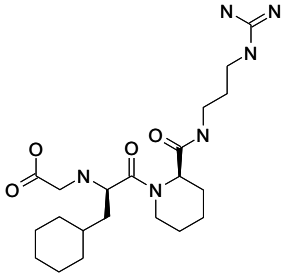


Erythromycin A N oxide
(EANO)

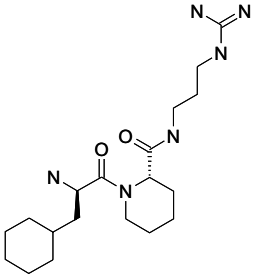
II. Structures of H 314/27 and related compounds (paper II).



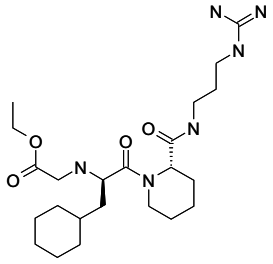
H 314/27 (k27)



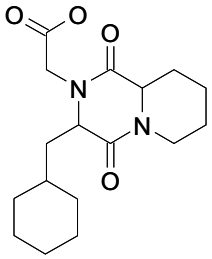
H 314/26 (k26)



H 314/21 (k21)

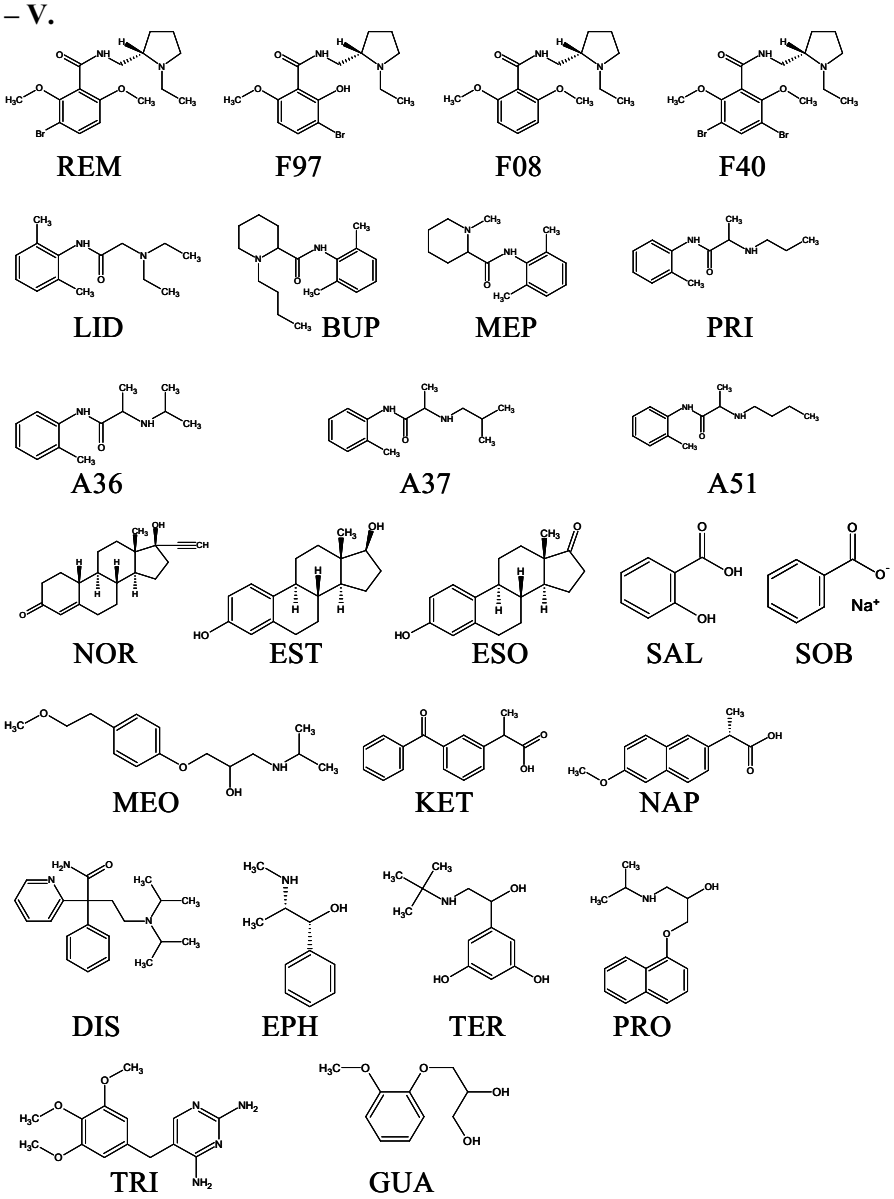


H 299/87 (k99)



DKP (DKP)

III. Structure of compounds (without hydrogen ions) used in papers III – V.



TRI: trimethoprim, LID: lidocaine, EST: estradiol, NOR: norethisterone, SAL: salicylic acid, NAP: naproxen, DIS: disopyramide, PRO: propranolol, PRI: prilocaine, SOB: sodium benzoate, MEO: metoprolol, ESO: estrone, MEP: mepivacaine, GUA: guaifenesin, TER: terbutaline, KET: ketoprofen, EPH: ephedrine, BUP: bupivacaine, REM: remoxipride, F97: FLA797, F08: FLA708, F40: FLA740, A36: AR-P016336, A37: AR-P016337, A51: AR-P017151. PRO was used in paper III; LID, TRI, PRO, NAP, EST, NOR, DIS and SAL were used in paper IV; and all of them were analysed in paper V. Structures not shown for confidential compounds R, M, G and P.

Acta Universitatis Upsaliensis

Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy

Editor: The Dean of the Faculty of Pharmacy

A doctoral dissertation from the Faculty of Pharmacy, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series *Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy.* (Prior to July, 1985, the series was published under the title "Abstracts of Uppsala Dissertations from the Faculty of Pharmacy".)

Distribution:
Uppsala University Library
Box 510, SE-751 20 Uppsala, Sweden
www.uu.se, acta@ub.uu.se

ISSN 0282-7484
ISBN 91-554-5778-9