UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA
IN
AREA DEL FARMACO E TRATTAMENTI INNOVATIVI

CICLO XXX

COORDINATORE Prof. Elisabetta Teodori

The development of new pseudostationary phases for capillary electrophoresis and the study of their mechanism of separation by chemometric techniques and magnetic resonance

Settore Scientifico-Disciplinare CHIM/01

Dottorando
Dott. Caprini Claudia
(firma)

Tutore
Prof. Furlanetto Sandra
(firma)

Coordinatore
Prof. Teodori Elisabetta
(firma)

Anni 2014/2017

INDEX

A te, mai come questa volta.

Abbreviations list

FDA: Food and Drug Administration

HPLC: High Performance Liquid Chromatography

UPLC: Ultra-high performance Liquid Chromatography

RP-UPLC: Reverse Phase UPLC

CE: Capillary Electrophoresis

OFAT: One Factor At Time

QbD: Quality by Design

DS: Design Space

BGE: Background Electrolyte

NMR: Nuclear Magnetic Resonance

MD: Molecular Dynamics Simulations

CCD: Central Composite Design

FCD: Face Centered Design

CCD: Central Composite Design

FFD: Full Factorial Design

FD: Fractional Design

ANOVA: Analysis of Variance

TD: Taguchi Design

MCs: Mixture Components

PVs: Process Variables

PCA: Principal Component Analysis

SVD: Singular Value Decomposition

LDA: Linear Discriminant Analysis

QDA: Quadratic Discriminant Analysis

DoE: Design of Experiment

CQAs: Critical Quality Attribute

CPPs: Critical Process Paramenter

KS: Knowledge Space

QRA: Quality Risk Assessment

RSM: Response Surface Methodology

MLR: Multi Linear Regression

PLS: Partial List Square

RSM: Response Surface Methododlogy

CS: Control Strategy

ICH: International Conference on Harmonization

Abstract

This thesis is embedded in the pharmaceutical analytical research field and has envisaged the development of different analytical methods by a chemometric approach. In this context, pseudostationary phases have been studied and applied on electrokinetic chromatography and analytical methods have been developed in order to be used in the field of quality assurance.

This thesis project has focused on multivariate strategy development of capillary electrophoresis methods facing up different pharmaceutical issues including those related to classification problems. The multivariate approach has made it possible to get the most amount of information with time and cost savings, obtaining additional information on the interactions between the factors studied.

Six projects focused on the optimization of methods for simultaneous determination of active ingredients and their impurities in pharmaceutical formulations.

The drugs under consideration were Glibenclamide, Zolmitriptan, Captopril, Hydrochlotothiazide, Ambrisentan, Diclofenac.

After the optimization phase, the methods studied were validated according to the ICH guidelines to ensure the appropriate performance of the method and reliable data for the specific application to which they were intended. Molecular Modeling and NMR techniques were investigated as the basis for the separation of analytes in the presence of complex pseudostationary phases such as microemulsions and cyclodextrins.

The pseudostation phases study was further investigated by conducting a multivariate study of a micellar electrokinetic chromatographic and microemulsion electrokinetic chromatography solvent modified studying the complexation and separation mechanisms amoung drugs and pseudostationary phases.

The method was applied for the determination of Sulpiride, Ambrisentan and Diclofenac. Multivariate classification methods were used to analyze some types of green tea of different botanical and geographical origin.

To conclude the Quality by Design approach was applied to investigate and study in depth the critical process parametes of dried blood spot sample preparation with liquid- chromatography mass spectrometry analysis.

1. Introduction

Quality is a benchmark of excellence for the health safety, is the process of vouching for integrity of products to meet the standard for the proposed goal.

Quality assurance is a wide concept that covers all aspects that collectively or individually impact the quality of the product. That is, the sum of organized arrangements that are made with the aim of ensuring pharmaceutical product satisfying the required quality for the intended use. Quality assurance play the role of sentinel, ensuring adherence to regulations as well as identifying possible shortfalls and making recommendations for improvements that may be made over time.

The quality control of active pharmaceutical ingredients is a prerequisite for the safe use of drug products. In the drug life cycle, drug impurity profiling plays a key role for assuring the safety and efficacy of a dosage form.

Since the world has gathered together to harmonize its practices and guides and the launching of the FDA current good manufacturing practices – the cGMP, for the 21st century, there has been a growing awareness for the significance of the quality of the pharmaceutical products.

Impurities within pharmaceutical products provide no benefits to the patient, and for many years regulators have wrestled with different strategies to address this problem [1]. Impurities will always be regarded as a ‘necessary evil’, but with the advent of more safety based limits to control related substances (ICH Q3A-D and M7) there is an increasing acceptance that control strategies based on process understanding will allow an appropriate tolerability of risk (patient needs versus commercial necessities) [1].

The control of impurities is a fundamental part and a critical analytical issue of the quality control of drug products. The development of suitable separation analytical methods for impurity assay is characterized by the need of adequate sensitivity and selectivity [2], which can only be reached by a comprehensive investigation of the analytical system.

In present days, analytical method failure is becoming more common especially during method transfer as well as in quality control departments.

Chromatographic methods are more commonly employed as right analytics at all the stages during the product life cycle. Common analytical methods for content uniformity, assay, impurity profile, and stability indicating assay are based on high performance liquid chromatographic (HPLC) or ultraperformance liquid chromatographic (UPLC) or rapid resolution liquid chromatography methods (RRLC). Regarding chromatography, due to complex parameters involved in the method development phase, low sensitivity, selectivity, and inadequate understanding between method performance and method parameters, always the revalidation protocol has been recommended in procedures [3].

Capillary Electrophoresis (CE) offers many advantages in analytical laboratories in spite of chromatographic techniques due to its low cost, small sample volume, high efficiency, high separation power, eco-friendly characteristics and possibility of performing chiral separations directly dissolving the chiral selector in the background electrolyte [4].

Currently the implemented analytical methods were based on one factor at a time (OFAT), in which one parameter alone is optimized for the expected response whilst others remained constant. This practice always yielded a narrow robust behavior of the method for instrumental variables used in method development phase. Hence the present strategy of analytical method (i.e., OFAT) development has high risk in method failure and always requires revalidation protocol after method transfer or alternative method development; there by it has been increasing the cost of the method [3].

Design Space (DS) concept finally marks the difference between the univariate optimization process approach, that allows only a single optimum point to be found, and experimental design approach that instead allows an optimum multidimensional zone to be obtained. In particular, regulatory documents are strongly recommending the implementation of Quality by Design (QbD) approach that implies the definition of a multidimensional zone, named Design Space, within which all combinations of the variables under study assure the quality of the process with defined probability.

QbD is a science risk-based approach that focuses on safety, efficacy, and quality throughout the product's life cycle.

Implementation of QbD concept is expected to strengthen the concept of "right analytics at right time" which plays significant role in drug product development cycle.

The process of developing and validating analytical methods parallel to product QbD benefits the quality of the product furthermore, with high degree of assurance, and it may be similarly benefitted. The dependence of pharmaceutical development and manufacture on robust analytical data intensifies the need for rigor in analytical method development and increasingly an analytical QbD.

CE is widely used as enantioseparation technique mainly due to its high efficiency and to the possibility of adding the chiral selector directly to the background electrolyte (BGE). In contrast to chromatography, either the binding constant difference between the enantiomers and the chiral selector or the mobility difference of the corresponding diastereomeric complexes may both result in enantioseparations [5].

Stereospecific recognition of chiral molecules is an important issue in various aspects of chemistry and life sciences [6].

Most often, the so called direct approach has been realized, in which stereoisomers, in most cases enantiomers, are separated in a chiral environment containing a chiral selector either fixed to an immobile support or added to the mobile phase or background electrolyte in chromatographic or electromigration techniques, respectively. This approach is based on the formation of transient diastereomeric complexes between the selector and the analyte enantiomers in thermodynamic equilibria [6].

As a matter of fact, CE does not provide any direct information regarding chemical and structural mechanism of analyte-pseudostationary phases interactions, therefore further techniques such as Nuclear Magnetic Resonance (NMR) spectroscopy and Molecular Dynamics Simulations (MD) have been used in combination in order to elucidate these intermolecular interactions. NMR is valuable for enabling conclusions about the spatial proximity of functional groups of host and guest, thus providing information on the involvement of different moieties of the molecules in the intermolecular complex formation [5]. MD constitutes an essential tool for exploring the mechanism of separation and spectroscopic methods in general.

This technique has been used to illustrate selector-selectand complexes. Systematic variations of the analyte structure were applied to deduce structure separation relationships [6]. NMR in particular, should be used to confirm the results of

MD simulations obtaining complementary information about intermolecular interactions.

In this thesis project different CE analytical methods were applied and validated facing up to various pharmaceutical challenges exploiting the implementation of QbD concept in particular, and Design of Experiment in general. MD and NMR were used for the investigation of intermolecular affinities, complexation and separation mechanisms involved in the interactions amoung drug and pseudostationary phases. Six studies regarded the method optimization for the simultaneous quantification of active principles and the main related substances in pharmaceutical formulations, respectively Zolmitriptan, Glibenclamide, Diclofenac, Captopril, Sulpiride, Ambrisentan. The methods were validated following the ICH guidelines in order to fulfill the needed requirements. Furthemore linear discriminant analysis and quadratic discriminant analysis were utilized as discrimination techniques for green tea samples origin discrimination. Thoroughly in some of these studies a comprehensive investigation was performed to deeply explore the intermolecular interactions between analytes and pseudostationary phases and their respective role in the electrophoretic migration by using MD and NMR.

Furthermore, during my abroad collaboration with the Inorganic and Chemistry Department of the University of Geneva the Quality by Desig workflow was applied to underline and study in depth the critical process parameters of Dried Blood Spots sample preparation with liquid chromatography-mass spectrometry.

2. Materials and Methods

The discussed techniques and developed methods presented in this thesis project are currently in press or already been published. Therefore, Materials and methods according each study are reported in chapter 10 “Appendix”.

3. Chemometric techniques

3.1. Design of Experiments (DoE)

Chemometrics, the science of relating measurements made on chemical systems or processes to their state via application of mathematical and statistical modeling has become a well-recognized sub-discipline in contemporary analytical chemistry [7].

Chemometric tools with suitable statistical analysis have become popular by the way, considering multiple advantages viz. reduction in the number of experiments and lower reagent consumption and less laboratory work.

The significance of chemometry has been well reflected by the fact that in the recent years numerous books, book chapters, review papers and innumerable research papers have been published describing its various applications in analytical chemistry. All these works emphasize chemometric approach as an emerging tool in the field of pharmaceutical analysis [7].

Experimental design is a classical statistical method that has found utility in pharmaceutical development, as well as in many other fields requiring determination of optimal conditions and modelling of response surfaces [8]. Statistically designed experiments are highly efficient since they give a fixed amount of information with much less effort than the classical one-variable-at-a-time approach, and many of them give additional information about interactions [9]. Information on interactions is an important clue in the search for optimum conditions, since their existence implies that the value assumed by one factor influences the effect on the response of another factor. With this technique the experimentation evolves in pre-planned stages. When a preliminary screening investigation has been completed, the increased knowledge about the system allows a more detailed study to be performed and a response surface describing the problem to be obtained [10].

Even though the number of variables is high, an empirical relationship between the responses and the individual factors can be obtained with a limited number of experiments. In particular, it was important to detect active factors: the variables whose level change determined a statistically significant variation of the response. Thus, a linear relationship was postulated, and the influence of factors was evaluated by means

of the model coefficients [8].

Linear methods reveal main effects and interactions, but cannot find quadratic (or cubic) effects. Therefore, they have limitations in optimization; the optimum is found in some edge point corresponding linear programming. They cannot model non linear systems; e.g. quadratic phenomena. In an industrial process even third-order models are highly unusual. Therefore, the focus will be on designs that are good for fitting quadratic models. Choice of appropriate types of experimental designs so as to either screen the most influential factors or optimize the selected factors' combination and the mathematical models in chemometry have been briefly recalled and the advantages of chemometric approaches have been emphasized [7].

Experimental designs are often written in terms of coded variables. For example, a design that requires only two values of a variable (so-called 'two-level' design) factor values are usually given as a series of +1 and −1 indicating whether one value (+1) or the other value (−1) is to be chosen. There are mathematical reasons for this practice, but also a practical use is that designs can be written independently of the particular factors under study [11].

Coded variables are calculated by the following equation after settling the *center* and the *step* of variation:

$$X_{ij} = \frac{U_{ij} - \overline{U}_j}{\Delta U_j}$$

where:

X_{ij} is the value of the coded variable *j* in *i-th* test

U_{ij} is the value of natural variable *j* in *i- th* test

$\overline{U}_j$ is the value of natural variable *j* in the *center* of the experimental domain

ΔU_j is the variation *step* of natural variable *j*.

The set of different experimental conditions is given by the *experimental matrix* which is a mathematical object that represents in coded form the experiments to be performed [12]. It consists of a data table consisting of *N* rows, corresponding to *n* experiments, and *K* columns, corresponding to the *k* variables studied. The *xij* element of the matrix thus formed corresponds to the value that assumes *the j*-th variable in the *i-th* experiment.

The *experimental plan* is the translation of an experimental matrix into natural variables: it is a table containing the data expressed in their units of measurement, directly usable by the researcher.

The *effect* of a factor is the change in response observed following level change.

The *response*, indicated by *y*, is the experimental result of an experience (time of analysis, reaction and so on), that is, a measurable manifestation of what is observed by varying the factors studied. The *response* variation is here mathematically schematized:

$$y = f(x_1, x_2, ..., x_k)$$

function *f* is called *response function* and is a polynomial mathematical function that describes the response as factor function. In general, it is possible to represent a problem of this kind in a matrix form:

$$y = X\beta + \varepsilon$$

where *y* is the column vector of the *n* responses, β is the vector of the coefficient estimates, *X* is the matrix *n* x *k* that defines the experimental points and ε is the error column vector.

The experimenter's goal is to find the relation between one or more experimental responses and the factors selected, that is, a mathematical model that relates the response to variables [12].

Because the expected outcome is no significant change of the response, allowing the claim of a rugged/robust method, many factors can be screened without concerns about interacting and non-linear effects. A main effects model will suffice.

$$R = \beta_0 + \sum_{i=1}^{i=k} \beta_i x_i + \varepsilon$$

And with interactions:

$$R = \beta_0 + \sum_{i=1}^{i=k} \beta_i x_i + \sum_{i=1}^{i=k} \sum_{j=i}^{j=k-1} \beta_{ij} x_i x_j + \varepsilon$$

For chromatographic separations it is important to have an acceptable response that meets minimum criteria and so the aim is often to locate that region (e.g. high pH, lower temperature) rather than find the absolute optimum. This makes DoE very powerful when the polynomial function does not fit the data perfectly, but does describe the response sufficiently to locate an acceptable region [11].

In order to have the possibility of a maximum (or minimum) quadratic terms are needed:

$$R = \beta_0 + \sum_{i=1}^{i=k} \beta_i x_i + \sum_{i=1}^{i=k} \sum_{j=i}^{j=k-1} \beta_{ij} x_i x_j + \sum_{i=1}^{i=k} \beta_{ii} x_i^2 + \varepsilon$$

The choice of the model is closely linked to the *experimental domain*: for a restrain domain a simple model is needed. The type of information obtained will depend on the model choice, the number and type of experiments to be conducted, and then the type of experimental planning. The experimental matrix must contain at least a number of experiments equal to the number of coefficients to be determined.

The *model matrix* is obtained from the experimental matrix using the equation of the model: this is made up of of lines coincinding with the number of experiments and a number columns coincinding with the terms present in the model equation. In the case of a linear model without interaction, the matrix of the model differs from the experimental matrix for the only column related to the known term b_0, which is constituted by all +1. In the case of a model with interactions, there is also the addition of columns related to each interaction obtained by multiplying the columns related to the variables for which the interaction is to be known.

The transposed **X '** is calculated from the *model matrix* **(**exchanging its rows with the columns) and then the *information matrix* given by **X'X**, whose inverse is the *dispersion matrix* $\mathbf{(X'X)^{-1}}$, which is useful for calculating the matrix characteristics.

The *dispersion matrix* also appears in the calculation of model coefficients (the β coefficients are not affected by error, while the *b* coefficients estimate the β are inclusive of experimental error), which are obtained by means of linear regression analysis multivariate by applying the following formula:

$$\boldsymbol{B = (X'X)^{-1}X'Y}$$

where B is the vector of estimates b_j, Y is the vector response containing the responses obtained from experiments carried out according to a suitable experimental matrix, $\boldsymbol{X'}$ is the transposition of the model matrix and $\boldsymbol{(X'X)^{-1}}$ is the dispersion matrix.

Once the experiments have been carried out and the coefficients calculated, the results can be analyzed and interpreted primarily through statistical tests. In particular, the application of statistical methods to collected data is made possible by the fact that experiments are performed according to a suitable initial design and therefore according to suitable matrices.

The importance of an accurate estimation of the coefficients consists in the fact that this allows to evaluate the effect each variable has on the response individually or, in the case of an interaction between variables, the effect of such interaction or even in the case of a quadratic term, of the quadratic effect of the variable. In particular are descibed as *main effects*, grouped in *linear* and *quadratic*, and of *interaction effects* between variables. The first give information about the weight applied by the variable to which the effect is related to the response (or curvature of the model if it is a quadratic term); the second are the index of the interaction between variables and hence the effect that a variable has on the effect exerted on the response from another variable. From the coefficient sign, however, you have information on which area of the experimental domain offers the best results [12]. Linear models with quadratic interactions or models can be represented graphically and this graphical representation of the model equation is named the *response surface*. The *response surface* represents the variation of the response in space by varying the different values assumed by the factors. The two-dimensional representation of the response surface constitutes the *isoresponse surface* and is represented by concentric lines along which the response is kept constant.

The *selection of the best experimental strategy* must be carried out as a result of the accurate and complete definition of the problem [13], is assessing the possible presence of experimental limitations, the risk of systematic error, the way in which experimental error is estimated, whether or not need to study all the factors from the beginning. Based on the objectives of an experiment, all the designs can be classified into two broad categories: Screening Designs, Response Surface Designs (optimization

design) here schematized in the Fig. 3.1:

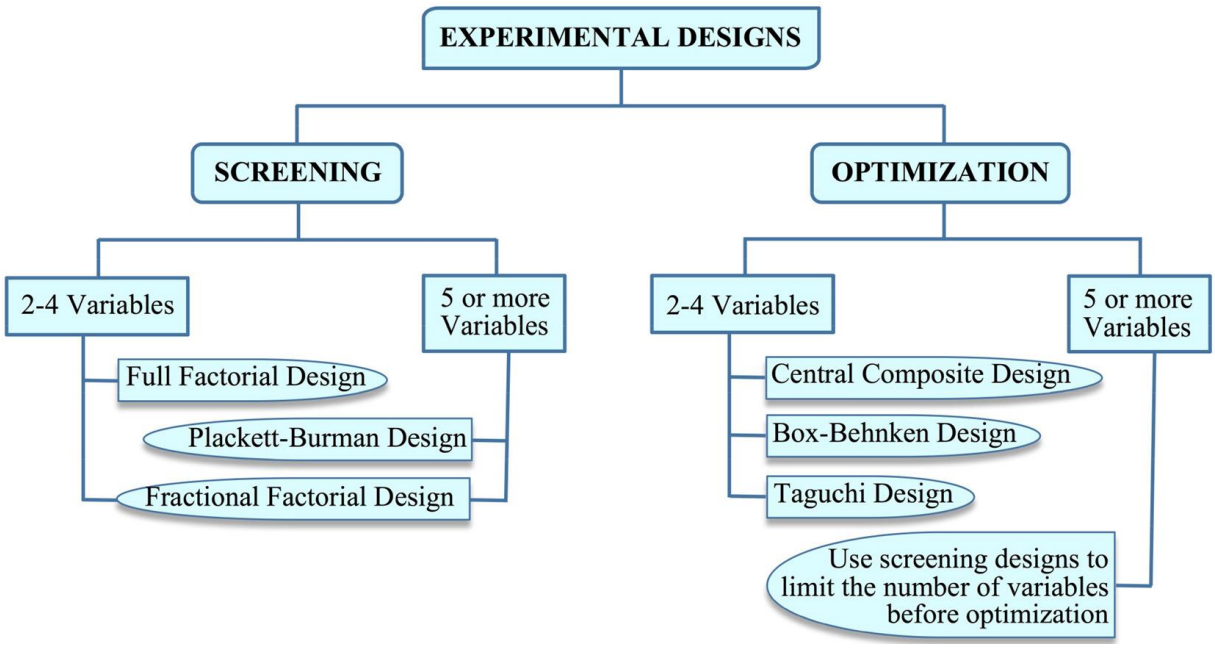


Fig. 3.1. Represantative models of *Screening* and *Optimization* design in the multivariated approach of the Design of Experiments.

3.1.1. Screening Designs

Since a huge number of factors influence the chromatographic process, some of them that do not have significant effect on it must be discarded. Screening of the most influential factors becomes the primary objective of employing experimental design in chromatography. These designs are used with a purport to identify the most important factors and their interactions from all potential factors. They are very useful to examine qualitative, quantitative and mixer-related factors simultaneously [7]. Two-level designs are chosen for screening factors and can give main and interaction effects, but not higher orders. Fractionation leads to designs that give main effects only with fewer runs. Calculation of effects in two level designs is easy and can be performed in a spread sheet. If the two levels are coded +1 and -1, then the column of +1 and -1 under each factor is multiplied by the response for each experiment. The product is summed and divided by half the number of experiments. This is the main effect for the factor. For an interaction effect a column is created that is the product of the level codes and the

procedure outlined above is applied to this column [11].

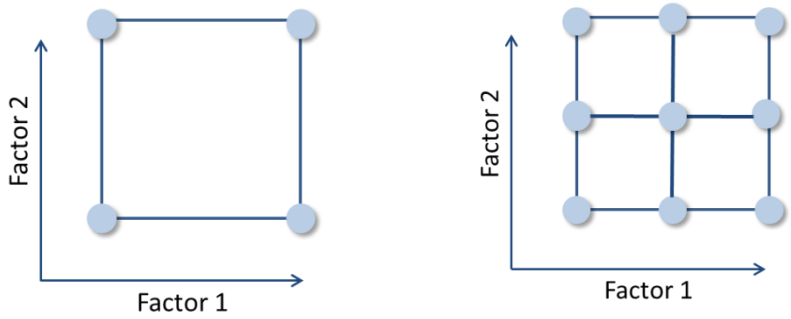


Fig. 3.1.1 Two-level designs for screening factors.

From the literature it is evidenced that Full Factorial designs (FFD), Fractional Factorial designs (FrFD) and Plackett-Burman designs (PBD) are frequently used as screening designs. Such two-level designs allow screening of high number of factors with fewer experiments. Analysis of variance (ANOVA) or regression analysis can be the basis for computing effect of the studied factors on a particular response. They are frequently applied for improvement of separation techniques, formulations, products or processes of quality control and robustness and ruggedness testing. The steps to be performed in such designs are identical to that of robustness or ruggedness test with the discrepancy in the intervals within the two levels of the factors. Several applications of three or more or mixed-level screening designs also has been evidenced from the literature.

3.1.2. Response Surface Designs

Optimization is an additional practice of chemometric approach that endorses the optimal condition or settings of a process.

Such approach usually preceedes with a screening design to select the potential factors. Response surface designs are of two types: symmetrical designs and asymmetrical designs.

Three-level FFD, Central Composite design (CCD) and Box-Behnken design (BBD), Taguchi design (TD), and Doehlert designs cover a symmetrical domain with a center point to estimate experimental error.

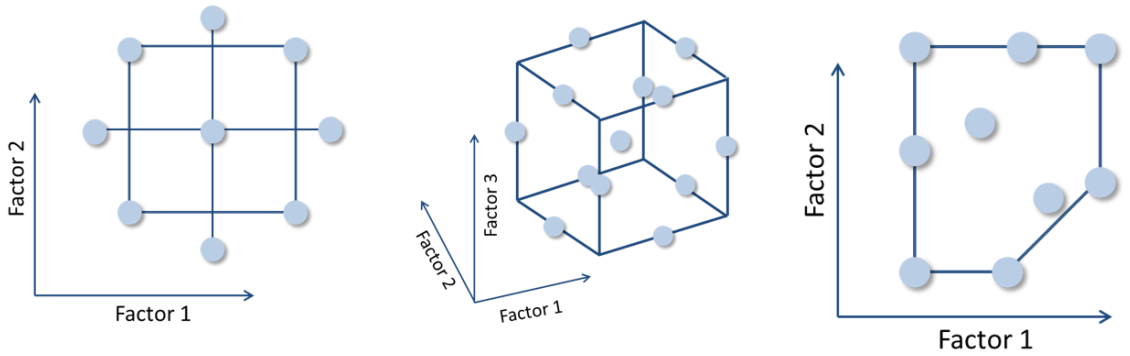


Fig. 3.1.2 Three-level FFD, Central Composite design (CCD) and Box-Behnken design (BBD), Taguchi design (TD), and Doehlert designs cover a symmetrical domain with a center point to estimate experimental error.

Asymmetrical designs such as D-optimal design form an asymmetrical shape when an asymmetrical experimental domain is examined. Such designs can also form a symmetrical shape in a symmetrical domain. Mixture designs are applied to study mixture variables only, i.e. to optimize the composition of mixture. ANOVA, signal-to-noise ratio and range analysis are the basis of the statistical analysis methods for response surface designs. Range analysis is used to find the effect of each factor and determine the optimal level of different factors. For a factor, the range of means is the difference of the maximum and minimum means of all levels. For a system, the factor with the largest range of means has the strongest influence on the performance. Range analysis can find the optimal value of different factors, but this method cannot clearly and quantitatively determine the significance of different factors.

In the ANOVA, the data are analyzed by a F-test. The F value of each factor implies the ratio of the variance for each factor to that of the experimental error. The percentage contribution of each factor is the percentage of the sum of square deviation due to that factor in the total sum of square deviation. It reflects the factor's influence. Regression analysis enables to estimate the relationships among variables via a regression function. Linear first order and second order models are quite common. A fruitful implementation of experimental design in chromatography can be executed

through four common stages; i.e.: (i) choosing the convenient design, (ii) suitable software, (iii) experimental trials, data analysis, and (iv) interpretation.

3.1.3. Analysis of Variance (ANOVA)

In an analysis of *v*ariance, ANOVA, the total variation of the response is defined as a sum of two components; a regression component (SS_{regr}) and a component due to the residuals (SS_{resid}). The sum of squares of the total variation, corrected for the mean (SS), can thus be written as

$$SS = SS_{regr} + SS_{resid}$$

Where $SS = \Sigma(y_i - \bar{y})^2$, $SS_{regr} = \Sigma(\hat{y}_i - \bar{y})^2$, $SS_{resid} = \Sigma(y_i - \hat{y}_i)^2$, y_i are the responses, $\bar{y}$ is the average of the responses, $\hat{y}_i$ are the predicted responses of the selected model. If there are replicates among the experiments, the residuals component are further divided into parts that are the sum of squares of lack of fit (SS_{lof}) and the sum of squares of pure experimental error (SS_{pe}):

$$SS_{resid} = SS_{lof} + SS_{pe}$$

where $SS_{pe} = \Sigma(y_i - \bar{y}_i)^2$ and $\bar{y}_i$ is the average of the obtained responses in the same operative conditions of the replicates [14].

In the ANOVA plot, the regression component of the total variation is compared to the residual component. If the standard deviation of the response explained in the model (SD regression) is larger than the standard deviation of the residuals multiplied by the square root of the critical *F* (RSD*), then the model is significant at the chosen probability level (usually *P*=0.05%) [14].

The lack of fit component of the residuals is compared to the pure experimental error component. If the standard deviation of lack of fit (SD LoF) is larger than the standard deviation of the pure experimental error, multiplied by the square root of the critical *F* (SD pe*), then the model suffers from a significant lack of fit [14].

The comparison of the bars in the ANOVA plots is similar to the ordinary variance ratio test, or *F*-test. In an *F*-test, the variance ratio between lack of fit (lof) and pure experimental error (pe), is being compared to tabled values of *F*-distribution. If $F_{lof/pe}$ exceeds the critical *F*, then there is a significant lack of fit at the probability level that is chosen (usually *P*=0.05) and the model is incorrect. In the ANOVA plots in Modde, the variances are simply exchanged for the standard deviations (SD=√variance), and the comparison SD_{lof}/SD_{pe} vs. the equally converted $\sqrt{F_{crit}}$ has been rearranged to SD_{lof} vs. $SD_{pe}*\sqrt{F_{crit}}$ [14].

3.1.4. Quality Criteria

The matrix quality criteria allow to know beforehand, ie before conducting the experiments, with what accuracy the coefficients will be estimated and then evaluating the quality of matrices to determine what is best suited to the purpose.

Criteria related to estimating model coefficients

The estimate of coefficients must be accurate and not correlated. The β_i coefficient is estimated by b_i and defined as accurate if the b_i -centered interval, with a certain probability of the true coefficient value, is as small as possible; variance of the estimate var (bi) measures that interval.

Therefore, it is possible to assign confidence limits to each coefficient and consider all factors simultaneously, so that the confidence limits of all the coefficients (joint confidence region) are defined by a hyperelliside in the coefficient space. This hyper-bissoid of confidence is centered on vector B, that is, on vector estimate of coefficients b. Hyperellissoid is reduced to an ellipse if the coefficients are only two. The characteristics of this hyperelliside, such as its orientation and extension, can be used as criteria for assessing the quality of the coefficient estimation. Such aspects depend in turn on the properties of the dispersion matrix $\mathbf{(X'X)^{-1}}$.

The *volume* of hyperelliside translates the overall accuracy of the coefficient estimation; for a given probability, smaller is the volume more accurated is the

estimatio. The form of hyperelliside is linked to the accuracy with which each estimate is determined. Because the accuracy in the estimate of the coefficients is the same for all coefficients, the shape of the hyperelliside must be that of a hypersphere. The hyperelliside orientation indicates the degree of correlation between coefficient estimates. Estimates of the coefficients are independent if the main axes of the hyperelliside are parallel to the axes of the coefficients [14].

In particular, as regards the volume of the hyperelliside and hence the accuracy in the estimation of the coefficients, this is proportional to the root of the dispersion matrix determinant $D(\xi) = \left|(XX')^{-1}\right|$.

$$V = a\left|(XX')^{-1}\right|^{1/2}$$

For a given model / experimental matrix, this criterion is defined as the determinant criterion: smaller is the dispersion matrix determinant smaller is the volume of the hyperelliside, and therefore the coefficients estimate is more accurate.

D-optimal design is based on the hyperelliside volume criterion.

The experimental matrix is called *D*-optimal if between all domain matrices is one that leads to the smallest value of the dispersion matrix determinant.

It is therefore the experimental matrix for which the confidence ellipsoid of the coefficients is as small as possible.

Criteria related to the quality of the predicted model

In order to obtain a good predicted model, a model that allows predicting the value of the answer at any point in the experimental domain, the objective is to obtain the best possible quality in predicting the response [14].

Among the quality criteria that indicate the predictive quality of the model, I-optimal criterion is based on the calculation of the average variance of model predictions in the experimental region.

The I-optimal design therefore, minimizing this average variance, produces a more accurate estimate of the above response. This in turn leads to a better placement of the optimal response [15]

3.2. Mixture-process Variable approach (MPV)

The Design Space is the region of the experimental space where multidimensional combinations of mixture components (MCs) and process variables (PVs) have been demonstrated to provide assurance of quality. MCs are the ingredients in a mixture, typically expressed as proportions. PVs are factors in an experiment that do not form any portion of the mixture, but whose settings (when changed) can affect the responses [16].

Using the data resulting from a mixture-process variable (MPV) experimental design, MPV models are developed to represent the relationship of system performance with MCs and PVs. Such data-based models provide an in-depth understanding of the problem and the basis for developing the design space and optimizing the quality of the analytical method [16].

Sometimes a two-stage approach is used to address MPV optimization problems. In the first stage, a mixture experiment is performed, mixture models for the responses are fit to the experimental data, and then the mixture models are used to develop an optimum mixture at fixed settings of the PVs (e.g., at standard or central values of PVs). In the second stage, the PVs are investigated using a response surface experiment, response surface models are fit to the resulting data, and the models are used to develop optimal settings of the PVs for the optimal mixture developed in the first stage. Unfortunately, this two-stage approach to MPV experiments does not provide for investigating interactions between MCs and PVs. This problem is avoided by (1) using a MPV approach that varies both MCs and PVs simultaneously in one experiment, (2) developing MPV models from the resulting data, and (3) using the MPV models to identify the design space, optimal subregion, and desirable MC proportions and PV settings within the optimal subregion [17].

MPV experimental designs and models (that approximate the true, unknown relationships between response variables and the MCs and PVs) can be very large as the number of MCs and/or PVs increases. For example, MPV designs and models are often formed by “crossing” separate mixture designs and models with separate PV designs and models:

$$y_{MC} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \varepsilon \quad (1)$$

Where β_i, β_{ij}, e β_{123} represent, respectively, the linear, quadratic, and special-cubic blending properties of the three MCs, and the x_i represent proportions of the MCs.

The MC proportions x_i satisfy the constraint $x_1+x_2+x_3=1$.

For PVs, the quadratic models is here reported:

$$y_{PV} = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 z_3 + \alpha_{12} z_1 z_2 + \alpha_{13} z_1 z_3 + \alpha_{23} z_2 z_3 + \alpha_{11} z_1^2 + \alpha_{22} z_2^2 + \alpha_{22} z_3^2 + \varepsilon \quad (2)$$

Where α_i, α_{ij}, e α_{ii} represent, respectively, the linear, quadratic, and special-cubic blending properties of the three PVs.

Crossing the models in (1) and (2) involves multiplying all terms in one model by all terms in the other model, yielding a (special-cubic mixture) x (quadratic PV) MPV model of the form:

$$\begin{aligned}
y_{SC-Q} &= g_1^0 x_1 + g_2^0 x_2 + g_3^0 x_3 + g_{12}^0 x_1 x_2 + g_{13}^0 x_1 x_3 + g_{23}^0 x_2 x_3 + g_{123}^0 x_1 x_2 x_3 \\
&+ (g_1^1 x_1 + g_2^1 x_2 + g_3^1 x_3 + g_{12}^1 x_1 x_2 + g_{13}^1 x_1 x_3 + g_{23}^1 x_2 x_3 + g_{123}^1 x_1 x_2 x_3) z_1 \\
&+ (g_1^2 x_1 + g_2^2 x_2 + g_3^2 x_3 + g_{12}^2 x_1 x_2 + g_{13}^2 x_1 x_3 + g_{23}^2 x_2 x_3 + g_{123}^2 x_1 x_2 x_3) z_2 \\
&+ (g_1^3 x_1 + g_2^3 x_2 + g_3^3 x_3 + g_{12}^3 x_1 x_2 + g_{13}^3 x_1 x_3 + g_{23}^3 x_2 x_3 + g_{123}^3 x_1 x_2 x_3) z_3 \\
&+ (g_1^{12} x_1 + g_2^{12} x_2 + g_3^{12} x_3 + g_{12}^{12} x_1 x_2 + g_{13}^{12} x_1 x_3 + g_{23}^{12} x_2 x_3 + g_{123}^{12} x_1 x_2 x_3) z_1 z_2 \\
&+ (g_1^{13} x_1 + g_2^{13} x_2 + g_3^{13} x_3 + g_{12}^{13} x_1 x_2 + g_{13}^{13} x_1 x_3 + g_{23}^{13} x_2 x_3 + g_{123}^{13} x_1 x_2 x_3) z_1 z_3 \\
&+ (g_1^{23} x_1 + g_2^{23} x_2 + g_3^{23} x_3 + g_{12}^{23} x_1 x_2 + g_{13}^{23} x_1 x_3 + g_{23}^{23} x_2 x_3 + g_{123}^{23} x_1 x_2 x_3) z_2 z_3 \\
&+ (g_1^{11} x_1 + g_2^{11} x_2 + g_3^{11} x_3 + g_{12}^{11} x_1 x_2 + g_{13}^{11} x_1 x_3 + g_{23}^{11} x_2 x_3 + g_{123}^{11} x_1 x_2 x_3) z_1^2 \\
&+ (g_1^{22} x_1 + g_2^{22} x_2 + g_3^{22} x_3 + g_{12}^{22} x_1 x_2 + g_{13}^{22} x_1 x_3 + g_{23}^{22} x_2 x_3 + g_{123}^{22} x_1 x_2 x_3) z_2^2 \\
&+ (g_1^{33} x_1 + g_2^{33} x_2 + g_3^{33} x_3 + g_{12}^{33} x_1 x_2 + g_{13}^{33} x_1 x_3 + g_{23}^{33} x_2 x_3 + g_{123}^{33} x_1 x_2 x_3) z_3^2 + \varepsilon
\end{aligned}$$

In models (1), (2) and (3), the ε term at the end of the model represents experimental and measurement error (uncertainty) in the response variable. When models are fit to data using ordinary least squares (OLS) regression, the ε errors are typically assumed to be statistically independent and identically distributed (i.e., with the same mean and variance). When the errors have a normal (Gaussian) distribution, many standard statistical data analysis methods can be applied [16].

The MPV model (3) contains 7×10=70 terms, a large number. An experimental design to support fitting this model would require at least 70 data points to fit the model,

plus extra data points to quantify “pure error” (experimental and measurement uncertainty) and to assess model lack-of-fit (LOF) [16]. A separate mixture design that supports fitting (1) consists of the four vertices, four edge centroids, and overall centroid of the irregular polyhedral region defined by the lower and upper bounds on the MCs. A separate central composite design in the three PVs that supports fitting (2) contains 8 vertices, 6 axial points, and the overall center point of the cube defined by the PV lower and upper bounds. Crossing these designs would yield a MPV design with 9×15=135 points.

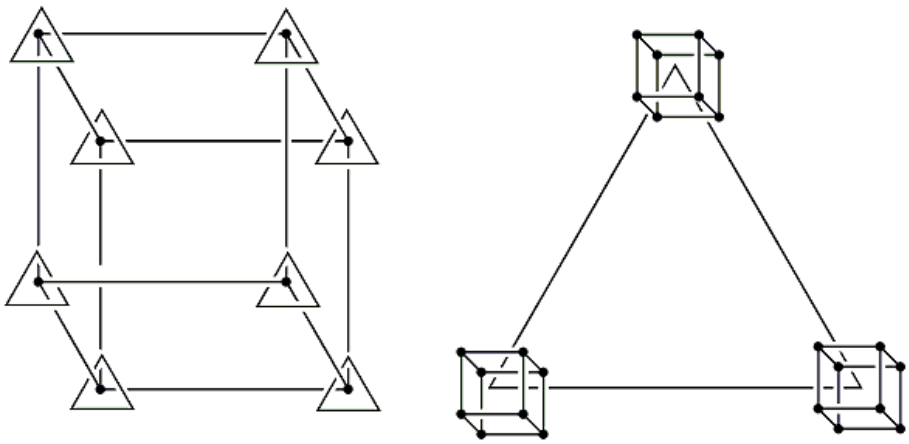


Fig. 3.2. MPV for 3 factors (2^3) with simplex design for a 3 components-mixture.

An alternative for generating MPV designs containing fewer points is to use an *optimal experimental design* approach [16]. With this approach, the researcher specifies a MPV model form that he/she believes will adequately approximate the relationships between the response variables and the MCs and PVs.

Then, software (e.g., Design-Expert [18], JMP [19], Minitab [20]) is used to generate a MPV design (containing a specified number of points) that optimizes a mathematical design criterion.

Atkinson et al. [15] discuss several optimal design criteria, including D-optimality, I-optimality, and others. D-optimality focuses on minimizing the uncertainties of the model coefficients, while I-optimality focuses on minimizing the average uncertainty of model predictions over the experimental region.

The number of points in an optimal MPV design must be at least as large as the number of terms in the MPV model selected for use with the optimality criterion

chosen. Further, the design should have at least 5 replicates to estimate experimental and measurement uncertainty for each response and 10 additional (non-replicate) points to assess the adequacy of the models fitted to the response data.

In summary, care must be taken in (1) selecting a MPV model that will adequately approximate the relationships between responses and the MCs and PVs and (2) generating a MPV experimental design that will support fitting the selected MPV model to experimental data and assessing model adequacy. After MPV models for the response variables have been developed and assessed using the experimental data, the fitted models can then be used to develop equations that specify the design space and determine the settings of the MCs and PVs to optimize the responses [16].

3.3. Principal Component Analysis (PCA)

Principal component analysis (PCA) is a multivariate technique that analyzes a data table in which observations are described by several inter-correlated quantitative dependent variables. Its goal is to extract the important information from the table, to represent it as a set of new orthogonal variables called principal components, and to display the pattern of similarity of the observations and of the variables as points in maps.

The quality of the PCA model can be evaluated using cross-validation techniques such as the bootstrap and the jackknife. PCA can be generalized as correspondence analysis (CA) in order to handle qualitative variables and as multiple factor analysis (MFA) in order to handle heterogeneous sets of variables.

Mathematically, PCA depends upon the eigen-decomposition of positive semi-definite matrices and upon the singular value decomposition (SVD) of rectangular matrices.

In PCA linear combinations of the measured chemical descriptors are derived, producing new variables called principal components (PCs), which are uncorrelated. The first PC (PC1) accounts for the largest portion of explainable data variability in the

measured data, and each successive PC accounts for as much of the remaining variability as possible. PCA cannot be used as a classification tool, but it can be applied with the aim of a better understanding of the discriminating efficiency of the descriptors and of visualizing the data trends, by generating score plots for groupings of the observations [4].

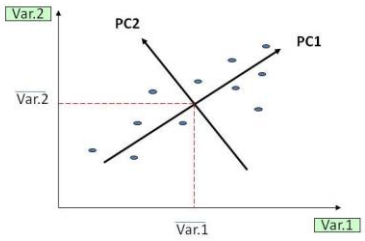


Fig. 3.3. Example of the case of a sample described by two variables by PCA.

In multivariate analysis, each variable *m* defines an axis and hence the variable *m* defines a M-dimensional space. Each object is located in this space by a point, whose coordinates are given by the value of the corresponding variables.

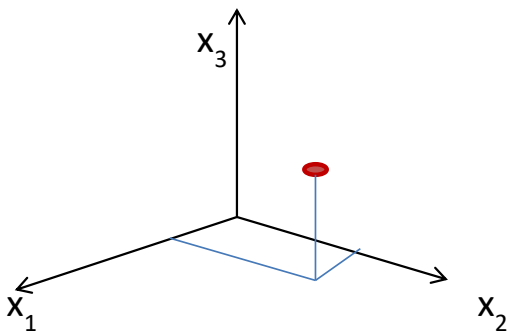


Fig. 3.3.1. Example of the case of a sample described by three variables by PCA.

The *n* objects in the variable m space define a set of points in a multidimensional space. If the objects are similar to each other, this set appears compact, whereas if there

is diversity, a dispersion of objects is observed. The following figure shows an example of a data set described by three variables.

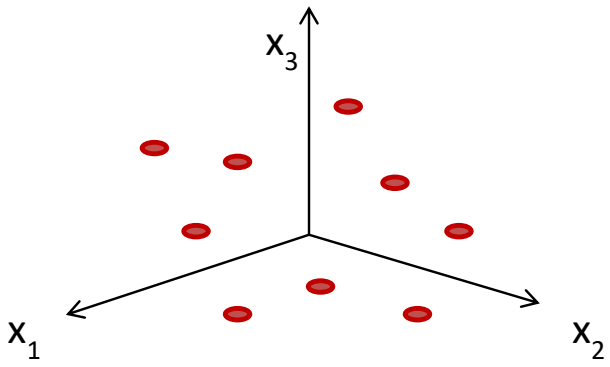


Fig. 3.3.2 Example of the case of a sample described by three variables by PCA..

PCA aims to decrease the number of original variables by operating a orthogonal rotation of Cartesian axes along the line of variation of the points plotted in the plane according to their natural variables. This rotation allows, starting from a large number of more or less correlated variables, to obtain a small number of unrelated variables.

The first PC1 component is then defined in such a way that it explains in its direction the maximum dispersion and hence the maximum variance existing between the data and around which the minimum dispersion is. Once this new axis is identified, PC2 is defined, ie the axis perpendicular to PC1. The two lines thus defined identify a planar pattern and the search for the subsequent components can proceed further until exhaustion of all variance contained in the data. However, the relevant information is usually contained in the first 2 or 3 main components and is therefore identifiable and visualizable.

Each autovector is obtained as a linear combination of the original variables, meaning that each variable gives its contribution more or less to the construction of each new axis. This contribution is expressed through a coefficient called loading. The scores instead represent the new coordinates of objects in the space of the main components. The PCA allows very good graphic representations of both plot scores and variable loads as well as objects and variables simultaneously (biplot). The loadings chart allows you to analyze the role of each variable in the different components and their importance. The score graph allows instead to analyze the behavior of objects in

different components and their similarity. Finally, the biplot graph enables objects and variables simultaneously to be represented in order to evaluate the relationships between them [22].

3.4. Multivariate optimization methods

Linear and quadratic discriminant analysis

Supervised pattern recognition refers to the techniques in which a priori knowledge about the category membership of samples is used for classification. The discrimination model is developed on a training set of samples with categories. The model performance is evaluated by means of some samples from a prediction set by comparing their predicted categories with their own true categories. Therefore, all samples were divided into a training set, used to built model, and an external set (test set), used to test the model. Two supervised pattern recognition techniques were applied: LDA (Linear Discriminant Analysis) and QDA (Quadratic Discriminant Analysis)

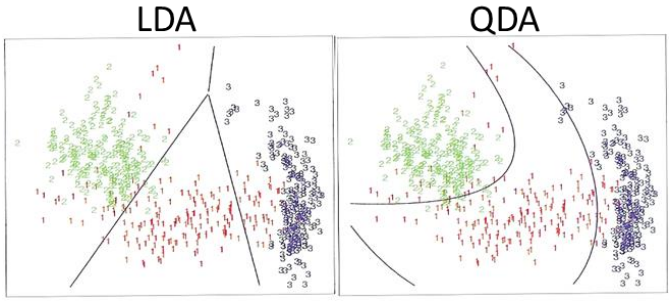


Fig. 3.4. LDA (Linear Discriminant analysis) and QDA (Quadratic Discriminant Analysis).

By using LDA, linear combinations of the selected descriptors can be performed, allowing the so-called discriminant functions to be obtained; the application of this approach best separates the classes according to the minimization of the ratio of within-class and between-class variances. The obtained latent variables are called canonical

variates and for k classes, $k-1$ canonical variates can be calculated [23]. QDA is identical to LDA, but class borders are quadratic curves instead of straight lines.

Prior to LDA and QDA the data were autoscaled and the normal distribution was assumed for the classes in both cases. LDA and QDA discrimination rates by cross-validation were used to optimize the number of variables, plotting the percentage of correct classification rate of the training set samples using cross-validation for the models with different number of variables as a function of the number of variables.

Both LDA and QDA can be derived from simple probabilistic models which model the class conditional distribution of the data

$$P(X|y=k)$$

for each class k. Predictions can then be obtained by using Bayes' rule:

$$P(y=k|X) = \frac{P(X|y=k)P(y=k)}{P(X)} = \frac{P(X|y=k)P(y=k)}{\sum_l P(X|y=l)\cdot P(y=l)}$$

and we select the class k which maximizes this conditional probability. More specifically, for linear and quadratic discriminant analysis

$$P(X|y)$$

is modelled as a multivariate Gaussian distribution with density:

$$p(X|y=k) = \frac{1}{(2\pi)^n|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(X-\mu_k)^t\Sigma_k^{-1}(X-\mu_k)\right)$$

To use this model as a classifier, we just need to estimate from the training data the class priors $P(y=k)$ (by the proportion of instances of class k), the class means μ_k (by the empirical sample class means) and the covariance matrices (either by the empirical sample class covariance matrices, or by a regularized estimator: see the section on shrinkage below).

In the case of LDA, the Gaussians for each class are assumed to share the same covariance matrix: $\Sigma_k = \Sigma$ for all k. This leads to linear decision surfaces between, as can be seen by comparing the log-probability ratios:

$$\log[P(y=k|X)/P(y=l|X)]$$

$$\log\left(\frac{P(y=k|X)}{P(y=l|X)}\right)=0 \Leftrightarrow (\mu_k-\mu_l)\Sigma^{-1}X=\frac{1}{2}(\mu_k^t\Sigma^{-1}\mu_k-\mu_l^t\Sigma^{-1}\mu_l)$$

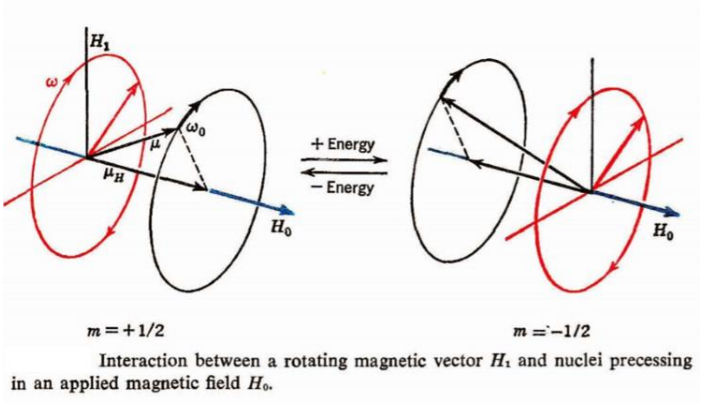
In the case of QDA, there are no assumptions on the covariance matrices Σ_k of the Gaussians, leading to quadratic decision surfaces.

4. Nuclear Magnetic Resonance (NMR) and Molecular Dynamics Simulation (MD)

4.1. Background

Over the past fifty years nuclear magnetic resonance spectroscopy, commonly referred to as NMR, has become the preeminent technique for determining the structure of organic compounds. Of all the spectroscopic methods, it is the only one for which a complete analysis and interpretation of the entire spectrum is normally expected. Although larger amounts of sample are needed than for mass spectroscopy, NMR is non-destructive, and with modern instruments good data may be obtained from samples weighing less than a milligram. Nuclear magnetic resonance (NMR) spectroscopy was discovered shortly after the Second World War, and since then its applications to chemistry have been continuously expanding. It was natural then that NMR took an important part in undergraduate chemistry education, being taught within various courses: physical chemistry, organic, inorganic and analytical chemistry. In recent years, the applications of NMR have been extended to biology and medicine, so they have also become an integral part of courses in these subjects [25]. NMR spectroscopy in depth is an analytical chemistry technique used in quality control and research for determining the content and purity of a sample as well as its molecular structure. For

example, NMR can quantitatively analyze mixtures containing known compounds. For unknown compounds, NMR can either be used to match against spectral libraries or to infer the basic structure directly.



Interaction between a rotating magnetic vector H_1 and nuclei precessing in an applied magnetic field H_0.

NMR technique is based on the fact that when a population of magnetic nuclei is placed in an external magnetic field, the nuclei become aligned in a predictable and finite number of orientations [25]. For ^{1}H there are two different orientations. The first one is the orientation represented by the concident alignment of protons with the external magnetic field (north pole of the nucleus aligned with the south pole of the magnet and south pole of the nucleus with the north pole of the magnet) and the second one is represented by the alignment of the nuclei against the field (north with north, south with south). The alignment with the field is also called the "alpha" orientation and the alignment against the field is called the "beta" orientation.

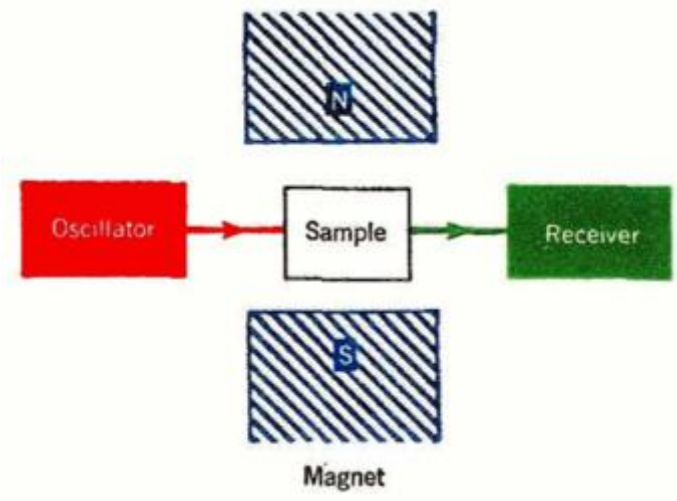


Fig. 4.1. Example of a NMR technique.

This angular moment (also called spin) has an associated magnetic moment. A few examples of magnetic isotopes are ^{13}C, ^{1}H, ^{19}F,^{14}N, ^{17}O, ^{31}P, and ^{33}S. If a nucleus is not magnetic, it can't be studied by nuclear magnetic resonance spectroscopy.

The nuclear magnetic resonance (NMR) spectroscopy experiment involves using energy in the form of electromagnetic radiation to pump the excess alpha oriented nuclei into the beta state. When the energy is removed, the energized nuclei relax back to the alpha state. The fluctuation of the magnetic field associated with this relaxation process is called resonance and this resonance can be detected and converted into the peaks recovered in an NMR spectrum. Once the basic structure is known, NMR can be used to determine molecular conformation in solution as well as studying physical properties at the molecular level such as conformational exchange, phase changes, solubility, and diffusion. In order to achieve the desired results, a variety of NMR techniques are available. NMR spectra are usually measured using solutions of the substance being investigated. It is important that the solvent itself doesn't contain any simple hydrogen atoms, because they would produce confusing peaks in the spectrum. In order to avoid this interference a solvent such as tetrachloromethane, CCl_4, which doesn't contain any hydrogen, can be used, or a solvent in which any ordinary hydrogen atoms are replaced by its isotope, deuterium - for example, $CDCl_3$ instead of $CHCl_3$. All the NMR spectra used on this site involve $CDCl_3$ as the solvent.

Deuterium atoms have sufficiently different magnetic properties from ordinary hydrogen that they don't produce peaks in the area of the desired spectrum.

4.2. Chemical Effects in NMR

4.2.1. Chemical Shift

Unlike infrared and uv-visible spectroscopy, where absorption peaks are uniquely located by a frequency or wavelength, the location of different nmr resonance signals is dependent on both the external magnetic field strength and the rf frequency. Since no two magnets will have exactly the same field, resonance frequencies will vary accordingly and an alternative method for characterizing and specifying the location of nmr signals is needed. The precise resonant frequency of the energy transition is dependent on the effective magnetic field at the nucleus. This field is affected by

electron shielding which is in turn dependent on the chemical environment. As a result, information about the nucleus' chemical environment can be derived from its resonant frequency. In general, the more electronegative the nucleus is, the higher the resonant frequency. Other factors such as ring currents (anisotropy) and bond strain affect the frequency shift. It is customary to adopt tetramethylsilane (TMS) as the proton reference frequency. This is because the precise resonant frequency shift of each nucleus depends on the magnetic field used. The frequency is not easy to remember (for example, the frequency of benzene might be 400.132869 MHz) so it was decided to define chemical shift as follows to yield a more convenient number such as 7.17 ppm.

$$\delta = (\nu - \nu_0)/\nu_0$$

The ^{1}H chemical shift, using this equation, is not dependent on the magnetic field and it is convenient to express it in ppm where (for proton) TMS is set to ν_0 thereby giving it a chemical shift of zero. For other nuclei, ν_0 is defined as Ξ ν_{TMS} where Ξ (Greek letter Xsi) is the frequency ratio of the nucleus (*e. g.*, 25.145020 % for ^{13}C). In the δ-scale more highly shielded protons appear at lower δ values, whereas

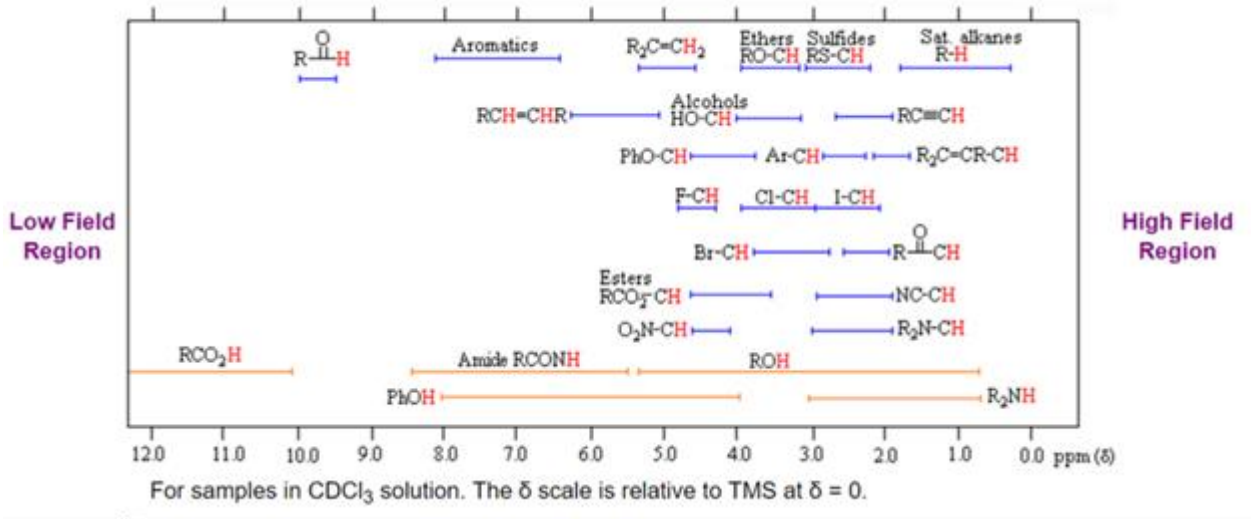


Fig. 4.2.1 The ^{1}H chemical shift for samples in $CDCl_3$.

relatively deshielded protons appear at higher δ values. Therefore, protons attached to carbon, which is less electronegative than oxygen, will be more shielded and

will typically appear around 1 ppm. Protons attached to oxygen, or those which are relatively close to more electronegative atoms, will be deshielded and will appear at higher δ values, as shown previously.

In the case of the ^{1}H NMR spectrum of ethyl benzene, the methyl (CH_3) group is the most electron withdrawing (electronegative) and therefore resonates at the lowest chemical shift. The aromatic phenyl group is the most electron donating (electropositive) so has the highest chemical shift. The methylene (CH_2) falls somewhere in the middle. However, if the chemical shift of the aromatics were due to electropositivity alone, then they would resonate between four and five ppm. The increased chemical shift is due to the delocalized ring current of the phenyl group.

4.2.2. Absorption Intensities

The magnitude or intensity of NMR resonance signals is displayed along the vertical axis of a spectrum, and is proportional to the molar concentration of the sample. Thus, a small or dilute sample will give a weak signal, and doubling or tripling the sample concentration increases the signal strength proportionally. If we take the NMR spectrum of equal molar amounts of benzene and cyclohexane in carbon tetrachloride solution, the resonance signal from cyclohexane will be twice as intense as that from benzene because cyclohexane has twice as many hydrogens per molecule [25]. This is an important relationship when samples incorporating two or more different sets of hydrogen atoms are examined, since it allows the ratio of hydrogen atoms in each distinct set to be determined. To this end it is necessary to measure the relative strength as well as the chemical shift of the resonance signals that comprise an NMR spectrum.

4.2.3. Spin-Spin Coupling

The source of signal splitting is a phenomenon called spin-spin coupling, a term that describes the magnetic interactions between neighboring, non-equivalent NMR-

active nuclei. The splitting patterns found in various spectra are easily recognized, provided the chemical shifts of the different sets of hydrogen that generate the signals differ by two or more ppm. The patterns are symmetrically distributed on both sides of the proton chemical shift, and the central lines are always stronger than the outer lines. The most commonly observed patterns have been given descriptive names, such as *doublet* (two equal intensity signals), *triplet* (three signals with an intensity ratio of 1:2:1) and *quartet* (a set of four signals with intensities of 1:3:3:1). The line separation is always constant within a given multiplet, and is called the *coupling constant (J)*.

The splitting patterns display the ideal or *"First-Order*" arrangement of lines. This is usually observed if the spin-coupled nuclei have very different chemical shifts (i.e. $\Delta\nu$ is large compared to *J*). If the coupled nuclei have similar chemical shifts, the splitting patterns are distorted (second order behavior). In fact, signal splitting disappears if the chemical shifts are the same.

The scalar coupling constant of two nuclei separated by *n* bonds is denoted as ^{n}J. As the number of the bonds that separate the two coupled nuclei increases, the interacting energy becomes lower and consequently the coupling constant becomes smaller.

On the other hand, ^{2}J and ^{4}J give smaller but still detectable effects in a spectrum. Couplings over larger number of bonds ($n \geq 5$) can generally be ignored [15]. An additional point is that *J* varies with the dihedral angle between the bonds. Thus, a $^{3}J_{HH}$ coupling constant is often found to depend on the angle φ according to the Karplus equation:

$$J = A + B\cos\varphi + C\cos 2\varphi$$

with *A, B, C* empirical constants taking values close to + 7 Hz, - 1 Hz, + 5 Hz respectively. Consequently, the measurement of ^{3}JHH in a series of related compounds can be used to determine their conformations

4.3. Carbon NMR Spectroscopy

The most important operational technique that has led to successful and routine ^{13}C NMR spectroscopy is the use of high-field pulse technology coupled with broad-band heteronuclear decoupling of all protons.

The results of repeated pulse sequences are accumulated to provide improved signal strength.

Also, for reasons that go beyond the present treatment, the decoupling irradiation enhances the sensitivity of carbon nuclei bonded to hydrogen. When acquired in this manner, the carbon NMR spectrum of a compound displays a single sharp signal for each structurally distinct carbon atom in a molecule (remember, the proton couplings have been removed). The dispersion of ^{13}C chemical shifts is nearly twenty times greater than that for protons, and this together with the lack of signal splitting makes it more likely that every structurally distinct carbon atom will produce a separate signal. The only clearly identifiable signals in the proton spectrum are those from the methyl groups.

The remaining protons have resonance signals between 1.0 and 2.8 ppm from TMS, and they overlap badly thanks to spin-spin splitting.

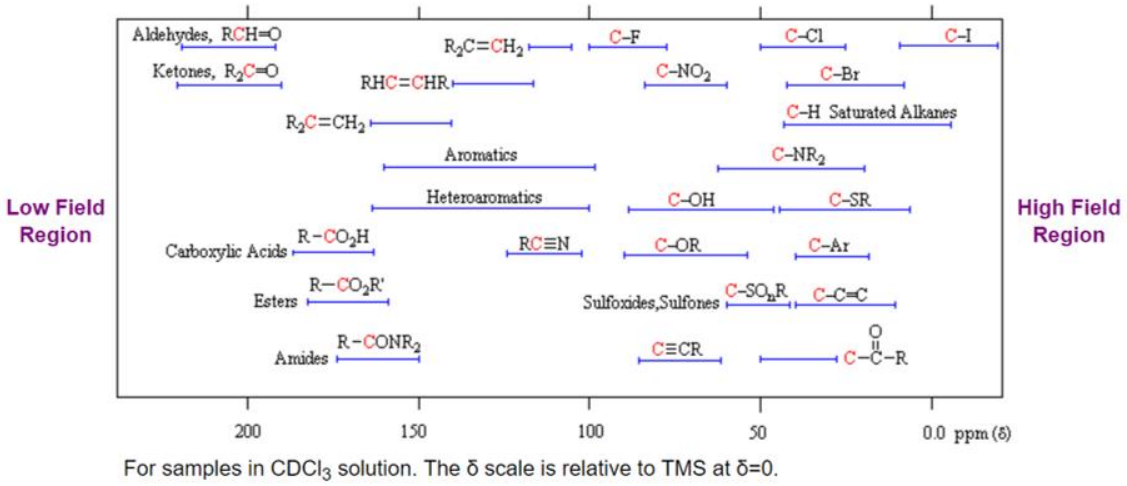


Fig. 4.3. The ^{13}C NMR spectroscopy for samples in $CDCl_3$.

Unlike proton NMR spectroscopy, the relative strength of carbon NMR signals are not normally proportional to the number of atoms generating each one. Because of this, the number of discrete signals and their chemical shifts are the most important pieces of evidence delivered by a carbon spectrum.

The general distribution of carbon chemical shifts associated with different functional groups is summarized in the Fig. 4.3.

Bear in mind that these ranges are approximate, and may not encompass all compounds of a given class.

Note also that the over 200 ppm range of chemical shifts shown here is much greater than that observed for proton chemical shifts.

4.4. Two-dimensional nuclear magnetic resonance

Two-dimensional (2D) NMR spectroscopy is of unprecedented value for structure determination by NMR.

In this paragraph, the general concepts of 2D NMR spectroscopy and the 2D experiments used in this thesis will be shortly introduced.

To other texts is referred for more in-depth information (100, 102, 104).. There are two frequency axes in a 2D NMR spectrum and it is possible to generate cross-peaks between connected resonances.

The nature of these cross-peaks and the structural information which can be extracted from such a spectrum depends on the type of 2D experiment used.

Traditional 2D NMR is a 2D map consisting of two axes of independent variables, both in frequency, denoting the chemical shifts or coupling constants.

The contour map of the signal intensity in the 2D domain represents the 2D spectroscopic correlation, which delineates the correlation of different atomic sites due to chemical exchange, diffusion, *J* coupling or quantum hopping.

The correlation between the chemical shift and the diffusion coefficient, called “diffusion ordered nuclear magnetic resonance spectroscopy (DOSYNMR),” has also been explored extensively [24].

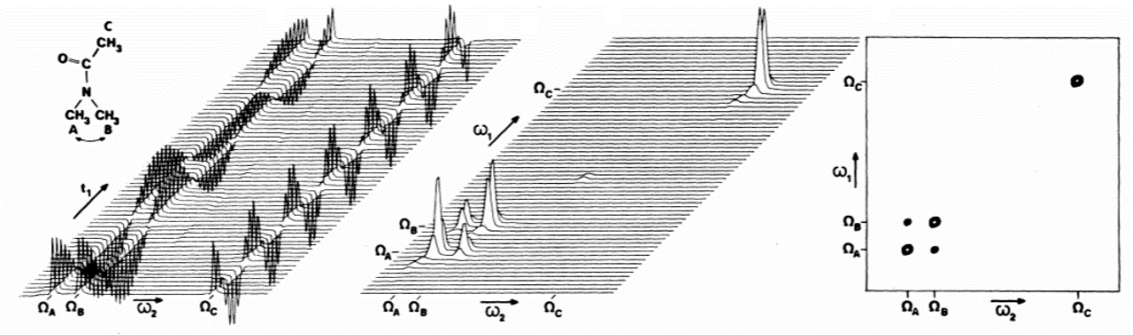


Fig. 4.4. The contour map of the signal intensity in the 2D domain represents the 2D spectroscopic correlation.

During the preparation time, a single 90° pulse can be applied to generate transverse magnetization, or, for example, a set of pulses is applied to transfer magnetization from one nucleus to another (e.g. INEPT (*105*)). The t_1-evolution period is incremented systematically in a series of separate experiments. During this period, magnetization evolves under influence of chemical shift and/or *J* coupling. The mixing period is applied to either transfer unobservable magnetization into observable magnetization (e.g. COSY) or to transfer magnetization from one nucleus to another by for example cross-relaxation (NOESY) or *J*-coupling (TOCSY). During the detection time t_2, magnetization evolves under influence of chemical shift and *J*-coupling and is detected [25].

Subsequent Fourier transformation in both dimension (t_1 and t_2) results in 2D NMR spectra with frequency domains in both dimensions. Cross peaks in Correlation Spectroscopy (COSY) spectra reveal the network of *J*-couplings in a molecule. This information is of great value for structural determination and 2D COSY is therefore one of the most commonly used 2D NMR experiments. Double Quantum Filtered COSY (DQF-COSY) experiments have the preference over the conventional COSY experiment, since diagonal and cross peaks have absorption line shapes, whereas in conventional COSY spectra, there is a 90° phase difference between the diagonal and cross peaks. This leads to long dispersion tails of the diagonal peaks which hide cross peaks resonating close to the diagonal peaks [25].

The only disadvantage of DQF-COSY experiments is the two-fold reduction in sensitivity compared to the conventional COSY experiment. Total Correlation Spectroscopy (TOCSY) is also a homonuclear 2D NMR experiment. TOCSY has large

similarities with COSY, since TOCSY gives also information about the network of *J*-couplings in a molecule. In TOCSY spectra, however, also cross peaks between spins which are indirectly coupled to each other are present. For example, if spin A is coupled to spin B and spin B is coupled to spin C, a crosspeak between A and C will be present in a TOCSY spectrum, even though there is no *J*-coupling between those spins.

TOCSY spectra have absorption mode lineshapes and are extremely valuable in identifying spins which belong to an extended network of couplings.

Nuclear Overhauser effect spectroscopy (NOESY) spectra show cross peaks for all spins for which cross relaxation has occurred. In other words, NOESY spectra show only correlations between spins which are reasonable close in space to each other (<5Å).

The intensity of a correlation in a NOESY spectrum is directly proportional to the distance between spins and NOESY spectra are therefore commonly used to obtain conformational and geometric information from biomolecules as well as small molecules.

The final 2D NMR experiments which will be introduced here are the Hetereonuclear Single Quantum Coherence (HSQC) and Hetereonuclear Multiple Bond Coherence (HMBC) experiments.

The resulting spectra of these experiments provide information about protons which are directly attached to a carbon (^{1}H-^{13}C HSQC) or between protons and carbons which are separated from each other with two or more chemical bonds (^{1}H-^{13}C HMBC).

The latter one is commonly used to assign quaternary, carbonyl or other carbons which are not bound to a proton. NMR experiments with higher dimensionalities (e.g. three and four dimensional) are based on the same principle as 2D NMR experiments, with the main difference being the additional evolution periods.

The spectra resulting from these experiments are simplified significantly, since spectral overlap is reduced. NMR experiments with higher dimensionalities are commonly applied to derive structural information of larger molecules like nucleic acids and proteins.

The most important requirement of these NMR experiments is the isotope labelling (^{13}C, ^{15}N and/or ^{2}H), since the natural abundance of these isotopes is in general too low to obtain spectra of sufficient quality [25].

4.5. Molecular Dynamics Simulations (MD)

The molecular dynamics method was first introduced by Alder and Wainwright in the late 1950's (Alder and Wainwright, 1957,1959) to study the interactions of hard spheres. Many important insights concerning the behavior of simple liquids emerged from their studies.

The next major advance was in 1964, when Rahman carried out the first simulation using a realistic potential for liquid argon (Rahman, 1964).

The first molecular dynamics simulation of a realistic system was done by Rahman and Stillinger in their simulation of liquid water in 1974 (Stillinger and Rahman, 1974).

The first protein simulations appeared in 1977 with the simulation of the bovine pancreatic trypsin inhibitor (BPTI) (McCammon, *et al*, 1977).

Today in the literature, one routinely finds molecular dynamics simulations of solvated proteins, protein-DNA complexes as well as lipid systems addressing a variety of issues including the thermodynamics of ligand binding and the folding of small proteins.

The number of simulation techniques has greatly expanded; there exist now many specialized techniques for particular problems, including mixed quantum mechanical - classical simulations, that are being employed to study enzymatic reactions in the context of the full protein.

Molecular dynamics simulation techniques are widely used in experimental procedures such as X-ray crystallography and NMR structure determination.

Simulation of systems having ~50,000–100,000 atoms are now routine, and simulations of approximately 500,000 atoms are common when the appropriate computer facilities are available [25].

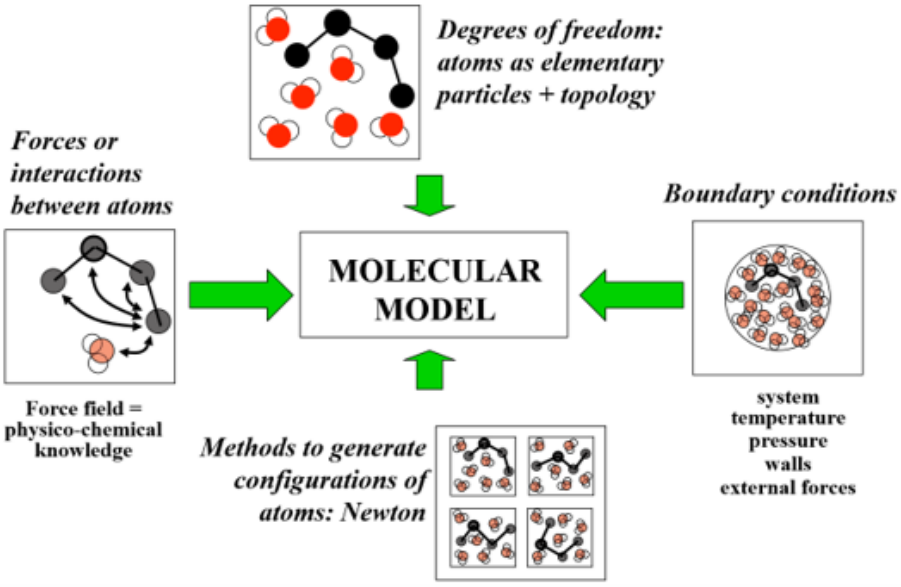


Fig. 4.5. Summary of the frequent applications of Molecular Dynamic Simulations.

This computational method calculates the time dependent behavior of a molecular system. MD simulations have provided detailed information on the fluctuations and conformational changes of proteins and nucleic acids. These methods are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes. They are also used in the determination of structures from X-ray crystallography and from NMR experiments.

Molecular dynamics simulations generate information at the microscopic level, including atomic positions and velocities. The conversion of this microscopic information to macroscopic observables such as pressure, energy, heat capacities, etc., requires statistical mechanics. Statistical mechanics is fundamental to the study of biological systems by molecular dynamics simulation.

Simulations can provide the ultimate detail concerning individual particle motions as a function of time. Thus, they can be used to address specific questions about the properties of a model system, often more easily than experiments on the actual system

The present generation of computers takes benefit of parallelism and accelerators to speed-up the process. The most popular simulation codes (AMBER [26], CHARMM [26], GROMACS [26], or NAMD [26]) have long been compatible with the messaging passing interface (MPI). When a large number of computer cores can be used simultaneously, MPI can greatly reduce the computation time. To benefit the locality of

interactions, the general strategy is to distribute the system to simulate among processors. This strategy is called spatial decomposition. Only a small fragment of the system has to be simulated in each processor. The most efficient division is not based in the list of particles, but in their position in space. Each processor deals with a region of space irrespective of which particles are present there [26].

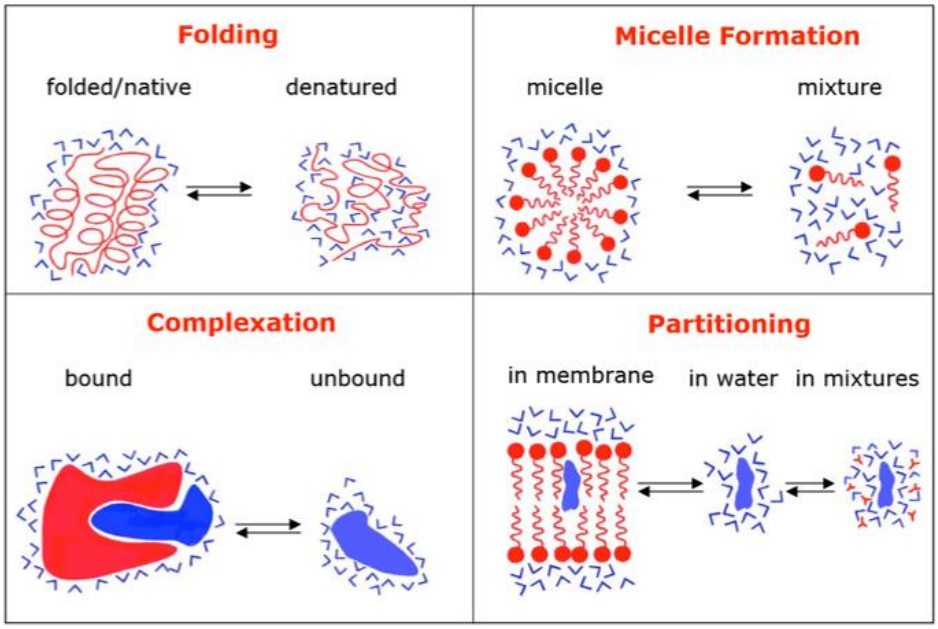
One of the most practical application of the concept of molecular recognition are docking strategies, either small molecule or protein docking. To understand how a ligand, typically a substrate or a regulator, binds to its macromolecular counterpart is a key issue in the understanding of function itself, and it is the basis of structurally driven drug design. The recognition process is by nature dynamic. Molecules are flexible entities, and the recognition process itself implies structural rearrangements, and this shape adjustment is part of the binding process not only from the structural point of view, but also from the energetics. Although this is a generally well-accepted idea, docking algorithms are far from considering dynamic effects as a routine [26].

In a molecular dynamics simulation, one often wishes to explore the macroscopic properties of a system through microscopic simulations, for example, to calculate changes in the binding free energy of a particular drug candidate, or to examine the energetics and mechanisms of conformational change. The connection between microscopic simulations and macroscopic properties is made via *statistical mechanics* which provides the rigorous mathematical expressions that relate macroscopic properties to the distribution and motion of the atoms and molecules of the N-body system; molecular dynamics simulations provide the means to solve the equation of motion of the particles and evaluate these mathematical formulas. With molecular dynamics simulations, one can study both thermodynamic properties and/or time dependent (kinetic) phenomenon.

An initial model of the system is obtained from either experimental structures or comparative modeling data. The simulated system could be represented at different levels of detail. Atomistic representation is the one that leads to the best reproduction of the actual systems [26].

There are three types of applications of simulation methods in the macromolecular area, as well as in other areas involving mesoscopic systems. The first uses simulation simply as a means of sampling configuration space. This is involved in the utilization of

molecular dynamics, often with simulated annealing protocols, to determine or refine structures with data obtained from experiments, as mentioned above. The second uses simulations to obtain a description of the system at equilibrium, including structural and motional properties (for example, atomic mean-square fluctuation amplitudes) and the values of thermodynamic parameters. For such applications, it is necessary that the simulations adequately sample configuration space, as in the first application, with the additional condition that each point be weighted by the appropriate Boltzmann factor.



The third area uses simulations to examine the actual dynamics. Here not only is adequate sampling of configuration space with appropriate Boltzmann weighting required, but it must be done so as to correctly represent the development of the system over time. For the first two areas, Monte Carlo simulations can be used, as well as molecular dynamics. By contrast, in the third area where the motions and their development with time are of primary interest, only molecular dynamics can provide the necessary information. The three sets of applications make increasing demands on simulation methods as to their required accuracy and precision [27].

Solvent representation is a key issue in system definition [26].

The essence of molecular modeling resides in the connection between the *macroscopic* world and the *microscopic* world provided by the theory of statistical mechanics.

Computational tools involved are:

- **Quantum Mechanics (QM),** *Electronic structure (Schrödinger)*

- **Classical Molecular Mechanics (MM),** *Empirical forces (Newton)*

- **Mixed Quantum/Classical (QM/MM)**

For an atomic system, the potential energy function consists of a set of equations that empirically describe bonded and non-bonded interactions between atoms. This energy function together with the set of its empirical parameters is referred to as the "force field." Molecular dynamics force fields usually consist of two major components. The first part describes interactions between atoms connected via covalent bonds, which typically includes bonds, bond angles, and dihedrals. The second part treats non-bonded interactions, typically as electrostatic interactions between the (partial) charges on each atom and a Lennard-Jones potential to model dispersive van der Waals interactions. Each molecule in the simulation is described by its 'topology', the combination of the set of all atoms with their non-bonded and bonded interaction parameters and the connectivity of those atoms in the molecule [28].

When a small molecule like a drug (for example, a ligand) approaches its target (for example, a receptor) in solution, it encounters not a single, frozen structure, but rather a macromolecule in constant motion.

MD studies are also limited by the short time scales typically simulated. To reproduce thermodynamic properties and/or to fully elucidate all binding-pocket configurations relevant to drug design, all the possible conformational states of the protein must be explored by the simulation. Unfortunately, many biochemical processes, including receptor conformational shifts relevant to drug binding, occur on time scales that are much longer than those amenable to simulation. With some important exceptions, simulations are currently limited to at most millionths of a second; indeed, most simulations are measured in billionths of a second.

5. Capillary Electrophoresis

5.1. Basic Principles

Capillary electrophoresis (CE) was born of the marriage of the powerful separation mechanisms of electrophoresis with the instrumentation and automation concepts of chromatography [29].

Capillary Electrophoresis (CE) had a very significant impact on the field of analytical chemistry in recent years as the technique is capable of very high-resolution separations, requiring only small amounts of samples and reagents. Furthermore, it can be readily adapted to automatic sample handling and real-time data processing.

Many new methodologies based on CE have been reported.

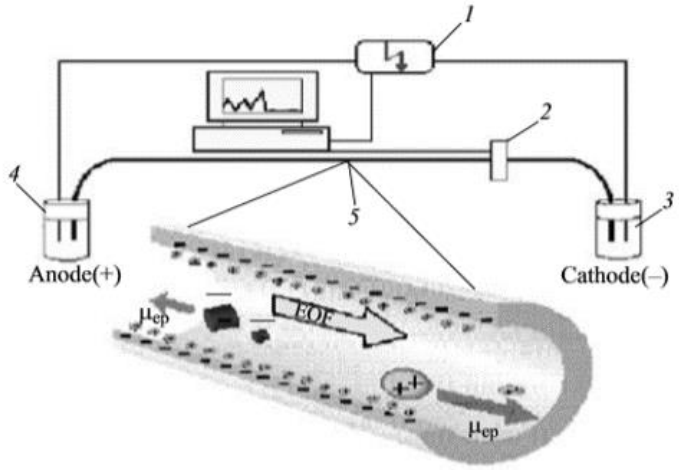
Rapid, reproducible separations of extremely small amounts of chemicals and biochemicals, including peptides, proteins, nucleotides, DNA, enantiomers, carbohydrates, vitamins, inorganic ions, pharmaceuticals and environmental pollutants have been demonstrated.

A wide range of applications have been developed in greatly diverse fields, such as chemical, biotechnological, environmental and pharmaceutical analysis [30].

Electrophoresis can be defined as the differential migration of charged species (ions) in an electric field, and was first described as a separation technique by Tiselius in 1937.

In its simplest form, capillary electrophoresis involves the separation of charged analytes, based on the difference in their electrophoretic mobilities, resulting in different migration velocities.

These separations are carried out in fused silica capillaries, typically 25–75 μm i.d. and 50–100 cm in length, filled with a background electrolyte (BGE) [30]. Capillaries were introduced into electrophoresis as an anti-convective and heat controlling innovation [30].



Schematic diagram of the CE system: (*1*) high voltage source; (*2*) detector; (*3*, *4*) compartments with electrolyte; (*5*) quartz capillary. The inset shows the electroosmotic flow (EOF) and the directions of the electrophoretic mobility μ_{ep} of positively and negatively charged particles.

When an electrical field is applied, the components begin to move in the field as previously described. The narrow capillary reduces lateral diffusion and insures that temperature differences between the center of the capillary and the wall are quite small. The geometry and other properties of the capillary electrophoretic separation also lead to a condition known as plug flow [29]. The capillary electrophoresis mechanism is based on two basic principles: electroosmotic flow and electrophoretic mobility.

Electroosmosis

The small diameter of the capillary contributes to another aspect of the separation process. This is the phenomenon known as electroosmosis, electroendoosmotic flow, or simply EOF. EOF exists in any electrophoretic system.

It will occur whenever the liquid near a charged surface is placed in an electrical field resulting in the bulk movement of fluid near that surface.

Because the surface to volume ratio is very high inside a capillary, EOF becomes a significant factor in CE.

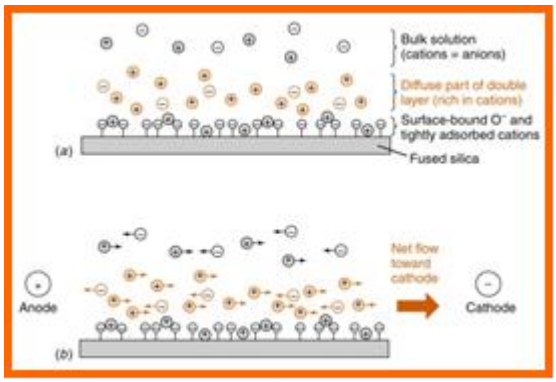


Fig. 5.1. The migration of the EOF inside the fused silica capillary by the application of a electric field.

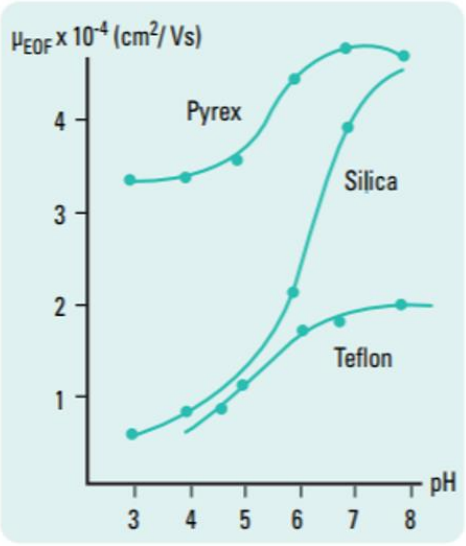
The velocity of the electroosmotic flow through a capillary is given by the Smoluchowski equation [31]:

$$\nu_{eof} = -(\varepsilon\zeta/4\pi\eta)E$$

where ε is the dielectric constant of the electrolyte, ζ is the zeta potential (Volts), η is the viscosity (Poise), and E is the potential applied (Volts/cm). In CE the zeta potential (ζ) is a measure of the charge on the wall of the capillary.

This charge arises from both the nature of the material that composes the capillary and the composition of the electrolyte (buffer). The most commonly employed capillary material is fused silica.

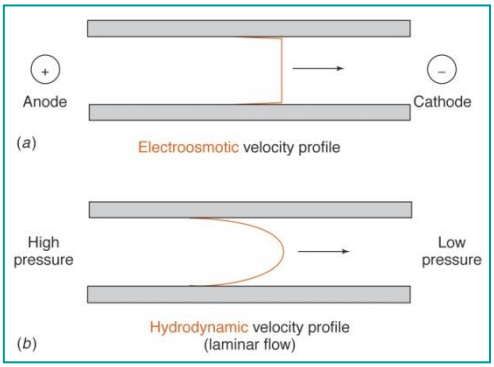
The surface of a fused silica capillary can be ionized to yield a negatively charged surface.



Effect of pH on electroosmotic flow mobility in various capillary materials.

The negatively charged wall attracts cations that are hydrated from the electrolyte solution, creating an electrical double layer. In an electrical field they migrate toward

the cathode, pulling water along and creating a pumping action. The zeta potential increases with the density of the charge on the surface. For fused silica and many other materials charge densities will vary with pH [31]. Bare fused silica behaves much like a weak acid with a pKa of 6.25. The relationship between EOF and pH is depicted in the graph above.



EOF also decreases with the square root of the concentration of the electrolyte, i.e., increasing buffer concentration decreases the velocity of EOF. Although the aforementioned discussion has focused on bare fused silica capillaries, other types of surfaces can be chemically created from fused silica.

The new surface may be positive, negative, or neutral. The direction of the electroosmotic flow will therefore depend on the sign of the charge on the wall of the capillary. Flow is always toward the electrode that has the same charge as the capillary wall. Thus, an uncharged wall will have, in theory, no EOF. In reality this is difficult to accomplish. In the narrow confines of the capillary the velocity of liquid is nearly uniform across the i.d. of the capillary resulting in what has been termed “plug flow”.

This is in contrast to the laminar flow exhibited by pumps systems, which creates a velocity profile across the diameter of the tube. On the other hand, plug flow greatly reduces the band broadening seen in system such as HPLC that are pumped by a pressure differential.

A comparison of laminar flow and plug flow is shown in the picture above. In practice it is very difficult to completely eliminate EOF, although it can be reduced to a value near zero. In many separations it is a significant factor, and it may affect the net movement of the analyte molecules more than does the electrophoretic force. The movement (vectors) due to the electrophoretic and electroosmotic forces may even be in opposite directions.

Electrophoretic migration in capillary tubes

The CE system consists of a buffer-filled capillary placed between two buffer reservoirs, and a potential field which is applied across this capillary.

In general, the flow of electroosmosis is towards the cathode, and hence a detector is placed at this end. Separation by electrophoresis is based on differences in solute velocity in an electric field. The velocity of an ion can be given by

$$v = \mu_e E \qquad (1)$$

where v = ion velocity; μ_e = electrophoretic mobility; E = applied electric field.

The electric field is simply a function of the applied voltage and capillary length (in volts/cm). The mobility, for a given ion and medium, is a constant which is characteristic of that ion. The mobility is determined by the electric force that the molecule experiences, balanced by its frictional drag through the medium reported in the equation below:

$$\mu e\ \alpha = \frac{Electric\ force\ (E_F)}{Frictional\ force\ (F_F)} \qquad (2)$$

The electric force can be given by $F_E = q\ E$ (3) and the frictional force (for a spherical ion) by $F_F = -6\ \eta\ \pi\ r\ v$ (4) , where q = ion charge; η = solution viscosity; r = ion radius; v = ion velocity. During electrophoresis a steady state, defined by the balance of these forces, is attained. At this point the forces are equal but opposite are expressed by this equation:

$$q\ E = 6\ \eta\ \pi\ r\ v \qquad (5)$$

Solving for velocity and substituting equation (5) into equation (1) yields an equation that describes the mobility in terms of physical parameters

$$\mu_e = \frac{q}{6\pi\eta r_i} \qquad (6)$$

The electrophoretic mobility usually found in standard tables is a physical constant, determined at the point of full solute charge and extrapolated to infinite dilution. This usually differs from that determined experimentally. The latter is called the effective mobility and is often highly dependent on pH (expressed by solute pKa) and composition of the running buffer [29].

Sample Injection

In CE only minute volumes of sample are loaded into the capillary in order to maintain high efficiency. These small volumes are, of course, proportional to the small volumes of the capillaries. With respect to sample overloading, the injection plug length is a more critical parameter than volume [29]. The two most common are hydrodynamic and electrokinetic.

Hydrodynamic sample injection is the most widely used method. It can be accomplished by application of pressure at the injection end of the capillary, vacuum at the exit end of the capillary, or by siphoning action obtained by elevating the injection reservoir relative to the exit reservoir.

The volume of sample loaded will be a function of the capillary dimensions, the viscosity of the buffer in the capillary, the applied pressure, and the time. This volume can be using the Hagen-Poiseuille equation:

$$V_c = \frac{\Delta P \pi d^4 t}{128\eta L_t}$$

Where ΔP = pressure difference across the capillary; d = capillary inside diameter; t = time; η = buffer viscosity; L = total capillary length [33].

Electrokinetic, or electromigration, injection is performed by replacing the injection-end reservoir with the sample vial and applying the voltage.

Usually a field strength 3 to 5 times lower than that used for separation is applied. In electrokinetic injection, analyte enters the capillary by both migration and by the pumping action of the EOF. A unique property of electrokinetic injection is that the quantity loaded is dependent on the electrophoretic mobility of the individual solutes. Discrimination occurs for ionic species since the more mobile ions are loaded to a greater extent than those that are less mobile.

The injected moles can be calculated by this equation:

$$moles = \mu_A \left(E \frac{k_s}{k_c} \right) t \pi r^2 C$$

where r is the capillary radius, C is sample concentration and k_s / k_c is conductivity BGE sample [32]. As described, sample loading is dependent on the EOF, sample concentration, and sample mobility. Variations in conductivity, which can be due to matrix effects such as a large quantity of an undetected ion such as sodium or chloride, result in differences in voltage drop and quantity loaded. Due to these phenomena electrokinetic injection is generally not as reproducible as its hydrodynamic counterpart. Despite quantitative limitations, electrokinetic injection is very simple, requires no additional instrumentation, and is advantageous when viscous media or gels are employed in the capillary and when hydrodynamic injection is ineffective [29].

Detection system

Detection in CE is a significant challenge as a result of the small dimensions of the capillary. Although CE requires only nanoliter volumes of sample, it is not a “trace” analysis technique since relatively concentrated analyte solutions or pre-concentration methods are often necessary.

A number of detection methods have been used in CE to meet this challenge, many of which are similar to those employed in liquid column chromatography. As in HPLC, UV Visible detection is by-far the most common [29].

UV-Visible absorption is the most widely used detection method primarily due to its nearly universal detection nature. With fused-silica capillaries, detection below 200 nm up through the visible spectrum can be used. The high efficiency observed in CE is due in part to on-capillary detection. Since the optical window is directly in the capillary there is no zone broadening as a result of dead-volume or component mixing. In fact the separation is still occurring while in the detection window [29].

As with all optical detectors, the width of the detection region should be small relative to the solute zone width to maintain high resolution. This is best accomplished with a slit designed for specific capillary dimensions. Since peaks in CE are typically 2 to 5 mm wide, slit lengths should be maximally one third this amount. Detector design is critical due to the short optical path length.

The optical beam should be tightly focused directly into the capillary to obtain maximum throughput at the slit and to minimize stray light reaching the detector. These aspects are important to both sensitivity and linear detection range [29].

5.2. Modes of operation

A CE system can be operated in several different modes. These modes offer the analyst a variety of ways to approach an analytical problem. The choice of mode will be based on the analytical problem under consideration [34].

5.2.1. Capillatry Zone Electrophoresis (CZE)

Capillary zone electrophoresis (CZE) is the most widely used mode due to its simplicity of operation and its versatility.

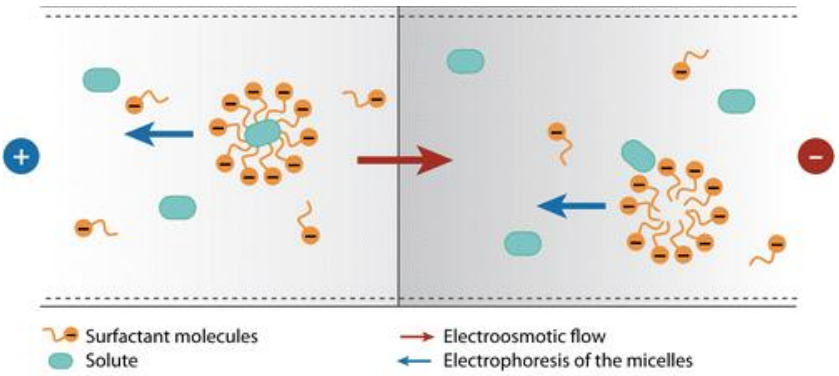
The application range of CZE is diverse. Application areas include the analysis of amino acids, peptides, ions, a wide range of enantiomers, and numerous other ionic species. CZE is fundamentally the simplest form of CE, mainly because the capillary is only filled with buffer. Separation of both anionic and cationic solutes is possible by CZE due to EOF.

Neutral solutes do not migrate and all coelute with the EOF. The name CZE is somewhat confusing in that it implies that it is the only mode in which “zonal” electrophoresis occurs [35].

In CZE, selectivity can most readily be altered through changes in running buffer pH or by use of buffer additives such as surfactants, co-surfactant or chiral selectors generally named *pseudostationary phases* (*PSP*). In this condition the mechanism of separation between analytes and PSP is based on their diffent mobilities and ripartition with PSP and BGE [36].

5.2.2. Micellar Electrokinetic Chromatography MEKC

Micellar electrokinetic chromatography (MEKC or MECC) is a hybrid of electrophoresis and chromatography.



Introduced by Terabe in 1984, MEKC is one of the most widely used CE modes [35]. Its main strength is that it is the only electrophoretic technique that can be used for the separation of neutral solutes as well as charged ones. The separation of neutral species by MEKC is accomplished by the use of surfactants in the running buffer. At concentrations above the critical micelle concentration, aggregates of individual surfactant molecules, micelles, are formed. Micelles are essentially spherical with the hydrophobic tails of the surfactant molecules oriented towards the center to avoid interaction with the hydrophilic buffer, and the charged heads oriented toward the buffer. A representation of micelles is here depicted. It is the differential interaction between the micelle and the neutral solutes that causes the separation. The surfactant and thus the micelles are usually charged and migrate either with or against the EOF (depending on the charge). Anionic surfactants such as SDS migrate toward the anode, that is, in the opposite direction to the EOF. Since the EOF is generally faster than the migration velocity of the micelles at neutral or basic

pH, the net movement is in the direction of the EOF. During migration, the micelles can interact with solutes in a chromatographic manner through both hydrophobic and electrostatic interactions. For neutral species, it is only partitioning in and out of the micelle that effect the separation [35].

The more the solute interacts with the micelle the longer is its migration time since the micelle carries it against the EOF. When the solute is not in contact with the micelle it is simply carried with the EOF. The more hydrophobic compounds interact more strongly with the micelle and are “retained” longer. The overall MEKC separation process is schematically represented in Fig. 5.2.2.

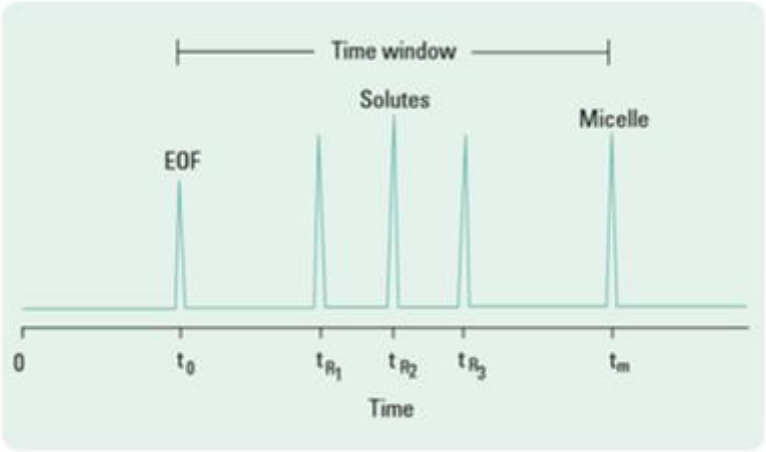


Fig. 5.2.2. The migration order in a MEKC separation process.

The separation mechanism of neutral solutes in MEKC is essentially chromatographic and can be described using modified chromatographic relationships. The ratio of the total moles of solute in the micelle (that is, the pseudostationaryphase) to those in the mobile phase, the capacity factor, *k'*, is given by:

$$k' = \frac{(t_R - t_0)}{t_0(1 - t_R/t_{mc})}$$

where t_R = retention time of the solute; t_0 = retention time of unretained solute moving at the EOF rate (or “dead time”); t_{mc} = micelle retention time. In the separation of neutral solutes, all solutes elute between t_0 and t_{mc}. Hydrophilic solutes that do not

interact with the micelle elute with the EOF and those that are totally retained by the micelles elute with the micelles. The selectivity can easily be manipulated in MEKC. In addition, MEKC can be performed using bile salts or microemulsions. In all cases, variations in buffer concentration, pH, temperature, or use of additives such as urea, metal ions, or chiral selectors can also be used to affect selectivity. Surfactants used for MEKC can also interact with the capillary wall and have dramatic effects on the EOF as well as solute-wall interactions. The direction of solute and micelle migration varies and depends on the micelle charge and the rate of EOF. Generally, high pH buffers are used to maintain reasonable EOF and ensure migration direction.

5.2.3. Microemulsion Electrokinetic Chromatography (MEEKC)

Microemulsion electrokinetic chromatography is a powerful technique employed in separation with neutral and charged analytes with a high range of solubility. The BGE consists in a microemulsion system [37].

Microemulsion, which is a dispersion of oil and water stabilized by surfactant molecules and/or short-chain alcohol (cosurfactant), was first introduced to electroseparation fields when microemulsion electrokinetic chromatography (MEEKC) was proposed by Watarai in 1991 [38].

MEEKC can separate analytes in a similar fashion as micellar electrokinetic chromatography (MEKC). Two effects determine the separation results of both MEEKC and MEKC systems. One is the difference of the partition coefficients of analytes in pseudostationary and aqueous phases (chromatographic effect), and the other one is the difference of migration velocities of analytes in electric fields (electrophoresis effect) [38]. Unlike the micelle that is regarded as pseudostationary phase for MEKC, the surfactant-coated oil droplets in microemulsion are employed as the pseudostationary phase for MEEKC.

Most papers which compared separations obtained by MEEKC and MEKC had concluded that highly hydrophobic compounds tend to be strongly retained by the micelles, thus MEEKC has a greater separation capability for highly hydrophobic compounds. In addition, since analytes are able to more easily penetrate the surface of

oil droplets when compared to micelles' more rigid surfaces, as a result MEEKC is applicable to a wider range of analytes than MEKC [38].

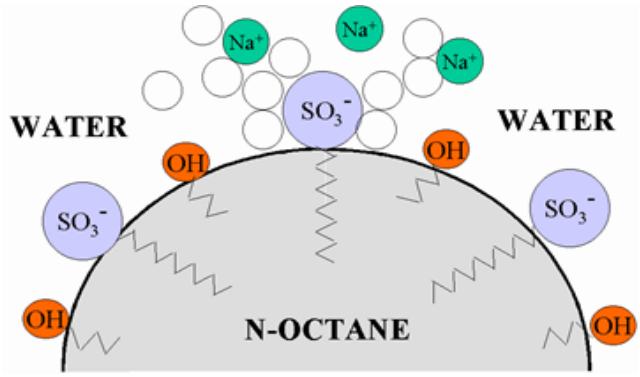
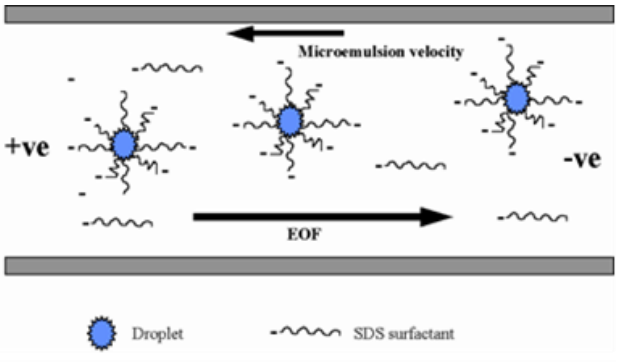


Fig. 5.2.3. Schematic representation of a microemulsion, which is a dispersion of oil and water stabilized by surfactant molecules and/or short-chain alcohol (cosurfactant).

Microemulsions solutions containing dispersed nanometre-sized droplets that allows chromatographic separation to be oblets of an immiscible liquid. Microemulsions are solutions containing dispersed nanometre-sized droplet of an immiscible liquid. In particular the microemulsions used in MEEKC are oil droplets dispersed in an aqueous buffer. The oil and water components are totally immiscible and do not mix together as there is a high surface tension between them [39].



The oil droplets are coated with a surfactant to reduce the surface tension between the two liquid layers which allows the emulsion to form. The surface tensions further lowered by the addition of a short-chain alcohol such as butan-1-ol which stabilies the microemulsion system. The microemulsion is therefore optically transparent as a larger droplet scatter white light.

Disintegration of the emulsion system actually represents an activation barrier, as the surface separationsion of the system is at an energy low. Therefore, if the combination of oil, surfactant and co-solvent is appropriate then the microemulsion

systems are highly stable and remain intact indefinitely. Use of a microemulsion containing ionic surfactant solutions allows chromatographic separation to be obtained as solutes can partition between the charged oil droplet and the aqueous buffer phase. Water insoluble compounds will favour inclusion into the oil droplet rather than into the buffer phase. This situation allows partitioning of the solute between the oil and water phases in a chromatographic fashion.

Solutes chromatographically interact with the micelles to achieve separation. Solutes are more easily able to penetrate the surface of the droplet than the surface of a micelle which is much more rigid. This ability allows MEEKC to be applied to a wider range of solutes. Sodium dodecyl sulphate (SDS) is the most widely used emulsifier surfactant in MEEKC. The oil droplet is coated with SDS surfactant molecules making the droplet negatively charged. The C_{12} alkyl chain of the surfactant penetrates into the oil droplet whilst the negatively charged hydrophilic sulphate groups resides in the surrounding aqueous phase. Charge repulsion of the negatively charged sulphate group on the SDS prevents highly efficient packing and prevents formation of an emulsion as the surface tension cannot be sufficiently reduced. A co-surfactant, usually a medium-chain-length alkyl alcohol such as butan-1-ol, is therefore essential in the formation and improved stability of the microemulsion. The co-surfactant bridges the oil and water interface and further reduces the surface tension of the system to zero. The Fig. 5.2.3 provides a schematic of the emulsion droplet showing the short-chain alcohol, SDS, the octane droplet and the sodium ions surrounding the droplet [39].

The microemulsion is a dynamic entity and it has a lifespan in the microsecond range.

The emulsion droplets exist in a variety of shapes with an average that is spherical. The range of droplet shapes that an individual emulsion system contains has been measured and is known as the polydisperity. A highly ordered microemulsion system has a low polydisperity and highly spherical droplets. The surfactant and co-surfactant act together to reduce the surface tension between the two liquid phases to zero. The presence of the co-surfactant helps the droplets to pack more effectively and create a more stable environment [39].

The options available in method development in MEEKC will be described which include type and concentration of surfactant, buffer pH and the type of oil. Selectivity

can also be affected by organic solvents, type of co-solvent, addition of cyclodextrins ion-pair reagents and the manufacture process for the microemulsion. MEEKC is a relatively recent technique and it has not been widely applied to a range of application types. However, there are currently sufficient applications to demonstrate the widespread potential uses of MEEKC. The reported application range will be summarised to provide an appreciation of the separation possibilities that MEEKC may offer [39].

Overall it is concluded that MEEKC can offer the possibility of highly efficient separations of a range of solute types. The technique can be equally applied to water-soluble and insoluble compounds and to charged or neutral solutes. It is predicted that use of MEEKC will increase rapidly as awareness of these possibilities increases.

Surfactant type and concentration

Surfactants are molecules with detergent properties, which are composed of a hydrophilic water-soluble head group and a hydrophobic water-insoluble hydrocarbon chain group. Although a large number of surfactants are commercially available, a limited number are widely used in MEKC separations. The surfactants suitable for MEKC must be soluble in the buffer solution to form micelles and the micellar solution must be homogeneous, UV transparent and also have a low viscosity [40].

There are four major classes of surfactants: anionic, cationic, zwitterionic and nonion. The most common and frequently used are listed below.

Surfactant	Type	CMC*	n
Sodium dodecyl sulphate (SDS)	anionic	8.1×10^{-3}	62
Sodium tetradecylsulphate (STS)	Anionic	2.1×10^{-3}	138
Sodium dodecanesulphate	anionic	7.2×10^{-3}	54
Sodium cholate	anionic	$13\text{-}15 \times 10^{-3}$	2-4
Cetyltrimethylammonium bromide (CTAB)	cationic	0.92×10^{-3}	61
Dodecyltrimethylammonium bromide	cationic	15×10^{-3}	56
Brij - 35	nonionic	0.1×10^{-3}	40
Sulfobetaine	zwitterionic	3.3×10^{-3}	55

Of these, ionic surfactants are generally used in MEKC. Every surfactant has a characteristic CMC (Critical Micellar Concentration) and aggregation number n (the number of surfactant molecules necessary to form a micelle). Another important parameter is the Kraft point, which represents the minimum value of temperature to whom the solubility of surfactants increases steeply due to the formation of micelles

The choice of surfactant has a marked effect on the separation achieved in MEEKC as it affects the oil droplet charge and size, the level and direction of the EOF, and the level of any ion-pairing with charged solutes. SDS is an anionic surfactant with a C_{12} alkyl chain, which penetrates into the oil droplet. Anionic bile salt surfactant such as sodium cholate has also been used to generate negatively charged droplets. Use of bile salt microemulsions gave different selectivity compared to when SDS was used to make the microemulsions [37].

Higher concentrations of the surfactant increase the capacity factor of neutral solutes as it increases the charge density on the oil droplet. If a mixture of charged and neutral solutes is employed, then altering the surfactant concentration will have an effect on peak migration order. Increasing the surfactant concentration also increases the ionic strength of the buffer which reduces the EOF level and increases analysis time.

Higher levels of surfactant reduce surface tension to a greater extent which generates more stable microemulsions.

Nonionic surfactants do not posses electrophoretic mobility and cannot be used, as “pseudostationary phase” in conventional MEKC, however can be useful for the separation of charged analytes. This technique using nonionic micelles can be classified as an extension of MEKC [40].

Micelles are amphiphilic aggregates of surfactants. Above a specific surfactant concentration, the surfactant molecules begin to self-aggregate, forming micelles, spherical aggregates that exhibit electrophoretic migration like any other charged particle. Micelles are a long chain molecules and are characterized as possessing a long hydrophobic chain and a hydrophilic head group. Generally, micelles are formed in aqueous solution with the hydrophobic tails oriented towards the center of the aggregated molecules and the hydrophilic heads pointing outward into the aqueous solution.

Micelle formation is a very dynamic process, as micelle disaggregate and reconstruct continuously, composing the “pseudostationary phase” which can include hydrophobic analytes.

Micellar solutions can solubilize hydrophobic compounds which otherwise would be insoluble in water. Micelles have the ability to interact with the analytes at molecular level based on hydrophobic and electrostatic interactions.

Even neutral analytes can bind to micelles due to the very strong solubilization power of the hydrophobic core [40].

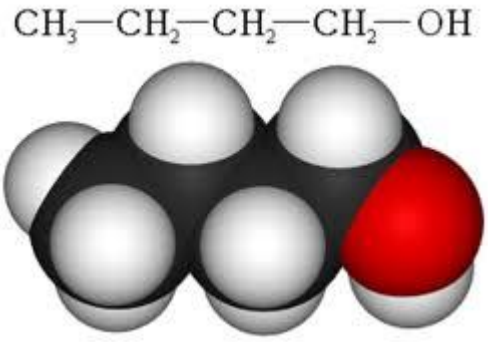
The micelles used in MEKC are charged on the surface, so an analyte with the opposite charge will strongly interact with the micelle through electrostatic forces while an analyte with the same charge will interact weakly due to the electrostatic repulsion. Therefore, the use of a cationic or an anionic surfactant will result in an entirely different result. The micellar phase can be modified by adding two different surfactants to form a mixed micelle; addition of an ionic and a nonionic surfactant can provide different selectivity in separation. A mixed micelle has a lower surface charge and a larger size; consequently, its electrophoretic mobility will be lower than the one of a simple ionic micelle [40]. Some surfactants like bile salts are chiral and can be used for enantiomers separation. Typically, SDS concentrations in the region of 110 m*M* (3%, w/w) are used which produces highly stable microemulsions with shelf lives exceeding several months. SDS concentrations as high as 6.5% (w/w) have been used [40].

Oil type

Generally, octane or heptane has been used to generate the oil droplet. Heptane may be preferred as the odd numbered alkanes have lower toxicity than the even numbered alkanes [41]. However, octane has been reported to give more repeatable microemulsions than heptane. Hexane, heptane and octane have been shown to give similar selectivity and migration times for separation of a range of neutral compounds [39].

Addition of organic solvents

In MEKC highly water insoluble solutes partition strongly into the micelles and are therefore highly retained with poor resolution of mixtures. In these instances, organic solvent are often added into MEKC buffers to reduce retention and improve resolution. Standard solvents used are acetonitrile, methanol and isopropanol. This approach has also been used in MEEKC [39].



The levels of solvent that can be added to MEKC buffers is generally limited to a maximum such as 30% (v/v). At levels greater than this the micelle are disrupted and selectivity is lost. It was found that there were also limits to the maximum solvent contents that could be used in MEEKC. When these limits were exceeded the microemulsion buffers disintegrated into a cloudy two-phase system, which could not be used for separation [39].

Co-surfactant type and concentration

Butan-1-ol is the most frequently employed co-surfactant. Studies have shown

that the separation selectivity is unaltered by variation of the butan-1-ol concentration. Higher butan-1-ol concentrations reduce migration times for water-insoluble solute but do not alter the capacity factors [42]. The migration times are altered with varying co-surfactant concentration as it affects the solution viscosity which in turn affects the EOF rate [39].

The size of the oil droplet increases with increased co-surfactant concentration which will affect the charge density on the droplet. Variation of co-surfactant concentration over a wide range affected the migration times of hop bitter acids [43].

This short chain alcohol is the most commonly applied cosurfactant for MEEKC, has a log P value (1-octanol/water partition coefficient) of 0.84, is considerably soluble in water [42], and due to medium chain length simultaneously exhibits properties, which are typical both for hydrophilic (propanol) and hydrophobic (pentanol) alcohols. In both MEKC and MEEKC, nBuOH is present in the aqueous buffer phase and in the Stern layer of the micelles, but also can penetrate into the micelles/microemulsion droplets, acting as a true class I modifier [42].

By using the two-dimensional NOE spectroscopy, it has been demonstrated that the nBuOH headgroup is located in contact with the SDS micellar headgroups, while the hydrocarbon chain is directed to the micelle core [40]. In particular, adding the alcohol to micelles, the size-to-charge ratio of the micelles increases because nBuOH reduces the charge density and promotes the formation of micelles with a more open structure [42].

Mixed SDS–nBuOH micelles have been recently reviewed in terms of micellization parameters and structure [41].

Normal Polarity

In normal polarity, the positive pole, is set at the beginning of the capillary (*inlet*) and the cathode, the negative pole, at the end (*outlet*). A negative PSP (Pseudostationary Phase), carrying a negative effective mobility, migrates to the anode, is necessary the use of an alkaline BGE in order to obtain a high EOF and carry analytes and PSP to the detector. In this condition the analytes are negative charged or uncharged.

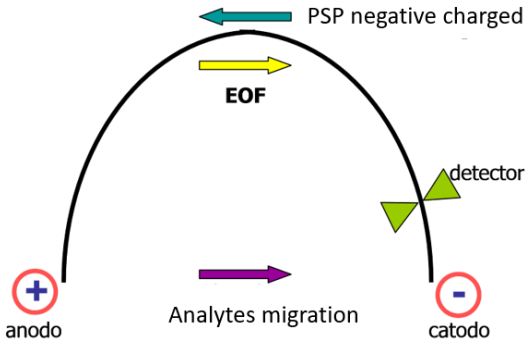


Fig. 5.2.3. The migration order in Normal Polarity.

Inverse Polarity

In inverse polarity the anode, the positive pole, is set at the end (*outlet*) of the capillary and the cathode, the negative pole, at the beginning (*inlet*).

In order to carry a negative PSP (Pseudostationary Phase), carrying a negative effective mobility, to the anode, is necessary to avoid the EOF (Electroosmotic Flow).

The use of BGE with low pH and high ionic force lead to the migration of analytes and PSP that are positive charged or uncharged

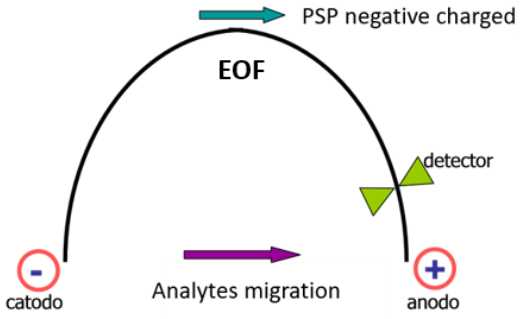
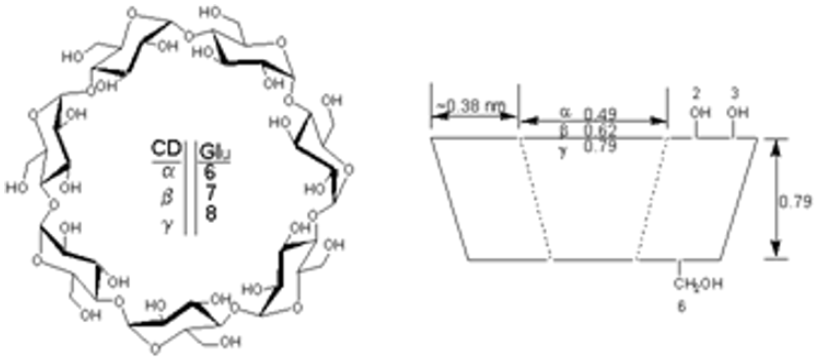


Fig. 5.2.3. The migration order inInverse Polarity.

5.3. Cyclodextrins in CE (CD)

Several forms of CE have been developed extensively over the last decade as powerful tools for rapid and efficient separation of charged and uncharged components present in small sample volumes.

Cyclodextrin-based complexation phenomena have received considerable attention in analytical separation science and particularly in capillary electrophoresis [44]. Cyclodextrins are oligosaccharides prepared by enzymatic degradation of starch and glycosyl transferases or cyclodextrinases producing a mixture of different CDs that can be separated by using, e.g., chromatography, crystallisation etc.



They have a shape of a truncated cone with an open cavity, relatively hydrophobic and an outside hydrophilic due to the presence of hydroxyl groups (positions 2, 3 and 6 of glucopyranose). Nevertheless, CDs with 6 to 12 glucose units have been separated, only those with six, seven and eight units (with the Greek names α-, β- and γ-CD, respectively) are currently used in analytical chemistry. The physical properties of the three native CDs are quite different, e.g., width of the cavity, solubility, molecular mass etc., however they possess the same depth [45].

Enhancement of selectivity using cyclodextrins in CE can be attributed to their ability to selectively include a wide variety of water-insoluble organic and inorganic molecules into their hydrophobic cavity.

The ability of cyclodextrins to act stereoselectively upon complexation with a guest molecule has also been well recognized in liquid chromatography, both as a mobile-phase additive and as a chiral stationary phase. The most frequent use for CDs in capillary electrophoresis has been as modifiers to effect chiral separation of enantiomers of charged compounds. Together with a micelle agent, neutral CDs can be extended for the chiral resolution of small neutral and hydrophobic compounds as a result of their partition between the micellar phase and the cyclodextrin-modified buffer.

Negatively charged CDs with their own electrophoretic mobility can also be prepared and used in CE to increase the separation window which enables better resolution or enantioresolution of analytes which weakly complex with or are poorly differentiated by neutral CDs [44]. Since two enantiomers possess similar physico–chemical properties, for a successful enantiomer resolution in CE, it is necessary to selectively modify the effective mobilities of the two analytes. This can be obtained on forming stereoselective complexes where hydrophobic, hydrogen, π–π, dipole, van der Waals interactions can be involved.

In CE the separation of chiral compounds is achieved mainly by the direct separation method where the chiral selector is simply added to the BGE or bonded to either the capillary wall or to a stationary phase. Using CDs, analytes fit the cavity either with the whole molecule or with their hydrophobic part on forming inclusion-complexes stabilised by secondary bonds between the rim of the chiral selector and substituent groups of the analyte asymmetric centre. The complex formed during the electrophoretic process, in equilibrium with the free analyte, possesses a different mass responsible for the change of the effective mobility [45].

While not as common, neutral and charged cyclodextrins have been used as additives in CE to optimize the separations of peptides, proteins, and small achiral molecules such as plant regulators, food colors, positional isomers, and toxic pollutants. Native CD and their nonionic derivatives are very effective in enhancing both selectivity and resolution through judicious partitioning of net charged guest molecules into their hydrophobic cavity. Together with a micelle agent, they have proven very useful for resolving numerous neutral and hydrophobic compounds in biological or environmental samples [44].

Neutral CD such as α, β, and γ and their derivatives are only useful to separate solutes with a net charge. In general, CD enhance both selectivity and resolution through judicious partitioning of guest molecules into the host hydrophobic cavity.

Formation of the inclusion complex is facile on the electrophoretic time scale and the apparent mean mobility of a solute can be related to the intrinsic mobility of the solute (μ_A) and the solute–cyclodextrin inclusion complex (μ_{ACD}):

$$\mu_{app} = \frac{\mu_{\Lambda} + \mu_{\Lambda CD} K[CD]}{1 + K[CD]}$$

The formation constant of the inclusion complex or *K* is defined:

$$K = \frac{[ACD]}{([A] + [CD])}$$

[ACD]/([A]+[CD]) where [CD], [A], and [ACD] are the concentrations of cyclodextrin, solute, and inclusion complex, respectively [44].

Therefore, the quality of a resolution and sensitivity are governed by the cyclodextrin concentration in the electrolyte solution, and the selectivity is dependent upon whether the solute will form a strong or weak complex with the cyclodextrin.

In pharmaceutical and drug analysis, several cyclodextrin modified CE methods have been developed for separation and quantitation of important drugs in biological samples.

The use of CE to determine drug-related impurities has also become established within industrial pharmaceutical analysis laboratories [46].

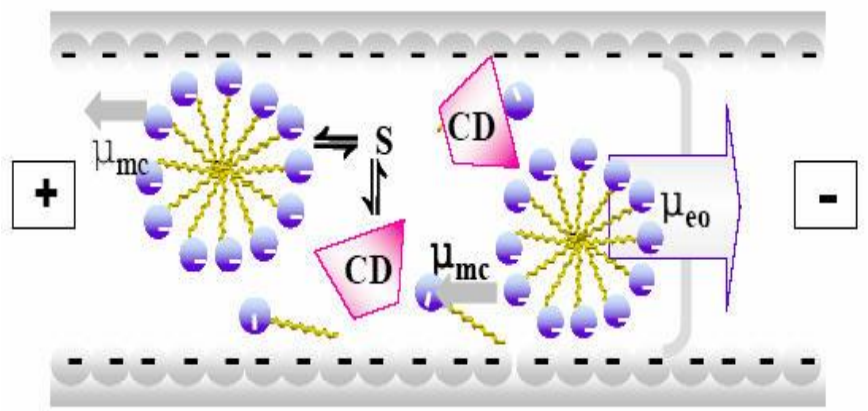


Fig. 5.3. Schematic representation of CD-MEKC operative condition.

Adding CD to the MEKC system has been shown to increase the versatility of the CE technique, particularly in separation of geometrical isomers of water-insoluble compounds. Sodium dodecyl sulphate (SDS) is the most used surfactant forming

micelles in CE for the enantiomers separation in presence of CDs. The method, called cyclodextrin-modified micellar electrokinetic chromatography (CD-MEKC) was firstly used by Terabe et al. [47] for the separation of hydrophobic achiral compounds such as chlorinated benzene congeners, polycyclic aromatic hydrocarbons, etc.

A micellar (SDS) and an aqueous phase containing the CD compose the system. A relatively high EOF transports the hydrophobic analytes to the detector (cathode) while SDS particles, negatively charged, possess their own mobility opposite to that of EOF. The analytes interact with SDS and CD modifying their mobility according to the different partition between the micellar and the aqueous phases.

Several types of CDs: α-CD, β-CD, 2,6-di-*O*-methyl-β-CD, 2,3,6-tri-*O*-methyl-β-CD and γ-CD were used together with SDS to manipulate the selectivity in CD-MEKC. Cyclodextrin-modified MEKC has found a variety of important applications in food, pharmaceutical, clinical analysis and biotechnology. Under normal electrophoretic conditions, the micellar mobility is of a smaller magnitude than the electroosmotic flow; thus, both the running buffer and micellar phases are transported at different velocities towards a common outlet of the separation capillary. Normally, the negatively charged micelle migrates at a significantly slower rate than the buffer phase, and functions as a pseudo-stationary phase. Both charged and neutral compounds can partition between the two phases, resulting in retention based on differential solubilization by the micelles. Neutral CDs can be added as an additive in micellar solutions for CD-MEKC separations to resolve lipophilic compounds which migrate close to the micelle and cannot be separated by the micellar solution alone. Since CDs are electrically neutral and their outside surface is hydrophilic, they do not interact with the micelle and move along with the buffer phase [46].

When a hydrophobic solute is introduced to the CD-MEKC system, it will be incorporated by either the micelle or CD. The capacity factor, k^*, of a hydrophobic solute in CD-MEKC is expressed as $\alpha V_{mc}/V_C$ where V_{mc} and V_{CD} are the volumes of micelle and CD, respectively. The distribution coefficient α indicates the relative affinity of the solute between the micelle and CD. The ratio of the solute incorporated in the micelle largely depends on its hydrophobicity, but the inclusion complex formation of the solute with CD will depend on the size of the solute and the internal cavity of CD in addition to the hydrophobicity. Therefore, the selectivity in CD-MEKC is mostly

governed by the tendency of the hydrophobic solute to form an inclusion complex with CD. A solute included more strongly in the CD cavity is expected to have a shorter migration time than a solute included less strongly in the CD cavity [44].

In food, pharmaceutical, and environmental analysis, measurement of trace levels of both charged and uncharged compounds is important and there is always a continuing need to develop better and faster separations using less and less materials. Besides the importance of cyclodextrins in chiral separation, there is little doubt that various forms of CE using cyclodextrins will be developed and become one of the most useful separation techniques to address `real world' problems in achiral separation and detection of small and neutral compounds [44].

5.4. Validation

The objective of any analytical measurement is to obtain consistent, reliable and accurate data. Validated analytical methods play a major role in achieving this goal. The results from method validation can be used to judge the quality, reliability and consistency of analytical results, which is an integral part of any good analytical practice. Validation of analytical methods is also required by most regulations and quality standards that impact laboratories.

The objective of validating an analytical procedure is to demonstrate “suitability for its intended purpose”.

The intent of analytical measurement is to generate accurate and reliable data. Method validation alone cannot guarantee this, but it should be part of integrated quality assurance for analytical measurement.

According to USP analytical methods should be validated through laboratory tests: “Validation of an analytical procedure is the process by which it is established, by laboratory studies, that the performance characteristics of the procedure meet the requirements for the intended analytical applications”.

The required laboratory tests for method validation have been defined in different working groups of national and international committees and are described in the literature.

Unfortunately, some of the definitions vary between the different organizations.

Therefore, laboratories should have a glossary with definitions on their understanding of the terms.

In an attempt to standardize, representatives from the industry and regulatory agencies from the United States, Europe and Japan defined parameters, requirements and methodology for analytical methods validation through the Internation Conference of Harmonization (ICH). The parameters, as defined by the ICH and other organizations and authors, are summarized in the table below:

Parameter	Comment
Specificity	USP, ICH
Selectivity	ISO 17025
Precision	USP, ICH
Repeatability	ICH, ISO 17025
Intermediate precision	ICH
Reproducibility	ICH, defined as ruggedness in USP, ISO 17025
Accuracy	USP, ICH, ISO 17025
Linearity	USP, ICH, ISO 17025
Range	USP, ICH
Limit of detection	USP, ICH, ISO 17025
Limit of quantitation	USP, ICH, ISO 17025
Robustness	USP, Included in ICH as method development activity, ISO
Ruggedness	USP, defined as reproducibility in ICH

Method validation is an essential component of the measures that a laboratory should implement to permit it to produce reliable analytical data [48]. At the same time, validation, following reputable international authorities has to be a cost-effective step of method development in order to encourage all scientists to make use of it. In fact, many academicians develop and optimise analytical methods using pure standards, leaving validation or application to those in the field [48]. These unvalidated methods are allowed to assume the aura of validity and authenticity, when, in practice, they may never be completely useful with samples that are typically encountered. A way to speed up the validation process consists of the use of experimental design, which can be very useful and advantageous for both the evaluation and the optimisation of some

performance parameters. Experimental design techniques are powerful tools for the exploration of multivariate systems [49]. In particular, statistical design is a way of choosing experiments efficiently and systematically to give reliable and coherent information. From a statistical standpoint, design means construction of experiments so that the analysis of results yields the maximum amount of information that can be extracted from the experiments. More specifically, experimental design helps the researcher to verify if changes in factor values produce a statistically significant variation of the observed response, and this approach can be used each time it is necessary to have this type of information. Typically, experimental design techniques are used to understand the effect of several variables on a system by a well-defined mathematical model [49]. The strategy is most effective if statistical design is used in most or all stages of development and not only for screening or optimising the process. A systematic use of statistical design in developing a method ensures traceability, supports validation, and makes the subsequent confirmatory validation much easier and more certain [49].

Experimental design strategies can be used in the pre-validation process (investigation of accuracy, precision and robustness) of analytical methods. In particular, the use of experimental design for the improvement of accuracy and precision is a very attractive goal, and the use of experimental design during validation constitutes a basic feature of multivariate optimisation, which, if appropriately used, can solve several problems and constitutes a powerful tool in the hands of researchers

6. Quality by Design

6.1. Introduction

At the beginning of the twentieth century, Sir Ronald Fisher introduced the concept of applying statistical analysis during the planning stages of research rather than at the end of experimentation. When statistical thinking is applied from the design phase, it enables to build quality into the product, comprising system thinking, variation understanding, theory of knowledge, and psychology [41].

The pharmaceutical sector is process-based and quality oriented, thus it would be expected to adopt the aforementioned paradigms early after their introduction [41].

However, the US Food and Drug Administration introduced quality by design (QbD) as a fundamental pharmaceutical quality model to be considered in the development of pharmaceutical products and processes [50, 51].

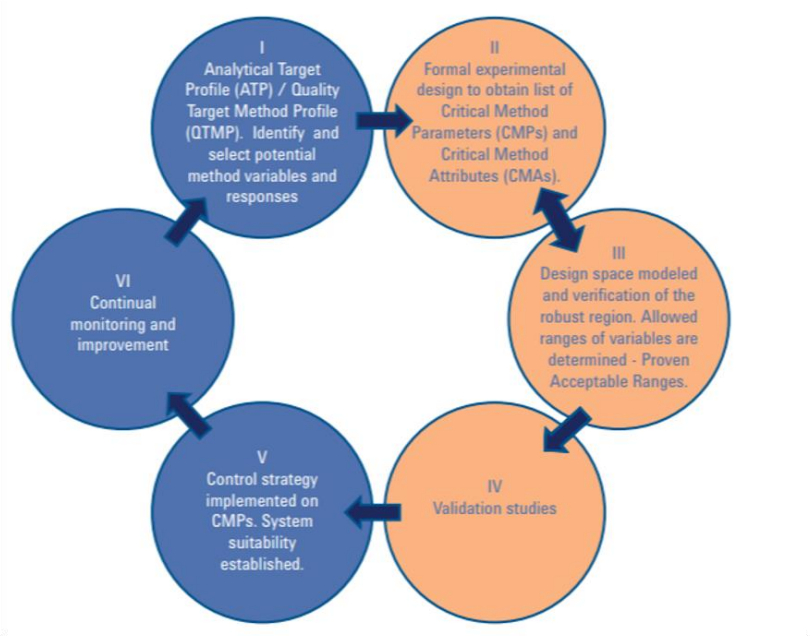
QbD principles have also been supported by International Conference on Harmonisation (ICH) guidelines Q8(R2), Q9, and Q10. ICHQ8(R2) on pharmaceutical development gives a basis for risk mitigation via the in-depth product and process understanding gained in pharmaceutical development, whereas ICHQ9 on quality risk management develops the principles and some of the tools of quality risk management for assessment, control, communication, and review of the risks of the quality of the medicinal product. ICHQ10 on pharmaceutical quality systems complements ICHQ8(R2) and ICHQ9 by establishing a model for a pharmaceutical quality-management system that would facilitate innovation and continuous improvement throughout the lifetime of the product.

Fortunately, with the development of the concept of QbD, there will be a significant transformation in pharmaceutical quality regulation, from an empirical process to a more scientific and risk-based approach [49].

QbD is defined as “a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management” [52].

By means of QbD, quality is designed into the process at the onset, thus counteracting the traditional quality by testing (QbT) approach, which tests product quality by checking it at the end of the manufacturing process. QbD leads to the establishment of the design space (DS), defined as the “multidimensional combination and interaction of input variables and process parameters that have been demonstrated to provide assurance of quality” [52].

According to this definition, DS should be characterized by multivariate techniques, therefore the use of design of experiments (DoE) has emerged as a fundamental activity for implementing QbD.



DoE is an excellent tool that allows pharmaceutical scientists to systematically manipulate factors according to a pre-specified design. A good design is based on sound cognition of product and effective management of whole process during manufacturing. DoE studies work together with mechanism-based studies to achieve better product and process understanding [48].

DoE is a reasonable method to determine the relationship between the inputs and outputs of a process. It can help identify optimal conditions, CMAs, CPPs, and, ultimately, the Design Space. It is wise to establish a Design Space through DoE for multivariate experiments [48].

QbD was originally intended for pharmaceutical product development, and has been mainly applied to small molecules, including formulation issues [53-54], synthesis of the active pharmaceutical ingredient (API) [55], and to biopharmaceuticals [55].

QbD has been attracting increased attention in the development of analytical separation methods, because these are intended to be used for quality control and analysis of API and drug products, to ensure product quality and thus patient safety [56].

Thus, analytical methods can be regarded as an integral part of the QbD concept and should meet their intended purpose similarly to the product requirements for a dosage form [56]. Many of the terms used in QbD are very familiar to analytical

chemists, including DoE, which has already been extensively used in the development of chromatographic and capillary electrophoretic (CE) methods [57, 58]. DoE incorporates a set of characteristics which are essential in QbD, for example the possibility of detecting interactions between the factors, of being used for both screening of factors and process characterization and optimization, and for simultaneously optimizing multiple responses [59].

Not all of the reviewed examples implement each step of QbD, and the different aspects of QbD may have been interpreted differently by the different authors, even if there are important commonalities. The general QbD guide for separation method development is here shown.

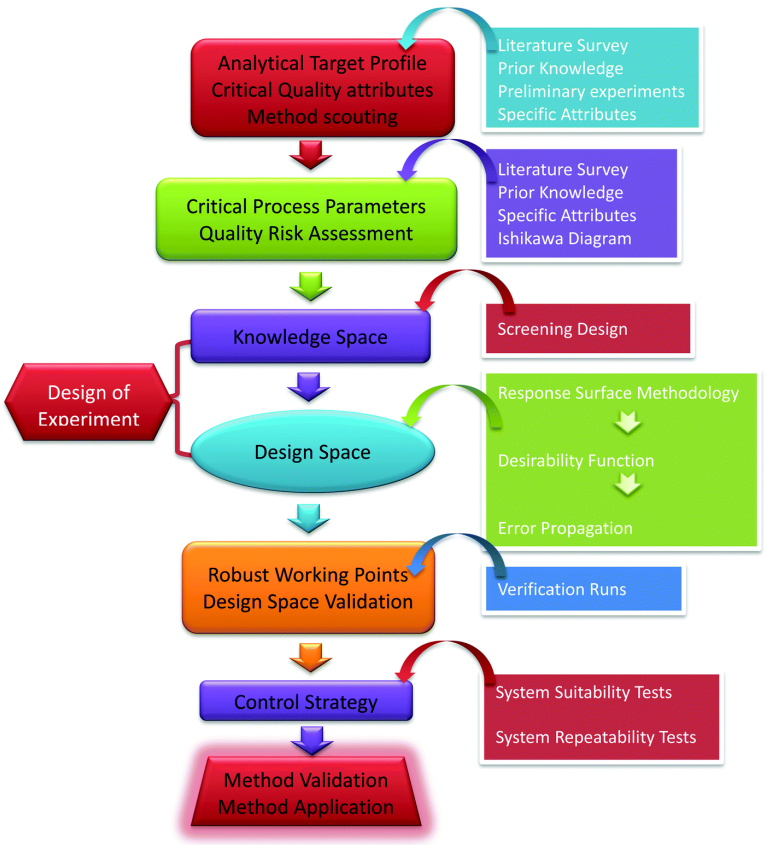


Fig. 6. 1. The general Quality by Design guide for separation method development.

6.2. Definition of quality target product profile and critical quality attributes

The analytical target profile (ATP) is a set of criteria that define what will be measured (e.g. the level of a specified impurity) and the performance criteria to be achieved by the measurement (e.g. accuracy, precision and range), but without specifying the method itself [59]. On the basis of ATP, different analytical methods and/or techniques are evaluated in a preliminary investigation to approach the method objective, in general with the purpose of achieving maximum selectivity with adequate efficiency, and improving the reproducibility and repeatability of measurements. Investigation of HPLC methods can involve selection of basic conditions, for example type of column (size and composition), pH, and organic modifier [59]. Investigation of CE methods can include selection from the different modes of operation on the basis of background electrolyte composition, which may consist of a plain buffer or may contain a pseudostationary phase mostly based on micelles or microemulsions. After these preliminary experiments, the QbD workflow can start, first by defining the quality target product profile (QTPP) and the critical quality attributes (CQAs).

Quality Target Product Profile (QTPP) is defined as a prospective summary of quality characteristics of a drug product to be achieved, taking into account dosage strength(s) and container closure system of the drug product, together with the attributes affecting pharmacokinetic characteristics (e.g., dissolution, aerodynamic performance) and drug product quality criteria (e.g., sterility, purity, stability and drug release) appropriate for the intended marketed product [48]. Applying this definition to analytical methods means that, first, the separation objectives should be well defined [60].

CQA is a "physical, chemical, biological or microbiological property or characteristic that should be within an appropriate limit, range, or distribution to ensure the desired product quality" [52]. In analytical methods, CQAs are key response variables, so the chromatogram or electropherogram will be related to a mathematical representation of quality [**61**]. CQAs can be classified into two major categories: "must have" and "intend to have". For instance, resolution of a pair of adjacent peaks must be not less than 1.5, and is intended to be not less than 1.8 [59]. Some examples of CQAs

are the critical resolution R_S, run time, efficiency and, in contrast with the QbT approach, robustness [61]. The separation criterion S was recently introduced [62, 63], and was defined as the difference between the retention times measured at the beginning of the second peak and at the end of the first peak of the critical pair. S was preferred to R_S to define the quality of the separation, because R_S modeling can lead to poor prediction, as a result of its non-linear and non-continuous behavior, when selectivity (order of the peaks) drastically changes. Moreover, even if S and R_S are highly correlated, computation of S is easier and its associated uncertainty is lower [62, 63].

6.3. Identification of critical process parameters by quality risk assessment

Critical process parameters (CPPs) are defined as "parameters whose variability have an impact on a CQA and therefore should be monitored or controlled to ensure the process produces the desired quality", and this statement can be fit perfectly to analytical methods.

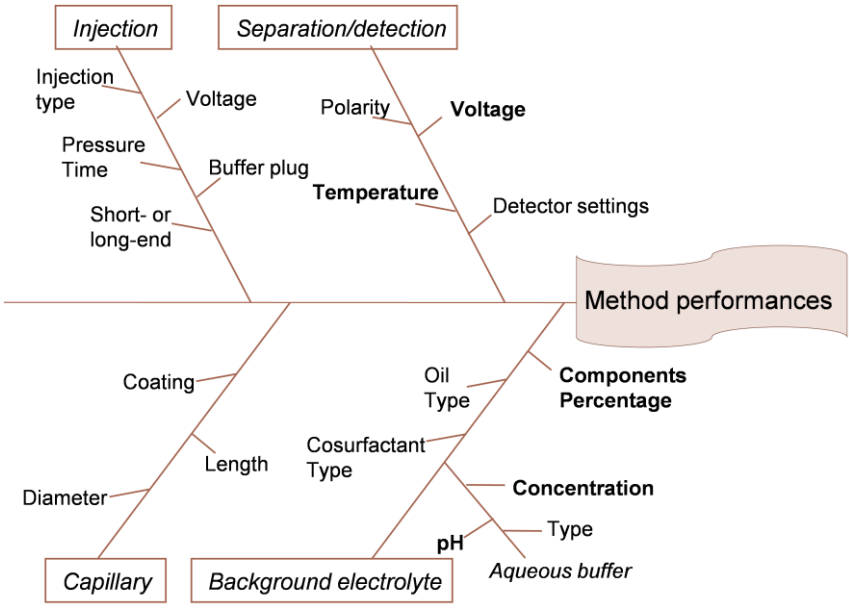


Fig. 6.3. Schematic representation of a Ishikawa diagram for the optimization of method performances.

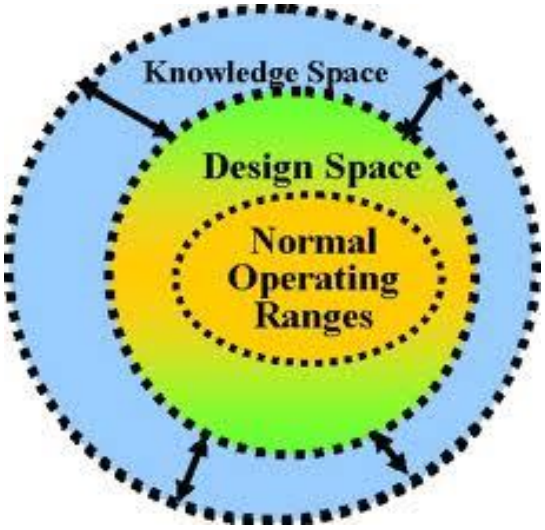
Lionberger et al. proposed classification of parameters into three categories: unclassified, critical or non-critical. The criticality of an unclassified parameter is undetermined or unknown, whereas a parameter is critical when a change in that parameter can cause the product to fail to meet the QTPP. Development studies should be able to move unclassified parameters to either non-critical or critical, otherwise they may need to be constrained at fixed values or narrow ranges because they might be critical [64]. The initial list of potential parameters which can affect CQAs can be quite extensive but can be reduced and prioritized by quality risk assessment (QRA). QRA is a sciencebased process [48] that can aid identification of CPPs and thus eliminating risk, resulting in high confidence that the analytical method will meet the QTTP under all conditions of use [48] Thus, a large number of parameters can actually be safely eliminated by use of QRA tools [4], for example failure mode effects analysis (FMEA) and Ishikawa diagrams [65] on the basis of prior knowledge and initial experimentation.

In FMEA the variables are ranked on the basis of the likelihood failure will occur (probability), affect on the analytical results (severity), and difficulty of detection (detectability), resulting in a risk priority number (RPN). Factors with an RPN above a cut-off level can then be evaluated by subsequent studies whereas factors with a lower RPN can be eliminated from further study [59]. Ishikawa diagrams segregate risks into different categories, for example those associated with instrumentation, materials, methods, measurements, laboratory climate, and human factors.

An Ishikawa diagram for an HPLC assay and impurities method was presented, highlighting the different sources of factors, followed by a CNX analysis to decide which factors should be controlled (C)—these were the potential noise (N) factors—and on which experiments (X) should be conducted to determine acceptable ranges [48]. In general, critical conditions in the overwhelming majority of HPLC separations are gradient time, temperature, pH of the aqueous phase, composition of organic modifier, and stationary phase [66]. The CPPs which are typically evaluated by DoE are shown; the other parameters are usually fixed by preliminary experiments and/or prior knowledge.

6.4. Investigation of knowledge space by Design of Experiments

After QRA, screening studies by means of DoE-effect analysis can be used to evaluate the effect of the higher-ranked CPPs and further reduce the number of CPPs to be studied in greater detail by selecting the factors that have a greater effect [67]. QRA tools, prior knowledge, and preliminary experiments are used to define knowledge space (KS), which is the experimental region investigated by screening DoE [67, 68]. Further preliminary experiments specifically aimed at proper selection of the investigation domain may be necessary, because the choice of a suitable KS minimizes the risk of not achieving good separation [63]. The object is to perform a minimum number of experiments on a maximum number of factors, so main-effects models are estimated mostly with fractional two-level and Plackett–Burman designs [11]. As an example, a Plackett–Burman design was used to screen the effects of mobile phase pH, flow rate, temperature, injection volume, and methanol concentration on peak resolution and USP tailing. Pareto ranking analysis by multivariate regression revealed that the statistically significant factors were pH, temperature and injection volume, which were further studied [69]. Screening is also typically used for discontinuous conditions, for example column, buffer and solvent choice; subsequent optimization by responsesurface methodology (RSM) is reserved for continuous conditions, for example temperature and gradient design [61]. Screening approaches can involve simple peak counting, followed by refinement of quality attributes reflecting resolution of specific components in RSM, thus automated peak tracking has recently become a valuable screening tool [61]. Sometimes DoE screening can be eliminated by conducting preliminary experiments and/or prior knowledge, enabling direct optimization by RSM [16, 60, 66]. In RSM, different designs, for example full factorial, fractional factorial, D-optimal, central composite, Box–Behnken, and Doehlert designs, can be used to evaluate the main, interaction and/or quadratic effects of the factors [11, 58]. Awotwe-



Otoo et al. applied a Box–Behnken design to study the effects of pH, temperature, and injection volume on peak resolution and USP tailing [69]. Debrus et al. used a factorialdesign to simultaneously optimize the method and estimate its robustness, considering gradient time, temperature, and pH as factors. They also suggested that if method optimization is the only objective, lighter designs, for example the fractional factorial or central composite designs, can be used [63]. The same type of design was applied by Molnár et al. to evaluate the effects of gradient time, temperature, pH, and ternary composition [60].

Another DOE approach has been proposed [71] in which the first phase of method development includes study of multiple analytical columns in combination with critical conditions, for example pH and organic solvent, and the second phase includes the remaining instrument factors. Data loss in the first step, as a result of both compound co-elution and peak exchanges, can be overcome by using DoE coupled with automatically computed trend responses, which can be of two general types: peak count-based and peak results-based [71]. Finally, mixture design, in particular, should be mentioned.

This is a special RSM in which the response is a function of the different proportions of the components [17]. A Scheffé 13-runs matrix with an hypothesized Scheffé special cubic regression model has been extensively used for optimization of microemulsion composition in CE methods for analysis of pharmaceuticals, followed by RSM for optimization of process factors such as cyclodextrin concentration, voltage, buffer concentration, and temperature [72, 73].

This matrix and this model have also been used to identify the range in which microemulsion formation is stable, instead of using traditional pseudo-ternary phase diagrams [43]. Piepel at al. joined the QbD initiative by using, as reported for the first time in the literature, a mixture-process variable (MPV) experimental design for simultaneously evaluating the effects of microemulsion composition (oil, surfactant and/or cosurfactant and buffer) and process conditions (voltage, buffer concentration, and pH) on peak efficiency in the development of and MEEKC method [16]. This new approach made it possible to identify significant interactive effects.

6.5. Definition of the Design Space (DS)

The DS for analytical methods has been defined as a multidimensional space which includes any combination of the variables that have been demonstrated to provide assurance of quality of the data produced by the method [59].

The DS is limited by the so-called edges of failure, outside which method performances are not acceptable [68]. Thus, the analytical method should be designed and validated not only under one fixed condition, but under a range of conditions [64, 66, 67]. Changes within the DS are not regarded as changes of the method and need no further regulatory approval [74, 64].

It is possible to either "establish independent Design Spaces for one or more unit operations, or to establish a single Design Space that spans multiple operations" [74]. This leads to the proposition by Monks et al. that DS in chromatography could be regarded as two entities: one column design space (CDS) and one eluent design space (EDS). When a robust column with a known equivalent column was found, establishment of the DS of all the other factors was carried out on the selected column [66]. DS can be presented graphically as a surface plot or can be explained mathematically by use of equations describing relationships between parameters for successful operation [74].

When multiple responses are involved, the most frequently used approaches are response overlay [68, 71] and desirability function [76, 69, 76].

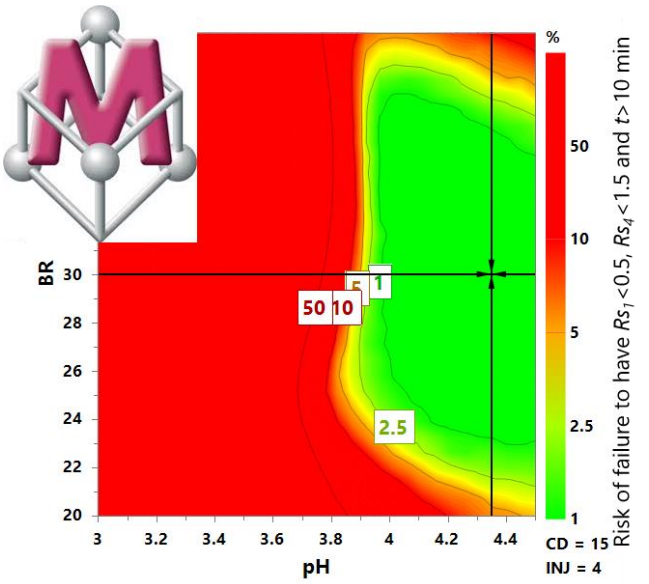
In response overlay, multiple response objectives are displayed on a graph, computed by software. In desirability studies, the global quality is summarized by use of a desirability index D.

This is obtained by assigning desired and accepted values to CQAs and by calculating the related partial desirabilities di, which have values between 0 and 1 where "1" corresponds to a highly desirable value and "0" to a unacceptable value.

Values which increase from 0 to 1 express an increase in the desirability of the criterion. Then, all di values are aggregated into D, which is to be maximized [76]. Piepel et al. specified an acceptable and an optimum limit for each response. On the basis of the calculated MPV models, the DS was defined by using the acceptable limits, and the optimum subregion within the DS by using the optimum limits.

Different papers by the same team of authors pointed out that when defining DS the model uncertainty should be taken into account, by estimating the errors in the criteria; otherwise no indication is provided about how well and how often the process can meet the specifications [76, 77, 75].

Thus, confidence in the predicted optimum conditions was investigated by means of a Monte-Carlo study of the propagation of the model's predictive errors [76, 77].



As an example, Debrus et al. estimated for each operating condition the expected probability that the CQAs are jointly within the specifications, and the results were presented as probability surfaces [77]. The probability that S is higher than a selected value ($\pi(S>0)$), known as quality threshold, was used to determine the DS. In this way quality in DS is not only predicted but also ensured by assessing the risk of not being within the acceptance limits. There are still no regulatory documents stating how to estimate the required quality level [7].

6.6. Selection of working points

There are several potential working points within the DS to choose from. The operating working point can be selected by choosing the optimum with a high desirability index [69], or the point giving the highest probability that the CQAs are within the specifications [77]. It can also be selected inside the design space and/or

optimum subregion on the basis of practical considerations, because the only requirement is that the point is included in the DS [16]. An important trend is to choose working points taking into account robustness [79], thus directly integrating this aspect of validation into method development. In this case, more than finding a single maximum of a function, the objective is to locate a response plateau. In some cases the optimum conditions can be directly selected from the CQA diagrams, which can help reveal areas of low and high robustness [61]. As an example, Monks et al. selected robust working conditions directly by observing the response surfaces, according to the CQA critical resolution, robustness range, and run time [66]. Swartz et al. suggested that methods at or near the edge of failures will only perform acceptably on average, thus these boundaries should be moved inside the mean performance design space to accommodate robustness and to define a "process operating space" [68]. A Monte-Carlo simulation was used to provide, for a given candidate point, a quantitative measure of the variation of a modeled response, e.g. the process capability index Cp. The robustness Cp values were modeled as additional response data sets and factor level settings were identified that would simultaneously meet mean performance and robustness objectives for all responses [68].

Robustness

In the past, robustness testing [74] was performed during validation, which often led to the late discovery of undesired surprises, with the method having to be re-optimized. In the QbD approach, robustness can be directly incorporated into the desired method qualities and evaluated during method development when selecting the DS and the working points [61, 78]. DS itself can be considered as a zone of theoretical robustness, because modification of the method conditions will not significantly affect separation quality [61]. Different approaches have been used to evaluate robustness. As previously mentioned, DoE with Monte-Carlo simulation was used to successfully integrate quantitative robustness modeling into LC method development [68].

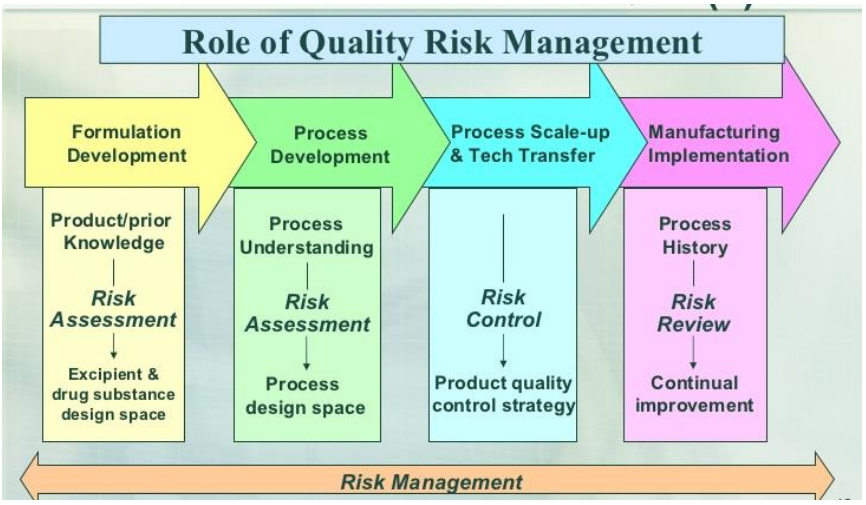
Another possibility is to perform a new DoE around the working point, forinstance with a Plackett–Burman design. Gavin et al. chose a two-level fractional

factorial design to consider the effect of mobile phase composition (pH, buffer concentration, organic solvent concentration, and ion-pair concentration) and column temperature on three critical aspects of performance (run time, tailing, and backpressure) [56]. Monks et al. visualized the region for method robustness as a geometric body within the resolution space in which critical resolution does not fall below a given value of RS. They then used a full factorial design centered on the working point to evaluate the robustness of the candidate points [78].

Molnár et al. discovered that the positions of critical pairs of peaks might not be constant between chromatograms, indicating the presence of several critical moving peaks. Thus, critical peak pairs within each individual resolution diagram were examined, and it was found that each robust region was made up of subregions corresponding to different critical peak pairs. Fixing the method in relation to a single critical band pair only was deemed both insufficient and impractical, because small changes in experimental conditions could change the critical peak pair. It was suggested it would be more reasonable to define the DS for several critical peak pairs above a certain value [70].

Method control

A control strategy (CS) should be designed to ensure that a product of required quality will be produced consistently [74]. ICH Q10 defines the CS as “a planned set of controls, derived from current product and process understanding, that assures process performance and product quality” [79].



With an analytical adaptation, CS has been defined as the controls on input factors that ensure the method meets both traditional system-suitability criteria and wider performance-related objectives [59, 48].

System suitability tests are a standard part of routine application and are typically established during method validation.

A Control Strategy (CS) generally includes appropriate system-suitability criteria to manage risks, thus QRA can also help to identify a specific control strategy.

An appropriate system-suitability test may be the only control element needed to ensure performance of the selected method [59], because it helps to identify failure modes and can prevent the generation of erroneous results [56].

An interesting example of CS enabled by a QbD approach was presented by Gavin et al. [56].

During robustness studies for an API impurity profile method, they found that resolution of the impurities of interest followed the same trend when method conditions such as n-propanol and temperature were varied. As a result, an early impurity pair was chosen for system suitability and became a convenient method CS, because the two compounds involved were easily obtained [56].

The system suitability test may need to be augmented with additional controls to properly manage risk. In this sense, a more common strategy could be the definition of accepted ranges for all the critical resolution values.

This can be achieved on the basis of the values measured during the system-repeatability study, because system precision provides valuable information about the

variability of the analytical system [10, 72, 73].

7. Results and Discussions

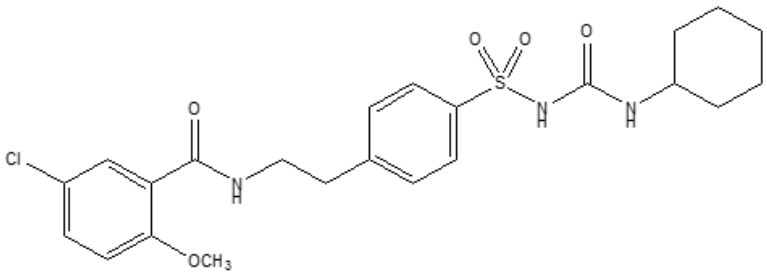
7.1. Fast analysis of glibenclamide and its impurities: quality by design framework in CE method development

The Quality by Design (QbD) concept is increasingly applied in the pharmaceutical industry to improve the quality of products, as recently recommended by International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) guideline Q8(R2) [52].

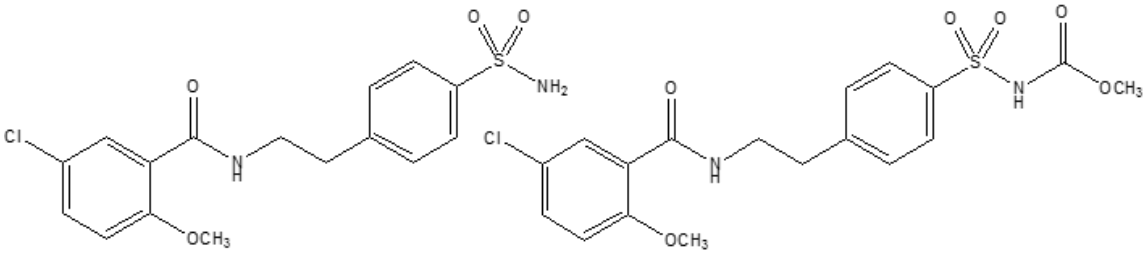
The fundamental tenet of QbD is a science- and risk-based approach to pharmaceutical development, where by quality is built into the product and not merely established by testing the end product.

Analytical QbD emphasizes the need to thoroughly understand the analytical system by an in-depth study of critical process parameters (*CPPs*) based on risk assessment and multivariate tools. The design space (*DS*) is determined as the multidimensional region of successful operating ranges for the *CPPs*, which lead to desired values for the critical quality attributes (*CQAs*). A fundamental aspect for the quality of the pharmaceutical product is the control of impurities, which often represents a critical analytical issue, owing to the similarity of the drug and the related substances in terms of chemical structure and behavior. Fast and validated methods should be available for the routine quality control of pharmaceutical dosage forms, which should allow the impurities to be determined at least at 0.1% w/w with respect to the main compound. For routine purity assessment, CE has been positioned as a complementary tool with respect to more commonly used high-performance liquid chromatography, owing to its several well-known benefits. Glibenclamide (GLI; CAS no. 10238-21-8), also known as glyburide, is a potent antidiabetic drug belonging to the second-

generation sulfonylurea class, and is administered by mouth in the treatment of type 2 diabetes mellitus [81]. Sulfonylureas appear to have several modes of action, apparently mediated by inhibition of ATP-sensitive potassium channels, from the initial increase of secretion of insulin from the β cells of the pancreas to the persistence of the hypoglycemic effect possibly due to inhibition of hepatic glucose production and to increased sensitivity to insulin [81]. In the European Pharmacopeia, GLI specified impurities are reported as A (I_A) and B (I_B) [82], and their structural formulas are reported below.



Glibenclamide (GLI)
1-[[4-[2-[(5-chloro-2-methoxybenzoyl)amino]ethyl]-phenyl]sulphonyl]-3-cyclohexylurea



Impurity A (I_A)
5-chloro-2-methoxy-*N*-[2-(4-sulphamoyl-phenyl)ethyl]benzamide

Impurity B (I_B)
Methyl [4-[2-[(5-chloro-2-methoxybenzoyl)-amino]ethyl]phenyl]sulphonyl]carbamate

The official chromatographic procedure described in the European Pharmacopeia for the analysis of GLI and its impurities has an analysis time of about 5 min.

Hence, the aim of this work was to develop a fast and simple CE method for the simultaneous analysis of GLI and its related substances for the quality control of GLI tablets. In this way, the DS was defined by consideration of the probability for the selected CQAs to fulfill their acceptance requirements, thus introducing a concept of assurance of quality

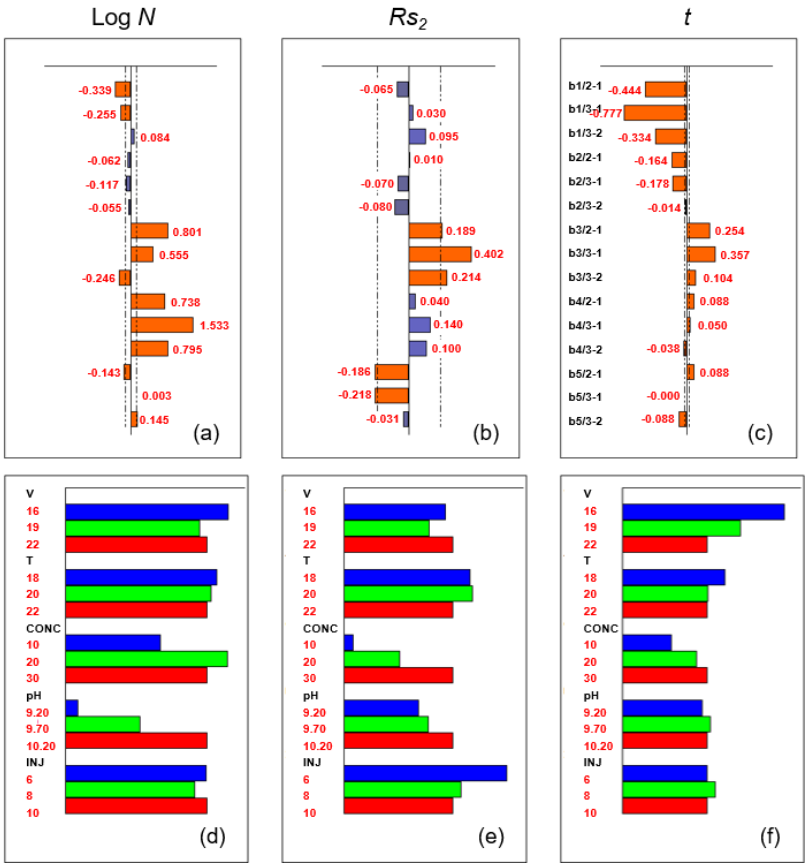
Method scouting and critical quality attributes

The analytical target profile of the method was represented by a CE method that baseline separates GLI and its impurities, allowing an accurate and precise determination of the analytes.

The analytes are weak acids (GLI, pK_a=5.3), thus leading to running preliminary experiments involving a simple capillary zone electrophoresis system based on borax buffer. The use of this separation system gave good results in terms of separation pattern, whereas the use of borate or borate/ phosphate buffer did not result in any improvement, and thus borax buffer was selected as the starting point for further method optimization. In these initial conditions, the migration order of the compounds was I_A, GLI, and I_B (with Rs_1 and Rs_2 indicating resolution values between the pairs I_A/GLI and GLI/I_B, respectively). Thus, the CQAs describing the global quality of the electropherogram were defined as I_A efficiency N, critical resolution Rs_2, and analysis time t. CQA requirements were set as $\log N \geq 4.50$, $Rs_2 \geq 0.40$, and $t \leq 2$ min. The target value for Rs_2 was selected as the value describing a baseline resolution between GLI and I_B peaks, which were very different for width characteristics [83]. The duration of the analysis was required to be equal to or less than 2 min.

Risk assessment, critical process parameters, and knowledge space

The CPPs were screened to investigate their effects on the selected CQAs by means of a 3^5//16 symmetric screening matrix [8], where each CPP was studied at three levels covering the knowledge space: voltage, 16, 19, and 22 kV; temperature, 18, 20, and 22 °C; buffer concentration, 10, 20, and 30 mM; buffer pH, 9.20, 9.70, and 10.20; injection time, 6, 8, and 10s. The highest voltage testedwas set at 22 kV to avoid a high developed current. The shortest injection time was fixed at 6 s to obtain adequate sensitivity, whereas values greater than 10 s were not considered in order to maintain good selectivity.



The bar length is proportional to the amplitude of the difference of the relative effects between two levels. The confidence intervals were calculated from the estimate of the experimental variance; the selected threshold of significance was set as 95% and is indicated by the dotted lines. In the second type of plot (d–f) the levels corresponding to the best results can be readily identified, as the longer bars correspond to levels which lead to higher values of the response. The highest level of each variable is represented by a red bar, of the same arbitrary length for all the factors. The indication of the presence of a statistically significant effect when passing from one level to another can be found in the first type of graph.

The maximization of logN was obtained with low voltages and medium–high buffer concentrations, and the higher increase of this CQA was undoubtedly noticed when switching from the lowest to the medium and to the highest level of buffer pH. As concerns Rs_2 (b-e), as expected a high buffer concentration and a short injection time

were preferred for maximization of this response. Finally, the most important effects on t were exerted by the change of voltage, temperature, and buffer concentration.

Response surface methodology and design space

Response surface methodology [8] was applied to conduct an in-depth investigation of the effects of the four CPPs on the CQAs throughout the new experimental domain. The following quadratic model was postulated to relate the CPPs with the CQAs:

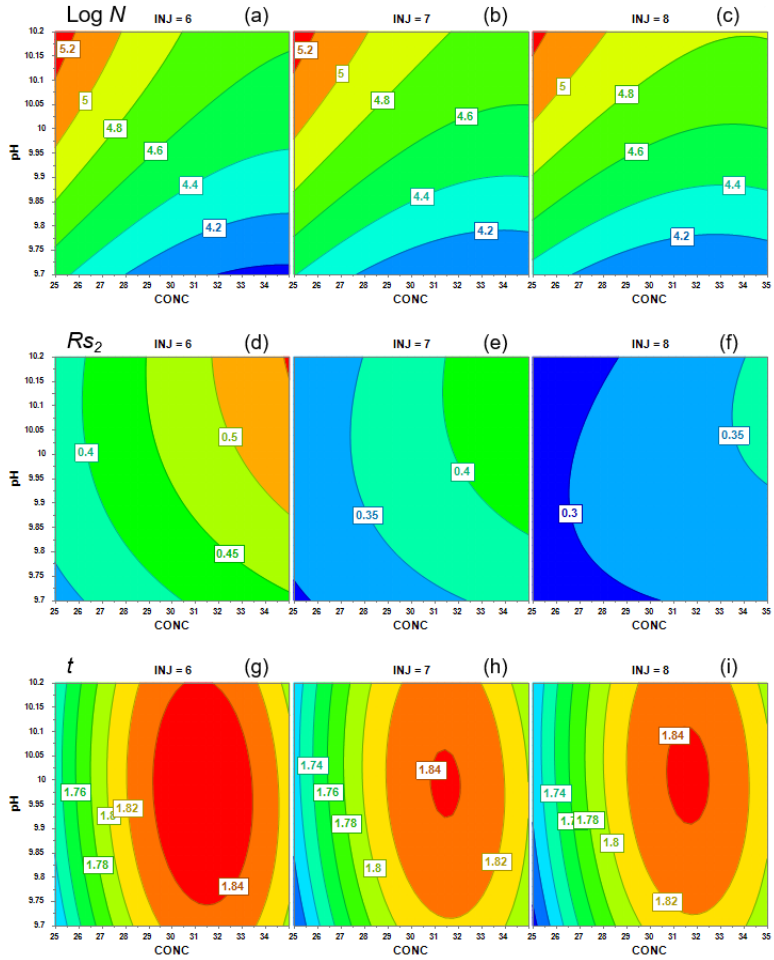
$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\beta_{11}x^2_1+\beta_{22}x^2_2+\beta_{33}x^2_3+\beta_{44}x^2_4+\beta_{12}x_1x_2+\beta_{13}x_1x_3+\beta_{14}x_1x_4 +\beta_{23}x_2x_3+\beta_{24}x_2x_4+\beta_{34}x_3x_4+\varepsilon$$

where y represents the experimental response, x_i the independent evaluated factors, β_0 the intercept, β_i the true coefficients, and ε the experimental error.

Thus, a Box– Behnken design, which requires three levels for each factor, was selected to calculate the coefficients of the model.

Box–Behnken designs are designs where the k variables are varied two at a time by 2^2 designs, while maintaining the remaining k-2 variables fixed at their middle level, and are usually very efficient in terms of the required runs.

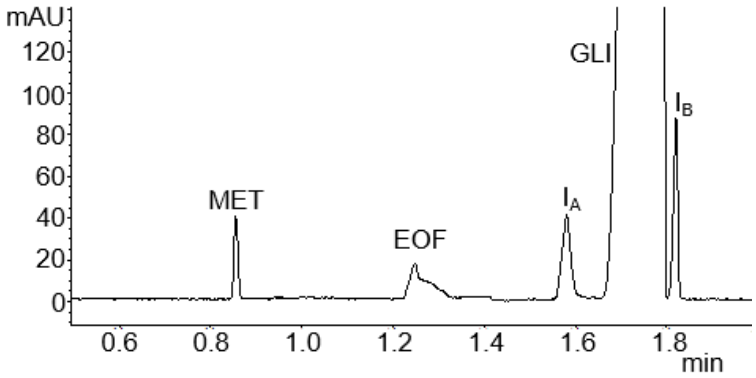
Four-dimensional contour plots were drawn reporting the calculated isoresponse curves in order to obtain detailed information on the behavior of the CQAs throughout the experimental domain, and are reported in the figure below for log N, Rs_2, and t, with the voltage maintained at its middle value.



For logN, the maximization of this response was obtained at low buffer concentrations and high levels of buffer pH, which were also shown to be the only significant factors. For Rs_2, the most important effect was obviously exerted by the injection time; however, buffer concentration and buffer pH also had a significant effect. Analysis time t was mainly influenced by voltage, but buffer concentration showed both linear and quadratic significant effects.

Monte Carlo simulations were performed by MODDE to take into account the model parameter uncertainty namely, propagating uncertainty from parameters to responses [85, 86, 87]. The selected level of probability for the CQAs to reach the desired values was set at $\pi \geq 90\%$. In this graph, the zone corresponding to the DS can be easily identified as the zone where the risk of error is 10% or less, and corresponds to the following ranges for the CQAs: voltage, 19-21 kV; buffer concentration, 26-34 mM;

buffer pH, 10.05-10.35. The original set point was selected as the working point, and the related electropherogram is shown, where all the compounds

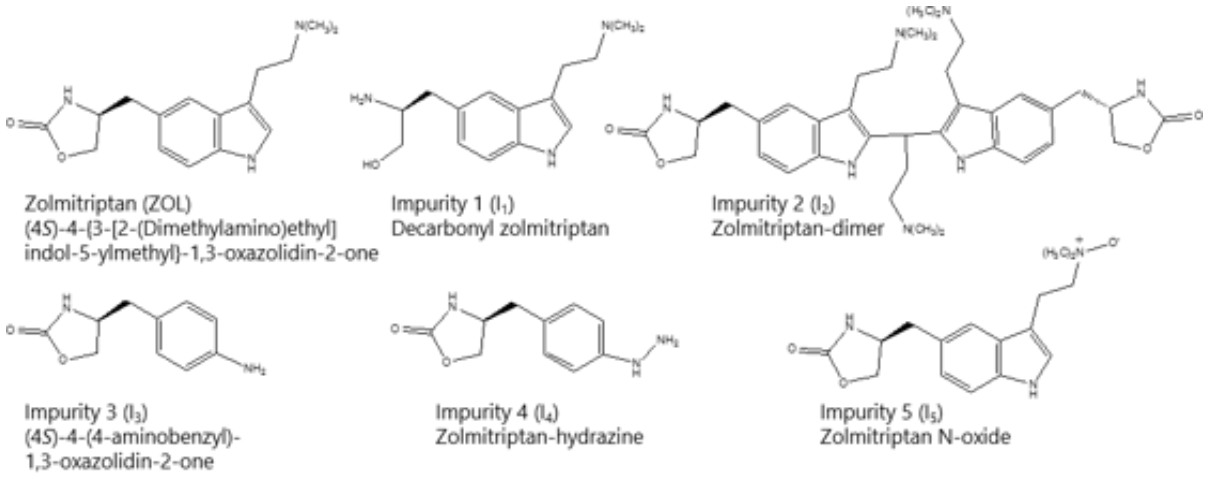


were baseline separated within 2 min with a generated current equal to 110 μA. Finally, a control strategy for the method was accomplished [52] on the basis of the data collected during method development. By means of a proper control strategy, the method performances can be monitored to ensure that the method remains in compliance with the defined analytical target profile. In this study, the extreme values of the CQAs measured during system repeatability studies were selected as limits for the system suitability intervals [88]: logN, 4.69-5.03; Rs_2, 0.44-0.51; t, 1.77-1.92 min.

7.2. Analytical Quality by Design in pharmaceutical quality assurance: Development of a CE method for the analysis of zolmitriptan and its impurities

Recently, pharmaceutical quality assurance had to face up to the newly introduced Quality by Design (QbD) principles concerning pharmaceutical development, described in International Conference on Harmonization (ICH) guideline Q8 [52]. Nevertheless, the advantages of QbD application to analytical methods is very attracting because analytical QbD supports the development of a robust analytical method with adequate performances and increases the regulatory flexibility of the analytical method itself [83,

85]. This flexibility consists in the freedom to change method operating parameters within the DS, which can be realized by demonstrating an enhanced knowledge of method performances by a suitable Design of Experiments protocol [52]. The present study was aimed to comprehensively apply QbD principles to the development of a CE method suitable for the simultaneous analysis of zolmitriptan (ZOL) and its impurities. ZOL (CAS139264-17-8) is a single enantiomer selective serotonine ($5HT_{1B/1D}$) receptor agonist, used for acute treatment of migraine attacks and of cluster headache [90]. ZOL belongs to the second generation triptans, which were developed with the goal of creating a more lipophilic, centrally active and rapidly absorbed drug than sumatriptan, the first triptan [91]. Different related impurities may be found in ZOL bulk samples and coated tablets [92]. The five main impurities considered in this paper were named zolmitriptan impurity 1 (I_1), zolmitriptan impurity 2 (I_2), zolmitriptan impurity 3 (I_3), zolmitriptan impurity 4 (I_4), zolmitriptan impurity 5 (I_5) according to their final migration order, and their structures are here shown.



Critical process parameters and investigation of knowledge space

The critical process parameters (CPPs) were selected as the classical ones which are generally known to have the strongest impact on CZE analysis and consequently on

the CQAs, namely buffer concentration (*CONC)*, pH (*pH*), *V*, and *T*. Other important factors as capillary length and injection time were fixed from scouting experiments, selecting the experimental conditions corresponding to the best compromise between selectivity and *t*.

A screening study was performed, considering three levels for each of the CPPs covering the knowledge space: *CONC*, 50-100-150 mM, *pH*, 2.5-3.0-3.5; *V*, 20-25-30 kV; *T*, 19-22-25°C. The responses were represented by the selected CQAs, and a 9-run screening symmetric matrix.

The value of pH was very important for the response Rs_4, and better separations were obtained at the lower pH values. A medium level of *T* was also preferable for maximizing this CQA. As for Rs_5, the changes of level of each CPP were significant on this response.

Medium and high levels of *CONC* were highly preferred in order to increase the separation, as well as low and medium pH values. An unexpected effect was the increase of Rs_5 with the increase of *V*, probably due to the general beneficial effect of high values of this factor on peak efficiency.

Maximization of this response was also obtained by medium-high levels of *T*. As expected, the most important effect on *t* was exerted by a change of *V* value, even if significant effects were also noticed by varying *CONC* and *T*.

Finally, as concerns efficiency, a drastic increase of this response was obtained by using medium-high levels of *CONC*. The other three factors had a comparable importance and for all of them an increase of Log *N* was observed at high levels.

Response surface methodology and design space

Response surface methodology was aimed to obtain detailed information on the effect of each CPP on the CQAs and to draw a predictive map for each of the CQAs covering the new experimental domain: *CONC*, 100–150mM; *pH*, 2.5–3.0; *T*, 21–25°C.

A quadratic model describing the relationship between the factors and the responses was postulated as the following:

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_{11}x^2_1+\beta_{22}x^2_2+\beta_{33}x^2_3+\beta_{12}x_1x_2+\beta_{13}x_1x_3+\beta_{14}x_1x_4+\beta_{23}x_2x_3+\beta_{24}x_2x_4+\beta_{34}x_3x_4+\varepsilon$$

where y is the experimental response, x_i the independent variables, β_0 is the intercept, β_i are the true coefficients, and ε is the experimental error. Box–Behnken design [8] was used to estimate the coefficients of the model.

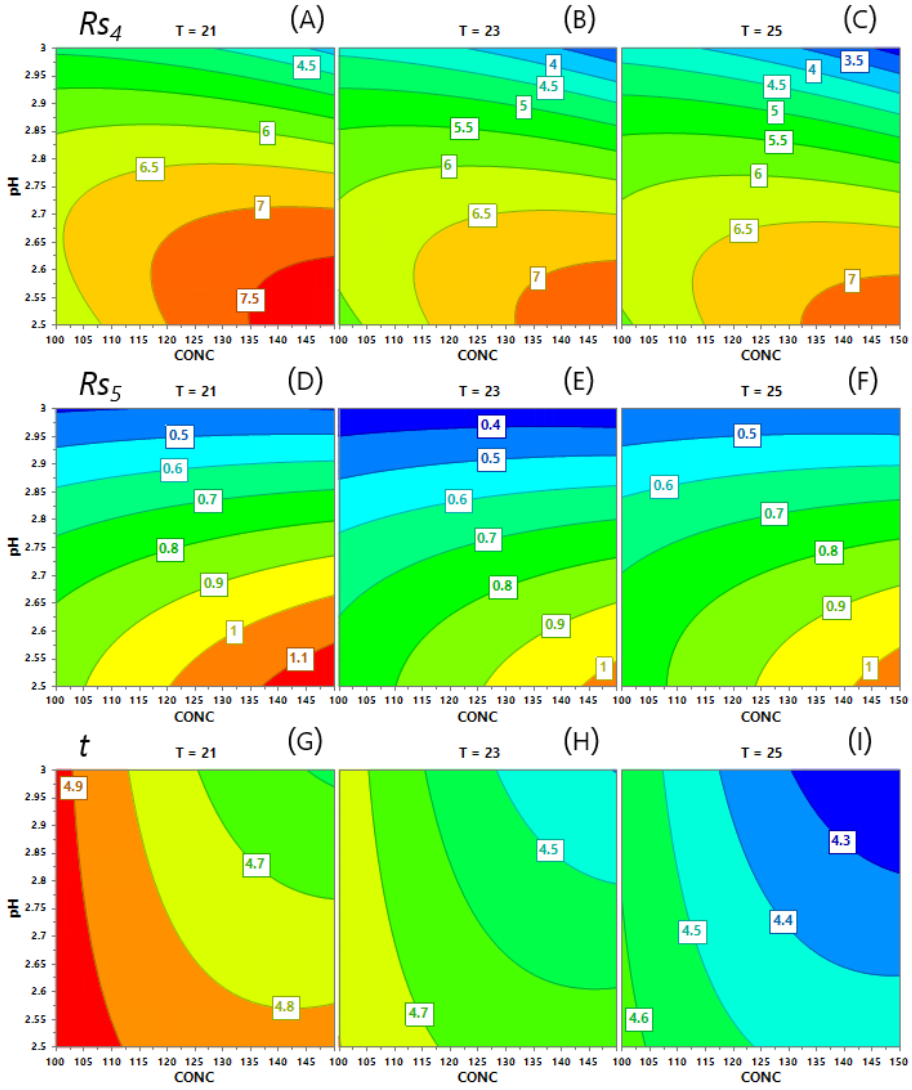
In this type of design, each factor is studied at three levels; the three-factor Box–Behnken design requires 12 runs plus the replicates at the center point, which were selected as three for a total of 15 runs.

The ANOVA demonstrated that all the models were valid and significant, apart from the model for Log *N*, which resulted to be valid but not significant.

This indicated that in the considered domain the value of I_2 efficiency was not significantly influenced by the change of the factors.

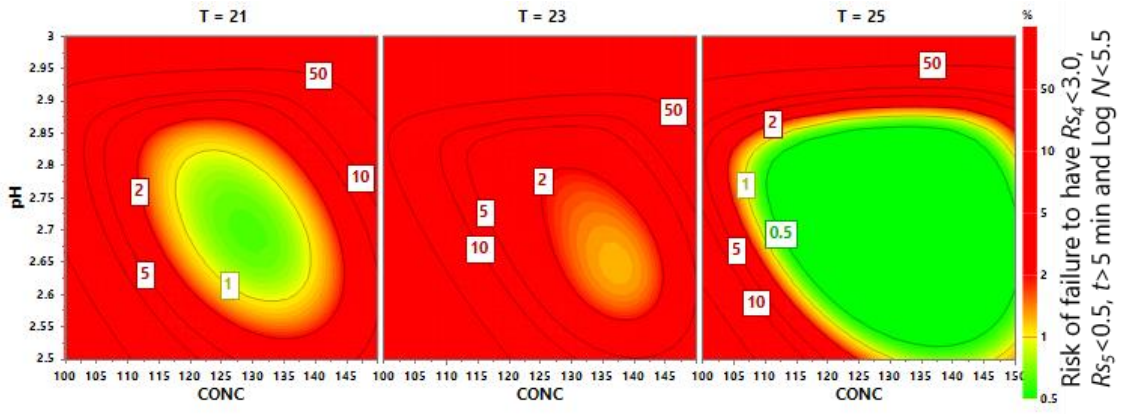
The obtained values of goodness of fit R^2 and goodness of prediction Q^2 for the other responses were, respectively: for Rs_4, R_2 =0.997, Q^2 =0.953; for Rs_5, R^2 =0.983, Q^2 = 0.810; for t, R^2 = 0.992, Q^2 = 0.908.

The resulting 4D contour plots are reported for Rs_4 (A–C), Rs_5 (D–F), and t (G–I).



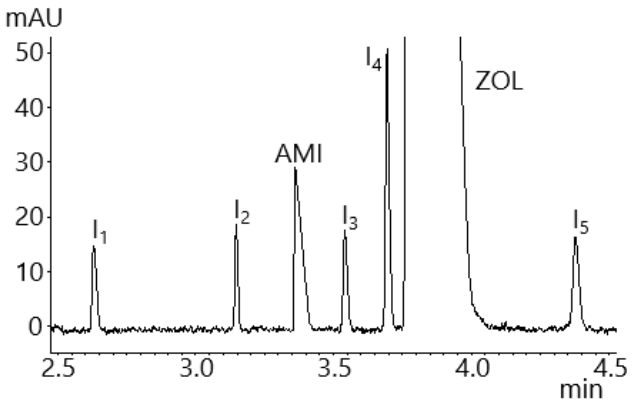
The isoresponse lines for Rs_4 and Rs_5 showed the same trend, pointing out that the best separations were obtained at low *T*, high concentration, and low pH, with the most important effect played by pH. Also, for both the responses graphic analysis of coefficients (not reported) evidenced quadratic effects of both *pH* and *T*, together with an important negative interaction between *CONC* and *pH*. As concerns *t*, the minimization of this CQA was obtained by applying high values of all the CPPs. A positive quadratic effect of *CONC* and a negative interaction between *CONC* and *pH* were also noticed.

The original set-point selected by the software for calculating DS was the following: *CONC*, 138 mM, *pH*, 2.74, *T*, 25°C. By setting a risk of error equal to 1% ($\pi \leq 99\%$), the probability maps reported were obtained, where the DS is colored in green.



The DS corresponded to the ranges: *CONC*, 115–150 mM; *pH*, 2.54–2.94; *T*, 24–25°C, where the selected interval for each factor was limited inside the domain tested by response surface methodology highly reduced by introducing a probability concept.

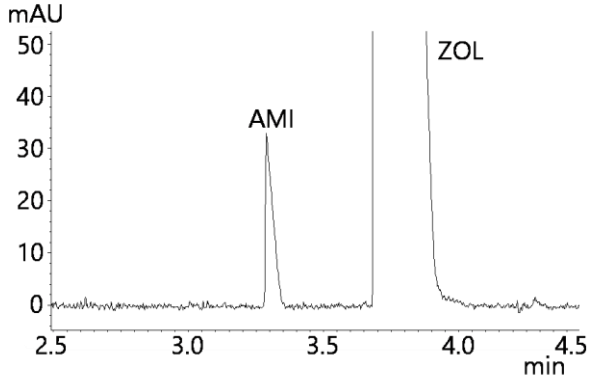
Applying the working point conditions, corresponding to the original set-point, an electropherogram was obtained where the baseline separation of the analytes was achieved in less than 5 min, with a generated current of about 133 µA.



DS was validated by using a Plackett–Burman matrix where the considered −1 and +1 levels for the factors corresponded to the extremes of the corresponding DS interval. The measured and calculated values for the CQAs showed a good agreement.

Validation and application

A control strategy [52] was finally planned in order to ensure that the method has adequate analytical performances and thus can be used to obtain reliable analytical data. The control strategy was based on the definition of system suitability limits, which corresponded to the lower and the higher CQAs values observed during system repeatability studies [88]. The CQAs accepted intervals were the following: Rs_4, 5.84–6.40; Rs_5, 0.66–0.86; *t*, 4.38–4.54; Log *N*, 5.55–5.72. The conclusions drawn from robustness test concerning the careful preparation of the BGE also constituted an important part of the control strategy of the method [52]. The developed method was validated according to ICH guidelines [88], and the performances were adequate for the method aim. Anyway, a validation issue concerning sensitivity was observed for I_2, in spite of the good efficiency of the related peak. As a matter of facts, the dimeric structure of this impurity did not allow the desired LOQ to be reached (0.1% with respect to the main compound) due to its low absorbance (I_2 LOQ: 0.0036 mg/mL, corresponding to 0.18%).



Analysis of a real sample of Zomig® coated tablets. Working point conditions: temperature, 25 °C; voltage, 30 kV; BGE, 138 mM phosphate buffer pH 2.74.

The developed method was finally applied to real samples of Zomig® coated tablets, labeled to contain 2.5 mg ZOL. Four analyses were performed, and the results were in agreement with the declared content ($\alpha/2 = 0.025$): percentage of label claim, 98.4 ± 1.1%; RSD, 0.7%. None of the five considered impurities was detected.

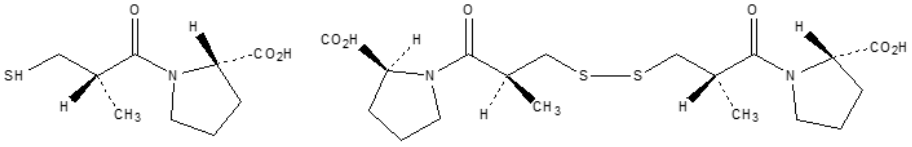
7.3. Cyclodextrin- and solvent-modified micellar electrokinetic chromatography for the determination of captopril, hydrochlorothiazide and their impurities.

Among the analytical methods applied for impurity assay of pharmaceuticals, HPLC is the main method recommended by different pharmacopeias, but capillary electromigration methods represent a significant alternative.

This is especially due to the inherent flexibility for the various available operative modes and to the orthogonality with respect to chromatographic principles, besides other well-known advantages including lower costs and higher eco-compatibility [2, 94].

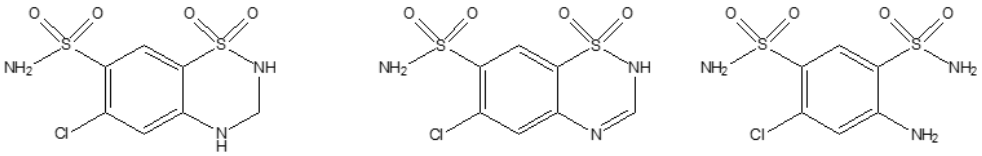
The aim of this study was to set up a fast and simple capillary electrophoresis (CE) method for the simultaneous determination of captopril (CPT), hydrochlorothiazide (HCT) and their related substances in the combined dosage form, following QbD tenets. CPT is a thiol-containing angiotensin-converting enzyme inhibitor used in the management of hypertension, in heart failure, following myocardial infarction and in diabetic nephropathy.

The structural formulas of the two drugs and of the related impurities considered in this thesis are here shown.



Captopril (CPT)
(2*S*)-1-[(2*S*)-2-Methyl-3-sulfanylpropanoyl] pyrrolidine-2-carboxylic acid
CAS 62571-86-2

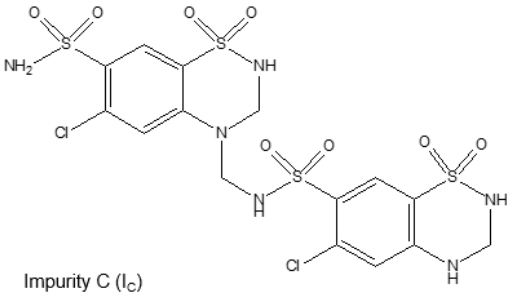
Captopril disulfide (I_{DIS})
1,1'-[disulfanediylbis[(2*S*)-2-methyl-1-oxopropane-3,1-diyl]]bis[(2*S*)-pyrrolidine-2-carboxylic] acid
CAS 64806-05-09



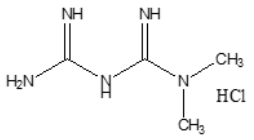
Hydrochlorothiazide (HCT)
6-Chloro-3,4-dihydro-2*H*-1,2,4-benzothiadiazine-7-sulfonamide 1,1-dioxide
CAS 58-93-5

Chlorothiazide (I_A)
6-Chloro-2*H*-1,2,4-benzothiadiazine-7-sulfonamide 1,1-dioxide
CAS 58-94-6

Salamide (I_B)
4-Amino-6-chlorobenzene-1,3-disulfonamide
CAS 121-30-2



Impurity C (I_C)
6-Chloro-*N*-[(6-chloro-7-sulfamoyl-2,3-dihydro-4*H*-1,2,4-benzothiadiazin-4-yl 1,1-dioxide)methyl]-3,4-dihydro-2*H*-1,2,4-benzothiadiazine-7-sulfonamide 1,1-dioxide
CAS 402824-96-8



Metformin hydrochloride (MET)
1,1-Dimethylbiguanide hydrochloride
CAS 1115-70-4
(internal standard)

According to the European Pharmacopeia [95], HCT specified impurities are reported as I_A, I_B and I_C. I_A and I_C are two main process impurities, while the primary HCT degradation pathway yields I_B and formaldehyde by hydrolysis. CPT may present several impurities, but in aqueous solution undergoes spontaneous oxidative degradation at the sulphydryl group, generating the dimer captopril disulphide (I_{DIS}), which represents the main degradation product [96, 97]. In general, CPT determination is very challenging mainly due to the aliphatic structure that lacks any strong chromophore, and to the vulnerability to oxidation of the sulphydryl group. The cis-trans isomerization of CPT can be observed during chromatographic or electrophoretic separation, since the isomerization occurs on the time order of minutes. This chemical behavior can give rise to peak splitting or to a typical pattern constituted by a plateau between the two conformer peaks [73, 98].

In this study, the CE method development was carried out implementing each step of analytical QbD approach [99]. A large part of the experiments of the scouting phase

was dedicated to the selection among different pseudostationary phases based on micelles, with or without additives, in order to overcome CPT detection and isomerization issues.

Analytical target profile, method scouting, critical quality attributes

From the first experimental runs, the main analytical issues were evidenced and were mainly attributable to: i) presence of analytes with ionizable moieties presenting different acid-base properties; ii) low UV absorbance value and efficiency of IDIS; iii) possible peak broadening or splitting of CPT due to the interference between electrophoretic migration and cis-trans reaction of isomerization. In particular, this interference can depend on several factors including the chemical nature of the system (BGE, electro-osmotic flow, electrophoretic mobility of the molecules), effective length of the capillary, applied electric field and temperature [100]. In this study, the possibility of using high temperature for avoiding peak splitting was discarded and ambient temperature was selected in order to provide more practical experimental conditions.

Hence, a large part of the experiments of the scouting phase was intended to the selection of a suitable operative mode enabling to overcome separation, detection and isomerization issues. In the scouting phase the concentration of all the compounds was kept low and equal to 0.06 mg ml^{-1}, in order to obtain clear information on the separation pattern and the migration behavior of the compounds. Capillary zone electrophoresis (CZE) with plain buffers as well as MEKC with different pseudostationary phases, with or without additives, were examined in order to find an useful separation system.

It was found that the use of γ-CD was fundamental for obtaining a great increase in IDIS peak efficiency, also producing a decrease in analysis time and an increase in selectivity. By adding *n*-butanol to this system, further beneficial effects on IDIS peak efficiency were observed, confirming that this organic modifier can exploit an active role in controlling the partitioning of the analytes, with a great potential for ameliorating the electrophoretic pattern [101, 102]. Hence, the sodium cholate-based MEKC operative mode with the addition of γ-CD and *n*-butanol enabled to overcome all the

main analytical issues which had been evidenced in the preliminary experiments, and was selected for further optimization steps. With this separative system, the migration order of the compounds was the following: MET (internal standard), I_B, HCT, I_C, CPT, I_A, I_{DIS}, with the resolution values indicated by Rs_1 (I_B/HCT), Rs_2 (HCT/I_C), Rs_3 (I_C/CPT), Rs_4 (CPT/I_A), Rs_5 (I_A/I_{DIS}). The responses which were selected as CQAs were critical resolution values Rs_1 and Rs_4, whose desired values were both set as ≥1.5, and analysis time (*t*), for which the accepted value was set as ≤3 min.

Critical process parameters and investigation of knowledge space

In the selected separative system, the separation mechanism was based on the different mobilities of the solutes, their partitioning in the sodium cholate micelles and on host-guest inclusion complexation in the cyclodextrin, with the consequence of a high number of experimental factors which could affect one or more CQAs. Seven CPPs were identified and were investigated by a multivariate screening procedure. The CPPs were both related to the BGE composition, *e.g.* borate concentration (*B CONC*), pH of the buffer (*pH*), sodium cholate concentration (*CHOL CONC*), *n*-butanol concentration (*BUT CONC*), γ-cyclodextrin concentration (*CD CONC*), and related to instrumental settings, *e.g.* temperature (*T*) and voltage (*V).*

The effect of the change of levels of the CPPs was studied by the symmetric screening matrix 3^7//16.

Exp. no.	*B CONC* (mM)	*pH*	*CHOL CONC* (mM)	*BUT CONC* (%v/v)	*CD CONC* (mM)	*T* (°C)	*V* (kV)	Rs_1	Rs_4	*t* (min)
1	80	8.40	60	8	15	24	26	1.31	0.01	2.84
2	120	8.00	60	6	20	21	26	0.74	0.01	2.92
3	120	8.80	50	6	15	24	22	1.85	0.98	3.56
4	100	8.80	70	4	15	21	26	2.18	1.10	2.41
5	120	8.40	70	8	10	21	22	1.87	0.01	3.97
6	100	8.80	60	8	20	18	22	2.16	1.71	4.00
7	100	8.40	70	6	20	24	18	1.47	0.01	3.07
8	100	8.40	60	6	15	21	22	1.38	0.78	3.71
9	120	8.40	60	4	15	18	18	1.32	0.86	4.24

10	80	8.80	60	6	10	21	18	1.85	1.88	4.75
11	80	8.00	70	6	15	18	22	0.01	0.01	3.87
12	100	8.00	50	8	15	21	18	0.01	0.01	5.27
13	80	8.40	50	4	20	21	22	0.01	0.01	3.20
14	100	8.00	60	4	10	24	22	0.01	0.01	2.93
15	100	8.40	50	6	10	18	26	1.63	0.88	2.76
16	100	8.40	60	6	15	21	22	1.45	0.82	3.55

Rs$_1$, resolution between I_B/HCT; *Rs*$_4$, resolution between CPT/I_A; *t*, analysis time.

Among the measured CQAs, for *Rs*$_1$ and *Rs*$_4$ a considerable number of experiments with a complete overlap of the peaks was reported, evidencing the particular criticism for selectivity responses; on the other hand, all the values obtained for *t* were lower than 5 min.

The graphs showed that *Rs*$_1$ was increased by medium-high levels of *B CONC*, *CHOL CONC* and *BUT CONC* and high levels of pH; for *Rs*$_4$, the preferred settings were high level for pH, medium level for *CHOL CONC*, medium level for *BUT CONC*, while the maximization of this response was obtained at low levels of *CD CONC*. For both *Rs*$_1$ and *Rs*$_4$, the change of levels of pH was undoubtedly of outmost importance, confirming the important role of this factor. As regards *t*, the only factor which exerted a significant effect was voltage.

Response surface methodology and design space

A five-factors FCD was employed for building a second order quadratic model for the selected CQAs. Each CPP was studied at three levels, covering the new experimental domain, with the star points located at the center of each face of the factorial space, with $\alpha=\pm1$. Twenty-nine experiments were required, including three replicates at the center point, and the experimental plan with the measured responses is here shown.

Exp. no.	*pH*	*CHOL CONC* (mM)	*BUT CONC* (%v/v)	*CD CONC* (mM)	*V* (kV)	Rs_1	Rs_4	*t* (min)
1	8.20	55	5	8	28	1.61	0.8	2.19
2	8.80	55	5	8	24	1.95	1.75	2.09
3	8.20	75	5	8	24	1.46	0.34	2.92
4	8.80	75	5	8	28	2.46	2.96	2.08
5	8.20	55	7	8	24	1.44	0.01	3.25
6	8.80	55	7	8	28	2.31	2.77	2.27
7	8.20	75	7	8	28	1.54	1.40	2.50
8	8.80	75	7	8	24	2.10	2.85	2.95
9	8.20	55	5	16	24	1.59	0.01	2.94
10	8.80	55	5	16	28	1.83	2.24	2.13
11	8.20	75	5	16	28	1.61	0.35	2.21
12	8.80	75	5	16	24	2.28	0.01	2.70
13	8.20	55	7	16	28	1.76	0.01	2.22
14	8.80	55	7	16	24	2.03	1.77	3.15
15	8.20	75	7	16	24	1.53	0.01	2.88
16	8.80	75	7	16	28	2.03	0.01	2.21
17	8.20	65	6	12	26	1.73	0.01	2.60
18	8.80	65	6	12	26	2.32	2.21	2.46
19	8.50	55	6	12	26	1.58	1.27	2.64
20	8.50	75	6	12	26	1.54	1.52	2.59
21	8.50	65	5	12	26	1.65	1.42	2.25
22	8.50	65	7	12	26	1.75	1.75	2.53
23	8.50	65	6	8	26	1.75	1.89	2.25
24	8.50	65	6	16	26	1.68	1.61	2.36
25	8.50	65	6	12	24	1.72	1.39	3.09
26	8.50	65	6	12	28	1.93	1.70	2.39
27	8.50	65	6	12	26	1.68	1.48	2.69
28	8.50	65	6	12	26	1.61	1.38	2.68
29	8.50	65	6	12	26	1.69	1.29	2.70

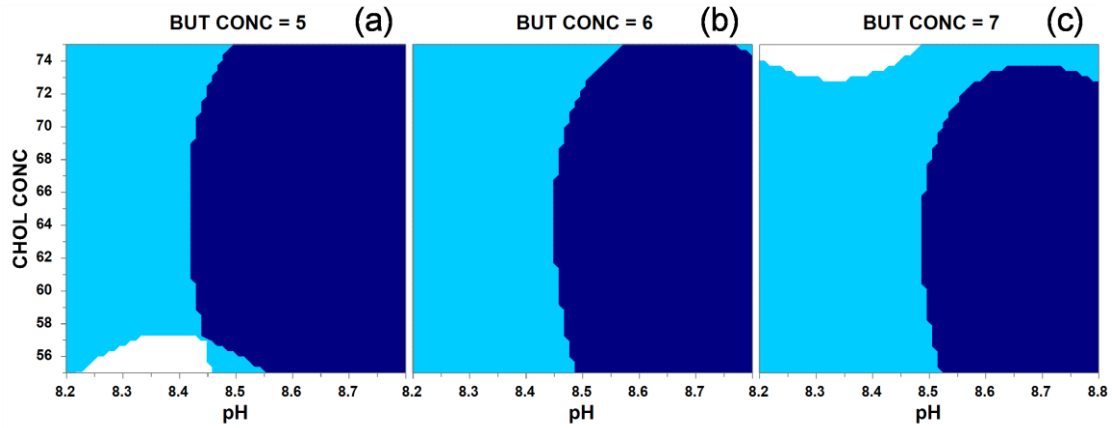
Rs_1, resolution between I_B/HCT; Rs_4, resolution between CPT/I_A; *t*, analysis time.

The calculated models were edited by removing some of the interaction or quadratic terms which did not exert a significant effect on the response, in order to ameliorate values of coefficient of determination R^2 and coefficient of predicted variation Q^2. The following final values were achieved: Rs_1, Q^2=0.923, R^2=0.983; Rs_4, Q^2=0.821, R^2=0.963; *t*, Q^2=0.840, R^2=0.976. The ANOVA demonstrated that the

regression models for Rs_1 and Rs_4 were statistically valid and significant, while the model for *t* was observed to be not valid due to the small experimental variance.

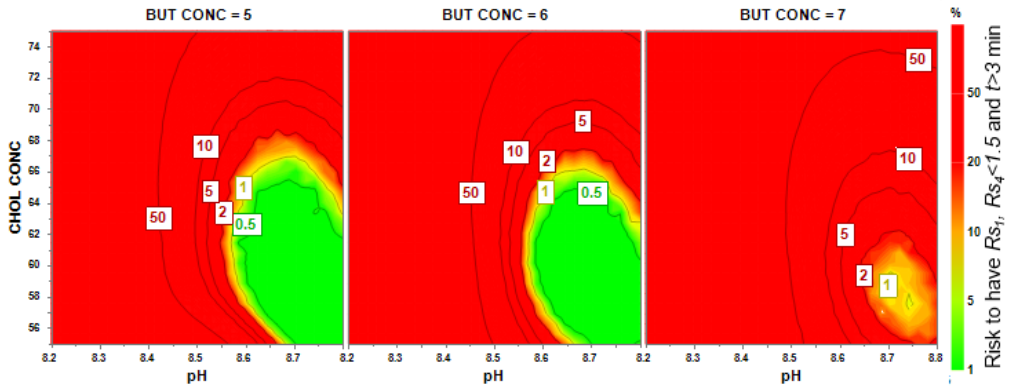
The sweet spot plots were drawn maintaining *CD CONC* and *V* at their central values.

These graphs made it possible to highlight by different colours the areas where one (white), two (pale blue) or three (dark blue) predicted CQAs fulfilled the requirements. The plots confirmed that it is necessary to set the pH value at medium-high levels, but wide zones leading to required CQAs values were obtained for all the values of *BUT CONC*.



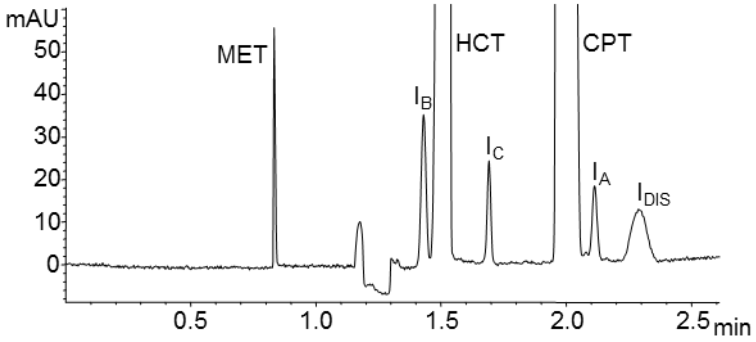
Monte-Carlo simulations were performed to propagate the error in prediction of the models from the factors to the response, giving access to the distributions of the responses for each operating condition of the RSM [103]. The set-point was the following: pH, 8.55; *CHOL CONC*, 64 mM; *BUT CONC*, 6.1 %v/v; *CD CONC*, 12 mM; *V*, 27 kV.

The DS was calculated from this set-point by using a search function [104] that expands the possible factor ranges to the largest possible range where all the response predictions are still within the specifications at a probability level of $\pi \geq 99\%$, and corresponded to the following intervals: pH, 8.48-8.62; *CHOL CONC*, 60-68 mM; *BUT CONC*, 5.4-6.8 %v/v; *CD CONC*, 11-13 mM; *V*, 26-28 kV.

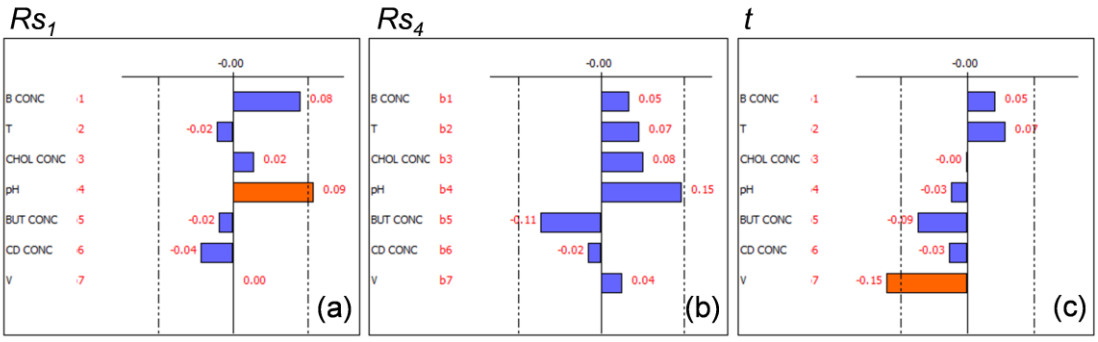


The DS is highlighted in green and is included in the line corresponding to 1% risk of failure. In contrast to what was observed in the sweet spot plots, the DS cannot be calculated when setting *BUT CONC* at 7 %v/v.

The working conditions corresponded to the set-point used for calculating the DS and made it possible to achieve the complete separation of the analytes in less than 3 min.



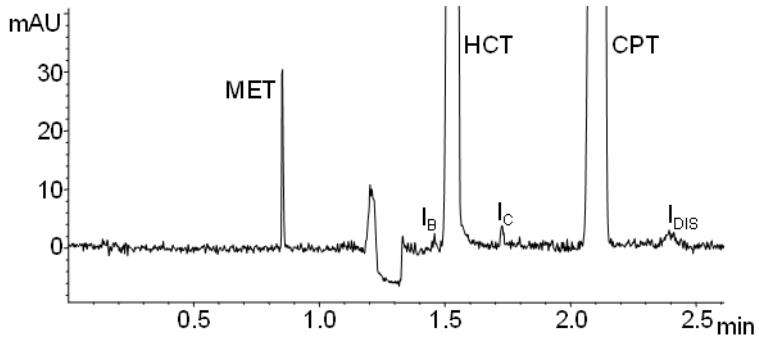
For robustness testing, the 8-run Plackett-Burman design was employed in order to calculate the main effects of small variations of all the seven CPPs considered in the screening phase, *e.g. B CONC* (98-102 mM), *T* (20-22 °C), pH (8.45-8.65), *CHOL CONC* (62-66 mM), *BUT CONC* (5.6-6.6 %v/v), *CD CONC*, 11-13 mM, *V*, 26-28 kV. The graphic analysis of effects revealed that pH was the only factor which exerted a significant effect on Rs_1, while only *V* was influent on *T*.



Graphic analysis of effects in robustness study. The orange bars exceeding the dotted line are related to the factors which exert a significant effect on the response: (a) Rs_1, resolution between I_B/HCT; (b) Rs_4, resolution between CPT/I_A; (c) *t*, analysis time.

Validation and application

After method validation, Acediur® tablets containing 50 mg CPT and 25 mg HCT were analysed and a typical electropherogram is here shown.



The percentage of claimed amount of the drugs was: CPT, 98.2±2.4% with a RSD of 1.5%; HCT, 98.7±1.9% with a RSD of 1.2% (n=4, α/2=0.025). As for the impurities, I_A was not detected, while the other impurities were detected at concentration values lower than the respective LOQ. The electropherogram of the real sample was also spiked with standard solutions of the impurities for verifying the

identity of the impurity peaks and for excluding the possibility of overlap of CPT and I_A peaks.

7.4. Chiral cyclodextrin-modified micellar electrokinetic chromatography and chemometric techniques for green tea samples origin discrimination

Green tea (GT) is the second most traditional style of tea product, representing 20-22% of world tea production, and is manufactured by steaming or drying fresh tea leaves and inactivating the enzymes to prevent the oxidation of the tea polyphenols [105].

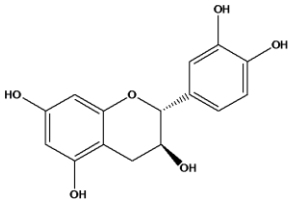
GT polyphenolic antioxidants are mostly represented by catechins, a group of compounds having a flavan-3-olic structure, which are easily extracted into water. Catechins may constitute up to 30% of the dry mass of tea leaves and are the major constituents of GT excluding carbohydrates, proteins and lignin [106].

GT has been extensively studied for its antioxidant activities and potential for reducing lifestyle-related diseases. The principal components of tea having biological effects have been identified as catechins and xanthines.

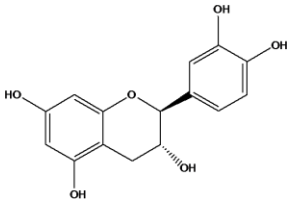
Xanthines are responsible for the stimulating effects; caffeine (CF) is a central nervous system and cardiac stimulant and has a diuretic effect, while theobromine (TB), which is present in lower amounts, has also a diuretic effect [107, 108].

Catechins are mainly responsible for the astringent and bitter taste of tea infusion, while CF can enhance observably tea flavor [109].

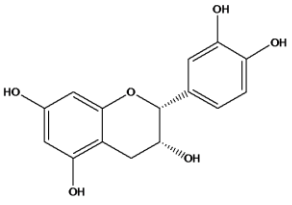
The composition of tea varies according to the species, season, climate, age of the leaf (plucking position) and cultivation and also to the technologies applied during extraction, concentration and preservation process. The most abundant catechins in GT are (+)-catechin ((+) C)), (-)-catechin ((-) C), (-)-epicatechin (EC), (-)-epigallocatechin (EGC), (-)-epicatechin gallate (ECG), (-)-epigallocatechin gallate (EGCG), whose structures, together with those of CF and TB, are here reported.



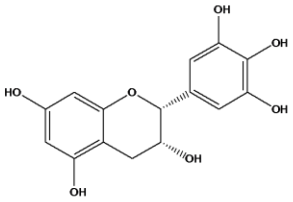
(+)-catechin ((+)C)
CAS 154-23-4



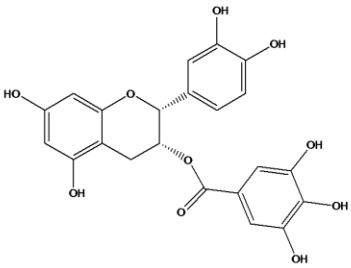
(-)-catechin ((-)C)
CAS 18829-70-4



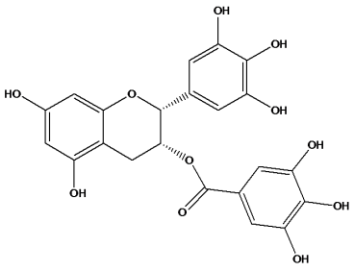
(-)-epicatechin (EC)
CAS 490-46-0



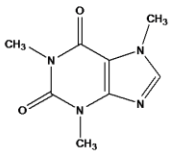
(-)-epigallocatechin (EGC)
CAS 970-74-1



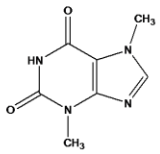
(-)-epicatechin gallate (ECG)
CAS 1257-08-5



(-)-epigallocatechin gallate (EGCG)
CAS 989-51-5



Caffeine (CF)
CAS 58-08-2



Theobromine (TB)
CAS 83-67-0

In the field of classification according to chemical composition, having at disposal a rapid and accurate analytical method is essential. The method was applied to the quantitation of catechins and methylxanthines in 92 GT samples of different geographical origin (Japan and China), having undergone different storage conditions and manufacturing processes.

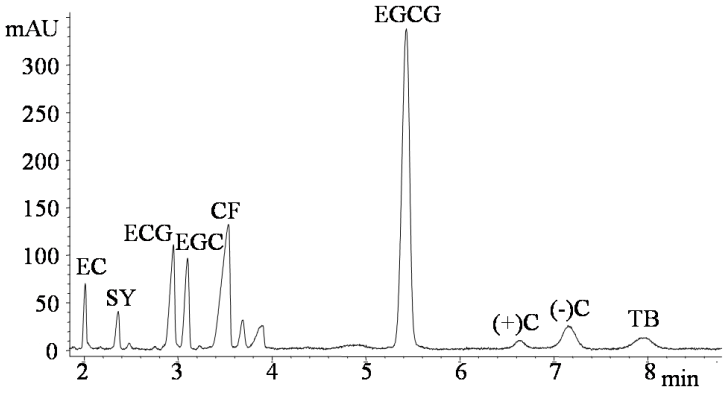
Analyte	*Range* ($\mu g\ mL^{-1}$)	*a*	s_a	*b*	s_b	R^2
EC	0.005-0.300	1592.49	74.50	5.53	13.63	0.9828
ECG	0.005-0.150	2530.60	63.55	-10.22	5.84	0.9950
EGC	0.010-0.300	1052.18	21.76	-2.91	3.98	0.9966
CF	0.010-0.500	1515.66	24.30	4.23	7.48	0.9979
EGCG	0.020-0.600	3279.77	139.82	-61.23	51.74	0.9857
(+)C	0.005-0.050	4681.74	109.22	6.91	3.44	0.9957
(-)C	0.004-0.038	4718.20	175.09	24.82	3.63	0.9864
TB	0.004-0.100	5649.80	188.78	6.53	9.09	0.9933

Regression equation, $y=ax+b$, where x is the analyte concentration (mg mL^{-1}), y is the ratio of area of the analyte to internal standard; s_a, standard deviation for the slope; s_b, standard deviation for the intercept, (n=5, k=2).

Unsupervised methods such as principal component analysis (PCA) and hierarchical cluster analysis (HCA) were used as exploratory techniques, while discrimination models were built using supervised methods such as linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) in order to obtain discrimination rules.

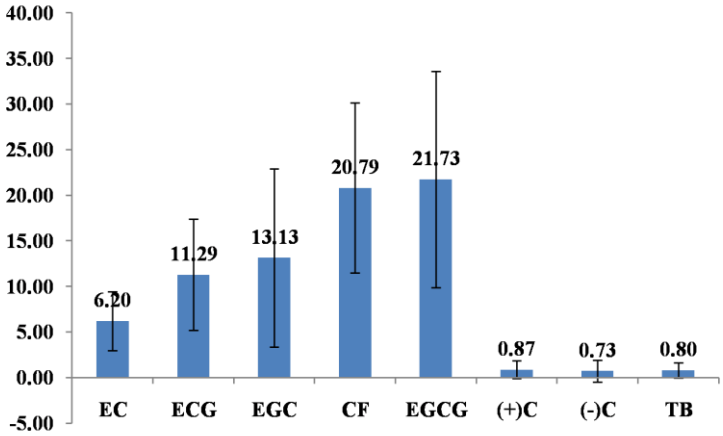
Catechins, caffeine and theobromine content

The CD-MEKC method was applied to the analysis of 92 commercially available GT samples, and a representative electropherogram of selected GT water extract of Chinese sample no. 38 is shown.



The average content of catechins and methylxanthines in the investigated GT samples (mg g^{-1}, dry basis) was calculated by the calibration.

The average content, the standard deviation and the range of each chemical descriptor are reported considering all the samples and the separate classes of Japanese GTs and Chinese GTs. Considering the entire data set, the total amount of the chemical descriptors follows the order EGCG>CF>EGC>ECG>EC>(+)-C>TB> (-)-C .



In general, with exception of ECG, the amount of all the compounds was found to be higher in Chinese GTs compared to the samples from Japan.

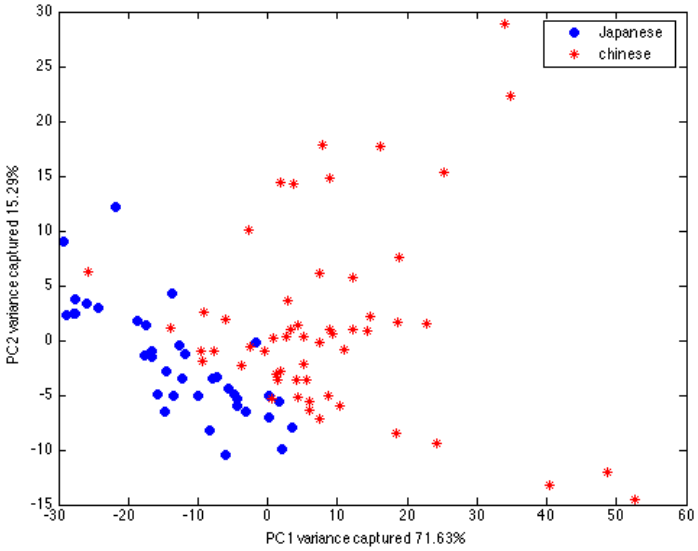
EGCG, which exhibits the highest antioxidant activity, is the most abundant catechin, with an average content of 27.14 mg g^{-1} in Chinese GT and 13.69 mg g^{-1} in Japanese GT, respectively, and values up to 68.10 mg g^{-1}.

Relatively high levels of EGC and ECG were also found, with average content of 17.50 and 6.63 mg g^{-1} for EGC and 10.90 and 11.87 mg g^{-1} for for ECG in Chinese and Japanese GTs, respectively.

Once the preliminary analysis of the data was carried out, pattern recognition procedures were then applied to the data matrix in order to achieve a more reliable GT differentiation.

Principal Component Analysis

In this study, PCA was performed after data centering, and the first 5 PCs were found to capture most of the variances needed to explain all variation in the data set. For visualising the information of the data set, a score plot using the first two PCs was chosen, which accounted for 86.92% of the total variance.



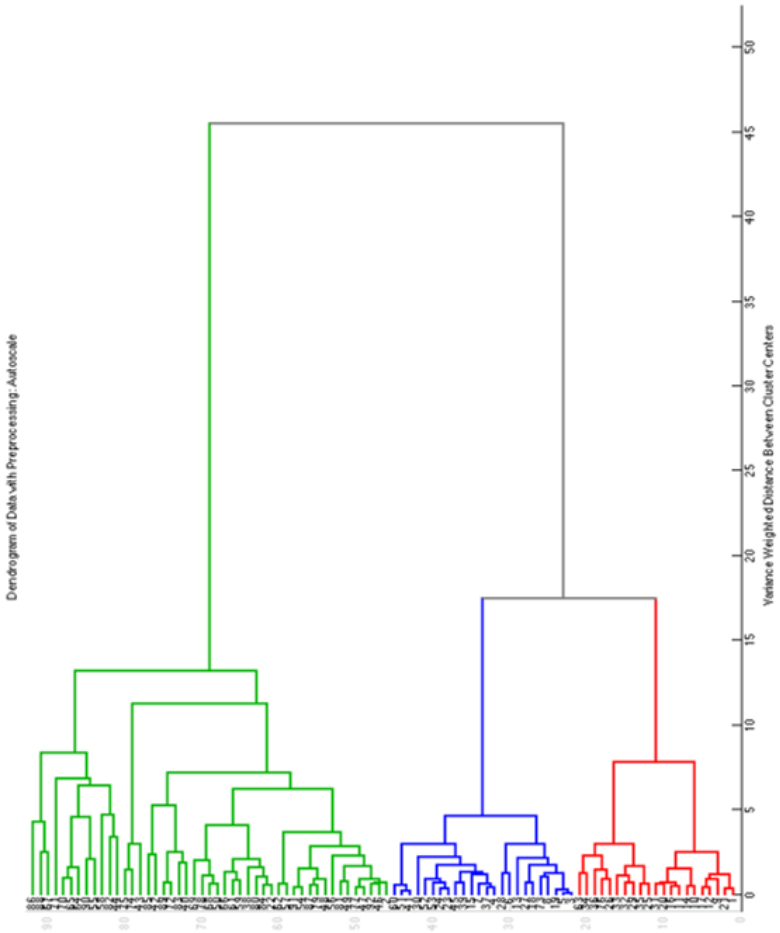
PC1 explained 71.63% and PC2 explained 15.29% of the total variance. From the scores, it was found that the first two PCs were able to evidence a good separation between Japanese and Chinese GT samples, even if a certain overlap was present and the complete separation between the classes was not obtained. The separation between the two groups was observed in particular along PC1. A larger dispersion of the Chinese teas appeared, whereas Japanese GTs formed a more compact group, located at the left side of the plot, at negative values of PC1.

All samples were then divided into a training set, used to perform a new PCA, and a test set, using the KS algorithm [110].

All samples in training test and test set were selected based on KS method; the assignment of the samples can be seen from Su, where the training set contained 75 samples, while the remaining 17 samples constituted the test set. It was possible to note that the test set was perfectly superimposed. It is worth mentioning that a selected training set should be sufficiently representative that a model is supposed to encounter. As it can be seen in the figure, the training set covers the test set both from the **X** and the **Y** space data. The test set was used as a completely independent set (unseen data) in order to be able to measure the performance of the built classification model

Hierarchical Cluster Analysis

All samples were gathered on the basis of similarities and the samples with the minimum variance were clustered preferentially. The dendrogram here shown was generated, finding that the samples were clustered in 3 different groups at a variance weighted distance between cluster centres of 20.



The first cluster was composed by 47 samples, of which only 2 were Japanese GTs, i.e., no. 22 and no. 2. The second cluster, made by 24 samples, was composed by 16 Japanese GTs and 8 Chinese GTs. In the third cluster, consisting of 21 samples, 19 of them were Japanese GTs and 2 were Chinese GTs (no. 63 and no. 76).

Thousands of bootstrap samples are generated by randomly sampling elements of the data, and bootstrap replicates of the dendrogram are obtained by repeatedly applying the cluster analysis to them [107]. *p*-value of a cluster is a value between 0 and 1, which indicates how strong the cluster is supported by data. Two types of *p*-values are provided: AU (Approximately Unbiased) *p*-value and BP (Bootstrap Probability) value. AU *p*-value, which is computed by multiscale bootstrap resampling, is a better

approximation to unbiased *p*-value than BP value computed by normal bootstrap resampling, which is the frequency that it appears in the bootstrap replicates. In Fig. S7, indicating the accuracy of classification, the values of AU/BP for the cluster dendrogram are reported. Red values at left branch are AU (Approximated Unbiased) *p*-values; green values at right branch are BP (Bootstrap Probability) *p*-values in percentage, and cluster labels are at the bottom. Clusters with high AU values are strongly supported by the data, meaning that they really are very similar units that form a natural cluster.

The variables are represented by the nodes and their correlations are represented as a set of edges that connect the nodes. The weights represent the intensity of a relationship: the thickest the edges between two nodes, the highest the correlation between the two connected variables, with green and red ties indicating a positive and a negative correlation, respectively. A strong positive correlation between the amounts of EC, EGC and (+) C and between EGCG and CF was found; only ECG shows negative correlations, which were evidenced with TB, (-)C and (+)C; anyway the negative correlations were much weaker with respect to all the other positive relationships.

Linear and Quadratic Discriminant Analysis

The classification model is developed on a training set of samples with categories. The model performance is evaluated by means of some samples from a prediction set by comparing their predicted categories with their own true categories. Therefore, all samples were divided into a training set, used to build model, and an external set (test set), used to test the model, using three different strategies. The first method was KS algorithm [111] as explained before. The second splitting was rationally performed in order to obtain a balanced distribution of the different Japanese types and Chinese zone in the different sets (the division was performed for each individual class separately, with 30% of the samples of each class in the test set). Finally, the third splitting consisted in a random splitting at half.

Two supervised pattern recognition techniques were applied: LDA and QDA. By using LDA, linear combinations of the selected descriptors can be performed, allowing the so-called discriminant functions to be obtained; the application of this approach best separates the classes according to the minimization of the ratio of within-class and between-class variances.

The obtained latent variables are called canonical variates and for *k* classes, *k-1* canonical variates can be calculated [112]. QDA is identical to LDA, but class borders are quadratic curves instead of straight lines.

To evaluate and validate the classification performances internally, a leave-five out Cross Validation method was used. Classification results are presented in the table below, reporting for both supervised methods the percentage of correct classification rate (%CCR) of the training set samples without cross-validation ($\%CCR_{TRAINING}$), of the training set samples using cross-validation ($\%CCR_{CV}$), and of the independent test set samples ($\%CCR_{TEST}$).

	%CCR			%Sensitivity			%Specificity		
KS splitting	TRAINING	CV	TEST	TRAINING	CV	TEST	TRAINING	CV	TEST
LDA	96.00%	94.67%	82.35%	92.59%	96.15%	90.91%	97.92%	93.88%	66.67%
QDA	92.00%	92.00%	76.47%	92.31%	88.46%	90.91%	91.84%	93.88%	50.00%
Type and zone splitting									
LDA	95.38%	90.77%	92.59%	100.00%	92.31%	90.91%	92.31%	89.74%	93.75%
QDA	92.31%	86.15%	85.19%	92.31%	84.62%	81.82%	92.31%	87.18%	87.50%
Random splitting									
LDA	91.11%	88.89%	87.23%	88.89%	88.89%	84.21%	92.59%.	88.89%	89.29%
QDA	100.00%	84.44%	76.60%	100.00%	77.78%	52.63%	100.00%	88.89%	92.86%

For each set, %sensitivity and %specificity were also calculated, which are related to the true-positive rate and to the true-negative rate, respectively.

The %sensitivity is the number of true positive predictions, divided by the number of true-positive and false-negative predictions. The %specificity is the number of true negative predictions divided by the number of true negative and false positive predictions [40]. Both LDA and QDA gave good results with each of the three splitting methods, evidencing that the model was not dependent on data structure, with very good

performances in terms of predictive ability (%CCR_{TEST} ranging from 92.59% to 76.47%).

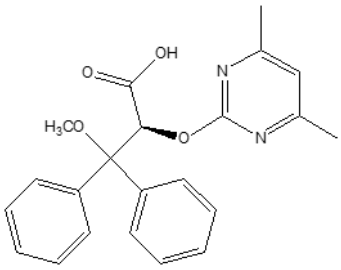
7.5. Enantioseparation and impurity determination of ambrisentan using cyclodextrin-modified micellar electrokinetic chromatography: visualizing the Design Space within Quality by Design framework

Ambrisentan (AMB) is an orally active and potent ET_A-selective ERA of the propanoic acid class, designated as an orphan drug for the treatment of PAH and chronic thromboembolic pulmonary hypertension. It is used in the treatment of patients with PAH classified as functional class II and III and in particular its efficacy has been shown in idiopathic PAH and in PAH associated with connective tissue disease.

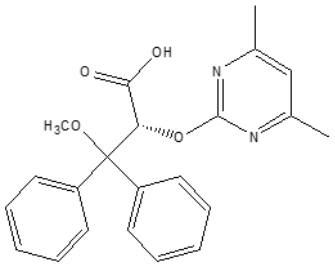
AMB is marketed as single (*S*)-enantiomer, representing the most active enantiomer, whose in vivo inversion to (*R*)-ambrisentan (*R*-AMB) is negligible [113]. *R*-AMB and other AMB impurities may be found in AMB bulk samples and dosage forms [92, 114, 115, 116].

The four main impurities considered were AMB impurity 1 (I_1), AMB impurity 2 (I_2) and AMB impurity 3 (I_3), which were named according to their migration order, and the impurity enantiomer (*R*-AMB). Their structures are here reported.

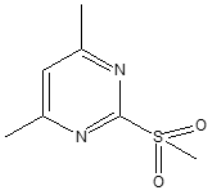
In this study, the CE method development was carried out within QbD framework, with the target of establishing a Design Space (DS) where the analysis performances meet predefined quality attributes with a selected degree of probability. QbD scouting made it possible to select Micellar ElectroKinetic Chromatography (MEKC) with the addition of γ-cyclodextrin (γ-CD) as operative mode



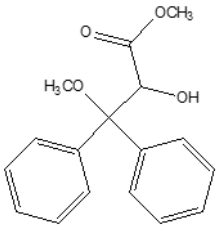
Ambrisentan (AMB)
(+)-(2*S*)-2-[(4,6-Dimethylpyrimidin-2-yl)oxy]-3-methoxy-3,3-diphenylpropanoic acid
CAS 177036-94-1



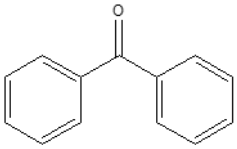
R-Ambrisentan (*R*-AMB)
(-)-(2*R*)-2-[(4,6-Dimethylpyrimidin-2-yl)oxy]-3-methoxy-3,3-diphenylpropanoic acid
CAS 1007358-76-0



Impurity 1 (I_1)
4,6-Dimethyl-2-methyl-sulfonyl-pyrimidine
CAS 35144-22-0



Impurity 2 (I_2)
2-Hydroxy-3-methoxy-3,3-phenylpropionic acid methyl ester
CAS 178306-47-3



Impurity 3 (I_3)
Diphenylmethanone
CAS 119-61-19

. This system involved complex separation mechanisms, including electrophoretic mobility, partitioning into the micelles and inclusion complexation into the cyclodextrin, and was essential for obtaining the separation of AMB, its *R*-enantiomer and the other impurities. The incorporation of QbD strategy in CE method development enabled dealing with optimization challenges in a rational and systematic way, providing the key for a better comprehension of the separation.

Method scouting and critical quality attributes

By employing a simple SDS-based MEKC with a plain buffer made by 100 mM borate buffer pH 9.20, the separation between I_2 and I_3 could not be achieved, not even tuning SDS concentration values (from 50 to 150 mM). Thus, the addition of a suitable CD was not only mandatory for the chiral separation, but also could be resolutive for obtaining a good separation of these neutral impurities, by exploiting the formation of inclusion complexes. Different CDs were added to standard MEKC conditions made by 100 mM SDS, 100 mM borate buffer pH 9.20, but only a few gave promising results. In

particular, α-CD was not useful neither for the enantioseparation nor for the separation of the neutral impurities. Instead, the derivatized β-CDs enabled the separation of I_2 and I_3, but did not allow the enantioseparation. The only CD that was decisive for the I_2/I_3 resolution and simultaneously made it possible to obtain a great improvement in the chiral separation was γ-CD. Hence, a CD-MEKC system based on SDS, borate buffer and γ-CD was selected for further optimization studies. In these experimental conditions, PABA was selected as suitable internal standard, being well separated from all the analytes, and the migration order of the compounds was the following: I_1, *R*-AMB, AMB, PABA, I_2 and I_3. The resolution between adjacent peak pairs was indicated by Rs_1 (I_1/*R*-AMB), Rs_2 (*R*-AMB/AMB), Rs_3 (AMB/PABA), Rs_4 (PABA/I_2), Rs_5 (I_2/I_3). The only critical resolution value was represented by Rs_2 between the enantiomers, whose target value was selected as $Rs_2 \geq 0.7$, corresponding to the value describing a baseline resolution between the enantiomers, which were very different for width characteristics [117]. Due the high complexity of the separation to be achieved, the target value for analysis time was deemed reasonable when equal or less than 20 min. Thus, the two critical quality attributes (CQAs) with their requirements were selected as $Rs \geq 0.7$ and $t \leq 20$ min.

Risk assessment, critical process parameters and screening phase

All BGE parameters were considered of high risk on method performances. The CPPs were represented by: capillary length (*L*), temperature (*T*), borate concentration (*B CONC*), pH, γ-CD concentration (*CD CONC*), SDS concentration (*SDS CONC*) and voltage (*V*). The settings of all the other factors, for instance sample parameters as injection type and volume, were selected on the basis of preliminary and scouting experiments. Thus, the experiments run in the scouting phase were fundamental for the selection of the levels to be considered for the CPPs: *L*, 48.5-64.5 cm; *T*, 18-22 °C; *B CONC*, 80-100-120 mM; *pH*, 8.20-9.20-10.20; *CD CONC*, 20-35-50 mM; *SDS CONC*, 80-100-120 mM; *V*, 20-25-30 kV. A Free-Wilson model was postulated for describing the effects of the CPPs on the CQAs, containing one constant term A_0 and for each CPP a number of coefficients equal to its number of levels minus one:

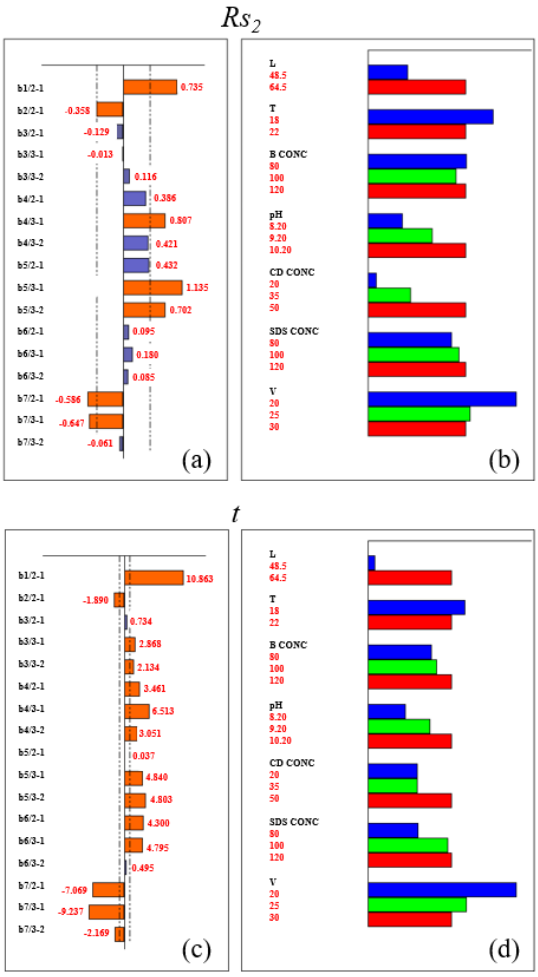
$$y=A_0+A_1A+B_1B+C_1C+C_2C+D_1D+D_2D+E_1E+E_2E+F_1F+F_2F+G_1G+G_2G$$

where *A* is capillary length, *B* is temperature, *C* is borate concentration, *D* is buffer pH, *E* is γ-CD concentration, *F* is SDS concentration, *G* is voltage. A 16-run asymmetric screening matrix was applied to estimate the coefficients and the corresponding experimental plan is reported in the table.

Exp. no.	*L* (cm)	*T* (°C)	*B CONC* (mM)	*pH*	*CD CONC* (mM)	*SDS CONC* (mM)	*V* (kV)	Rs_2	*t* (min)
1	48.5	22	100	10.20	35	120	30	0.00	9.05
2	48.5	18	100	9.20	50	100	30	0.96	17.51
3	48.5	18	80	9.20	35	120	25	0.40	10.23
4	64.5	18	120	8.20	35	100	30	0.42	16.28
5	48.5	22	120	10.20	20	100	25	0.00	15.93
6	64.5	18	100	10.20	50	80	25	1.78	23.33
7	64.5	22	120	9.20	50	120	20	1.95	34.59
8	64.5	22	100	9.20	35	100	25	0.42	20.19
9	48.5	22	100	8.20	35	80	20	0.00	10.09
10	48.5	18	100	9.20	20	100	20	0.00	14.39
11	48.5	18	120	9.20	35	80	25	0.00	9.60
12	64.5	18	80	10.20	35	100	20	2.02	33.57
13	48.5	22	80	8.20	50	100	25	0.00	8.28
14	64.5	18	100	8.20	20	120	25	0.15	21.18
15	64.5	22	80	9.20	20	80	30	0.00	12.85
16	64.5	22	100	9.20	35	100	25	0.50	19.99

Rs_2, enantioresolution between *R*-AMB/AMB; *t*, analysis time.

ANOVA showed that the models for both Rs_2 and *t* were significant and graphical analysis of effects was carried out, as reported.



The change of buffer concentration was not significant on Rs_2, while the low and medium levels were preferred for reducing *t*, thus this CPP was fixed at 100 mM. By increasing pH and *CD CONC*, both Rs_2 and *t* increased, with a prevalent effect on Rs_2; the experimental domain for both the CPPs was moved towards higher values for maximizing resolution. *SDS CONC* showed no significant effect on resolution; this CPP was fixed at the medium level in order to maintain a good balance with respect to *CD CONC*. Finally, voltage had the highest impact on *t* apart from capillary length; thus, due to the long analysis time observed during the screening, up to about 35 min, the experimental domain for *V* was moved to higher levels. Hence, the new domains for the factors to be further studied by RSM were the following: *pH*, 9.20-10.20; *CD CONC*, 36-50 mM; *V*, 24-30 kV.

Response Surface Methodology and Design Space

A Face Centered Design (FCD) was employed for estimating the second-order polynomial equation relating the CQA and the CPPs, which included linear, quadratic and interaction effects:

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_{11}x_1^2+\beta_{22}x_2^2+\beta_{33}x_3^2+\beta_{12}x_1x_2+\beta_{13}x_1x_3+\beta_{23}x_2x_3+\varepsilon$$

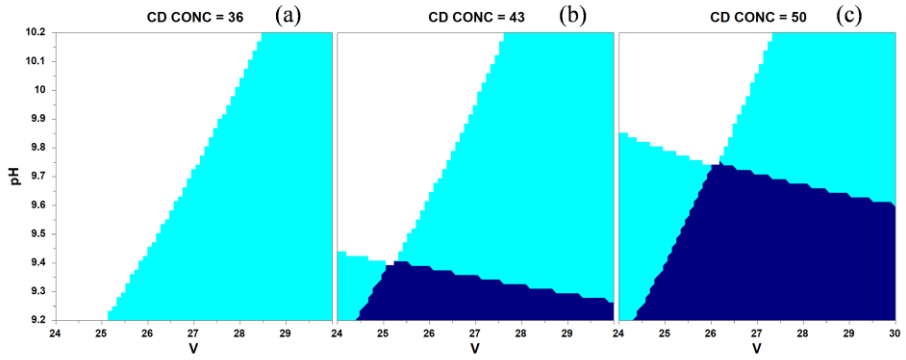
where y represents the experimental response, x_i the independent evaluated factors, β_0 the intercept, β_i the true coefficients and ε the experimental error. The experimental region of the FCD for three factors represents a cube, where the star/axial points are located on the center of the faces of the cube, and only three levels are required for each factor. The experimental plan with the measured responses is shown and the statistical significance and validity of quadratic models was verified by ANOVA.

Exp. no.	*V* (kV)	*pH*	*CD CONC* (mM)	Rs_2	*t* (min)
1	24	9.20	36	0.45	21.29
2	30	9.20	36	0.29	15.26
3	24	10.20	36	0.55	27.06
4	30	10.20	36	0.38	18.03
5	24	9.20	50	1.74	20.33
6	30	9.20	50	1.47	14.69
7	24	10.20	50	0.79	25.45
8	30	10.20	50	0.44	16.71
9	24	9.70	43	0.49	23.89
10	30	9.70	43	0.32	16.23
11	27	9.20	43	0.55	17.57
12	27	10.20	43	0.55	21.46
13	27	9.70	36	0.32	20.42
14	27	9.70	50	0.38	18.88
15	27	9.70	43	0.41	17.67
16	27	9.70	43	0.51	18.30
17	27	9.70	43	0.52	18.73

Rs_2, enantioresolution between *R*-AMB/AMB; *t*, analysis time.

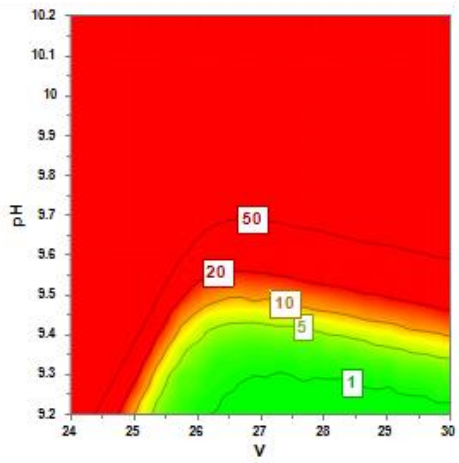
New reduced models were obtained by removing some of the coefficients (interactions or quadratic terms) which did not have a significant effect, in order to obtain higher values of coefficient of determination R^2 and coefficient of predicted variation Q^2. Values obtained after model refining were the following: *Rs*$_2$, Q^2=0.690, R^2=0.861; *t*, Q^2=0.958, R^2=0.981.

The predicted and accepted values for the CQAs were taken into consideration for drawing the sweet spot plots. In these graphs, the zone where the requirements of both the CQAs were fulfilled is depicted in dark blue, where the zone where only one CQA is fulfilled is depicted in pale blue. It is possible to note that when *CD CONC* was maintained at 36 mM no combination of the other two variables led to satisfactory performances.

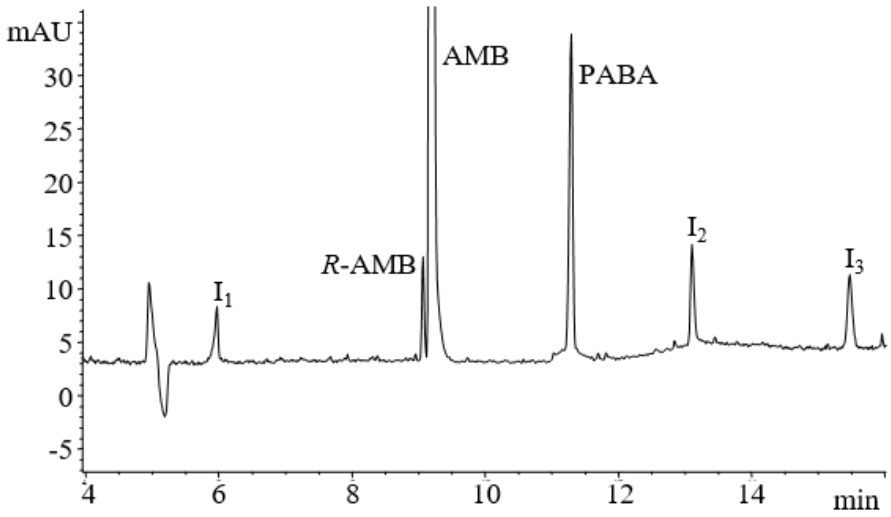


Sweet spot plots obtained by plotting pH *vs.* voltage, maintaining γ-CD concentration at (a) 36 mM; (b) 43 mM; (c) 50 mM. The zones where the requirements for both the CQAs are fulfilled are colored in dark blue, while the zones where the requirement for only one CQA is fulfilled are colored in pale blue.

These could be achieved only by setting medium-high values of *CD CONC*, low pH values and medium-high voltage values. For defining the DS, Monte-Carlo simulations were carried out, in order to propagate the predictive error by using the model equation to the CQAs, so that the probability of reaching the desired objective was computed and shown as probability surfaces. The DS was established for a specified quality level of $\pi \geq 90\%$, resulting enclosed within the following intervals: *V*, 26-30 kV; *pH*, 8.80-9.60; *CD CONC*, 43-50 mM. The DS was graphically represented as the green zone in the probability, drawn maintaining CD CONC at 50 mM.



The DS was validated by performing experiments at its edges, selected by a Plackett-Burman design [8] where the -1 and +1 levels corresponded to the extremes of the corresponding DS interval for each CPP [101], and verifying the fulfilment of the requirements. The working point was selected as 30 kV, pH 9.20 and 50 mM *γ*-CD, corresponding to the original set-point from which the DS was calculated. The related electropherogram is here reported showing the baseline separation of all the peaks in about 15 minutes with a generated current of about 60 μA.



A multivariate approach was applied for testing robustness, considering all the CPPs taken into account when performing the screening study, apart from capillary length, to which small variations could not be technically applied. The effect of the

CPPs on the CQAs was evaluated in a small interval around the working point, with the exception of voltage, which could not be set above the optimized value of 30 kV due to instrumental limitations. For testing robustness, in general linear models are advisable for the small experimental domain and for the reduction in the number of experiments. A Plackett-Burman design [100] was applied and the results of the eight runs are indicated.

Exp. no.	*T* (°C)	*V* (kV)	*pH*	*B CONC* (mM)	*SDS CONC* (mM)	*CD CONC* (mM)	Rs_2	*t* (min)
1	23	30	9.3	96	102	48	1.38	15.56
2	21	30	9.3	104	98	52	0.88	15.61
3	21	29	9.3	104	102	48	1.35	17.85
4	23	29	9.1	104	102	52	1.37	16.53
5	21	30	9.1	96	102	52	0.93	15.48
6	23	29	9.3	96	98	52	1.12	15.92
7	23	30	9.1	104	98	48	1.18	15.08
8	21	29	9.1	96	98	48	1.13	15.91

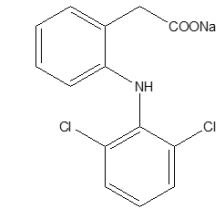
Rs_2, enantioresolution between *R*-AMB/AMB; *t*, analysis time

The treatment of the responses by ANOVA demonstrated that no one of the CPPs exerted a significant effect on the CQAs in the small interval considered, pointing out the high robustness of the method.

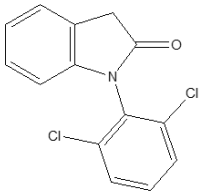
Information for planning an adequate system control were obtained by the lower and the higher values of enantioresolution and analysis time measured during system repeatability studies [101] and corresponded to the following system suitability criteria: Rs_2, 0.99-1.25; *t*, 13.28-15.92.

7.6. A comprehensive strategy in the development of a cyclodextrin-modified microemulsion electrokinetic chromatographic method for the assay of diclofenac and its impurities: mixture-process variable experiments and quality by design

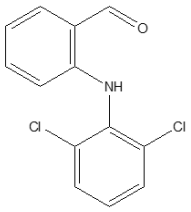
In the European Pharmacopeia (Ph. Eur.) [113] DIC specified impurities are reported as I_A, I_B, I_C, I_D, I_E and their nomenclature with the structures and CAS is shown below.



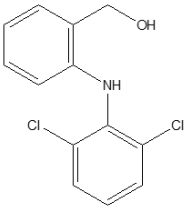
Diclofenac sodium (DIC)
Sodium 2-[(2,6-dichlorophenyl)-amino]phenyl]acetate
CAS [15307-79-6]



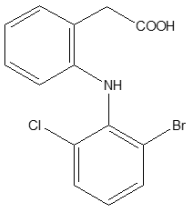
Impurity A (I_A)
1-(2,6-Dichlorophenyl)-1,3-dihydro-2*H*-indol-2-one
CAS [15362-40-0]



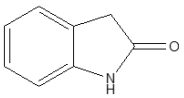
Impurity B (I_B)
2-[(2,6-Dichlorophenyl)amino]-Benzaldehyde
CAS [22121-58-0]



Impurity C (I_C)
[2-[(2,6-Dichlorophenyl)amino]-phenyl]methanol
CAS [27204-57-5]



Impurity D (I_D)
2-[2-[(2-Bromo-6-chlorophenyl)-amino]phenyl]acetic acid
CAS [127792-23-8]



Impurity E (I_E)
1,3- Dihydro-2*H*-indol-2-one
CAS [59-48-3]

The official procedure in the Ph. Eur. for the test of related substances reports a HPLC method with a run time of about 40 min.

The selected CE operative mode was MicroEmulsion ElectroKinetic Chromatography (MEEKC) with the addition of methyl-β-cyclodextrin (MβCD).

This technique uses an aqueous buffer containing nanometer-sized oil droplets, a surfactant and a cosurfactant to form a microemulsion background electrolyte (BGE) [118]. MEEKC can claim a great flexibility because the high number of involved variables makes it possible to finely modulate the resulting electrophoretic pattern.

The critical process parameters (CPPs) include both the mixture components (MCs) and the process variables (PVs). MCs are the ingredients in the microemulsion, expressed as proportions that sum to 1, while PVs are independent factors related to instrumental parameters and buffer. The critical quality attributes (CQAs) may depend on both the proportions of the MCs and on the settings of the PVs, with the consequence of the need of an effective support by multivariate techniques during the optimization [119].

Risk assessment, critical process parameters and screening phase

Several operating parameters have to be considered when optimizing a MEEKC method, including both MCs of the microemulsions and PVs, as recently highlighted in the Ishikawa diagram proposed for the risk assessment of a generic microemulsion method [120].

In this study, the selected CPPs were seven, including the PVs, *e.g.* voltage (V), buffer concentration (BC), buffer pH (pH), MβCD concentration (CD) and the MCs of the microemulsion, *e.g.* aqueous phase (W, borate buffer), oil (O, *n*-heptane), surfactant/cosurfactant (S, SDS/*n*-butanol in 1:2 ratio).

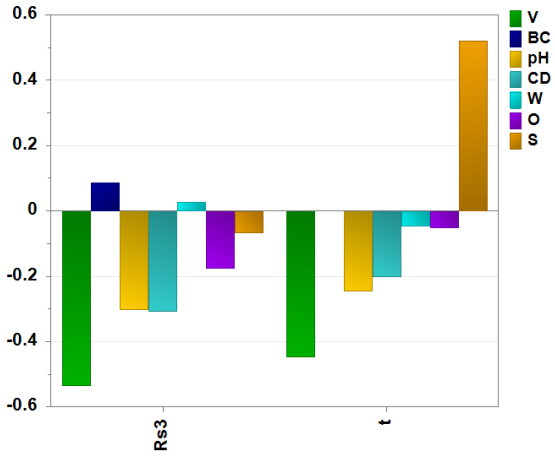
Other factors can have an important impact on the selected CQAs, including capillary length, injection time and pressure, temperature, type of buffer, but they were fixed at optimal levels according to the experiments performed in the scouting phase, in order to obtain a good compromise between selectivity, sensitivity and analysis time.

A linear MPV model for MCs and PVs was postulated and a 16-run D-optimal Design [65] was generated including 3 replicates at the centre of the domain.

Exp. No.	*Process variables*				*Mixture components*				
	V (kV)	BC (mM)	pH	CD (mM)	W	O	S	Rs_3	*t* (min)
1	24	10	8.20	4.0	0.8600	0.0150	0.1250	1.00	24.00
2	24	30	8.20	4.0	0.9200	0.0090	0.0710	1.48	14.3
3	30	30	8.20	4.0	0.8660	0.0090	0.1250	0.74	12.9
4	30	10	10.20	4.0	0.9200	0.0090	0.0710	0.79	6.99
5	24	30	10.20	4.0	0.8660	0.0090	0.1250	0.99	22.72
6	30	30	10.20	4.0	0.9150	0.0150	0.0700	1.04	8.44
7	24	10	8.20	8.0	0.8660	0.0090	0.1250	1.29	21.74
8	30	10	8.20	8.0	0.9150	0.0150	0.0700	0.72	8.04
9	24	30	8.20	8.0	0.9150	0.0150	0.0700	1.06	12.93
10	30	30	8.20	8.0	0.8600	0.0150	0.1250	0.71	12.71
11	24	10	10.20	8.0	0.9150	0.0150	0.0700	0.59	8.87
12	30	10	10.20	8.0	0.8600	0.0150	0.1250	0.64	9.63
13	24	30	10.20	8.0	0.8600	0.0150	0.1250	1.03	20.21
14	30	30	10.20	8.0	0.9200	0.0090	0.0710	0.72	7.71
15	30	30	10.20	8.0	0.8962	0.0116	0.0922	0.57	9.23
16	30	30	10.20	8.0	0.8962	0.0116	0.0922	0.66	9.46
17	30	30	10.20	8.0	0.8962	0.0116	0.0922	0.63	9.60

Rs_3, resolution between DIC/I_D; *t*, analysis time.

The models fitted by PLS resulted significant and the PLS solution was converted into the plot of the regression coefficient overview of the PLS model [65], shown below, which can give the best overview of the effect of each factor on each response.



Regression coefficient overview of the screening PLS model.

V and S were found to be the variables with the most important effect on the responses; an increase in V led to a decrease of both the responses, while a low percentage of S was important for decreasing *t*. pH and CD had a negative effect on both the CQAs, with the consequence that the levels necessary for optimizing the two CQAs were opposite: low levels for maximizing Rs_3 and high levels for minimizing *t*.

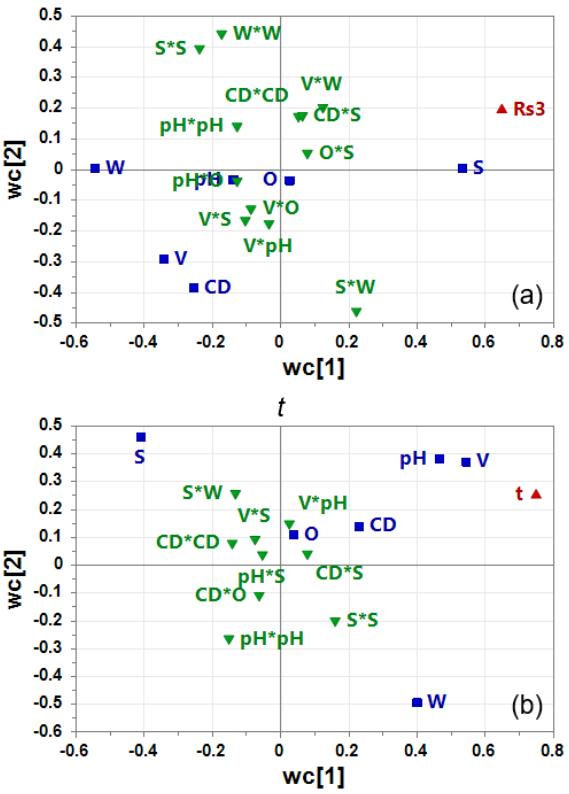
Response Surface Methodology and design space

In RSM a second order MPV model for PVs and MCs was postulated, where all the main, quadratic and interaction effects were considered. The experiments were run in two days and a blocking factor was taken into account. A 36-run MPV D-optimal Design was generated, with 4 experiments at the centre of the domain for obtaining an estimation of the experimental variance. The D-Optimal Design with the measured responses, had a G-efficiency value of 53.2% and a condition number of 14.29, respectively [65].

Exp. No.	*Process variables*				*Mixture components*				
	V (kV)	BC (mM)	pH	CD (mM)	W	O	S	Rs_3	*t* (min)
1	24	10	8.20	4.0	0.8600	0.0150	0.1250	1.00	24.00
2	24	30	8.20	4.0	0.9200	0.0090	0.0710	1.48	14.3
3	30	30	8.20	4.0	0.8660	0.0090	0.1250	0.74	12.9
4	30	10	10.20	4.0	0.9200	0.0090	0.0710	0.79	6.99
5	24	30	10.20	4.0	0.8660	0.0090	0.1250	0.99	22.72
6	30	30	10.20	4.0	0.9150	0.0150	0.0700	1.04	8.44
7	24	10	8.20	8.0	0.8660	0.0090	0.1250	1.29	21.74
8	30	10	8.20	8.0	0.9150	0.0150	0.0700	0.72	8.04
9	24	30	8.20	8.0	0.9150	0.0150	0.0700	1.06	12.93
10	30	30	8.20	8.0	0.8600	0.0150	0.1250	0.71	12.71
11	24	10	10.20	8.0	0.9150	0.0150	0.0700	0.59	8.87
12	30	10	10.20	8.0	0.8600	0.0150	0.1250	0.64	9.63
13	24	30	10.20	8.0	0.8600	0.0150	0.1250	1.03	20.21
14	30	30	10.20	8.0	0.9200	0.0090	0.0710	0.72	7.71
15	30	30	10.20	8.0	0.8962	0.0116	0.0922	0.57	9.23
16	30	30	10.20	8.0	0.8962	0.0116	0.0922	0.66	9.46
17	30	30	10.20	8.0	0.8962	0.0116	0.0922	0.63	9.60

Rs_3, resolution between DIC/I_D; *t*, analysis time

The calculated models were refined by deleting some of the interaction and/or quadratic terms in order to increase the values of coefficient of determination R^2 and goodness of prediction Q^2.

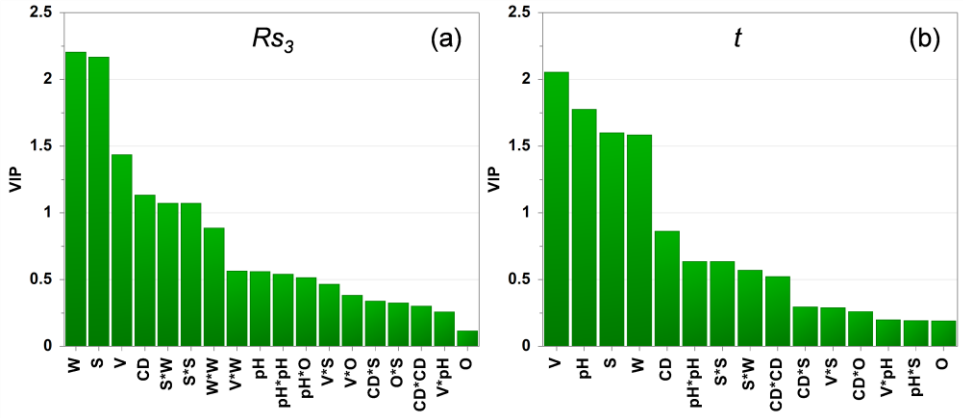


The obtained PLS models were characterised by high values of both R^2 and Q^2: Rs_3, R^2=0.921, Q^2=0.784; t, R^2=0.964, Q^2=0.847. Analysis time t was inversely transformed by t^{-1} and ANOVA evidenced that both the PLS models for Rs_3 and t were significant and valid and could be visualized by drawing PLS loading plots and contour plots.

The plots of the PLS loadings of the second model dimension w*c2 *vs.* the loadings of the first model dimension w*c2 made it possible to visualize the impact that each variable has on the model considering the distance to the plot origin [65].

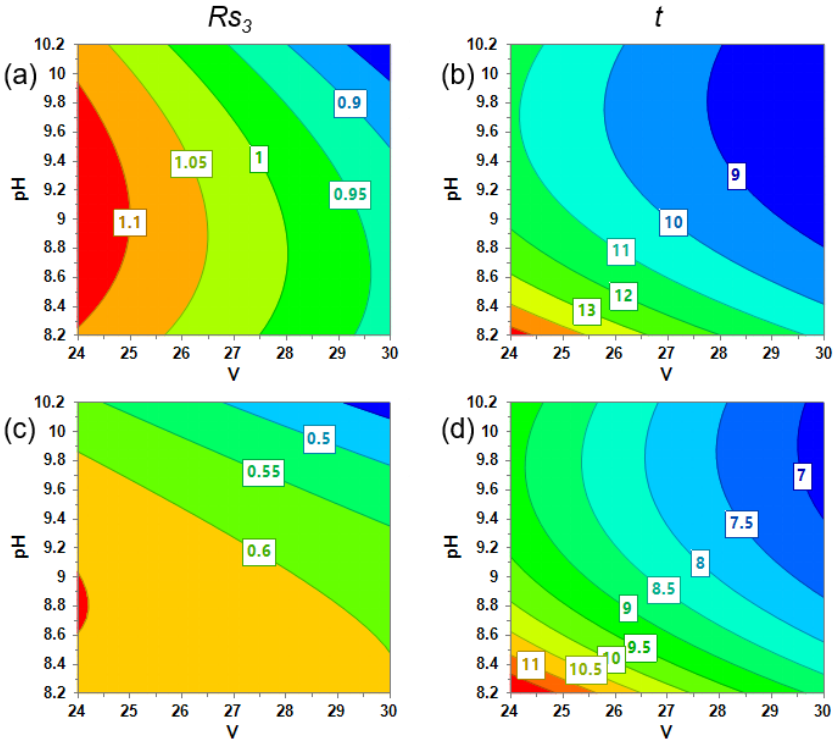
From the plot (a) abreast appeared that W, V and CD had an important negative effect on Rs_3, while S had an important positive effect. Moreover, some quadratic and interaction terms had an important influence on Rs_3, in particular a quadratic effect of S and interaction effects between S and W but also between MCs and PVs, such as V and W, pH and O. From the plot (b), which should be interpreted keeping in mind that the response t was transformed to t^{-1}, V, pH and CD had a negative effect on t, while S had a positive effect.

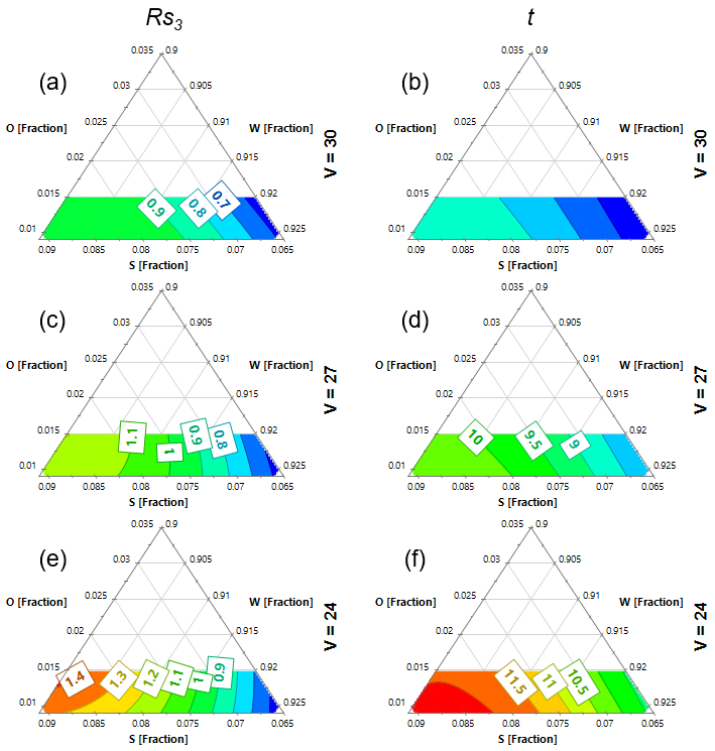
Quadratic effects were noticed for pH and S and the interactions with the higher impact were between S and W and between CD and S. The variable importance can be better clarified from the variable importance in projection (VIP plot) shown below, which is useful for revealing which are the dominating terms.



VIP plot in Response Surface Methodology. (a) Rs_3, resolution between DIC/I_D; (b) *t*, analysis time.

The trend of the predicted responses throughout the experimental domain could be easily visualized by means of the contour plots, which were drawn for both PVs and MCs. In the figure below the isoresponse surfaces obtained by plotting pH *vs.* V and maintaining CD at its central value of 6 mM are shown for Rs_3 (a,c) and for *t* (Fig. b,d). In figure (a) and (b) refer to the reference microemulsion, made by 91.18% W, 1.13% O, 7.68% S, while figure (c) and (d) refer to a microemulsion with the lower possible value of S, namely composed by 92.17% W, 1.33% O and 6.50% S.





The maximization of Rs_3 was in general obtained by setting low levels of V and medium values of pH, but the predicted values with the two microemulsions were different. By comparing figure (a) and (c) it was also evidenced that the trend of the response Rs_3 can be different when MCs settings are varied: as a matter of facts, in (a) there is a marked curvature due to the quadratic effect of pH, but the curvature is much less evident in (c). By analyzing the plots related to *t*, the same trend could be observed for the two microemulsion compositions; the minimization of this response was obtained in the same zone for the PVs, *e.g.* at high levels of V and pH. Anyway, the predicted values were lower in the case of the microemulsion with lower S content (d).

The plots show the MCs contour plots for Rs_3 (a,c,e) and *t* (b,d,f), calculated at different values of V: 24, 27 and 30 kV. The minimization of the triangular shape of the investigated domain was adjusted according to the imposed constraints and the isoresponse lines were obtained in function of the different percentages of the MCs.

In agreement with the information from the loading plots, the zone which led to the maximization of Rs_3 was characterised by high levels of S, low levels of W and low V, while the opposite conditions were required for minimizing *t*.

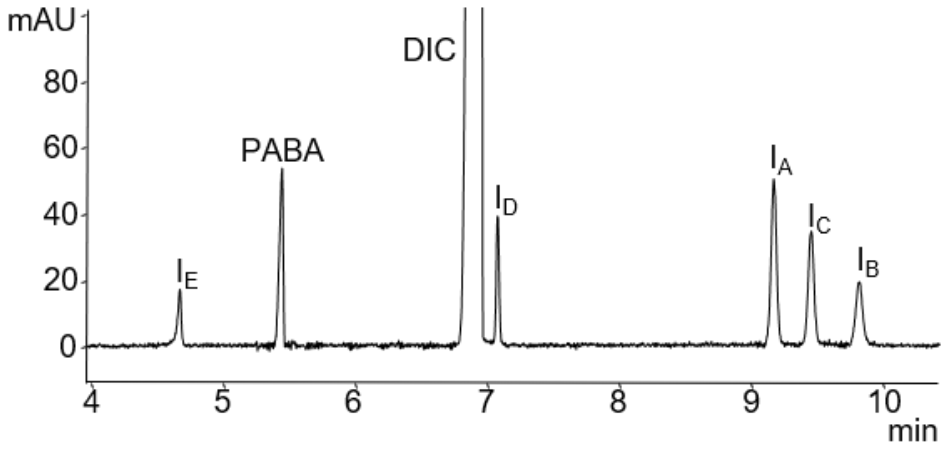
It is worthwhile to note that at high levels of V a curvature was observed, thus confirming that the simultaneous study of MCs and PVs is fundamental for obtaining a detailed and complete information on the system.

Hence, the DS was calculated by MODDE software by means of Monte-Carlo simulations [87], using a search function which expands the possible CPP ranges from

an optimum set-point to the widest possible zone where all the CQA predictions are within the specifications at the selected probability level of 99%.

Working points, robustness and method control

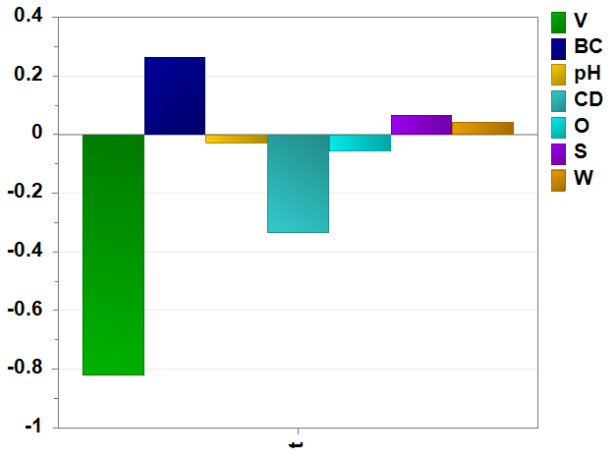
The working point coincided to the original set-point for calculating the DS. The typical electrophoretic pattern obtained applying the working point conditions is reported, evidencing that the baseline separation of all the peaks was obtained in about 10 min and the generated current was about 60 μA.



Ten additional verification points, selected according to a D-optimal design for a linear model, and using the limits of the DS as extremes of the investigated experimental domain, were tested [101], confirming that all the measured CQAs fulfilled the requirements inside the DS.

A linear MPV model was postulated and a 17-run D-optimal Design [65] was generated including one experiment in triplicate. The model was found not significant for Rs_3, but it resulted significant for *t*.

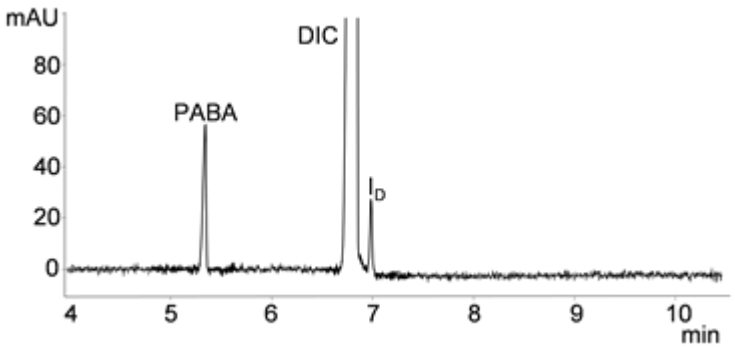
For this CQA an important effect of V was evidenced, as it is possible to see from the plot of the regression coefficient overview of the PLS model [65], shown below.



Robustness testing: regression coefficient overview of the PLS model for *t*.

Actually, this factor is often observed to exert a significant effect on analysis time when performing CE robustness studies, and also in this case the need of a careful control of V value was confirmed. Method control was then established by defining an accepted range for the CQAs, corresponding to the system suitability criteria determined from system repeatability study [101]: Rs_3, 0.93-1.12; *t*, 9.43-10.21.

The method has been validated according to ICH guideline Q2(R1) [121]. Real samples of tablets of Dicloreum®, labeled to contain 50 mg DIC, were analyzed by applying the developed method and performing four replicates (n=4, $\alpha/2$=0.025). The percentage of label claim was in agreement with the declared content (98.2±1.8%, RSD 1.2%). Among the impurities, only I_D was detected and was quantified at 0.67±0.04% with respect to the main compound, with a RSD value equal to 3.8%.



7.7. Combination of capillary electrophoresis, molecular modeling and NMR to study the enantioselective complexation of sulpiride with double cyclodextrin systems

Chiral recognition using CDs is based on the different inclusion interaction affinity between the chiral selector and the enantiomers and on the different electrophoretic mobilities of the complexes with the enantiomers [122]. The chemical modification of the native CDs leads to significant changes in their physicochemical properties and in their chiral recognition ability [123]. Several types of forces are involved in the inclusion complex formation, including hydrophobic interactions, reduction of conformational strain, hydrogen bonding, dipole-dipole and electrostatic interactions, van der Waals and dispersion forces, and their relative contribution depends on the guest and CD type [124].

Charged CDs have been shown to be superior to neutrally substituted CDs in several studies and they often have chiral resolving capabilities at lower concentration value [126, 123]. The high effectiveness of the charged CDs towards ionizable solutes is generally due to the opportunity of additional electrostatic interaction between the charged CD and oppositely charged enantiomers, thus resulting in the opposite electrophoretic mobility of the chiral selector with respect to that of the analytes. Moreover, in contrast with neutral CDs, charged CD derivatives have a self electrophoretic mobility, which allows their use as carriers in electrokinetic chromatography. Among the many charged CDs investigated, studies with sulfated CDs repeatedly demonstrated the widest application range [127]. In particular, sulfated-β-CD (SβCD) is known to be a powerful chiral selector, which has been used to separate many chiral compounds. SβCD is ionized at all pH values and the multiple ionogenic groups (and the counter ions) on each cyclodextrin contribute significantly to the conductivity of the running buffer [128]. The developed CE method was able to adequately discriminate between the enantiomers, but could not provide any information on the stability and structure of the inclusion complexes. Other techniques have to be applied to investigate the involved enantioseparation mechanism; in this context, Molecular Modeling constitutes an essential tool for exploring the mechanism

of molecular recognition. In particular, this approach provides good insights into host-guest interactions between CDs and enantiomers [129], making it possible to obtain the possible geometrical structures of the inclusion complexes. Spectroscopic methods in general, and NMR in particular, should be used to confirm the results of molecular dynamics (MD) simulations and to provide complementary information about the intermolecular interactions that take place [129].

Chiral separation of sulpiride by CE

In the previously developed CE method, the scouting phase was carried out with the aim of finding a chiral selector which could allow both a baseline resolution between *S*-SUL and its enantiomeric impurity in a reasonable analysis time.

The scouting experiments were run at a concentration of both the enantiomers of 0.0200 mg mL^{-1}, a value which is one hundred fold lower with respect to the test concentration of *S*-SUL, in order to obtain clear information on the enantioseparation ability of different CDs used as chiral selectors.

Thus, further CE experiments were carried out using the negatively charged SβCD in the presence of acidic BR (pH=3.0) buffer and operating in reverse polarity mode. In these conditions, the electroosmotic flow is negligible, and the complexed analytes were found to migrate at the anode.

Different SβCD concentration values in the range 1.8-10 mM were tested, discarding higher concentration values which led to extremely high electric current and very noisy baselines. When increasing SβCD concentration, a decrease of both analysis time (*t*) and resolution (*Rs*) were observed, and a good compromise was obtained at 10 mM.

From these experiments, SβCD was demonstrated to be a suitable chiral selector for enantiorecognition, but the resolution values obtained could not be sufficient to maintain a baseline resolution in the case of enantiomeric excess determination, where the test sample is characterized by an overloading of *S*-SUL (test value, 2 mg mL^{-1}).

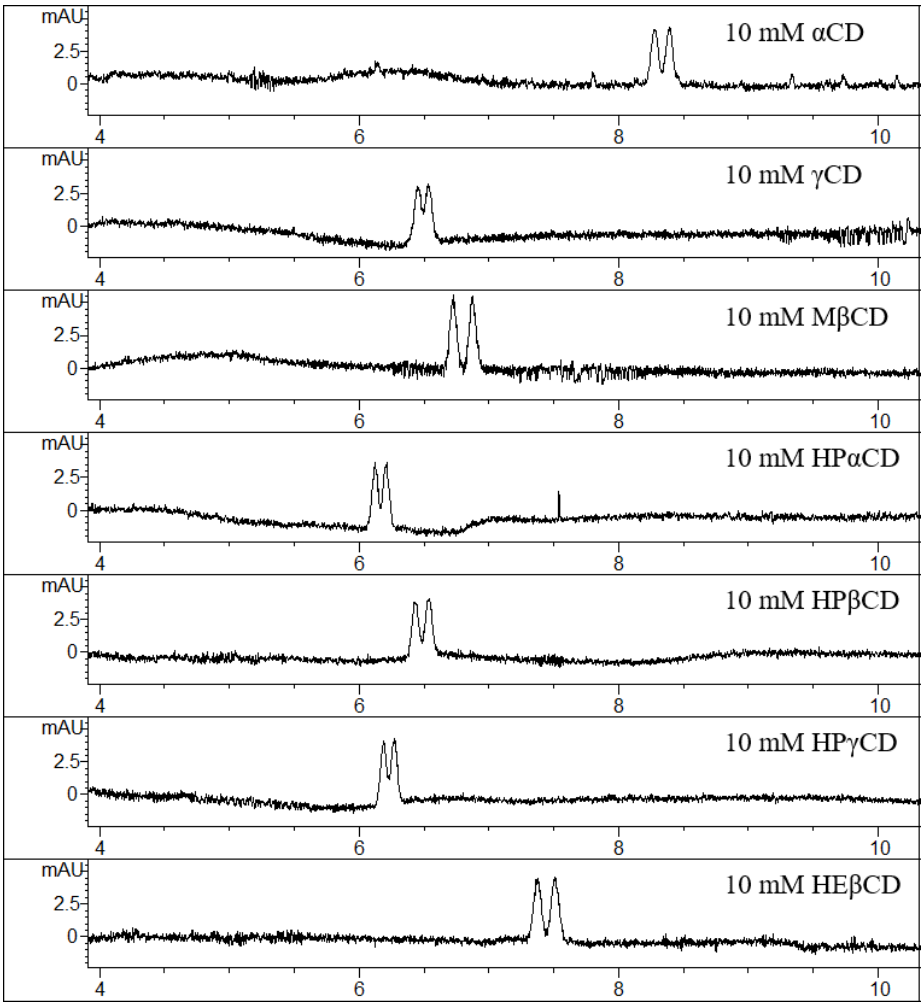
Hence, a dual system using SβCD in combination with a neutral CD, was considered to enhance enantioseparation. Several dual CD systems listed in the table

below were tested, keeping constant SβCD concentration at 10 mM and setting the neutral CD at two concentration values (10 and 30 mM).

Neutral CD	*Rs*	*t*	μ_R ($cm^2\ V^{-1}\ s^{-1}$)	μ_S ($cm^2\ V^{-1}\ s^{-1}$)	α
10 mM αCD	1.04	8.39	-1.64×10^{-4}	-1.63×10^{-4}	9.33×10^{-3}
30 mM αCD	0.78	5.46	-2.21×10^{-4}	-2.19×10^{-4}	8.89×10^{-3}
10 mM γCD	0.70	6.53	-1.87×10^{-4}	-1.85×10^{-4}	9.84×10^{-3}
30 mM γCD	0.92	7.44	-1.67×10^{-4}	-1.66×10^{-4}	1.06×10^{-2}
10 mM MβCD	1.50	6.87	-1.86×10^{-4}	-1.83×10^{-4}	1.69×10^{-2}
30 mM MβCD	2.39	13.32	-1.23×10^{-4}	-1.12×10^{-4}	9.61×10^{-2}
10 mM HPαCD	0.93	6.21	-1.92×10^{-4}	-1.90×10^{-4}	1.19×10^{-2}
30 mM HPαCD	0.93	7.19	-1.62×10^{-4}	-1.60×10^{-4}	1.18×10^{-2}
10 mM HPβCD	0.88	6.27	-2.28×10^{-4}	-2.26×10^{-4}	1.21×10^{-2}
30 mM HPβCD	1.37	8.45	-1.82×10^{-4}	-1.79×10^{-4}	1.30×10^{-2}
10 mM HPγCD	0.88	6.27	-1.80×10^{-4}	-1.78×10^{-4}	1.11×10^{-2}
30 mM HPγCD	0.80	7.41	-1.47×10^{-4}	-1.46×10^{-4}	1.09×10^{-2}
10 mM HEβCD	1.19	7.51	-1.95×10^{-4}	-1.93×10^{-4}	1.25×10^{-2}
30 mM HEβCD	1.52	16.87	-1.03×10^{-4}	-1.00×10^{-4}	2.80×10^{-2}

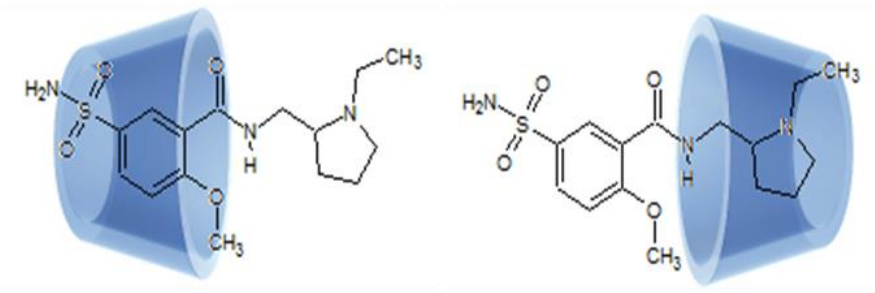
The increase of the concentration of the neutral CDs caused an increase of both enantioresolution (*Rs*) and analysis time (*t*). Representative electropherograms with the tested neutral CDs set at 10 mM are reported**.**

The selection of MβCD was made on the basis of the possibility of achieving slightly higher values of *Rs* and the final optimized conditions were: BGE, 5 mM BR buffer pH 3.45, 10 mM SβCD, 34 mM MβCD, voltage 14 kV.



Molecular Modeling

Inclusion complexes of *R*- and *S*-SUL with all the considered neutral CDs and with SβCD were drawn. *R*- and *S*-SUL were drawn with the protonated pyrrolidinic nitrogen ring, while SβCD was drawn with all the sulfated groups in anionic form. Two types of inclusion complexes were considered according to the moiety included into the CD. The complex type 1 was simulated with the aminosulfonyl group into the cyclodextrin cavity, while the complex type 2 was simulated with the pyrrolidinic group into the cyclodextrin cavity, as shown.



Type 1 Type 2

The complexes were submitted to 3 ns of MD simulations in implicit solvent at constant pH. Thirty conformations were collected and "minimized" during the simulation and the potential and the gain energy of the inclusion complexes were calculated on the minimized conformations. The mean values of potential energy E_p for the inclusion complex between *R*-, *S*-SUL and CDs are reported in the table.

CDs complexes	*Edock average (Kcal/mol) R-SUL*	*Edock SD (Kcal/mol) R-SUL*	*Edock average (Kcal/mol) S-SUL*	*Edock SD (Kcal/mol) S-SUL*
αCD-1	-10.1	2.2	-9.3	2.5
αCD-2	-14.0	3.3	-9.4	2.5
γCD-1	-12.6	3.5	-14.8	3.0
γCD-2	-14.9	3.1	-12.8	4.1
MβCD-1	-19.4	2.7	-22.9	3.8
MβCD-2	-19.8	4.4	-20.6	3.0
HPαCD-1	-10.1	3.1	-9.4	2.2
HPαCD-2	-7.0	3.7	-9.8	2.7
HPβCD-1	-15.5	3.6	-13.2	4.1
HPβCD-2	-13.7	3.3	-14.4	3.1
HPγCD-1	-19.0	4.4	-21.0	3.7
HPγCD-2	-15.1	3.5	-20.9	4.6
HEβCD-1	-18.6	4.6	-16.4	2.9
HEβCD-2	-13.6	3.8	-18.7	4.7
SβCD-1	-23.3	3.6	-25.3	7.5
SβCD-2	-27.6	2.2	-18.9	2.6

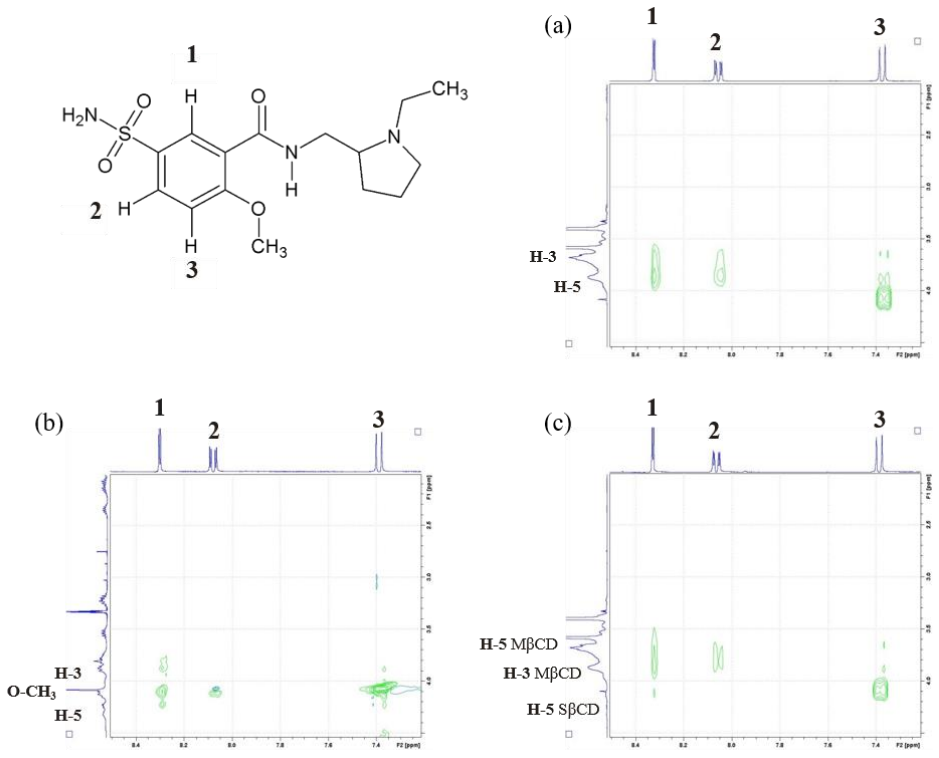
Edock values (average and standard deviation) of complexes type 1 and complexes type 2 between *R*- and *S*-SUL and CDs.

With the exception of αCD, the most favored complexes using all of the studied neutral CDs were found to involve *S*-SUL; conversely, using the charged SβCD, the energetically favored complex was calculated to be that with *R*-SUL. Even though the energy differences between the complexes are not statistically significant, this result was found to be in agreement with the migration order observed in CE experiments.

The docking average energies (Edock) of both types of complexes between *R*- and *S*-SUL with CDs, calculated as the gain in potential energy.

NMR

A set of 2-D ROESY experiments was performed providing information on relative inter- and intramolecular proton distance, confirming the formation of complexes of *S*-SUL with the CDs. Samples containing *S*-SUL and either MβCD or SβCD or MβCD/SβCD dual CD system were analyzed and the related spectra are reported in the figures below.



The 2-D NMR spectra related to MβCD (a) and SβCD (b) showed several intermolecular cross peaks between H3 and H5 protons of both CDs and aromatic ring

protons close to the aminosulfonyl moiety of *S*-SUL, demonstrating the inclusion of this group in the CDs cavity. In the spectrum of the solution where both CDs are present (c), there was a greater intensity in the signals that relate to the NOE interaction with internal protons of MβCD.

This can be explained by the higher molar concentration, about 3 times, for MβCD with respect to SβCD.

From the results obtained by MD, two types of complexes could form according to the moiety included in the hydrophobic cavity. NMR studies were able to confirm the formation of complexes, also suggesting predominance of the complexes of type 1, characterized by the inclusion of aminosulfonyl moiety of *S*-SUL in CD cavity.

Under the applied electrophoretic conditions involving the strong suppression of the EOF, the addition of a neutral CD to the buffer containing the enantioselective carrier SβCD, is expected to increase the migration time of the cationic solute, merely because of the effect of the inclusion into neutral CD, on the analyte mobility.

One of the variables affecting the migration time values in the studied systems is the stability of the complex of the analyte with the neutral CDs.

This result was shown to be in general agreement with the experimental CE data and in particular with those obtained when either MβCD and HEβCD were used in combination with SβCD, suggesting that the use of MD in the early phase of method development should be encouraged, as it could support the selection of the suitable chiral selector approaching the analytical target profile.

7.8. Combined approach using capillary electrophoresis, NMR and molecular modeling for ambrisentan related substances analysis: investigation of intermolecular affinities, complexation and separation mechanism

CyD-modified MEKC (CyD-MEKC) is an operative mode of CE where surfactants above the critical micelle concentration (CMC) and CyDs are added to the BGE, and can be effectively used both in the enantioseparation of chiral compounds and in the separation of compounds of similar structure such as a drug and its impurities [119].

SDS is commonly used as the micellar system [118], and CyDs can form host-guest inclusion complexes, providing a hydrophobic environment for a variety of guest molecules [130].

CyDs and surfactants are able to form complexes with most surfactants in 1:1 $(CyD)_1(surfactant)_1$ or 2:1 $(CyD)_2(surfactant)_1$ stoichiometries with high binding constants by including surfactants' hydrophobic tails into CyD cavities [131, 132]. However, since a surfactant is a long and fine molecule, its 1:2, 1:3, 1:4 complexes with CyD in theory may also occur. In the specific case of SDS as a guest, the driving force for complex formation is the hydrophobic effect associated to three main contributions, *i.e.* the dehydration of the guest molecule chemical groups that are inserted into the CyD cavity, the interaction of those groups with the interior of the CyD, and the release of water molecules from the CyD cavity to bulk water [130].

In particular, $(\gamma CyD)_1(SDS)_2$ complexes with head-by-head conformation of SDS are very stable, probably due to Na^+ mediated salt bridges between sulfate groups [130]. Moreover, water molecules released by the formation of 1:1 complexes make the cavity more hydrophobic and promote further inclusion of another compound [133] with the formation of ternary 1:1:1 complexes. It has been also demonstrated that in ternary complexes the two guests can present a cooperative effect [133].

As a matter of facts, CE does not provide any direct information regarding chemical and structural mechanisms of analyte-CyD interactions, therefore further

techniques such as NMR spectroscopy and molecular dynamics (MD) simulation have been used in combination in order to elucidate these intermolecular interactions.

NMR is valuable for enabling conclusions about the spatial proximity of functional groups of host and guest, thus providing information on the involvement of different moieties of the molecules in the intermolecular complex formation.

Moreover, CyDs can be studied in solution under comparable conditions by NMR as well as CE, allowing direct correlations between the complex structure and the electrophoretic behavior of the guest molecules. On the other hand, MD can clarify different aspects of the dynamics of complex formation as well as structural aspects of the formed complexes.

Molecular modeling

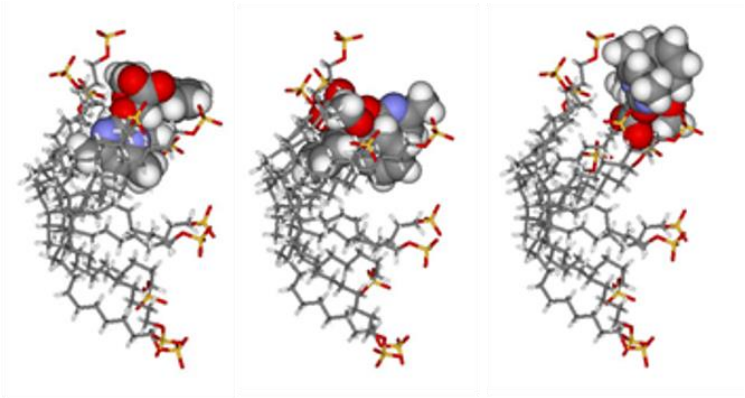
Different aggregates and inclusion complexes of the analytes were drawn involving all the species present in the BGE.

All the tested CyDs and SDS micelles were included in the study, with a special focus on the complexes with γCyD.

Aggregates between SDS micelle and analytes

Simulations of the aggregates between the analytes and SDS micelles were performed by building SDS aggregates corresponding to 1/6 of the micelle in spherical form, in order to decrease the calculation time.

The SDS micelles were anchored to a rigid structure in order to avoid the modification of the system geometry. For each of the analytes, complexes in three different conformations were built, as reported below for the aggregate between the micelle and *S*-AMB.



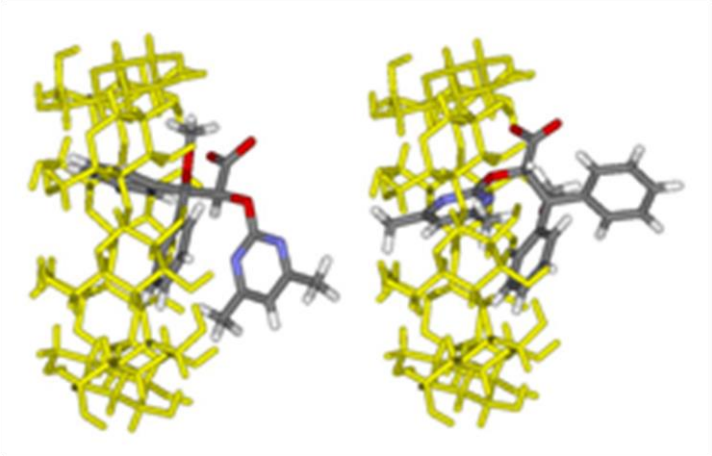
The average of potential energy (*Ep*) with the respective SD values and the gain in potential energy (E_{GAIN}) for the aggregates of all the considered compounds. From the obtained data, the formation of aggregates with the micelles by *S*-AMB and *R*-AMB was not energetically favorable, and this could be explained by the repulsion between the negative charge on the analytes and the sulphate polar head of SDS. All the other compounds showed a significant E_{GAIN} when included in the SDS micelles, thus indicating a high tendency of forming aggregates

	Ep average (Kcal mol^{-1})	*Ep* SD (Kcal mol^{-1})	E_{GAIN} (Kcal mol^{-1})
I_1	-941.9	3.2	-7.5
R-AMB	-1109.5	2.8	-1.2
S-AMB	-1110.7	4.5	-1.1
I_2	-876.5	2.8	-6.2
I_3	-884.6	3.4	-8.1

Potential energy (*Ep*) and gain in potential energy (E_{GAIN}) of the aggregates of analytes and micelles (*n*=90).

Inclusion complexes between γCyD and analytes

Simulations of the 1:1 inclusion complexes between γCyD and each compound were drawn considering for each complex two conformations, according to the molecule moiety located inside γCyD cavity. The two conformers of the complex $(\gamma CyD)_1$(S-AMB)$_1$ are shown.



The complex between γCyD and SDS seemed to be favored with respect to the complex between γCyD and any other compound.

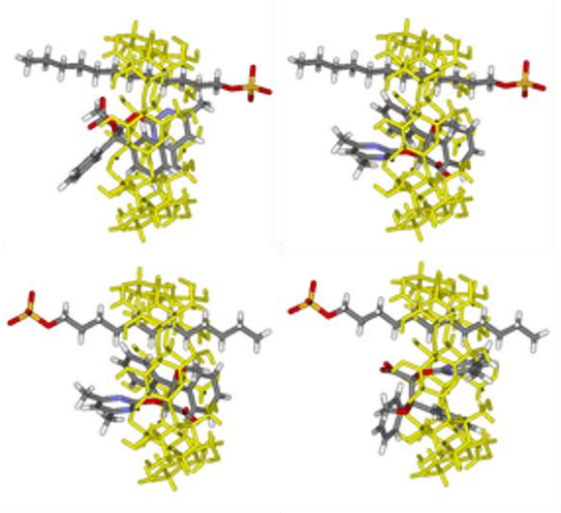
As a matter of facts, the average E_{GAIN} values for all the analytes were statistically different ($p<0.05$) from the SDS value, with the consequence that SDS could be able to successfully compete with the other analytes for the formation of the 1:1 complexes with γCyD.

Mixed inclusion complexes

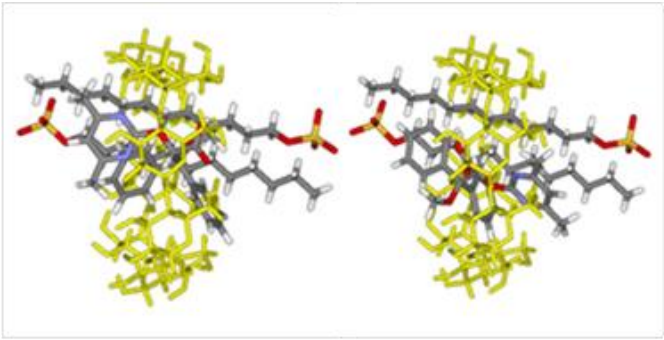
The γCyD cavity is wide enough to suggest the possibility of the formation of mixed inclusion complexes.

Each analyte could be inserted in γCyD together with 1 or 2 molecules of SDS, thus forming ternary complexes of different stoichiometry, *i.e.* $(\gamma CyD)_1(SDS)_1(analyte)_1$ or $(\gamma CyD)_1(SDS)_2(analyte)_1$.

For each mixed complex 1:1:1 and 1:2:1 four conformations were drawn, by combining all the possible orientations of the molecules included in the cavity.



The four conformers of the complex $(\gamma CyD)_1(SDS)_1(S\text{-}AMB)_1$ are reported. two conformers of the complex $(\gamma CyD)_1(SDS)_2(S\text{-}AMB)_1$ are shown.



The *Ep* and the E_{GAIN} values of the mixed complexes 1:1:1 and 1:2:1 are shown in the table. In general, by comparing the E_{GAIN} of 1:1 and 1:1:1 complexes, it appears that the formation of mixed complexes is preferred.

The probability of formation of the mixed complex is significant for the one between $(\gamma CyD)_1(SDS)_1(I_1)_1$, with the average E_{GAIN} (-19.7 Kcal mol^{-1}) but not statistically different (p>0.05) with respect to the E_{GAIN} for the complex $(CyD)_1(SDS)_2$ (-19.9 Kcal mol^{-1}).

	1:1:1 complexes				*1:2:1 complexes*			
	Ep average (Kcal mol^{-1})	*Ep* SD (Kcal mol^{-1})	E_{GAIN} (Kcal mol^{-1})	*t* test (*P* value)	*Ep* average (Kcal mol^{-1})	*Ep* SD (Kcal mol^{-1})	E_{GAIN} (Kcal mol^{-1})	*t* test (*P* value)
I_1	116.4	3.4	-19.7	0.813	28.3	5.2	-22.8	0.031
R-AMB	-54.4	4.8	-14.8	<0.001	-144.8	4.2	-22.0	0.003
S-AMB	-48.6	10.8	-8.2	<0.001	-144.6	2.9	-20.6	<0.001
I_2	188.6	3.1	-11.6	<0.001	91.1	3.5	-24.1	0.146
I_3	176.8	4.1	-17.3	0.015	83.9	3.0	-21.4	0.531
SDS	63.0	4.0	-19.9		-27.2	2.9	-25.2	

In general, by comparing the E_{GAIN} of 1:1 and 1:1:1 complexes, it appears that the formation of mixed complexes is preferred.

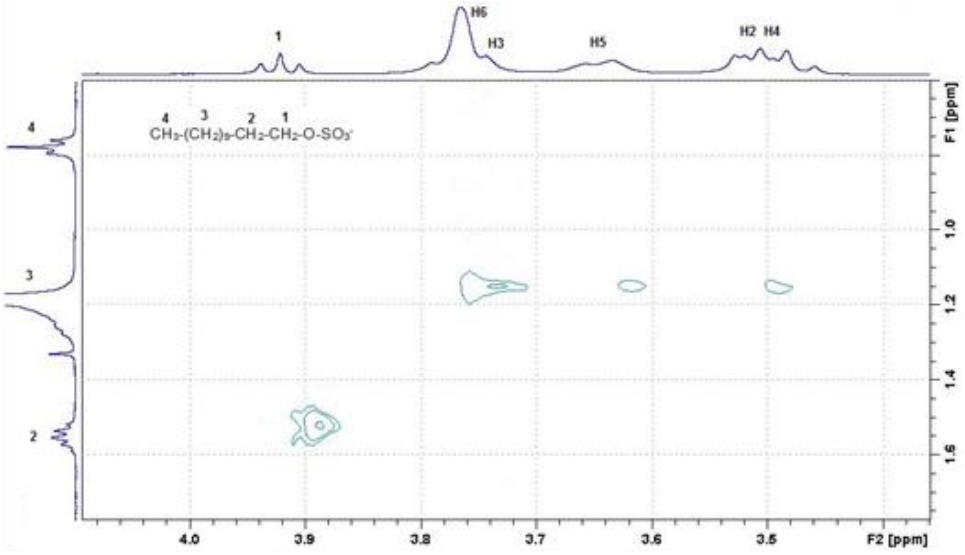
The probability of formation of the mixed complex is significant for the one between $(\gamma CyD)_1(SDS)_1(I_1)_1$, with the average E_{GAIN} (-19.7 Kcal mol^{-1}) but not statistically different (p>0.05) with respect to the E_{GAIN} for the complex $(CyD)_1(SDS)_2$ (-19.9 Kcal mol^{-1}).

For all the other compounds the probability of formation of the mixed complex is lower. As for 1:2:1 stoichiometry, the calculated E_{GAIN} values indicate that there is a high probability of formation for I_2 and I_3 complexes, which present an average E_{GAIN} not statistically different with respect the complex $(\gamma CyD)_1(SDS)_3$ (p>0.05). *R*-AMB and *S*-AMB have a similar and lower tendency of formation of this type of complexes.

NMR

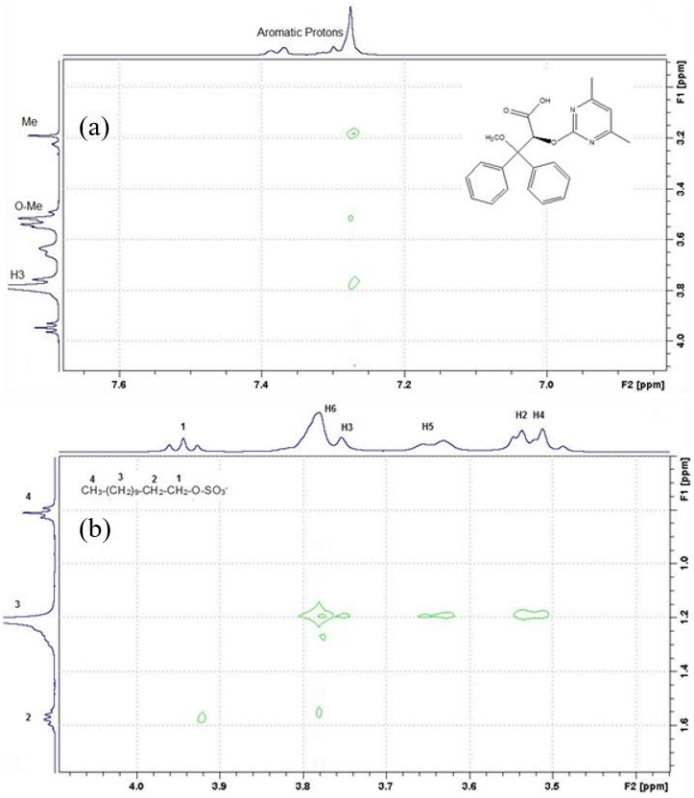
2-D NOESY experiments were carried out analyzing different samples, each containing γCyD in combination with SDS, and/or *S*-AMB or I_1, either as two-components or three-components samples.

The spectra of the samples made of γCyD:*S*-AMB and γCyD:I_1 in 1:1 molar ratio showed no evident interaction between the compounds. Instead, by observing the spectra of γCyD:SDS in 1:2 ratio, reported in the figure, cross peaks were found between the SDS aliphatic chain and the γCyD internal (H3,H5) and external (H2,H4) protons.

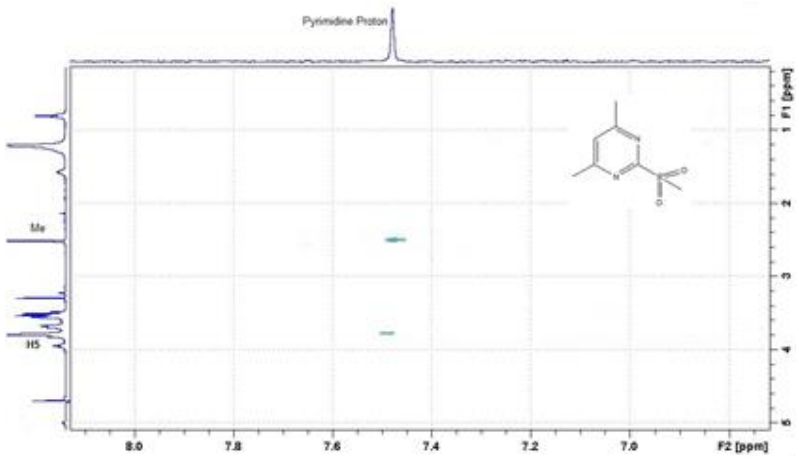


The possible existence of mixed complexes was investigated by means of samples containing γCyD:SDS:*S*-AMB and γCyD:SDS:I_1 in 1:2:1 ratio, for elucidating the interactions between the different compounds.

In the case of *S*-AMB, shown below, cross peaks were evidenced between *S*-AMB aromatic protons and γCyD internal proton H3 (a), and cross peaks between SDS and all γCyD protons were also found (b).



As for I_1, evident cross peaks were present between the proton of I_1 pyrimidine ring of and γCyD internal proton H5.



Also in this case, cross peaks were found between the protons of SDS aliphatic chain and both the γCyD internal and external protons. These findings led to confirm the existence of mixed complexes in the involved equilibrium.

Discussion

The E_{GAIN} values of binary 1:1 complexes (γCyD)$_1$(analyte)$_1$ highlighted the high propensity of SDS to be included in γCyD with respect to the other compounds, indicating that the most favoured complexes were in the order SDS>I_1>I_3>I_2>*S*-AMB>*R*-AMB. Up to 3 SDS molecules could be included in the cavity (E_{GAIN} -25 kcal mol^{-1}) and the interactions between the SDS aliphatic chain and both the internal and the external γCyD protons were demonstrated by NMR spectra. These results suggested the possible existence of a competition between SDS and the different compounds to be included in the γCyD cavity and this could be a possible explanation of the fact that γCyD was necessary for obtaining the separation of I_2 and I_3, which could not be resolved in a simple SDS micellar system. Unexpectely, NMR spectra of γCyD:*S*-AMB e γCyD:I_1 did not reveal any interactions, but the CZE results clearly showed that γCyD could be able to separate the enantiomers even without SDS addition.

Finally, the E_{GAIN} values of the mixed 1:1:1 and 1:2:1 complexes between γCyD, SDS and the different compounds showed the much greater tendency of forming ternary complexes with respect to the formation of binary 1:1 complexes. Focusing on 1:1:1 complexes, SDS and I_1 showed the highest stability, with the enantiomers characterized by a lower but significant propensity of forming this type of complexes. On the other hand, for 1:2:1 complexes the higher stability was observed for I_2, I_3 and SDS. The formation of mixed complexes for *S*-AMB and I_1 was confirmed by the NMR spectra.

7.9. Exploring the intermolecular interactions acting in solvent-modified MEKC by Molecular Dynamics and NMR: the effect of n-butanol on the separation of diclofenac and its impurities

In both MEKC and MEEKC, nBuOH is present in the aqueous buffer phase and in the Stern layer of the micelles, but also can penetrate into the micelles/microemulsion droplets, acting as a true class I modifier [4, 130, 134]. By using two-dimensional NOE

spectroscopy (2-D NOESY) it has been demonstrated that the nBuOH headgroup is located in contact with the SDS micellar headgroups, while the hydrocarbon chain is directed to the micelle core. In particular, adding the alcohol to micelles, the size-to-charge ratio of the micelles increases because nBuOH reduces the charge density and promotes the formation of micelles with a more open structure [135].

nBuOH shows a complicated influence on the structure of the micelles and can either decrease or increase the size of the aggregates.

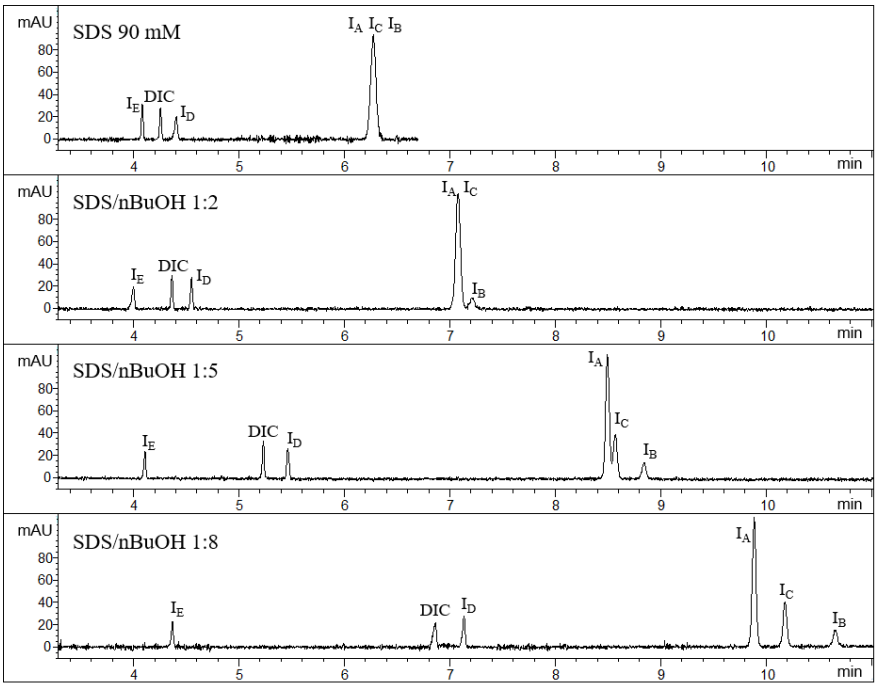
The present study was aimed to providing insight into the separation mechanism and into the interactions between the analytes and the components of the separation system (SDS, nBuOH, MβCyD). Due to the high complexity of the MEEKC system, and due to the fact that for obtaining the separation the use of nBuOH resulted to be compulsory, in the present study an analogue solvent-modified MEKC system was taken into consideration, with similar composition of the optimized microemulsion but made only by SDS and nBuOH, without the *n*-heptane oil phase.

The combination of NMR and Molecular Dynamics (MD) has been already demonstrated to be anuseful tool for increasing the knowledge on the mechanisms controlling the separation in electrophoretic analyses.

A special interest was focused on elucidating the important role of nBuOH, whose presence in the BGE was compulsory for obtaining the separation. MD simulations of SDS micelles, nBuOH aggregates, mixed micelles and aggregates of the analytes with the latter supramolecular structures were performed. Inclusion complexes of the analytes, SDS, nBuOH with MβCyD were also simulated.

Capillary Electrophoresis

The important role of nBuOH for obtaining the separation was confirmed by carrying out a series of experiments on a test sample where all the analytes were at low concentration values (0.040 mg ml^{-1}), as shown here.

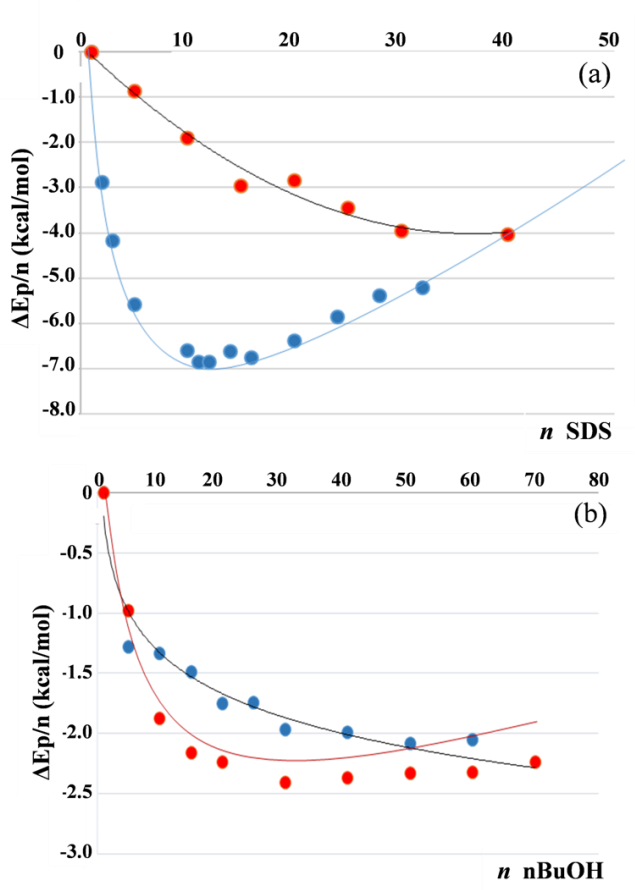


The working point conditions were applied for pH, concentration of borate buffer and percentage of SDS (2.63%, corresponding to about 90 mM), while the concentration of nBuOH was varied so that the SDS/nBuOH molar ratios tested were 1:2, 1:5, 1:8. The latter value corresponded to the same molar ratio present in the optimized microemulsion (1:2 weight ratio).

By applying these separation systems, it was clear that the increase of nBuOH concentration led to an increase of analysis time and to an increase of the separation between I_E and DIC, and that the use of a 1:8 molar ratio as in the MEEKC optimal conditions could enable a complete separation of I_A, I_B and I_C.

Information on the extent of solute association with micelles and mixed SDS/nBuOH micelles was then obtained by calculating the capacity factors for neutral and weakly ionized solutes (I_A, I_B, I_C, I_E) and corrected capacity factors for charged analytes (DIC, I_D).

Information on the extent of solute association with micelles and mixed SDS/nBuOH micelles was then obtained by calculating the capacity factors for neutral and weakly ionized solutes (I_A, I_B, I_C, I_E) and corrected capacity factors for charged analytes (DIC, I_D). For charged analytes the migration phenomenon does not only involve the partition mechanism between the aqueous phase and the micellar phase as for the neutral compounds, but also their electrophoretic mobility.



Three BGEs were prepared with a SDS/nBuOH molar ratio equal to 1:1, 1:4 and 1:8, and three analyses were run for each BGE. The extent of association of the charged compounds is much lower than that of the neutral solutes, due to the repulsion of the negative charges. The association of DIC and I_D is higher for the SDS micelle than for the mixed micelle 1:1 and then increases with the increase of nBuOH concentration. As concerns the most hydrophobic compounds, I_A, I_C and I_B, the affinity of these solutes for the mixed micelles increases when increasing nBuOH concentration, but then undergoes a decrease from the molar ratio SDS/nBuOH 1:4 to the ratio of 1:8. I_E has a much lower affinity than the other neutral compounds for both the micelles and the mixed micelles.

	90 mM SDS		SDS/nBuOH 1:1		SDS/nBuOH 1:4		SDS/nBuOH 1:8	
k'_{corr}		*sd*		*sd*		*sd*		*sd*
DIC	1.03±0.55	0.22	0.30±0.22	0.09	1.37±0.23	0.09	2.48±0.11	0.04
I_D	1.25±0.84	0.34	0.32±0.23	0.09	1.61±0.27	0.11	2.78±0.05	0.02
k'								
I_E	1.51±0.09	0.04	1.60±0.21	0.09	1.24±0.23	0.09	0.92±0.02	0.01
I_A	11.29±1.90	0.77	13.81±6.98	2.81	22.61±6.12	2.46	14.24±2.17	0.87
I_C	11.29±1.90	0.77	13.81±6.98	2.81	22.61±6.12	2.46	18.19±3.21	1.29
I_B	11.29±1.90	0.77	13.81±6.98	2.81	34.07±17.34	6.98	28.09±6.86	2.76

SDS micelles and n-butanol aggregates

The preliminary step of MD study was focused on analyzing different types of micelles and aggregates, made by SDS only, by nBuOH only and by SDS with increasing concentration values of nBuOH.

The SDS micelle formation was simulated in both spherical and rod-like shape, with a parallel orientation of the hydrocarbon tails in both the dispositions.

The progressive introduction of a number (*n*) of SDS molecules in these two forms involved changes in terms of potential energy (Ep) gain.

The trend of Ep for the spherical shape is shown in red, while the trend for rod-shape is shown in blue. Initially, with a low number of SDS molecules ($n<15$), the Ep gain for the rod-like shape is greater than for the spherical one, and it can assumed that SDS aggregates with a small number of molecules presumably tend to organize in rod-like shape; but by increasing *n*, and therefore when increasing SDS concentration, the Ep gain decreases. On the contrary, for the spherical form, the Ep gain increases up to 30-40 units of SDS and then tends to decrease, but less markedly than for the rod-like shape. Further increase of *n* over 70-80 units does not entail a gain in Ep, therefore for the MD simulations spherical SDS micelles, formed by 80 monomers, were taken as reference.

As for nBuOH aggregates, the same criterion for SDS micelle was employed, and aggregates of rod-like and of spherical shape were simulated with random orientation of hydroxilic groups in both forms. The trend of Ep gain due to the progressive increase of number of molecules *n* of nBuOH is shown in (b).

Both the values and the trend of Ep gain were quite similar for both the shapes (with a maximum gain of about 2.0-2.5 kcal/mol).

Aggregates of SDS, BuOH and analytes

MD simulations were then directed to evaluate the Ep gain due to the formation of aggregates between the considered analytes and SDS micelle. For each aggregate, three different conformations were built in the case of DIC.

MD simulations were then performed considering a quarter of SDS/nBuOH mixed micelle in 1:2 molar ratio, where the analyte was inserted.

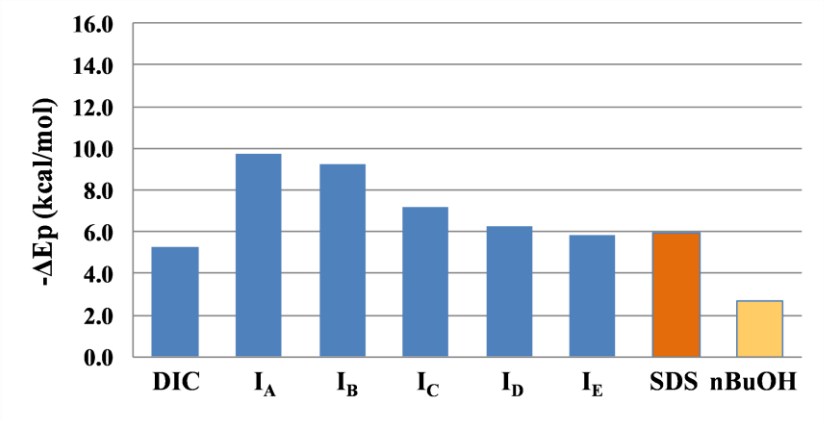
The Ep gain obtained due to the formation of an aggregate of the analyte and the mixed micelle. From the histogram it can be noticed that in general the Ep gain of the analyte-mixed micelle aggregates is generally lower than the Ep gain calculated for the aggregates with only SDS, but the general trend is similar.

From the electrophoretic experiments, the addition of nBuOH increases the resolution among the neutral species I_A, I_B and I_C, explaining the trend of migration order and playing a fundamental role for obtaining the electrophoretic separation.

The number *k* was estimated based on the surface of the molecules exposed to the solvent, with the equation $k = (Sa/(Sb/2))$, where Sa is the value of the surface exposed to the solvent of the analyte and Sb is the value of the surface exposed to the solvent of nBuOH (this is halved because only one face is able to interact).

Inclusion complexes with MβCyD

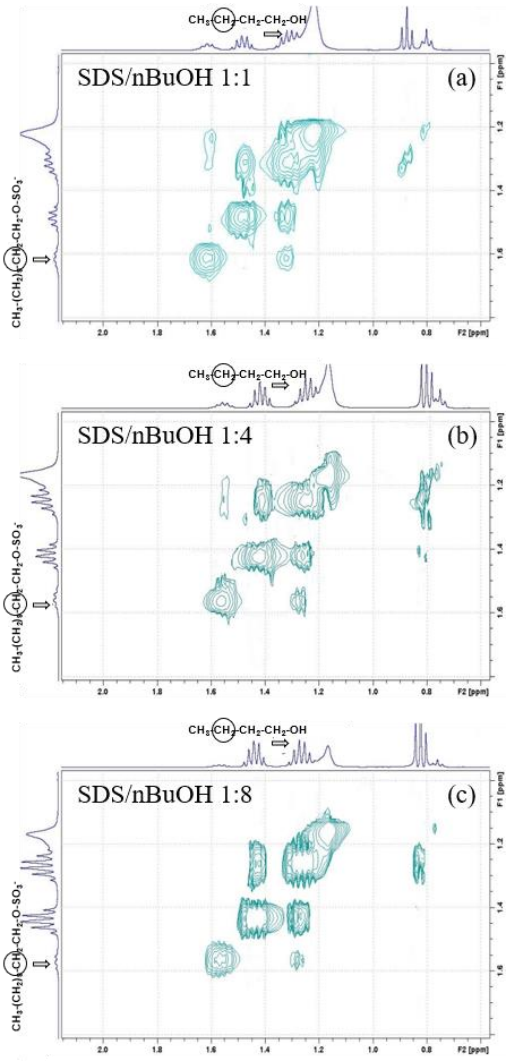
The MD simulations of inclusion complexes with MβCyD were performed for each analyte, but also for SDS e nBuOH, assuming the formation of 1:1 complexes and considering two different conformations according to the side of entrance of the molecule inside the cavity. The average values of obtained Ep gain for the formation of the inclusion complexes are reported in form of histogram.



MβCyD is capable of forming inclusion complexes with all the analytes, with an Ep gain on average higher than that calculated for the formation of aggregates with mixed micelles, and this explains the increase in migration velocity, due to the fact that MβCyD acts as a second pseudostationary phase and competes with the mixed micelle to include the analytes. The Ep gain calculated for the formation of the inclusion complex between MβCyD and SDS is of the same order of magnitude of the Ep gain for the formation of the SDS micelles, thus the latter is not prevented by the presence of MβCyD.

NMR

In order to support the findings of MD simulations, a set of 2-D NOESY spectra was performed on samples of mixed micelles in different SDS/nBuOH ratios (1:1, 1:4, 1:8), as reported.

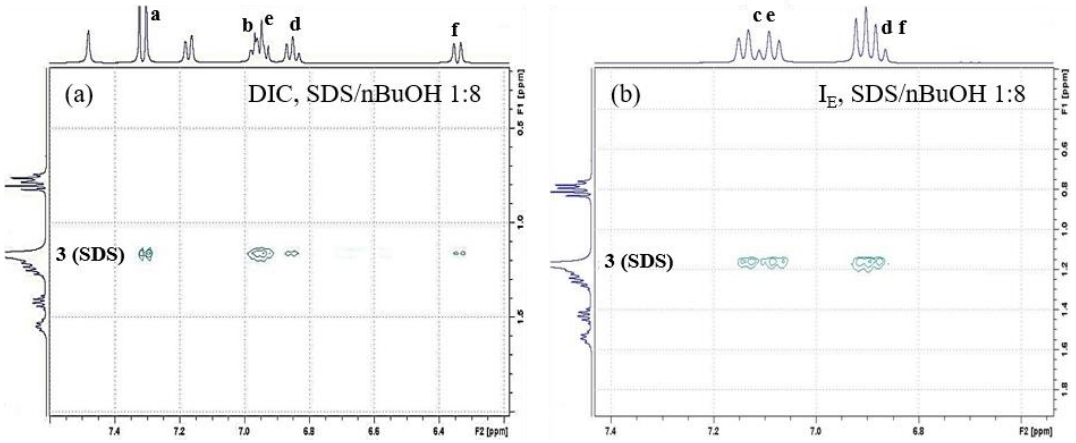


Some cross peaks between methylene group in position 3 of nBuOH and methylene group in position 2 of SDS were evidenced, highlighting the insertion of the short chain of the alcohol in the lipophilic portion, keeping the hydroxylic group outside the micelle.

Taking as a reference the peak of the SDS methylene group in position 2, the integral of the cross peak measured in the different molar ratios between SDS and BuOH presented approximately a similar value (0.09 for 1:1 ratio, 0.08 for 1:4 ratio and 0.06 for 1:8 ratio). Hence, it can be hypothesized that the composition of the micelles does not vary when increasing nBuOH, confirming that a SDS/nBuOH ratio between 1:1 and 1:2 is the most probable for mixed micelles.

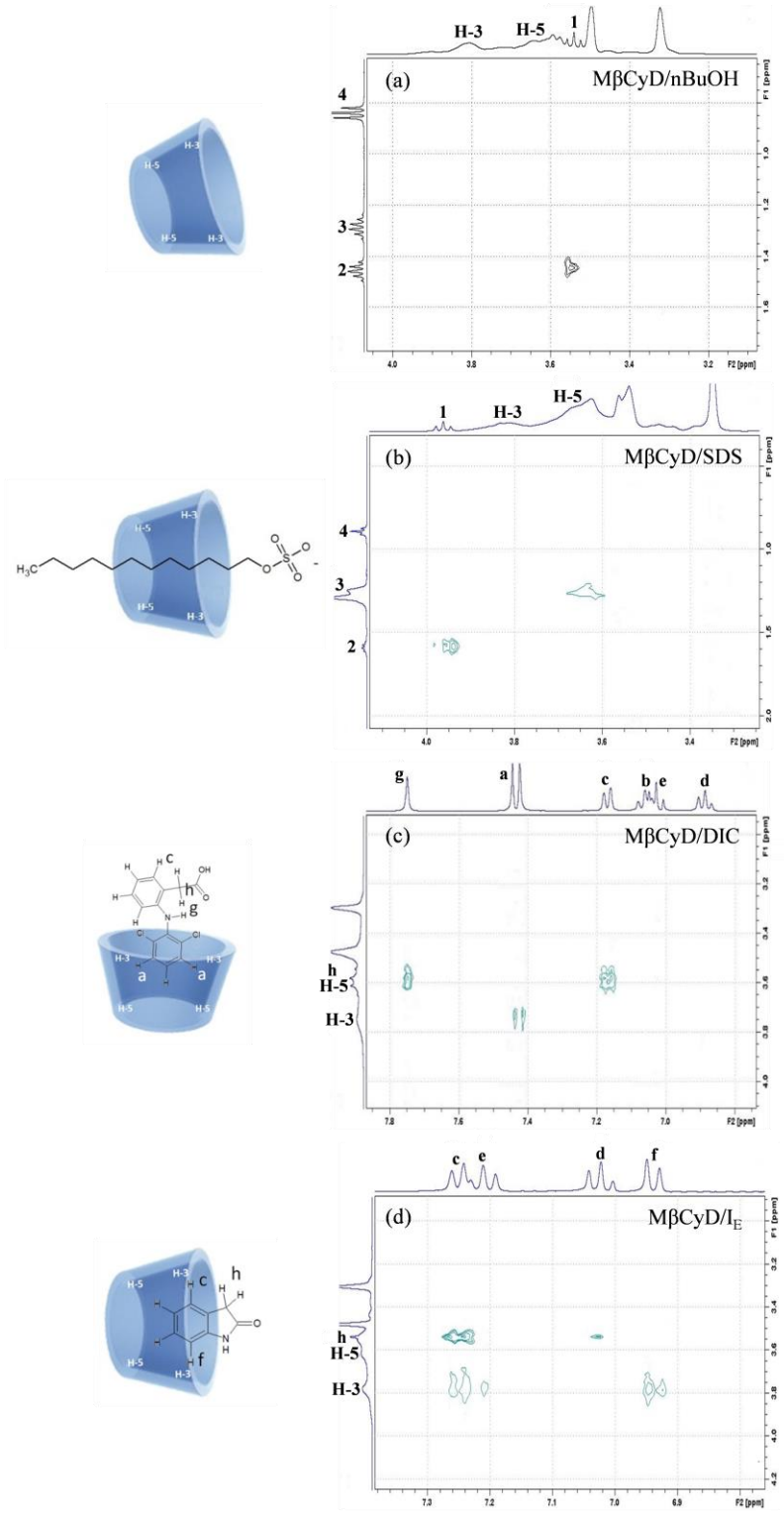
The NOESY spectra of DIC and I_E in presence of SDS/nBuOH mixed micelles (1:8 molar ratio), unexpectedly only showed interactions of the analytes with SDS and not with nBuOH. This suggests the presence of a competition within the micelle between the analytes and nBuOH, where the insertion of the analyte seems to be preferred.

This behavior confirmed the MD simulations of aggregates including micelles and analytes, where the Ep gain was demonstrated to be lower in mixed aggregates.



Finally, NOESY experiments were performed providing information on the formation of complexes between MβCyD and either nBuOH, SDS, DIC or I_E.

The obtained spectra are reported below, together with the most probable conformation of the complex. It is already well known that SDS can exist in free, complexed and micellized states and the spectrum confirmed the presence of inclusion complexes with MβCyD.



For both DIC and I_E inclusion complexes could form, as evidenced by some cross-peaks with H3 proton, located in the wider portion of the internal cavity of the CyD. Instead, for nBuOH no cross peak was evidenced, maybe because the possible inclusion

complex with MβCyD is of low stability due to the small dimensions of nBuOH and its solubility in water, as also indicated by MD histogram.

From the results of CE experiments, it was evident that the presence of SDS micelles in the BGE could make it possible the separation of acidic analytes (DIC, I_D) and I_E, but the presence of nBuOH was necessary in order to obtain the separation of I_A, I_B and I_C. The addition of nBuOH to SDS micellar solution led to similar results both in MEEKC and in solvent-modified MEKC, pointing out that the presence of the oil *n*-heptane was not necessary for the separation. Furthermore, the addition of MβCyD to the BGE was useful for reducing analysis time.

Further CE experiments showed that when keeping the percentage of SDS constant and adding increasing amounts of nBuOH, the molar ratio between SDS/nBuOH which led to the best results in terms of selectivity was 1:8, namely the ratio corresponding to that present in the optimized microemulsion (1:2 in terms of weight). Thus, it was hypothesized that the high resolving power of the separation buffer was due to the combination of SDS and nBuOH in adequate proportions. The 1:8 ratio also corresponded to significant values of capacity factors for DIC and I_D and high values for I_A, I_B and I_C, with the exception of I_E.

MD simulations and NMR experiments contributed to a possible explanation of the effects of the addition of nBuOH to the SDS micellar solution.

The results indicated that nBuOH could be present not only in the mixed micelles but also as monomer and nBuOH aggregates in the aqueous buffer. MD simulations also investigated the insertion of the analytes into the SDS micelle and in the mixed micelle with molar ratio 1:2.

Indeed, NMR highlighted the presence of inclusion complexes with SDS but not with nBuOH, probably due to the fact that nBuOH presents a lower affinity for the lower lipophilicity and the smaller dimensions with respect to SDS. The presence of MβCyD in the BGE may lower the availability of both SDS and nBuOH and mostly could compete with the mixed micelle for the insertion of the analytes, thus reducing their migration time.

7.10. Quality by Design applied to Dried Blood Spot liquid chromatography-mass spectrometry analysis.

In the sixties, Guthrie and Susi reported the blood collection on adsorbent papers in order to determine phenylalanine to detect phenylketonuria in newborns. This was the early days of Dried Blood Spot (DBS), a sample preparation technique which became, in the nineties, the method of choice to screen various metabolic disorders in newborn.

The DBS collection is simple; spots of capillary blood from a finger are spiked onto a cellulose-based filter paper and dried. The analytes are then extracted from the DBS with solvent. Clean extracts are commonly obtained for laboratory analyses, even if it seems to be necessary to develop an extraction protocol depending of the matrix and the target analytes (pharmaceutical susbtances, metabolites…) regarding the type and volume of solvent or the extraction duration, in order to ensure good recovery rates and minimize the matrix effects, for a maximum number of analytes from the same punch [137].

Several parameters such as spot homogeneity, blood-to-plasma ratio of the drug, sample volume or hematocrit effects (Htc) (it is the volume fraction of blood occupied by erythrocytes) can have an influence on quantitative analyses. Theses parameters and especially Htc need to be considered for Therapeutic Drug Monitoring (TDM), in order to have precise quantitative analysis to adapt the therapeutic treatment. This phenomenon influences blood viscosity and affects spreading of the matrix onto the filter paper. It is possible to override this drawback by spotting a precise volume of blood and an innovative approach has recently been developed to fix the taken blood volume independently from the Hct: Volumetric Absorptive Microsampling (VAMS) [138]. On another hand, an alternative is to use plasma spot instead of blood spot to counter the possible rupture of erythrocytes.

DBS extraction reaches several prerequisites for a sample preparation technique such as easy storage, shipments and logistics because the DBS has been reported to be stable under ambient conditions and do not need any freezing; it is thereby cheap and the other advantage it that it limits biohazard risks [139]. Moreover, this technique can be automated in order to reach high throughput extraction method [139]. The spot is usually made of blood but plasma or even urine and cerebrospinal fluid can also be

spiked. Another advantage which could be taken in consideration is the low sample consumption and micro-sampling, especially for pharmaceutical and newborn clinical field.

Micro-sampling helps the pharmaceutical companies to reach the 3R's (Reduce, Refine, Replace) doctrine, being more ethical. In clinic, it allows a newborn screening (NBS) of several metabolic disorders thanks to the less painful and less invasive sample collection, since a really low blood volume is collected (of the order of the microliters), contrary to venipuncture [139].

For Therapeutic Drug Monitoring (TDM) of adult patients, this micro-sampling is not really a challenge, except for phlebitis, cancer and weak patients, but the DBS sample preparation has the advantage to allow self-testing. [140].

Nowadays, the increasing affordability of sensible LC–MS/MS techniques enables precise drug quantification in matrices incorporating even vast interfering compounds.

The liquid chromatography/tandem mass spectrometry (LC/MS/MS) technique can be considered as the method of choice for TDM of various critical dose drugs including PI and NNRTI.

The significant sensitivity improvement of current MS instruments is adequate for routine analysis of xenobiotics in DBS, with some applications also reported from hospitals for therapeutic drug monitoring (TDM) [140].

The present work investigates the effect of the different parameters such as drying time, extraction duration, extraction volume or extraction solvent on the extraction efficiency from DBS of protease inhibitors, since these analytes are of interest in TDM, but we wanted to investigate steroids also, in order to have another class of molecules and because these endogenous compounds can also be of interest, in order to get a better understanding of what are the critical parameters of DBS for future developments.

Chemical and reagents

Methanol (MeOH) was purchased from Fischer Scientific (Loughborough, UK), ethyl acetate (EtOAc) from Acros Organics (Morris Plains, NJ, USA) and acetonitrile (ACN) from Biosolve B.V. (Valkenswaard, Netherlands). Formic acid (FA) (>98%) was provided from Merck Millipore Corporation (Darmstadt, Germany). Amprenavir (AMP) was obtained from Moravek Biochemicals (Brea, CA, USA), atazanavir (AZV) was obtained from American Radiolabeled Chemicals (St. Louis, MO, USA), lopinavir (LPV) and ritonavir (RTV) were obtained from Ontario Chemicals Inc. (Ontario, Canada), darunavir (DRV), amprenavir-d4 (APV-d4), darunavir-d9 (DRV-d9), atazanavir-d5 (AZV-d5), ritonavir-13C3 (RTV-13C3) and lopinavir-d8 (LPV-d8) were obtained from Santa Cruz Biotechnology, Inc. (Dallas, TX, USA), propranolol, estrone, progesterone, testosterone were purchased from Sigma-Aldrich (St. Louis, MO, USA), bosentan (I) was obtained from F. Hoffman-La Roche Ltd. (Basel, Switzerland).

LC analysis

For the screening phase, LC analyses were carried out on a Nexera HPLC system (Shimadzu Corporation, Kyoto, Japan) equipped with a CBM-20A$_{lite}$ controller module, a quaternary pump LC-30AD with low pressure gradient mode, a SIL-30AC auto sampler and a CTO-30A oven. LC separations were achieved on a XB C18 column (Kinetex, Phenomenex, 2.1 x 100 mm, 2.6 µm) at 40°C, with a mobile phase constituted of A (H_2O with 0.1% formic acid) and B (ACN with 0.1% formic acid) at a flow-rate of 0.3 mL/min, with a gradient from 40 to 76% of B in 5 minutes. For the Response Surface Methodology (RSM), the Design Space (DS) and the Robustness, LC analyses were carried out on a UltiMate 3000 HPLC system (Dionex) equipped with a TCC 3200 controller module, a HPG 3400 M quaternary pump and a WPS 3000 TSL auto sampler. LC separations were achieved on an Luna C18(2) column (Phenomenex, 2.1 x 100 mm, 3 µm) at 40°C, with a mobile phase constituted of A (H_2O with 0.1% formic acid) and B (ACN with 0.1% formic acid) at a flow-rate of 0.3 mL/min, with a gradient from 40 to 76% of B in 5 minutes.

MS conditions

The Nexera system was coupled to a triple quadrupole mass spectrometer LCMS-8050 (Shimadzu) through an electrospray ion source operated in positive mode.

MS conditions were as follows: capillary voltage = 4 kV, temperature of desolvation line = 300°C, temperature of heating block = 400°C, interface temperature = 400°C, nebulizing gas = 3 L/min (nitrogen), drying gas = 10 L/min (nitrogen), heating gas = 10 L/min, collision induced dissociation (CID) gas = 270 kPa (argon). For quantitation analysis, selected reaction monitoring (SRM) mode was used with a dwell time of 50 ms for each transition. SRM transitions and collision energies (CE) were optimized with the LabSolutions software. The voltages at Q1 prebias, collision cell and Q3 pre bias were optimized for each analyte to obtain the highest sensitivity.

The UltiMate 3000 system (U3000) was interfaced with a Turbo-V IonSpray electrospray source to a 4000 Q TRAP (AB Sciex, Concord, ON, Canada). MS source conditions were as follows: curtain gas = 40 psi, gas 1 = 30 psi, gas 2 = 25 psi, electrospray voltage = 5200V, source temperature = 350°C. Collision gas (N2) setting set to high. Experiments were performed in the Selected Reaction Monitoring (SRM) mode with 10 ms dwell time transitions.

Software for data acquisition and processing

The Nexera system and the LCMS-8050 mass spectrometer were controlled and the data acquisitions were collected by the LabSolutions software (Shimadzu).

The U3000 was controlled by the Chromeleon software (Thermo Fisher), and the 4000 Q TRAP was controlled, and the data acquisitions were collected by the Analyst software (Sciex).

Nemrod-W software was used for generating the asymmetric screening matrix and the Plackett-Burman designs. MODDE10 software was employed for generating Central Composite Centered design (CCD) and to draw risk of failure maps by Monte-Carlo simulations. Projection to Latent Structures regression (PLS) was employed as regression technique. The definition of DS was based on the obtained PLS models and Monte-Carlo simulations, by taking into account the error in prediction of the models.

Preparation of stock and spiking solutions and sample collection

Stock solutions of AMP, AZV, DRV, LPV, NFV and RTV were prepared by dissolving them in methanol at concentration of 1 mg/mL, and a spiking solution containing all the protease inhibitors was prepared at 50 µg/mL for AMP, AZV, DRV, LPV, NFV and RTV, respectively. Organic bovine blood was provided by the slaughterhouse of Meinier (Meinier, Switzerland). Sample preparation was performed on dried blood spot (DBS) according to a published procedure. 15 µL of a solution constituted of 980 µL of plasma and 20 µL of spiking solution of sample were spiked onto a paper disk. After spiking, DPS was dried for 30 to 120 minutes and 100 to 900 µL of extraction solvent were added, following by 5 to 15 minutes of sonication. Then, filter paper was removed, and extraction solvent was evaporated by centrifugation**.** The sample was reconstituted in 200 µL of injection solvent (ACN/water), containing 20 ng/mL of the labelled protease inhibitors, used as internal standards. This was followed by the mixing at 1400 rpm for 5 minutes (Thermomixer comfort, Eppendorf). The samples were let 5 minutes at 4°C and after a centrifugation step at 12000 rpm of 2 minutes, 150 µL of supernatant were transferred in vial prior to injection (injection volume was of 5 µL).

Critical process parameters (CCPs) and investigation of knowledge space

QbD emphasizes the need to thoroughly understand the analytical system by an in-depth study of CPPs and their interactions using DoE to find cause-and-effect relationships. To get a better understanding of the criticality of the parameters involved in the sample preparation procedure for DBS-LC-MS/MS assay for specific compound, several parameters, such as volume of solvent, sonication time, the times of dry and type of solvent were evaluated.

The selected CPPs to be investigated by DoE were represented by volume of solvent (SOLVENT VOL), time of sonic (SONIC TIME), time of dry (DRY TIME), type of solvent (SOLVENT). The considered CQAs were the peak area ratios

(analyte/IS) of the five protease inhibitors Rs1, Rs2, Rs3, Rs4, Rs5, and the three steroids Rs6, Rs7, Rs8. The aim of a sample preparation is to get the highest extraction recovery, that is why peak area ratios were chosen as representative of extraction quality. The volume of the solvent, the time of sonic and the time of dry were studied at three levels, instead the type of solvent was studied at five levels in order to compare the extracting power and reach a truly understanding on sample preparation process. The type of solvent variable was chosen as the most pertinent to investigate at five levels because it was a qualitative variable and it allowed covering more physico-chemical properties of solvent. Indeed, this parameter effectively affects the extraction depending on the solvent / analyte affinity.

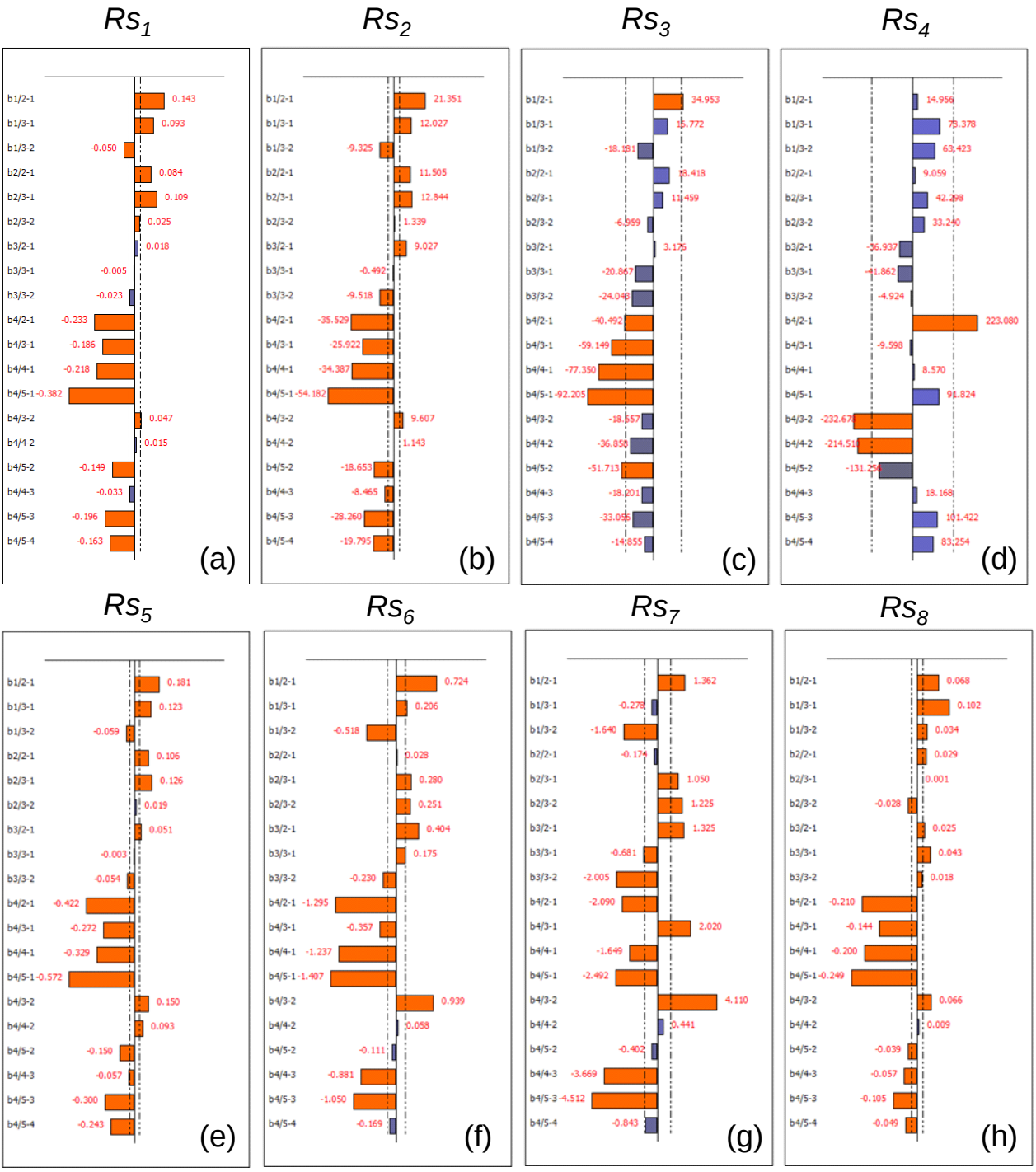
Nemrod-W software was employed to postulate the screening asymmetric matrix used to estimate the coefficients of the following Free-Wilson model:

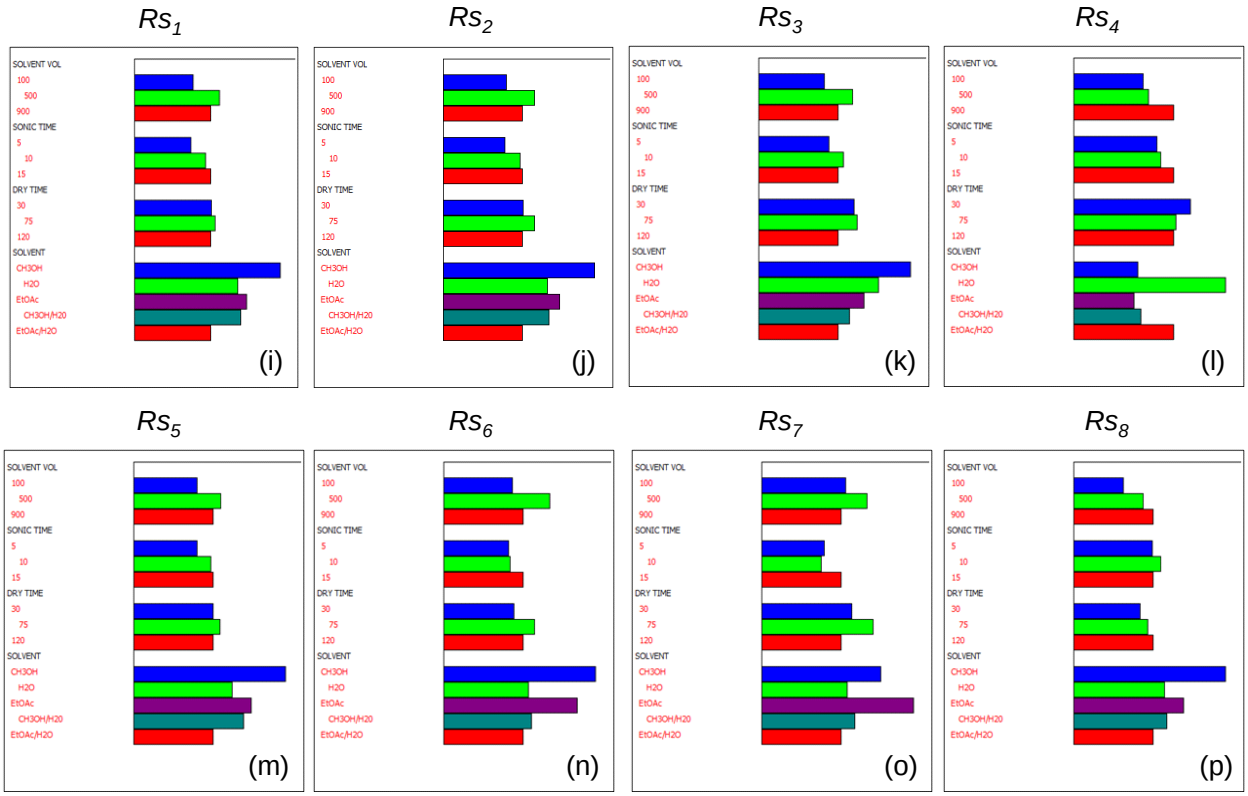
$$y = b0 + b1A * (X1A) + b1B * (X1B) + b2A * (X2A) + b2B * (X2B) + b3A * (X3A) + b3B * (X3B) + b4A * (X4A) + b4B * (X4B) + b4C * (X4C) + b4D * (X4D)$$

where X1 is SOLVENT VOL, X2 is SONIC TIME, X3 is DRY TIME, X4 is SOLVENT, b0 is the constant term and bi the linear coefficients of each factor.

The 18-run experimental plan is here reported with the measured responses; each analysis was repeated five times in order to obtain a reliable estimate of the experimental variance.

ANOVA showed that the models for all the responses were significant and graphical analysis of effects was carried out, as reported in Fig. 1-2.


Rs1
Rs2
Rs3
Rs4
Rs5
Rs6
Rs7
Rs8
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)



The graphic analysis of effects made it possible to obtain two types of plots; the first type of plot, reported in Fig. 1(a-h), the bars are related to the effects on the CQAs of the changes of level of the CPPs and the bar length is proportional to the amplitude of the difference of the relative effects between two levels.

The low level and the high level for the CPPs studied at three levels are indicated by 1,2 and 3, respectively. Instead for the type of solvent, investigated at five levels, are indicated with 1 to 5 respectively.

Thus, the bar indicated by b1/2-1 is related to the effect on the considered response of the change from level 2 to level 1 of the first factor, namely solvent volume. As another example, the bar indicated by b4/5-4 is related to the effect of the change level from level 5 to level 4 of the type of solvent.

The significant effects correspond to the orange bars, exceeding the reference line, calculated by means of the experimental error.

The second type of plot, shown in Fig. 2 (i-p), the bars correspond to the effects of the different levels of the factors, giving the possibility of identifying which are the levels which maximize or minimize the response, and thus which are the best levels for each considered CQAs.

From the investigation of these plots, we can underline that all our responses were maximized by a medium-high level volume of solvent except for Rs4, the effect of the change of level of this parameter was not significant, and Rs7, on the contrary was maximized by a medium level of this factor. Medium or high level of the time of sonic maximized all the responses without a marked difference, however the changing of this factor wasn't significant for Rs4.

The change of the time of dry wasn't significant for both Rs1 and Rs4, and medium-low level of this factor were preferred for the remaining CQAs.

The change of levels of type of solvent was undoubtedly of outmost importance, confirming the important role of this factor. All the responses were maximized by methanol except for Rs4.

The information from the two types of plots was matched, mediating the best values and ranges of the CPPs with respect to the different CQAs.

Hence, the final outcome of the screening phase consisted in fixing methanol as election exctracting solvent, moving to high values for volume of solvent and time of sonic and defining a new experimental domain for time of dry, in order to be focused on by RSM.

Futher experiments will consequentially carry out in order to obtain the restrain of the domain for each critical process parameters by RSM and finally reach the DS, where all the combinations of the variables lead to optimal values of the responses with a certain level of probability.

8. Conclusions

The projects developed during my PhD thesis have thoroughly highlighted the potential of capillary electrochemical chromatography in pharmaceutical analysis through the study and development of methods exploiting and fully investigating its different operating modes.

Capillary electrophoresis has been confirmed to be a flexible analytical technique since it allows the modulation of different operating conditions in order to obtain the desired separation and short analysis time and thus solving different analytical issues.

The wide range of applications for which the use of this method has proved to be successful includes identification of pharmaceutical substances, assay of drugs, determination of drug-related impurities, physicochemical measurements of drug molecules or chiral separations.

For pharmaceutical analysis, the range of applications for which CE can be used is extensive, possibly eclipsing the applications of the more frequently used HPLC. MEKC is a branch of CE, which has become during the years one of the most popular techniques in CE due to its high resolving power and capability of separating both ionic and neutral analytes.

MEKC solvent modified in particular is the most flexible of all CE techniques, offering the greatest selectivity to the widest range of compounds and can be considered the separation method of choice when performing CE analysis for pharmaceutical substances, relating its power to combine the separation mechanism of chromatography with the electrophoretic and electroosmotic movement of analytes and solutions for the separation of constituents in a sample.

Molecular Dynamics Simulations as a complement to Nuclear Magnetic Resonance provides claryfing informations for the investigation of intermolecular affinities, complexation and separation mechanisms involved in capillary electrophoresis analysis.

The systematic use of a multivariate approach has confirmed its potential especially in the presence of complex systems, not only for obtaining final optimized conditions, but also for the initial research of the correct experimental domain to be explored.

Chemiometry has proved to be complementary to analytical techniques to ensure the reliability of experimental data and a greater understanding of the phenomena in the study.

However, it is important to point out that the correct use of chemiometric techniques is of great help both in pharmaceutical analysis, leading to significant benefits in terms of time spent, sustained costs and results obtained, but also in any other study where the aim is to know the relationships between the variables studied and the optimization of the experimental system.

Regulatory documents, indeed, are strongly recommending the implementation of QbD approach, a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding.

9. Bibliography

[1] R. Holm, D. P. Elder, *Eu J Pharm Sci* 87 (2016), pp. 118-135.

[2] A. Jouyban, E. Kendle, *Electrophoresis* 29 (2008), pp. 3531-3551.

[3] R. Peraman, K. Bhadraya, Y. Reddy, *Int J Anal Chem* 15 (2015), pp.1-9.

[4] B. Pasquini, S. Orlandini, M. Goodarzi, C. Caprini, R. Gotti, S. Furlanetto, *Talanta* 150 (2016), pp. 7-13.

[5] B. Pasquini, F. Melani, C. Caprini, M. Del Bubba, S. Pinzauti, S. Orlandini, S. Furlanetto, *J Pharm Biomed Anal* 144 (2017), pp. 220-229.

[6] G. K. E. Scriba, *J Chromatogr A* 1467 (2016), pp. 56-78.

[7] P. K. Sahu, N. R. Ramisetti, T. Cecchi, S. Swain, C. S. Patro, J. Panda, *J Pharm Biomed Anal*, in press, 2017.

[8] G.A. Lewis, R. Phan-Tan-Luu, *Pharmaceutical Experimental Design*, Marcel Dekker, New York, 1999.

[9] R. Carlson, Design and Optimization in Organic Synthesis, Elsevier, Amsterdam, 1992.

[10] S. Furlanetto, S. Orlandini, I. Giannini, G. Beretta, S. Pinzauti, *Electrophoresis* 30 (2009), pp. 633-643.

[11] D. B. Hibbert Experimental design in chromatography: A tutorial review. *J Chromatogr B* (2012).

[12] A. Peissik, Methodologie de la recherche experimentale: proprietes et caracteristiques des matrices d'experiences pour les modeles polynomiaux du second degre, Tesi di Dottorato in Scienze, Marseille, 1995.

[13] J.L. Goupy, Methods for Experimental Design, Elsevier, Amsterdam, 1992.

[14] T. Lundstedt, E. Seifert, L. Abramo, B. Thelin, Å. Nyström, J. Pettersen, R. Bergman, *Chemometr. Intell. Lab. Syst.* 42 (1998), pp. 3-40.

[15] A. Atkinson, A. Donev, R. Tobias, Optimum Experimental Designs, With SAS, Oxford University Press, 2007, Oxford.

[16] G. Piepel, B. Pasquini, S. Cooley, A. Heredia-Langner, S. Orlandini, S. Furlanetto, *Talanta* (2012).

[17] J.A. Cornell, Experiments with Mixtures (3rd Edn.), John Wiley & Sons, New York, 2002.

[18] Design-Expert V8, Minneapolis, MN.

[19] Design-Expert V8, Minneapolis, MN.

[20] JMP V9, SAS Institute, Inc., Cary, NC.

[21] Minitab V16. Minitab Inc., State College, PA.

[22] R. Todeschini, Introduzione alla Chemiometria, Edises, Napoli, 1998.

[23] L. Song, Q. Ou, W. Yu, G. Li, J. Chromatogr. A 699 (1995) 371-382.

[24] B. Sunn, K. Dunn, *Magnetic Resonance Imaging*, 23 (2015), pp. 259-262.

[25] A. J. Kolkman, NMR spectroscopy in drug discovery 1985.

[26] A. Hospital, J. R. Goñi, M. Orozco, J. L. Gelpí, *Adv Appl Bioinform Chem.* 8 (2015), 37–47.

[27] M. Karplus, A. McCammon, *Nature Structural Biology* 9 (2002), pp. 646 – 652.

[28] C. Kandt, W. Ash, P. Tielman, *Methods* 41 (2007), pp. 475-488.

[29] D.N. Geiger, High Performance Capillary Electrophoresis-An Introduction, Hewlett-Packard, Waldbronn, 1992.

[30] S. F. Y. Li, *Capillary Electrophoresis*, Elsevier, Amsterdam, 1992.

[31] M. V. Smoluchowski, *Elektrosche kataphorese. Physik* 6 (1905), pp. 529.

[32] D.C. Harris, Chimica Analitica Quantitativa, 2ª Ed., Zanichelli, Bologna, 2005.

[33] F. Tagliaro, F.P. Smith, *TRAC-Trends Anal. Chem.* 15 (1996), pp. 513-525.

[34] H. Whatley, *Clinical and Forensic Applications of Capillary Electrophoresis*, (2001), pp. 21-58.

[35] C.F. Poole, The Essence of Chromatography, Elsevier, Amsterdam, 2003.

[36] Li, Y., D.Q., Yang, S., Sudini, R., McGuire, M.A., Bhanushali, D.S., Kord, *A.S.J. Pharm. Biomed. Anal*. 52 (2010), pp. 493-507.

[37] H. Watarai, *Chem. Lett.* 3 (1991), pp. 391-394.

[38] Huang J, Kaul G, Cai C, Chatlapalli R, Harnandez-Abad P, Ghosh K, Nagi A *Int J Pharm* 382 (2009), pp. 23-32.

[39] K.D. Altria, *J. Chromatogr. A* 892 (2000), pp. 171-186.

[40] G. Hancu, B. Simon, A. Rusu, E. Mircia, Á. Gyéresi, *Adv Pharm Bull* 3 (2013), pp.1–8.

[41] X. Fu, J. Lu, A. Zhu, *J. Chromatogr. A* 735 (1996), pp. 341-344.

[42] S.H. Hansen, C. Gabel-Jensen, D.T.M. El-Sherbiny, S. Pedersen-Bjergaard, *TrAC-Trends Anal. Chem.* 20 (2001), pp. 494-499.

[43] N. Gorski, M. Gradzielski, H. Hoffmann, Ber. Bunsenges, *Phys. Chem.* 100 (1996), pp. 1109-1117.

[44] J. H. T. Luong, A. L. Nguyen, *J. Chromatogr. A* 792 (1997), pp. 431-444.

[45] S. Fanali, *J. Chromatogr. A* 875 (2000), pp. 89-122.

[46] K.D Altria, *J. Chromatogr. A*, 735 (1996), p. 53.

[47] S. Terabe, Y. Miyashita, Y. Ishihama, O. Shibata, *J. Chromatogr.* 636 (1993), pp. 47-55.

[48] P. Borman, P. Nethercote, M. Chatfield, D. Thompson, K. Truman, *Pharm Tech 31* (2007), pp. 142-152.

[49] S. Furlanetto, S. Orlandini, P. Mura, M. Sergent, S. Pinzauti, *Anal. Bioanal. Chem.* 377 (2003), pp. 937-944.

[50] Pharmaceutical CGMPs for the 21st century – A Risk-Based Approach. Final report (2004) U.S Food and Drug Administration.

[51] Guidance for Industry. PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance (2004) U.S Food and Drug Administration.

[52] International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (2009) ICH harmonised tripartite guideline, Q8(R2), pharmaceutical development. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, Geneva.

[53] S. Verma, Y. Lan, R. Gokhale, D. J. Burgess, *Int J Pharm* 377 (2009), pp. 185-198.

[54] P. Lebrun, F. Krier, J. Mantanus, H. Grohganz, M. Yang, E. Rozet, B. Boulanger, B. Evrard, J. Rantanen, P. Hubert, *Eur J Pharm Biopharm* 80 (2012), pp. 226-234.

[55] Wu H, White M, Khan MA (2011) *Int J Pharm*, 405:63-78.

[56] Gavin PF, Olsen BA (2008*) J Pharm Biomed Anal* , 46:431-444.

[57] D. B. Hibbert, *J Chrom B* 910 (2012), pp. 2-13.

[58] G. Hanrahan, R. Montes, F. A. Gomez, *Anal Bioanal Chem* 390 (2008), pp. 169-179.

[59] F. G. Vogt, A. S. Kord, *J Pharm Sci* 100(2011), pp. 797-812.

[60] Molnár, I., Rieger, H.-J., Monks, K. E., *J. Chromatogr. A* 1217 (2010), pp. 3193-3200.

[61] M. McBrien *Chromatogr Today* 20 (2010), pp. 330-34.

[62] B. Debrus, P. Lebrun, A. Ceccato, G. Caliaro, E. Rozet, I. Nistor, R. Oprean, F. J. Rupérez, C. Barbas, B. Boulanger, P. Hubert *Anal Chim Acta* 691 (2011), pp. 33-42.

[63] P. Debrus, P. Lebrun, J. Mbinze Kindenge, F. Lecomte, A. Ceccato, G. Caliaro, J. Mavar Tayey Mbay, B. Boulanger, R. D. Marini, E. Rozet, P. Hubert, *J. Chromatogr A* 1218 (2011), pp. 5205-5215.

[64] R. A. Lionberger, S. L. Lee, L. M. Lee, A. Raw, L. X. Yu, *AAPS* 10 (2008), pp. 268-276.

[55] Ishikawa K (1985) What is Total Quality Control? The Japanese Way, Prentice-Hall, Englewood Cliffs, NJ, 63-64.

[66] K. Monks, I. Molnár, H. J. Rieger, B. Bogáti, E. Szabó, *J Chromatogr A* (2012) 1232, pp. 218-230.

[67] H. Y. Huang, C. W. Chiu, Y. C. Chen, J. M. Yeh, *Electrophoresis* 26 (2005), pp. 895-902.

[68] M. Swartz, J. Turpin, P. H. Lukulay, R. Verseput *LCGC Am* 26 (2009), pp. 1190-1197.

[69] D. Awotwe-Otoo , C. Agarabi, P. J. Faustino, M. J. Habib, S. Lee, M. A. Khan, R. B. Shah, *J Pharm Biomed Anal* 62 (2012), pp. 61-67.

[70] I. Molnár, K. E. Monks *Chromatographia* 73 (2011), pp. S5-S14.

[71] I. Krull, M. Swartz, J. Turpin, P. H. Lukulay, R. Verseput, *LCGC Am* 26 (2008), pp. 1190-1197.

[72] I. Giannini, S. Orlandini, R. Gotti, S. Pinzauti, S. Furlanetto, *Talanta* 80 (2009), pp. 781-788.

[73] S. Orlandini, R. Gotti, I. Giannini, B. Pasquini, Furlanetto S *J. Chromatogr. A* 1218 (2011), pp. 2611-2617.

[74] ICH Harmonised Tripartite Guideline, Q8(R2), *Pharmaceutical Development* 2009.

[75] B. De Backer, B. Debrus, P. Lebrun, L. Theunis, N. Dubois, L. Decock, A. Verstraete, P. H. Hubert, C. Charlier, *J Chromatogr B* 877 (2009), pp. 4115-4124.

[76] P. Lebrun, B. Govaerts, B. Debrus, A. Ceccato, G. Caliaro, P. Hubert, B. Boulanger *Chemom Intell Lab Syst* (2008) 91, pp. 4-16.

[77] Debrus P, Lebrun P, Mbinze Kindenge J, Lecomte F, Ceccato A, Caliaro G, Mavar Tayey Mbay J, Boulanger B, Marini RD, Rozet E, Hubert Ph (2011) *J. Chromatogr A* 1218, pp 5205-5215.

[78] K. E. Monks, H. J. Rieger, I. Molnár, *J Pharm Biomed Anal* 56 (2011), pp. 874-879.

[79] ICH Harmonised Tripartite Guideline. Pharmaceutical Quality Systems Q10 (2008) International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use.

[80] S. Orlandini, I. Giannini, M. Villar Navarro, S. Pinzauti, S. Furlanetto, *Electrophoresis* 31 (2010), pp. 3296-3304.

[81] SC. Sweetman, Martindale: the complete drug reference, 34th edn. Pharmaceutical Press, 2005 London.

[82] Council of Europe (2010) European pharmacopoeia 7.0, vol 2. Council of Europe, Strasbourg.

[83] S. Orlandini, B. Pasquini, M. Del Bubba, S. Pinzauti, S. Furlanetto, *J Chromatogr A* 1380 (2015), pp. 177–185.

[84] Lewis G, Mathieu D, Phan-Tan-Luu R (1999) Pharmaceutical experimental design, Marcel Dekker, New York.

[85] E. Rozet, P. Lebrun, B. Debrus, B. Boulanger, P. Hubert, *Trends Anal Chem,* 42 (2013), pp. 157–167.

[86] B. De Backer, B. Debrus, P. Lebrun, L. Theunis, N. Dubois, L. Decock, A. Verstraete, P. Hubert, C. Charlier, *J Chromatogr B* 877 (2009), pp. 4115– 4124.

[87] M. A. Herrador, A. G. Asuero, A. G. Gonzalez, *Chemom Intell Lab Syst*, 79 (2005), pp. 115–122.

[88] J. Ermer, J. H. McB. Miller (Eds.), *Method Validation in Pharmaceutical Analysis-A Guide to Best Practice*, Wiley-VCH, Weinheim 2004.

[89] S. Orlandini, S. Pinzauti, S. Furlanetto, *Anal. Bioanal. Ch*em. 405 (2013), pp.443-450.

[90] Mathew, N. T., Loder, E. W., *Am. J. Med.* 118 (2005), pp. 28S-35S.

[91] Bigal, M. E., Bordini, C. A., Antoniazzi, A. L., Speciali, J. G., *Arq. Neuropsiquiatr*. 61 (2003), pp. 313-320.

[92] Douša, M., Gibala, P., Rádl, S., Klecán, O., Mandelová, Z., Břicháč, J., *J. Pharm. Biomed. Anal. 58 (*2012), pp. 1-6.

[93] J. Ermer, J. H. McB. Miller (Eds.), *Method Validation in Pharmaceutical Analysis-A Guide to Best Practice*, Wiley-VCH, Weinheim 2004.

[94] S. El Deeb, H. Wätzig, D. Abd El-Hady, C. Sänger-van de Griend, G.K.E. Scriba, *Electrophoresis*.

[95] European Pharmacopoeia 7.0, Vol. 2, Council of Europe, Strasbourg, 2010.

[96] J.A.L. Souza, M.M. Albuquerque, S. Grangeiro Jr, M.F. Pimentel, D. Pereira de Santana, S.S. Simões, *Vib. Spectrosc*. 62 (2012), pp. 35-41.

[97] A.R. Tôrres, S. Grangeiro Jr, W.D. Fragoso, *Microchem. J*. 118 (2015), pp. 259-265.

[98] A.S. Rathore, C. Horváth, *Electrophoresis* 18 (1997), pp. 2935-2943.

[99] Orlandini, S., Pinzauti, S., Furlanetto, S., *Anal. Bioanal. Chem.* 405 (2013), pp. 443-450.

[100] S. Ma, F. Kálmán, A. Kálmán, F. Thunecke, C. Horváth, *J. Chromatogr. A* 716 (1995), pp. 167-182.

[101] S. Furlanetto, S. Orlandini, B. Pasquini, M. Del Bubba, S. Pinzauti, *Anal. Chim. Acta* 802 (2013) 113-124.

[102] B. Pasquini, S. Orlandini, M. Del Bubba, E. Bertol, S. Furlanetto, *Electrophoresis* 36 (2015), pp. 2650-2657.

[103] B. Debrus, D. Guillarme, S. Rudaz, *J. Pharm. Biomed. Anal*. 84 (2013), pp.215-223.

[104] MODDE v. 10, MKS Umetrics AB, Sweden.

[105] T. Takeo, Green and semi-fermented teas, in: K.C. Willson, M.N. Clifford (Eds.), Tea: Cultivation to Consumption, Chapman & Hall, London, 1992, pp. 413-457.

[106] M. T. Nelson, W. Humphrey, A. Gursoy, et al, *Int J Supercomput Appl High Perform Comput.* (1996), pp. 251–268.

[107] R. Suzuki, H. Shimodaira, Pvclust: an R package for assessing the uncertainty in hierarchical clustering, Bioinformatics 22 (2006), pp. 1540-1542.

[108] T.S. Pilgrim, R.J. Watling, K. Grice, *Food Chem.* 118 (2010), pp. 921-926.

[109] J.S. McKenzie, J.M. Jurado, F. de Pablos, *Food Chem.* 123 (2010), pp. 859-864.

[110] R.W. Kennard, L.A. Stone, *Technometrics* 11 (1969), pp. 137-148.

[111] R. Gotti, S. Furlanetto, S. Lanteri, S. Olmo, A. Ragaini, V. Cavrini, *Electrophoresis* 30 (2009), pp. 2922-2930.

[112] R Core Team, A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2013. http://www.R-project.org. Accessed 31 Aug 2015.

[113] European Medicines Agency, Assessment report for Volibris, Doc. Ref.: EMEA/123999/2008. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/000839/WC500053063.pdf, 2008 (accessed on 7th March 2016).

[114] M.B.V. Narayana, K.B. Chandrasekhar, B.M. Rao, *J. Chromatogr. Sci.* 52 (2014), pp. 818-825

[115] N.R. Ramisetti, R. Kuntamukkala, *New J. Chem.* 38 (2014), pp. 3050-3061.

[116] M. Nazeerunnisa, L. Garikapati, S.S. Bethanabhatla, *Malays. J. Anal. Sci.*19 (2015), pp. 595-602.

[117] S. Orlandini, I. Giannini, R. Gotti, S. Pinzauti, E. La Porta, S. Furlanetto, *Electrophoresis* 28 (2007), pp. 395-405.

[118] M. Silva, *Electrophoresis* 34 (2013), pp. 141-158.

[119] S. Orlandini, B. Pasquini, C. Caprini, M. Del Bubba, M. Douša, S. Pinzauti, S. Furlanetto, *J. Chromatogr. A* 1467 (2016), pp. 363-371.

[120] L. Eriksson, E. Johansson, N. Kettaneh-Wold, C. Wikström, S. Wold, Design of Experiments – Principles and Applications, MKS Umetrics AB, Umeå, Sweden, 2008.

[121] ICH Harmonised Tripartite Guideline. Quality Risk Management Q9 (2005) International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use.

[122] B. Chankvetadze, W. Lindner, G.K.E. Scriba, *Anal. Chem.* 76 (2004), pp. 4256-4260.

[123] M. Fillet, Ph. Hubert, J. Crommen, *J. Chromatogr. A* 875 (2000), pp. 123-134.

[124] P. Mura, *J. Pharm. Biomed. Anal.* 101 (2014), pp. 238-250.

[125] M. Fillet, L. Fotsing, J. Crommen, *J. Chromatogr. A* 17 (1998), pp. 113-119.

[126] M. Fillet, B. Chankvetadze, J. Crommen, G. Blaschke, *Electrophoresis* 20 (1999), pp. 2691-2697.

[127] C.E. Evans, A.M. Stalcup, *Chirality* 15 (2003), pp. 709-723.

[128] M.A. Rodriguez, Y. Liu, R. McCulla, S.W. Jenks, D.W. Armstrong, *Electrophoresis* 23 (2002), pp. 1561-1570.

[129] F. Melani, I. Giannini, B. Pasquini, S. Orlandini, S. Pinzauti, S. Furlanetto, *Electrophoresis* 32 (2011), pp. 3062-3069.

[130] P. Brocos, N. Díaz-Vergara, X. Banquy, S. Pérez-Casas, M. Costas, Á. Piñeiro, *J. Phys. Chem. B* 114 (2010), pp. 12455-12467.

[131] L. Jiang, Y. Yan, J. Huang, *Adv. Colloid Interfac.*169 (2011), pp. 13-25.

[132] M.H. Maeso, S.P. González, C. Bravo-Díaz, E. González-Romero, *Colloid Surface A* 249 (2004), pp. 29-33.

[133] N. Funasaki, H. Yodo, S. Hada, S. Neya, *Chem. Soc. Jpn.* 65 (1992), pp. 1323-1330.

[134] E.D. Vega, K. Lomsadze, L. Chankvetadze, A. Salgado, G.K.E. Scriba, E. Calvo, J.A. López, A.L. Crego, M.L. Marina, B. Chankvetadze, *Electrophoresis* 32 (2011), pp. 2640-2647.

[135] D. A. R. Rubio, D. Zanette, F. Nome, C.A. Bunton, *Langmuir* 10 (1994), pp. 1151-1154.

[136] S.H. Hansen, C. Gabel-Jensen, S. Pedersen-Bjergaard, *J. Sep. Sci.* 24 (2001) 643-650.

[137] L. Pedersen, K. Andersen-Ranberg, M. Hollergaard, M. Nybo, *Clin Biochem.* 50 (2017), pp. 703-709.

[138] P. M. De Kesel, S. Capiau, W. E. Lambert, C. P. Stove, *Bioanalysis.* 6 (2014), pp. 1871-4.

[139] U. Duthaler, B. Berge, S. Erb, M. Battegay, E. Letang , S. Gaugler, S. Krähenbühl, M. Haschke, *J Mass Spectrom.* 52 (2017), pp. 534-542.

[140] M. Wagner, D. Tonoli, E. Varesio, G. Hopfgartner, *Mass Spectrometry Rewiers* 2014.

10. Appendix

PUBLISHED PAPERS

- **C. Caprini**, B. Pasquini, F. Melani, M. Del Bubba, A. Giuffrida, S. Orlandini, S. Furlanetto, Exploring the effect of cosurfactant on separation selectivity in solvent-modified MEKC by Molecular Dynamics and NMR: the case of diclofenac and its impurities, ***J. Pharm. Biomed. Anal***. 149 (2018), pp. 249-257.

- B. Pasquini, F. Melani, **C. Caprini**, M. Del Bubba, S. Pinzauti, S. Orlandini, S. Furlanetto, Combined approach using capillary electrophoresis, NMR and molecular modeling for ambrisentan related substances analysis: Investigation of intermolecular affinities, complexation and separation mechanism, ***J. Pharm. Biomed. Anal.*** 144 (2017), pp. 220-229.

- S. Orlandini, B. Pasquini, **C. Caprini**, M. Del Bubba, M. Dousa, M. S. Pinzauti, S. Furlanetto, Enantioseparation and impurity determination of ambrisentan using cyclodextrin-modified micellar electrokinetic chromatography: Visualizing the design space within quality by design framework, ***J. Pharm. Biomed. Anal.*** 1467 (2016), pp. 363-371.

- S. Orlandini, B. Pasquini, **C. Caprini**, M. Del Bubba, L. Squarcialupi, V. Colotta, S. Furlanetto, A comprehensive strategy in the development of a cyclodextrin-modified microemulsion electrokinetic chromatographic method for the assay of diclofenac and its impurities: Mixture-process variable experiments and quality by design, ***J. Chromatogr. A*** 1466 (2016), pp. 189-198.

- B. Pasquini, S. Orlandini, **C. Caprini**, M. Del Bubba, M. Innocenti, G. Brusotti, S. Furlanetto, Cyclodextrin- and solvent-modified micellar electrokinetic chromatography for the determination of captopril,

hydrochlorothiazide and their impurities: A Quality by Design approach, ***Talanta*** 160 (2016), pp. 332-339.

- B. Pasquini, S. Orlandini, M. Goodarzi, **C. Caprini**, R. Gotti, S. Furlanetto, Chiral cyclodextrin-modified micellar electrokinetic chromatography and chemometric techniques for green tea samples origin discrimination, ***Talanta*** 150 (2016), pp. 7-13.

- S. Orlandini, B. Pasquini, **C. Caprini**, M. Del Bubba, S. Pinzauti, S. Furlanetto, Analytical Quality by Design in pharmaceutical quality assurance: Development of a capillary electrophoresis method for the analysis of zolmitriptan and its impurities, ***Electrophoresis*** 36 (2015), pp. 2642-2649.

- S. Furlanetto, S. Orlandini, B. Pasquini, **C. Caprini**, P. Mura, S. Pinzauti, Fast analysis of glibenclamide and its impurities: quality by design framework in capillary electrophoresis method development, ***Anal. Bioanal. Chem.*** 407 (2015), pp. 7637-7646.

- F. Melani, B. Pasquini, **C. Caprini**, R. Gotti, S. Orlandini, S. Furlanetto, Combination of capillary electrophoresis, molecular modeling and NMR to study the enantioselective complexation of sulpiride with double cyclodextrin systems, ***J. Pharm. Biomed. Anal.*** 114 (2015), pp. 265-271.

11. Scientific Acknowledgments

Department of Chemistry, University of Florence:

Professor Sandra Furlanetto, for the possibility to start this adventure making a fantastic wish come true, to work on these ambitious and challenging projects and for her believing in my competences before anyone else.

Professor Fabrizio Melani, for sharing his constant passion and his practical professionalism with a clever irony.

Professor Sergio Pinzauti, for his continued support and his precious advices at any time of the day.

Dr Serena Orlandini, for her extreme competence, patience and constructive corrections in every project.

Dr Benedetta Pasquini, for sharing the goals of quality in every aspect during this three years, for having taught me an excellent working philosophy and a powerful way of thinking.

Department of Chemistry, University of Geneve:

Professor Gerard Hopfgartner, to have given me the opportunity to make a living experience and to have deepened new analytical techniques transversal to mine.

Dr Laura Akbal, for supporting me in our project, taught excellently and patiently supported in our daily challenges.

Dr Sophie Veyrat, for giving me new ways to deal with the issues, for showing me great potential behind a daily smile.

Dr Thomas Sticker, for his excellent ability to listen, his bad fake Italian and fantastic afternoons of work.