UNIVERSITATEA DE MEDICINĂ ŞI FARMACIE "IULIU HAŢIEGANU" CLUJ-NAPOCA

ŞCOALA DOCTORALĂ

UNIVERSITÉ DE LIÈGE

CLUJ-NAPOCA 2012

EDITURA MEDICALĂ
UNIVERSITARĂ
IULIU HAŢIEGANU
CLUJ NAPOCA

DOCTORAL THESIS

Optimization of separation methods by a design of experiment – design space methodology

PhD student **Iolanda Nistor**

Scientific supervisors **Radu Oprean** / **Philippe Hubert**

Mamei mele

Motto

"If a problem is fixable, if a situation is such that you can do something about it, then there is no need to worry. If it's not fixable, then there is no help in worrying. There is no benefit in worrying whatsoever."
— Dalai Lama XIV

Acknowledgements

This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

I would like to start by expressing my deepest gratitude to my thesis co-supervisors Prof. Radu Oprean and Prof. Philippe Hubert who gave me the opportunity to initiate and achieve this research work by joining their excellent research groups. Their confidence, invaluable help and unfailling support have provided a good basis for the present thesis.

My sincere thanks go to Dr. Eric Rozet, Dr. Benjamin Debrus, Dr. Pierre Lebrun and Frederic Lecomte for their key contributions to my studies and priceless lessons in science and in life. I would also like to thank all the members of the Laboratory of Analytical Chemistry from Liège for making me feel welcome right from the start.

The financial support of Wallonie-Bruxelles Internationl is gratefully acknowledged.

I would like to offer special thanks to the members of my thesis committee for their supervision along with the members of the jury who granted me the honour of assessing this work.

I grace all my friends for their love and support.

Last but surely not least, I am thankful to my late father for watching me from up above and to my mother for her endless love, supporting my dreams and always believing in me.

LIST OF PUBLICATIONS

1. Benjamin Debrus, Pierre Lebrun, Eric Rozet, **Iolanda Nistor**, Attilio Ceccato, Gabriel Caliaro, Radu Oprean, Bruno Boulanger, Philippe Hubert, Nouvelle méthodologie pour le développement automatisé de méthodes analytiques en chromatographie liquide pour l'analyse de mélanges de composés inconnus, Spectra Analyse 268 (2009) 28-33.

2. Benjamin Debrus, Pierre Lebrun, Attilio Ceccato, Gabriel Caliaro, Eric Rozet, **Iolanda Nistor**, Radu Oprean, Francisco J Rupérez, Coral Barbas, Bruno Boulanger, Philippe Hubert, Application of new methodologies based on design of experiments, independent component analysis and design space for robust optimization in liquid chromatography, Analytica Chimica Acta 691 (2011) 33-42. *ISI factor 4.310*

3. **Iolanda Nistor**, Martine Cao, Pierre Lebrun, Frédéric Lecomte, Luc Angenot, Michel Frederich, Radu Oprean, Philippe Hubert, Application of a new optimization strategy for the separation of tertiary alkaloids extracted from *Strychnos usambarensis* leaves, Journal of Pharmaceutical and Biomedical Analysis 56 (2011) 30-37. *ISI factor 2.733*

4. Martine Cao, Raymond Muganga, **Iolanda Nistor**, Monique Tits, Luc Angenot, Michel Frederich, LC–SPE–NMR–MS analysis of *Strychnos usambarensis* fruits from Rwanda, Phytochemistry Letters 5 (2012) 170-173. *ISI factor 1.364*

5. **Iolanda Nistor**, Pierre Lebrun, Attilio Ceccato, Frederic Lecomte, Ines Slama, Radu Oprean, Eduard Badarau, Fabien Dufour, Katina Sourou Sylvestre Dossou, Marianne Fillet, Jean-François Liégeois, Philippe Hubert, Eric Rozet, Implementation of a design space approach for enantiomeric separations in polar organic solvent chromatography, Journal of Pharmaceutical and Biomedical Analysis (submitted)

Table of Contents

ABBREVIATIONS

CPP	Critical Process Parameter
CQA	Critical Quality Attribute
CSP	Chiral Stationary Phase
DAD	Diode Array Detector
DEA	Diethylamine
DoE	Design of Experiments
DS	Design Space
HPLC	High Performance Liquid Chromatography
HTLC	High-Temperature Liquid Chromatography
ICA	Independent Component Analysis
ICH	International Conference on Harmonization
LC	Liquid Chromatography
MeCN	Acetonitrile
MeOH	Methanol
OM	Organic Modifier
OVAT	One-Variable-At-a-Time
QbD	Quality by Design
R_S	Resolution
RSA	Response Surface Analysis
S	Separation criterion
SFC	Supercritical Fluid Chromatography
TFA	Trifluoroacetic acid
UHPLC	Ultra High-Performance Liquid Chromatography
UV	Ultraviolet

INTRODUCTION

Liquid chromatography is the analytical technique of choice in the pharmaceutical field. It is mainly used to separate, identify and quantify the compounds of interest and related substances. Whether for quality control of drugs, the study of their stability or to provide qualitative and quantitative informations to researchers in disciplines as varied as synthetic chemistry, toxicology, pharmaceutics, liquid chromatography has become a necessity. Research seeking innovative strategies for the development of chromatographic methods is therefore ever more important in the analytical field.

In the current environment guided by the concepts of quality by design (QbD) and design space (DS) recommended by the International Conference on Harmonization (ICH), the methodologies for optimization of chromatographic methods should provide analysts along with more robust chromatographic conditions, a good understanding of process optimization and increased credibility of the predictions.

Therefore, as part of this work, a methodology combining both chromatographic behavior modeling compounds and optimization of analytical conditions was tested. This methodology, based on design of experiments, the inclusion of the prediction error and the propagation of this error, from the modeled responses to the separation criterion, has allowed the formalization and the identification of design spaces.

The novelty of the studies presented here resides in the use of a methodology which is in perfect agreement with the QbD concept on complex samples. This approach can be applied to other analytical techniques such as capillary electrophoresis or near infrared spectroscopy.

Similarly, it is not unreasonable to think that this methodology could be used for process optimization going out of the analytical framework and more particularly into production processes of the pharmaceutical technology.

The combination of design of experiments and design space methodology presented in this work allowed for this innovative global approach to pave the way towards the automated development of chromatographic methods.

REVIEW OF THE LITERATURE

1. LIQUID CHROMATOGRAPHY

Liquid chromatography is a technique suited for analyzing non-volatile and thermally fragile molecules including high molecular weight compounds. High performance liquid chromatography (HPLC) life science applications focus on the separation, quantitation, and purification of biomolecules in medical research and drug discovery, including purifying both small molecules and macromolecules derived from chemical synthesis or natural processes.

Throughout the years, HPLC has acquired a high degree of versatility which is a result of its ability to easily separate a wide variety of chemical mixtures. As stated by a new research report on HPLC systems and accessories, this market is saturated in most of the technologies and instead of new product development, the growing trend is sustained by the replacement sector, continual technology refinement and improvements [1]. This includes the lab automation segment, software improvement, accurate and high speed analysis.

The high demand for HPLC devices is linked to the pharmaceutical and biotechnological industry, followed by food products industry and chemical industry. Recently, HPLC systems are also included in various industries such as petrochemicals, pesticides, cosmetics and environmental testing.

As presented by another report, in 2011, the spectrometry market which is driven by the liquid chromatography – mass spectrometry (LC-MS) hyphenated technique had the largest share (33.8%), closely followed by chromatography (22%), but overall this report which provides a comprehensive review of the global life science and chemical instrumentation market estimates a growth that will reach $ 45.2 billion in the year 2016 starting from $ 30.2 billion in 2011 [2].

1.1. Historical aspects

Modern HPLC has its roots in the separation and purification of mixtures of plant pigments into pure constituents undertaken by a Russian botanist named Mikhail S. Tswett. This experiment represents the first rudimentary form of liquid chromatography [3].

From the form of liquid-liquid chromatography presented in 1903, Martin and Synge published in 1942 a paper on a form of liquid-liquid chromatography using gravity-fed silica columns and thus introducing partition chromatography [4].

The invention of gas chromatography in 1952 together with its successful applications provided the basis for the development of LC.

By the 1980s, when two major impediments represented by the lack of high-sensitivity detectors and reliable injectors were surpassed, HPLC began to immerge making it today an indispensable technique in pharmaceuticals as well as other industries.

1.2. Basic principles

Chromatographic process can be defined as a separation technique involving mass-transfer between the stationary and the mobile phase.

The rate of chromatographic zone migration is determined by the interaction and partitioning of the different components of the sample between the liquid mobile phase and the stationary phase. In chromatography, by convention, distribution coefficients are always given with reference to the stationary phase. The distribution of analytes between the two phases is determined by the equilibrium constant, K, termed also partition coefficient, defined as the molar concentration of analyte in the stationary phase (C_S) divided by the molar concentration of the analyte in the mobile phase (C_M) as presented in Eq. 1 [5].

$$K = \frac{C_S}{C_M} \qquad \text{Eq. 1}$$

The substances distributed preferentially in the moving phase will pass through the chromatographic system faster than those that are distributed preferentially in the stationary phase. As a result, differences in the retention times will be observed. The components will be eluted from the chromatographic system in the inverse order of the magnitude of their distribution coefficients with respect to the stationary phase. If strong interactions are present between the solute and the stationary phase, C_M will have a smaller value than C_S meaning that the molecule will move through the column more slowly and elute at later retention times. Because each solute has its own equilibrium between adsorption onto the surface of the solid and solubility in the solvent, the least soluble or best adsorbed ones travel more slowly.

As seen above, in HPLC, components are dissolved in a solvent and then is injected and delivered by a high-pressure pump through a chromatographic column to a detector. Usually, HPLC instrumentation includes the following modules: a pump system for mobile phase delivery, an autosampler, a column oven, a detector and a chromatographic data handling system.

1.3. HPLC modes

In this section, a classification of the four major separation modes of HPLC is presented starting with normal phase chromatography where dipol-dipol (polar) interactions are dominant. This traditional separation mode is based on repeated adsorption-desorption steps of the analyte on the stationary phase which is an adsorbent (typically silica or alumina) [5].

The second one is reversed-phase chromatography which results from the partition coefficient of hydrophobic molecules onto a hydrophobic solid support in a polar mobile phase. The main interactions encountered for this mechanism of separation are the solvophobic or hydrophobic ones [5].

As for the ion-exchange chromatography, the chromatographic separation is based on the ion interactions between the ion on the surface of the stationary phase and oppositely charged sample ions [5,6].

Size-exclusion chromatography is a technique concentrated solely on the molecular size. The stationary phase has precisely controlled pore sizes so that a large molecule is excluded from the pores and migrates quickly, whereas a small molecule can penetrate the pores and has a slower migration [5,6].

Besides the four separation modes described above, there are several others encountered in HPLC such as affinity chromatography, a technique very useful in separation of biological samples. The stationary phase contains functional groups with very specific reaction to the analytes. Another one is represented by chiral chromatography which dominates in the pharmaceutical industry and is oriented towards the exclusive separation of chiral substances [6].

1.4. Chromatographic parameters

The four fundamental chromatographic parameters that the kinetic and thermodynamic aspects of the HPLC theory are described below:

1. Retention factor (k)

The retention factor (k) is a unitless measure describing the degree of retention of the sample component in a particular chromatographic system. It is also defined as the time the solutes resides in the stationary phase relative to the time it resides in the mobile phase, as described in Eq. 2, where t_R is the retention time (the time between the sample injection and the apex of the peak) and t_0 is the retention time of the unretained species or the void time [7].

$$k = \frac{t_R - t_0}{t_0} \qquad \text{Eq. 2}$$

It is convenient to use the retention factor because it is independent of the column size and the solvent flow rate.

2. Selectivity (α)

Selectivity is the ability of a chromatographic system to discriminate between two different analytes as presented in Eq. 3:

$$\alpha = \frac{k_B}{k_A} \qquad \text{Eq. 3}$$

A value of selectivity higher than 1 is an indication of peak separation and that compound B is more strongly retained by the stationary phase, while compound A has a faster elution time [5-7].

Compared to the retention factor, selectivity is dependent on various factors such as the nature of the stationary phase and the mobile phase composition.

3. Resolution (R_S)

Although selectivity describes the separation of two sample components, it does not take into consideration peak widths (ω). Another measure of the degree of separation of two adjacent analytes is provided by the resolution (R_S) which is defined in Eq. 4 as [8]:

$$R_S = 2\frac{t_{R2} - t_{R1}}{\omega_2 + \omega_1} \qquad \text{Eq. 4}$$

A value of 1 for the resolution (R_S=1) indicates a partial separation, while values higher than or equal to 1.5 ($R_S \geq 1.5$) show that the HPLC method has achieved baseline separation for the analytes [9].

If a further step is taken, the resolution can be related to both thermodynamic factors (retention factor and selectivity) and kinetic ones (peak width and column efficiency which will be described at a later time). This relationship is expressed in Eq. 5 [7-9]:

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

Retention Selectivity Efficiency

To achieve higher resolution, it is often found that controlling the retention factor (k) will lead to an improved chromatographic separation. Also maximizing selectivity can have a positive effect on resolution as well as increasing the column efficiency [10].

4. Efficiency (N)

Column efficiency is dependent of N that represents the number of theoretical plates or plate number. This term measures the degree of peak distortion in a particular column and can be expressed as shown in Eq. 6.

$$N = 16\left(\frac{t_R}{\omega}\right)^2 \qquad \text{Eq. 6}$$

Given that from the initial injection band as a compound passes through the column, it slowly diffuses away makes band broadening inside the column fundamental to all chromatographic separations [8,11].

Another possibility of describing the efficiency of a given column is offered by the height equivalent to a theoretical plate (HETP or H) which is related to the ratio of column length (L) and number of theoretical plates (N) as presented in Eq. 7.

$$N = \frac{L}{H} \qquad \text{Eq. 7}$$

Generally, a better separation can be achieved if column lenght (L) is increased which also results in an increase for number of theoretical plates (N), but one must keep in mind that the pressure drop and the analysis time [12].

Among the factors defining the efficiency of a particular column reside the geometry, uniformity and density of the column packing.

The best known expression that describes the quantitative relationship between the height equivalent to a theoretical plate (H) and flow rate is the Van Deemter function presented in Eq. 8 [7]:

$$H = A + \frac{B}{u} + Cu \qquad \text{Eq. 8}$$

Where u is the linear flow velocity and A, B and C represent constants for the various mechanisms that contribute to band broadening as illustrated by the graphic of the Van Deemter curve in (fig.1) [11,12].

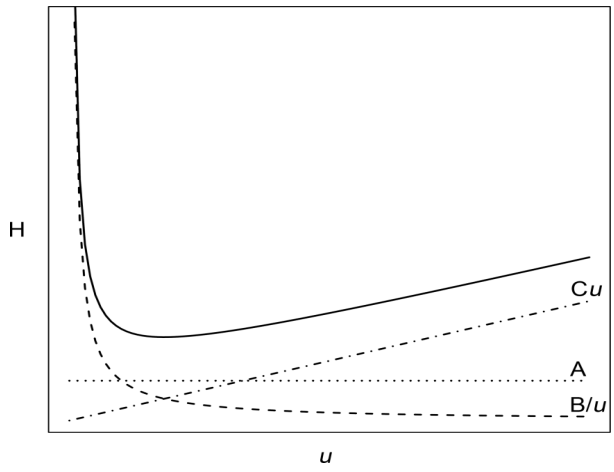


Fig.1. The plot of the plate height versus the average linear velocity of the mobile phase as described by the Van Deemter equation.

One of the processes leading to dispersion is caused by eddy diffusion (A) or the multipath effect. This term is independent from the mobile phase linear flow velocity and is proportional to the diameter of particles. In order to diminish this effect, smaller

particles as well as smaller column diameters should be chosen [7-11]. By doing so, the inhomogeneity of flow path velocities around the stationary phase particles is reduced.

A second contribution to dispersion is added by molecular diffusion (B) which defines the effect arisen from the browninan movement of analytes in the mobile phase. The term B is becomes more significant at very low flow rate because if the velocity of the mobile phase is high, the analyte spends less time on the column, thus the effect of longitudinal diffusion decreases [11,12].

An additional contribution to dispersion is provided by mass transfer (C) which is a component related to the amount of time a certain analyte takes to achieve a dynamical equilibrium between the stationary phase and the mobile phase. This kinetic effect is important at higher flow rates because if the velocity of the mobile phase is high and the analyte has a strong affinity for the stationary phase, band broadening is increased [7-11].

1.5. Contemporary HPLC

In the last decade, in modern chromatography, the need for even more efficient and faster analysis led to new developments based on the use of silica-based monolithic supports, elevated mobile phase temperatures, sub-3 μm superficially porous particles (fused core), sub-2 μm porous particles for ultra high-performance liquid chromatography (UHPLC) and supercritical fluid chromatography (SFC) for chiral separations. With all the possibilities available on the market, when selecting the best analytical strategy, a choice has to be made on the basis of the analytical problem (e.g. isocratic vs. gradient, throughput vs. efficiency) and the properties of the analyte [13].

The silica-based monolithic supports present highly porous rods of silica with a mesopore/macropore structure which allows smaller analysis time through low column backpressure when using a conventional chromatographic system [14]. High flow rates are also achieved because of faster transfer kinetics giving monoliths outstanding properties.

The principle behind high-temperature liquid chromatography (HTLC) which follows the trend of green chemistry is that an increase in temperature leads to an important reduction of mobile phase viscosity, improving mass transfer. Thus, an increase in optimum linear velocity and a column backpressure reduction are achieved. Although these positive aspects, a temperature range between 60°C and 200°C entails specific instrumentation and depends on the stability of the analytes studied.

Increased mobile phase temperature is usually used for large molecules, while UHPLC and fused-core technologies are utilized when small-molecular weight compounds are involved.

Fused-core technology is based on 2.7 μm superficially porous particles formed by a 1.7 μm solid inner core and a 0.5 μm porous outer core which decreases resistance to mass transfer and thus reducing peak broadening. Along with other features, such as narrow particle size distribution and high packing density, these sub-3 μm superficially porous particles provide the high speed and high efficiency of sub-2 μm particles, but at approximately half the backpressure for the same column length [13].

Columns packed with sub-2 μm particles can help achieve high efficiency while shortening analysis time because the mass-transfer term is low for these columns. The pressure limitation that was reached for conventional chromatographic systems by using these types of columns packed with smaller particles was overcome by new generations of systems capable of withstanding pressures as high as 1000 bar.

Together with the four trends presented earlier, SFC which employs columns similar to those utilized in HPLC has gained a renewed interest for scientist interested in chiral separations. Mobile phases have low viscosities and high diffusion coefficients as to those of traditional HPLC resulting in the use of higher mobile phase flow rates and longer columns, thus achieving rapid analyses and high efficiency separations.

1.6. Method development

When developing a HPLC method, three general steps are usually demanded as follows:

Step 1 represents setting suitable objectives for method development. Step 2 is optimizing the method to define method parameters which achieve the desired separation. This step includes performing screening experiments to identify the critical process parameters that are going to be optimized. And last, but not least is step 3 which corresponds to method validation [5,6,17].

Goals for method development have a wide range starting from improving its sensitivity and/or specificity, achieving a better recovery of drugs degree, enhancing the degree of accuracy and precision to simplifying or increase its affordability. The choice always depends on the needs that have to be met [7,9].

During the optimization step, the initial set of parameters and conditions that have evolved from the first stage of development and the screening sub-step which is a part of optimization, are improved in terms of resolution, peak shape, elution time and

other parameters that meet the set goals. Two optimization strategies are available for the scientist at this point in time: a univariate and a multivariate approach [15].

The classical univariate procedure represented by the one-variable-at-a-time (OVAT) approach has only one factor at a time which is varied and optimized, but there are some inconveniences to this procedure, some of them being that the interactions between the factors are not taken into consideration and a global optimum might not be found [15]. Up to now, commercially available software such as Drylab [16], Chromsword [17], ACD/LC simulator [18] and Osiris [19] come with a drawback regarding the number and type of parameters to be optimized. This limitation is to two or three general factors and shows a lack in specificity for techniques like ion-pairing chromatography and chiral separations.

On the other hand, a multivariate approach which can be formed by sequential or simultaneous procedures has the advantage of varying several factors simultaneously.

After the optimization step, robustness testing is required as part of method validation. This is applied with the intent to evaluate the method capacity to remain unaffected by slight deliberate variations in its parameters. Recent trends in pharmaceutical development and regulations lean towards a simultaneously step for method optimization and robustness evaluation which have a starting point based on design of experiments (DoE) that will be detailed in section I.3.1.

2. Regulatory support for biopharmaceutical development

Within a distinctively rapid pace industry, regulatory bodies around the globe have initiated a more modern philosophy regarding biopharmaceutical development. This new approache involves the utilization of numerous concepts such as quality by design (QbD) and design space (DS) that along with quality risk management principles serve as a definitive reference source in the step of modernizing biopharmaceutical development.

2.1. Quality by Design

The concept of QbD promotes a more scientific, risk-based, holistic and proactive approach to pharmaceutical development. The Q8 (pharmaceutical development), Q9 (quality risk management) and Q10 (pharmaceutical quality system) guidelines issued by the International Conference on Harmonization (ICH) provide the overall framework for implementation of QbD in biopharmaceutical development [20-23].

For QbD, process knowledge encompasses an understanding of variability and the relationship between critical process parameters (CPPs) and critical quality attributes (CQAs) [24]. In order to implement the QbD paradigm to biopharmaceutical method development, after identifying the potential CQAs, the next step relates to using the previously identified CPPs to establish an experimental design. The results thus retreived, along with tools like design of experiments (DoE), response surface analysis (RSA) and risk assessment, can then facilitate the difinition of the DS [25].

2.2. Critical quality parameter

CQA is defined as “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” [20]. In general, prior knowledge that is key to the risk assessments recommended by ICH Q9 leads to the identification of possible CQAs [21]. Usually, CQAs can be used to describe a quantitative measure of the process quality. Thus, for an analytical method development, among the potentials CQAs one can find the resolution of one or more components in the mixture, the analysis time and a recently introduced separation criterion [26].

2.3. Critical process parameters

A CPP is defined as "a process parameter whose variability has an impact on a CQA and therefore should be monitored or controlled to ensure the process produces the desired quality" [20].

CPPs are determined through a risk assessment process by using a design of experiment approach. As for the CQAs, when evaluating the potential CPPs, experimental knowledge as well as practical experience is taken into consideration. The range of each process parameters classified as critical will ensure the operating space and will further help establish the DS.

Translated to analytical method development, usually, the CPPs vary in function of the CQAs starting from the composition of mobile phase and the nature of the stationary phase to the type of elution, column temperature and flow rate.

2.4. Design space

ICH Q8 addresses the concept of the DS, which is currently defined as "the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality" [20]. The idea that quality has to be built in by design is therefore emphasized.

Robustness of a chromatographic method development is a fundamental criteria sustained by the QbD approach. This modern paradigm attempts to assess all the CPPs that influence the CQA, allowing thus, along with a comprehensive knowledge of all the sources of variability in the process, to construct the DS [27].

By including robustness testing in the early steps of method development, DS provides the flexibility of operating within the experimental domain formed by the factor ranges determined with the help of DoE.

The DoE methodology represents an efficient manner to simultaneously test for variable effects and interactions establishing causative relationships between CPPs and CQA.

2.5. Quality risk management

As presented in ICH Q8, information from method development studies can be a basis for quality risk management that is a systematic process for the assessment, control, communication and review of risks [20].

"Risk management can be used to design a quality product and its manufacturing process to consistently deliver the intended performance of the product." [21]. The quality risk management system requires identifying the CPPs along with their ranges in order to ensure quality for the method developed.

3. Chemometrics applied to analytical method development

An actual definition of chemometrics is: "the science that develops or applies mathematical methods or methods based on formal logic for the extraction of chemically relevant information from chemical data" [28].

In the fast-growing pharmaceutical industry, the use of chemometric tools for developing analytical methods as well as for determining optimal experimental conditions is a powerful way to overcome, solve or diminish serious issues as the treatment, evaluation and interpretation of multidimensional and very complex analytical data obtained [29].

In the analytical chemistry domain, both descriptive and predictive issues can be resolved by applying chemometrics. In predictive applications, chemometric tools help understand data patterns and build predictive models. The models capture the relationships among the CPPs to allow assessment of risk related with a single set of decisions or CQAs.

Among the techniques employed such as multivariate calibration, pattern recognition and clustering, experimental design has been widely adopted within the chemometrics community [29,30].

3.1. Design of experiments

DoE is a powerful tool compared to the univariate or One-Variable-At-a-Time (OVAT) approach that requires large resources to obtain a limited amount of information [15]. By involving the smallest possible number of useful experiments, DoE provides maximum information about the process factors effects and interactions, something not obtainable through a trial-an-error manner, thus increasing the system knowledge.

An experimental design is a systematic approach to simultaneously evaluate the influence the factors and their levels have on the responses of interest [20]. The responses obtained by applying a DoE strategy are fitted to mathematical equations that serve as models for prediction [26]. The most common empirical models are linear or quadratic ones. For example, a linear model with two factors (X_1 and X_2) can be written as the one presented in Eq. 9:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \varepsilon \quad \text{Eq. 9}$$

Where Y is the response that can be modelised with the help of Eq. 9 for given levels of the main effects X_1 and X_2 and the possible interaction X_1X_2. β_0, β_1, β_2 and β_{12} are the parameters of the model and ε represents the experimental error.

Typically, a quadratic or second-order model is employed in response surface designs meaning that the factors are evaluated at two levels and a curvature effect is also taken into consideration [31].

Roughly, experimental designs can be classified in screening designs such as full factorial and fractional factorial designs, response surface designs and mixture designs [30-34].

In fonction of the number of factors that are to be tested, a screening design can be applied before the optimisation step. This will permit screening a large number of factors in a small number of experiments [32].

Response surface designs are useful for establishing desirable operating conditions by finding the optimal levels for the CPPs [33].

Mixture designs are based on response surface designs, the only difference is related to the factors examined that are part of a mixture, meaning that they represent the percentage of a certain component in the mixture [34].

Basically, performing an experimental design will facilitate identifying the key factors in the process or the CPPs, the main and interaction effects for this process as well as the optimal condition that delivers acceptable process performance. The choice for the type of design used is strongly related to the number of the factors and the objectives that are to be atteined during method development.

3.2. Response surface analysis

When employing a response surface methodology two major objectives related to finding the optima and describing the response are met [35].

The regression model is an algebraic representation that helps understand the nature of the relationship among the factors as independent variables and the response as the dependent variable.

The first objective of determining the optimal conditions is usually derived from combining RSA along with DoE. This methodology is beneficial for robust design optimization.

The second objective is achieved by a simulation application that shows how the response varies in fonction of the different factors analyzed. This response surface can be vizualized graphically. The inconvenient appears if more than three factors are to be analyzed because graphs can illustrate only parts of the entire response surface [35].

3.3. Tolerance intervals

Tolerance intervals are tools based on prediction that include at the same time risk management principles in accordance with ICH Q9 guidelines, as to bring confidence and to demonstrate that the analytical method developed is suited for its intended purpose [21, 36].

B-expectation tolerance intervals or accuracy profile represent the likelihood that the expected proportion of the future measurements lies within the acceptance limits [37].

The general formula describing tolerance intervals that can be also be used as prediction intervals is presented in Eq. 10 [37]:

$$E\left(P\left[L \le x_i \le U\right]\right) = \beta \qquad \text{Eq. 10}$$

Where L and U are the lower and upper limits and E and P represent the estimators for the mathematical expectancy and for the probability, respectively. If β=0.95, this means that 95% of the results are included in the interval *[L; U]*.

If the tolerance intervals fall within the acceptance limits [-λ, λ] then this represents a guarantee that the method will provide accurate results according to the previously defined objectives of the analytical procedure [36].

PERSONAL CONTRIBUTIONS

1. Objectives

Liquid chromatography is currently the most common analytical technique in the pharmaceutical industry and in many other areas. Modern methodologies for development of chromatographic methods now make use of commercial computer software based on the theory of liquid chromatography. These methodologies can predict and optimize chromatographic separations. However, in the softwares currently available, no estimation of the prediction error is performed.

The current trend in analytical sciences is to estimate the risk of not achieving the initial objectives. Thus, for example, the validation of analytical methods is evolving more and more towards the possibility of assessing the proportion of future results found in the acceptance limits.

As part of pharmaceutical development and in particular in the development of chromatographic methods domain, validation is formalized by the concept of Quality by Design.

The first objective of this work is to identify and to establish a new methodology capable of optimizing chromatographic separations while estimating the quality of the development process and its predictions. In the optimization step, the method which takes into acount the correlation structure of the joint distribution of the variables will be prefered to the marginal or univariate approach. The interpretation of the model parameters based on the factors defining the experimental conditions will also be considered in order to highlight the relationship between predicted behavior and that described by the chromatographic theory.

In the last two parts of this work, the capacity of this global methodology to deliver precise predictions and a robust development of chromatographic methods will be tested. Because sample complexity not only makes separations difficult, but makes liquid chromatography method development an ardous task, chiral separations and plant extract complex samples will be chosen to demonstrate the powerfull methodology. Concurrently, identifying the optimal chromatographic conditions adapted to achieve the desired separation in the shortest possbile time, while maintaining the critical quality attributes, will be identified by implementing this new methodology.

2. The design of experiment – design space methodology

2.1. Introduction

The optimization of LC operating conditions for the separation of several compounds in complex samples is often intricate. Indeed, it can be tricky to separate compounds due to their similar chromatographic behaviours or fussy to elute all of them well separated when some have widely distinct physico-chemical properties (e.g. polarities, pKa, log P). Since past decades and mainly those last years, a lot of improvements in the fields of method development have been done [12,38,39] and it is thus possible to foresee different strategies that could be applied to find optimal separations in a rather automated way.

The main difficulty is that these strategies need to give accurate and robust predictions prior to the validation and transfer of these analytical methods. ICH Q8 (R2) guideline [20] provides a harmonized guidance to improve the robustness and reliability of pharmaceutical development. In this guideline the DS is defined as "the multidimensional combination and interaction of input variables (e.g. material attributes) and process parameters that have been demonstrated to provide assurance of quality".

To decrypt this definition, it can be assumed that, the "multidimensional combination and interaction of input variables and process parameters" is a multidimensional space (or subspace) whose dimensions are the factors used during the method development.

Afterwards, the guideline mentions "that have been demonstrated to provide assurance of quality" which means that the size of this space is defined by the set of combinations of factors ranges wherein a process provides quality results. In the LC field, and more precisely when the separation optimization is the main aim, the DS predicts a space wherein the separation is achieved taking into account the measurements, process and models uncertainties.

The separation quality can be evaluated by a chromatographic criterion such as the resolution ($R_S = 2\cdot(t_{R,2} - t_{R,1})/(w_{b,1} + w_{b,2})$; with $t_{R,2} > t_{R,1}$ and $w_{b,1}$, $w_{b,2}$ are the peaks widths at baseline; $t_{R,1}$ and $t_{R,2}$ being the retention times of the critical pair peaks) and the method uncertainty is estimated by the probability to reach a given criterion threshold (e.g. the probability for R_S to be higher than 1.5) in future uses of the method. It also states that "working within the design space is not considered as a change".

Therefore, in the framework of separation methods development, DS can be clearly considered as a zone of theoretical robustness since modifications of the method parameters will not significantly affect the separation quality. Thus, no decrease of quality in the separation should be observed while working in DS. Consequently, to optimize HPLC separations and advisedly compute DS, DoE is one of the most appropriate strategies. Examples of DS for HPLC assays can be found in references [40,41].

Thus, based on DoE strategy, a recent methodology [26] was used to model and predict the retention times according to selected chromatographic factors (e.g. mobile phase composition, gradient time, mobile phase pH) and subsequently optimize the separation. Afterwards, the prediction error was also estimated in order to allow the DS computation.

However, DoE involves the recording of chromatograms at very different operating conditions. By this way, very distinct selectivities are obtained. Peak detection and identification in each chromatogram can thus become tedious and time-consuming. Hence, an additional methodology based on independent component analysis (ICA) has been developed to detect and match peaks among the chromatograms resulting from DoE [42].

Two test samples were selected to evaluate these complementary methodologies. The first sample (sample 1) was sent from Eli Lilly and Company and was considered as an unknown sample. A similar test mixture was already used by Biswas et al. to conduct a method screening study [43]. However, in this work, the amount of compounds and their nature were deliberately occulted to test out both methodologies.

The second sample (sample 2) is a common-cold pharmaceutical formulation. Studies involving some of these compounds were previously published: the separation of the three active drugs using a poly(ethyleneglycol) column [44], the separation of the

three active drugs using cyano columns [45], the comparison between five HPLC columns [46], the comparison between electrophoresis and liquid chromatography separation [47] and the validation of a HPLC method for the quantification of the active drugs [48]. Furthermore, a review of analytical methods published for the separation of some of these substances can be found in the works of Marín et al. [48]. These abundant publications highlight the interest that still remains in separating and quantifying these compounds. Nevertheless, as the sweeteners of a new sugar-free formulation were not taken into account in the aforementioned methods, a new method development was necessary.

2.2. Materials and methods

2.2.1. Chemicals

Methanol (MeOH) (HPLC grade) was purchased from Sigma (St-Louis, MO, USA). Ultra–pure water was obtained with a Millipore (Billerica, MA, USA) Milli-Q Academic A10. Formic acid (>98%) was purchased from Merck (Darmstadt, Germany), ammonium formate (99%) was purchased from Alfa Aesar (Karlsruhe, Germany) and ammonium hydrogen-carbonate (99.7%) was purchased from VWR (Fontenay-sous-Bois, France). The unknown sample (sample 1) was provided by Eli Lilly & Co (Indianapolis, IN, USA) and was ready for injection. The pharmaceutical formulation (sample 2) came from the Cinfa Laboratory (Huarte, Spain). It consisted in a common-cold formulation which contained phenylephrine hydrochloride, acetaminophen, chlorpheniramine maleate, sodium saccharin, sodium cyclamate, mannitol, orange flavor, sunset yellow FCF and polyvinylpyrrolidone (PVP-K30).

2.2.2. Sample preparation

Sample 1 – unknown sample

This sample was provided in solution which was a mixture of water-acetonitrile (MeCN) (50:50, v/v). The solution was then filtered with a 0.20 µm syringe filtration disk to a vial for injection in the HPLC system. The injection volume was 0.5 µL. The rather low injection volume was selected to prevent any peak distortion such as peak fronting or retention times shift due to high elution strength of this solubilization mixture (i.e. 50% of acetonitrile) compared to the elution strength at the beginning of the gradient tested in the DoE (i.e. 5% of methanol). In addition, for 0.5 µL injections, injected volume repeatability was tested and the resulting estimated precision was very good (CV% = 0.5%).

Sample 2 – pharmaceutical formulation

800 mg of the formulation powder were dissolved in a 10 mL volumetric flask with a mixture of water-acetonitrile (80:20, v/v). After 10 min of magnetic stirring an aliquot was filtered with a 0.20 µm syringe filtration disk to the vial for injection in the HPLC system. In accordance with the theoretical formulation content, the concentrations of the active drugs were: acetaminophen (13 mg mL^{-1}), phenylephrine (0.16 mg mL^{-1}) and chlorpheniramine (0.06 mg mL^{-1}). The injection volume was 1.0 µL.

2.2.3. HPLC experiments

Both separations were carried out on an Alliance 2695 separation module coupled with a UV-DAD 2996 detector from Waters (Milford, MA, USA). The analytical column was an XBridge C18 (100x2.1mm i.d.; particle size 3.5 µm) from Waters. A C18 column was selected because this kind of columns usually offers relatively good retention properties for most organic compounds generally having a hydrophobic character. XBridge columns also sustain highly acidic and alkaline conditions (pH 1–12) thanks to the trifunctionally bonded C18 and the ethylene bridges within the silica matrix. In order to keep suitable column lifetime, the pH range was slightly narrowed from 2.6 to 10 (see Table I). The experiments were carried out at a flow rate of 0.25 mL min^{-1} at 30°C. The buffers consisted in 10 mM pH 2.6 formic acid, pH 4.45 ammonium formate, pH 6.3 ammonium formate, pH 8.15 ammonium hydrogencarbonate and pH 10.0 ammonium hydrogencarbonate. The pH was adjusted to the selected value with concentrated formic acid or ammonia 35% aqueous solution. The buffers concentration (10 mM) was adjusted to avoid pH modification and to minimize modification of the chromatographic behaviour due to the change of buffer composition. Linear gradients were carried out from 5% to 95% of MeOH and an isocratic step at 95% was hold during 10 min before column equilibration for the next injection. This final isocratic step was kept this long to avoid compound elution after the gradient and to ensure compound detection in a given time window for most unknown mixtures. If a mixture composition is known, the final isocratic step could obviously be reduced by a factor 5.

Table I. Full factorial design used in the present study.

Factors	Levels				
pH	2.6	4.4	6.3	8.1	10.0
T_G (min)	10	20	30		

All the chromatograms were recorded between 210 nm and 400 nm with an estimated step of 1.2 nm and an acquisition frequency of 2 Hz. The peaks integration

was carried out at 280 nm and 215 nm for sample 1 and 2, respectively. For coeluting peaks, the integrations were made on their respective independent components resulting from ICA numerical separation.

2.2.4. HPLC robust separation optimization

2.2.4.1. Responses and criteria

The first critical step is the selection of the response. Indeed, the choice of the factors directly depends on this response which also needs to be modelled by a multiple linear function. The use of resolution (R_S) of the critical pair (i.e. the two most proximate peaks in a chromatogram) as the response of the model can be hazardous due to its non-linear and discontinuous behaviour [26]. Thus, the retention times at the beginning, the apex and the end of each peak (respectively t_B, t_R, t_E) were measured and were transformed into retention factors ($k_{tR} = (t_R-t_0)/t_0$), with t_0 equal to the column dead-time. For each compound, the modelled responses were the logarithm of the retention factor (k_{tR}) and the logarithm of the semi-widths (i.e. $w_l = k_{tR} - k_{tB}$ and $w_r = k_{tE} - k_{tR}$). Based on these responses, the computation of the predicted value of some chromatographic criteria (resolution, total analysis time, etc.) only involved simple parameters (i.e. t_B, t_R and t_E). As the further computation of error propagation can be delicate when R_S is selected as a criterion (see section 2.2.4.5.), a simpler one was introduced: the separation criterion, S, defined as the difference between t_B of the second peak and t_E of the first peak of the critical pair. Even if S and R_S are highly correlated, the computation of S and its associated uncertainty is easier. Regardless of the critical pair peaks size ratio, a value of S equal or higher than 0 means that the peaks are baseline resolved which is similar to a value of R_S approximately equal or higher than 1.5.

2.2.4.2. Experimental factors, ranges and levels

The choice of the factors and their respective range is primordial. These factors have to affect the selected response in order to modify the selectivity enough to achieve a total separation of all the compounds. Furthermore, the investigated experimental domain (i.e. the factors ranges multidimensional space or workspace) needs to be large enough to create a response variation which allows encountering acceptable criteria values (i.e. $S > 0$ or $R_S > 1.5$). As the response variation is not known a priori, the selection of a large experimental domain minimizes the risk of not finding any good separations. Furthermore, the number of levels determines the polynomial order (or polynomial degree) of the multi-linear equation used to model the response. For example, if two levels are selected, the modelling equation can only be linear. Generally, if no prior information about the response variation is known, preliminary experiments should be carried out to estimate the range and the order of variation of each factor.

In the present study, the selected factors were the pH of the aqueous part of the mobile phase (pH) and the gradient time (T_G) to change the organic modifier proportion from 5% to 95%. In order to orthogonally modify the selectivity, it is common use to change the column or the organic modifier. However, these qualitative factors (i.e. which can only take some given values) considerably increase the number of experiments in DoE. Hence, selectivity modification was mainly based on pH effect. Besides, even if T_G did not affect the selectivity as considerably as pH, this factor was selected to mainly adjust the retention times in order to get acceptable analysis times. This consideration is of first importance because the ICA-DoE-DS methodology is intended to be as generic as possible.

Generally, one can observe a sigmoidal shape of the theoretical relationship between t_R and pH for monoprotic molecules. So five pH levels were selected to use a polynomial order of four (i.e. until pH^4). Furthermore, three levels of gradient time were chosen to estimate its quadratic effect (i.e. T_G^2).

2.2.4.3. Choice of design

The selected DoE needs to have good statistical properties as orthogonality and/or rotatability and to maintain the number of experiments as low as possible. DoE can be split up in two categories. Screening designs allow estimating the factors effect on the selected response. When too many factor seems to affect the response (i.e. four factor or more), screening designs are mainly used to select factors that have the higher effects (i.e. the factors that create the higher response variation). In this category, a well-known design is the Plackett and Burman design. In liquid chromatography, Plackett and Burmann are also used to estimate the robustness of an optimal separation [49-53]. The second category corresponds to optimization designs (i.e. used to model the response and to optimize response and/or criteria). These DoE are mainly full factorial (i.e. complete combination of all factors levels), fractional factorial designs (i.e. a statistically selected subpart of full factorial design), D-optimal (i.e. tend to minimize the parameter estimates covariance then maximizing the D factor) and central composite designs (i.e. tend to place the experiments on a circular/spherical shape around the experimental domain centre to obtain a good rotatability property) [54]. Nowadays, several softwares can be used to determine the experimental designs according to the analyst's choices.

In the present study, one full factorial design was repeated for each sample (see Table I). The central operating condition (i.e. pH 6.3 – T_G=20 min) was further independently repeated twice (i.e. preparation of new buffers and mobile phases). The DoE thus counted a total of 17 experiments including 2 independent repetitions at the centre (i.e. the central point was performed in triplicates as usually recommended).

2.2.4.4. Response modelling

The response modelling is achieved by fitting a multiple linear equation (i.e. polynomial model) whose mathematical terms depend on the selected factor level number (e.g. linear for two terms, quadratic for three terms). Nevertheless, at least one degree of freedom must theoretically be kept to estimate the error. Practically, two or more degrees of freedom are kept to accurately estimate this error. For example, for a simple linear model, two experiments are mandatory to estimate the equation parameters but a third one is needed to approximate the model error which depends on parameters error. On the other hand, the addition of too much high order terms or crossed terms can lead to overfitting (i.e. noise/random error modelling). For this purpose, stepwise regression was applied to maximize the adjusted coefficient of determination (R^2_{adj}) [30]. In the present case, two factors were selected (pH and T_G) and the model used for multiple linear regressions for both samples is shown with Eq. 11.

$$\log(k_{tR}) = \beta_0 + \beta_1 \times pH + \beta_2 \times pH^2 + \beta_3 \times pH^3 + \beta_4 \times pH^4 + \beta_5 \times T_G + \beta_6 \times T_G{}^2 + \beta_7 \times pH \times T_G + \varepsilon \quad \text{Eq. 11}$$

with $\beta_0 \ldots \beta_7$ the model parameters and ε the estimated error.

In order to evaluate the models adequacy, lack of fit tests were also performed. Lack-of-fit tests consist in comparing experimental and predicted variances. A normal zero-centred residuals distribution was already a good clue for adequate models (see section 2.3.1.2. and 2.3.2.2.). χ^2 and anova tests were performed and confirmed the models adequacy (i.e. p-values > 0.05).

At this step, it is also important to mention that the models could be used to highlight or to confirm some information about the physico-chemical properties (e.g. pKa, logP). However, this kind of interpretations was not the aim of the present study. The main objective was to demonstrate the applicability of the ICA-DoE-DS methodology for robust LC method optimization.

2.2.4.5. Quality criteria, error propagation, design space computation and robustness estimation

Following the ICH Q2(R1) guideline [55], "the robustness of an analytical procedure is a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters". In a separative framework, it means that the criteria – that quantify the separation quality – will remain unchanged around the optimal separation area. Robustness can be evaluated after the optimization process [49,50]. Using DoE-DS methodologies, robustness assessment is concurrently performed during the optimization phase. The error was assumed normal and was defined by a Gaussian law as shown in Eq. 12.

$$\log(k_{tR,i}) = \log(\hat{k}_{tR,i}) + N(O,\sigma_i) \quad \text{Eq. 12}$$

with $\log(k_{tR,i})$, the estimate distribution of the ith response (corresponding to the ith compound); $\log(\hat{k}_{tR})$, the mean predicted value for the ith modelled response and $N(0,\sigma i)$, a normal distribution centred on 0 and with a standard deviation equal to σi obtained as computed by the ith model. In practice, the Monte-Carlo method [56] was then used to propagate the Gaussian estimation of $\log(k_{tR,i})$ error to the separation criterion to assess robustness. For this purpose, 2500 simulations were carried out. Prediction error computation was carried out in a similar way for w_l and w_r. The distribution of $t_{R,i}$ (i.e. the distribution of t_R for the ith compound) can be obtained following Eq. 13. The distribution of $t_{B,i}$ and $t_{E,i}$ can also be obtained by Eq. 14 and 15, respectively.

$$t_{R,i} = (\exp(\log(k_{tR,i})) + 1 \times t_0 \quad \text{Eq. 13}$$

$$t_{B,i} = t_{R,i} - \exp(\log(w_l)) \quad \text{Eq. 14}$$

$$t_{E,i} = t_{R,i} + \exp(\log(w_r)) \quad \text{Eq. 15}$$

Finally, the error was propagated to the separation criteria, S, with Eq. 16.

$$S = t_{B,2} - t_{E,1} \quad \text{Eq. 16}$$

Where, S is the estimated distribution of the separation criterion, $t_{E,1}$ is the end of the first peak of the critical pair and $t_{B,2}$ is the begin of the second peak of the critical pair, considering that this computation was carried out at a given operating condition to simplify notations. Similar computations can also be used to find the distribution of R_S or other selected quality criteria.

If and only if the criteria error is estimated, the computation of the DS can be done. If R_S is selected as criterion, the division taking place in its calculation makes its error more complex than for S which is computed by a simple subtraction [26]. Following the ICH Q8(R2) guideline, the DS can be defined as the set of experimental conditions where the criterion meet acceptance limits (e.g. S > 0.2 min) with a given probability, known as the quality level (e.g. π = 60%), as shown with Eq. 17.

$$DS = \left\{ x_0 \in \chi \middle| E_{\hat{\theta}} \left[P(S > \lambda | \hat{\theta} \right] \geq \pi \right\} \qquad \text{Eq. 17}$$

Where x_0 is a point in the experimental domain, χ. λ is the acceptance limit for criterion S, π is the quality level and is the set of estimated parameters of the model. P and E respectively correspond to the estimators of probability and mathematical expectation. The DS is thus assumed to be a zone of theoretical robustness as the predicted criterion values are higher than an advisedly selected acceptance limit (threshold) together with a high probability to at least reach this acceptance limit. In the present study, π was selected as 90% of the optimal probability to get S > 0 min. For instance, for sample 1, P(S > 0) at the optima was equal to 94.4%. The quality level was settled at 85% (i.e. 94.4%×90%) (see section 2.3.1.3. and fig. 5). On one hand, a low π does not mean that unacceptable separations could be obtained (i.e. S < 0). The peak could be well separated concurrently with a relatively high error tarnishing the t_R predictions leading to predictions of low probability while peaks could be well separated. On the other hand, a high π could induce obtaining a good separation.

Up to now, it is important to stress that there is still no regulatory documents stating how to conveniently compute or estimate (nor define) the required quality level. Nevertheless, ICH Q8(R2) remains the most appropriated guideline when discussing about DS [20].

2.2.5. Automatic reading of chromatograms

Before modelling responses, the retention times (t_B, t_R and t_E) have to be obtained for each peak in each tested experimental conditions. Three successive steps can be identified: peak detection (i.e. effective presence of peaks in the chromatogram, corresponding to the investigated compounds) which is not obvious when coelutions are present, peak matching (i.e. concordance between the detected peaks in a given

chromatogram and the peaks of another chromatogram obtained in another analytical condition of DoE) and peak identification (i.e. attribution of a given peak to a clearly identified compound). The chromatograms coming from DoE depict various selectivities and therefore complicate peak detection and peak matching. This step is highly time-consuming and its automation can be very useful.

ICA is a mathematical process which maximizes the statistical independence between estimated independent components. ICA has already been used on UV-DAD chromatograms [42] and further details can be found in the works of Hyvärinen et al. [57,58]. In the present study, ICA numerical separation was carried out on each chromatogram to accomplish the integration on the independent components rather than on the original peaks that can result from coeluting compounds. Indeed, when two compounds coelute, the determination of the integration limits is biased. If the coelution is slight ($1.0 < R_S < 1.5$; $S < 0$), the integration limits are usually obtained using a vertical separator at the minimum of the valley between the two peaks (also called the valley drop-line). This approximation restricts the accuracy of this measurement and prevents obtaining mathematical model that reflects, as accurately as possible, the chromatographic behaviour of compounds in the sample. In the case of a major coelution ($0 < R_S < 1$; $S << 0$), estimation of peak widths is highly biased. These inaccurate integration limits generate inaccurate predictions on the whole experimental domain. Therefore, the limits of integration (i.e. the beginning and the end of a peak, t_B and t_E) were automatically determined on the independent components. Starting from the top of a peak and moving left or right, the first point whose height is less than twice the noise was respectively selected as the beginning or end of the peak. By this way, ICA methodology also provides an accurate estimation of the peak number in the sample and ease peak matching [42].

2.2.6. Software

In-house computer software were developed to run the ICA algorithm (i.e. fastICA R package) and to perform the multiple linear regressions. The coding was carried out with the academic version of Revolution R 1.3.0 (including R 2.7.2), freely distributed at http://www.revolutionanalytics.com/.

2.3. Results and discussion

To avoid column equilibration at each mobile phase pH for each injection, experiments at a same pH were carried out in a row. Experiments within a pH level were however conducted in a random order. In fact, it can be assumed that the mean relative error performed during buffer pH measurements is about 0.3% (i.e. corresponding to 3 times the specified pH-meter precision). Therefore, the retention time variability component coming from pH measurement can be considered as negligible – in the framework of the present feasibility study – in comparison to other variability sources coming from the LC equipment (e.g. t_R variability coming from gradient slope or mobile phase composition, etc.). Examples of chromatograms for both samples are depicted in fig. 2.

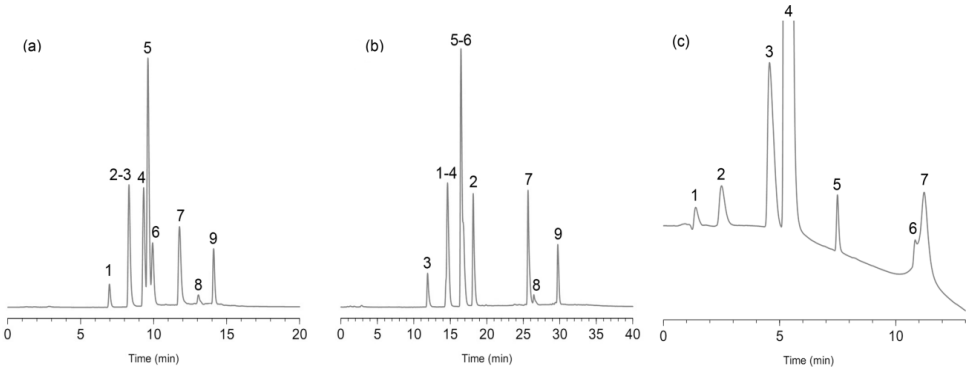


Fig.2. Examples of chromatogram obtained during the experimental design. (a) Chromatogram of sample 1 at pH 8.1 and T_G = 10 min displayed at 280 nm, (b) chromatogram of sample 1 at pH 10 and T_G = 30 min displayed at 280nm and (c) chromatogram of sample 2 at pH 6.3 and T_G = 10 min displayed at 215 nm. Sample 2 peak numbering: 1 =maleate, 2 = phenylephrine, 3 = saccharin, 4 = acetaminophen, 5 = sunset yellow FCF, 6 = chlorpheniramine, 7 = PVP-K30.

2.3.1. Sample 1 – unknown sample

As sample 1 was unknown, peaks were numbered in accordance with their elution order for the first tested experimental condition (i.e. pH 6.3 – T_G=10 min).

2.3.3.1. Automatic reading of chromatograms

When the experimental design was carried out, a maximum of nine peaks was observed. At this stage, sample 1 should then contain at least nine compounds. Fig. 3 presents two examples of numerical separation of coeluting peaks obtained when carrying out the experimental design. ICA numerical separation also confirmed the presence of nine compounds in sample 1. Finally, the retention times (i.e. t_B, t_R and t_E) were automatically obtained for each independent component.

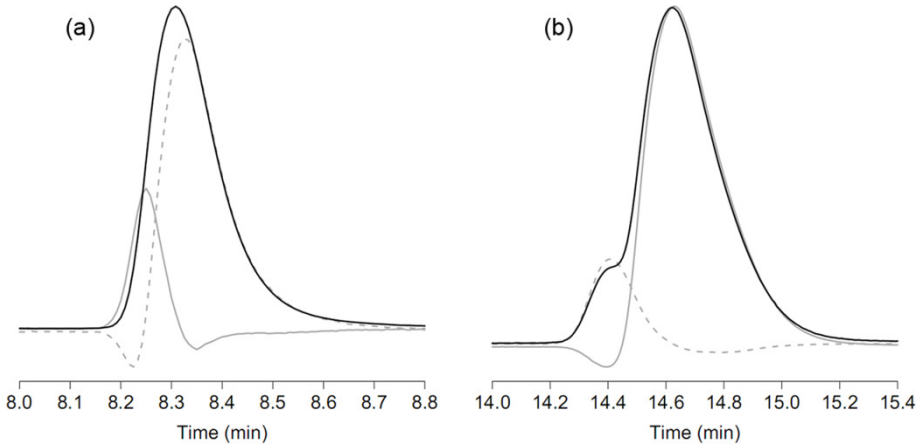


Fig. 3. Examples of a numerical separation of coeluted peaks carried out by ICA. The black line represents the recorded signal. (a) Numerical separation of compound 2 (dashed line) and compound 3 (grey line) for the chromatogram depicted in fig. 2a. (b) Numerical separation of compound 1 (dashed line) and compound 4 (grey line) for the chromatogram displayed in fig. 2b.

2.3.1.2. Responses modelling and criterion computation

In order to assess the polynomial fit quality, the adjusted coefficients of determination (R^2_{adj}) are presented in Table II. The computed R^2_{adj} resulted from the models (see Eq. 11) rather than the correlation between experimental and predicted retention times on fig. 4a.

Table II. Adjusted coefficient of determination (R^2_{adj}) and residuals standard deviation (σ) for the logarithm of the retention factors ($\log(k_{tR})$), for the left half-width ($\log(w_l)$) and for the right half-width ($\log(w_r)$) for sample 1.

Compound N°		1	2	3	4	5	6	7	8	9	Mean
R^2_{adj}	$\log(k_{tR})$	0.991	0.993	0.999	0.997	0.999	0.998	0.999	0.998	0.999	0.997
	$\log(w_l)$	0.901	0.801	0.453	0.959	0.643	0.797	0.935	0.512	0.303	0.700
	$\log(w_r)$	0.955	0.696	0.444	0.780	0.942	0.833	0.768	0.491	0.372	0.698
σ	$\log(k_{tR})$	0.023	0.020	0.003	0.016	0.008	0.011	0.007	0.004	0.006	0.011
	$\log(w_l)$	0.117	0.220	0.450	0.196	0.213	0.172	0.093	0.279	0.173	0.212
	$\log(w_r)$	0.296	0.318	0.414	0.553	0.330	0.363	0.388	0.189	0.205	0.339

Additionally, the adequacy between the retentions times predicted by the models and those observed also allowed confirming fits quality (fig. 4). This latter figure clearly illustrates the good relationship between the experimental responses and the predicted ones as well as the corresponding residuals. In addition, residuals Gaussian distributions also corroborated the quality of the multiple linear regressions and obtained models. For each response, a Shapiro-Wilk test was carried out on the residuals distribution to verify its normality. The p-values were equal to 0.111, 0.129 and 0.080 for $\log(k_{tR})$, $\log(w_l)$ and $\log(w_r)$, respectively. The residuals were thus normally distributed (p-values > 0.05). Once responses were correctly modelled, the separation criterion computation was carried out. Using a grid search method, the separation criterion and its corresponding error were computed over the whole experimental domain. Furthermore, using Monte-Carlo simulations, the probabilities for S to be greater than 0 min (i.e. baseline resolved peaks) were also calculated as it can be seen in fig. 5.

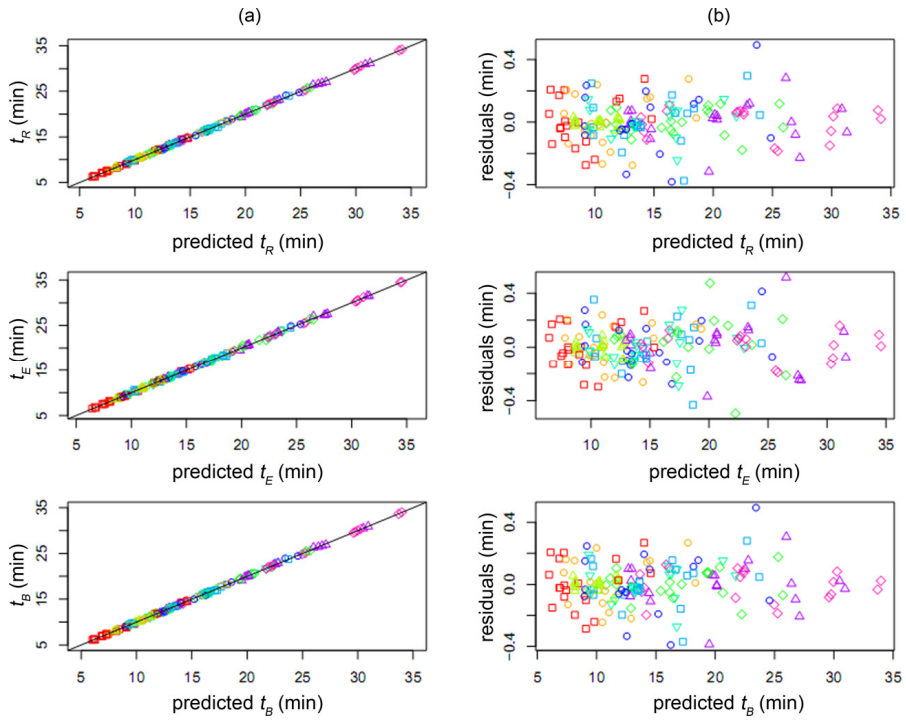


Fig. 4. Modelling results for sample 1. (a) Predicted versus experimental values for t_R, t_E and t_B. (b) Corresponding residuals plots. Compound assignation: (red circle) compound 1, (orange circle) compound 2, (green triangle) compound 3, (blue circle) compound 4, (blue triangle) compound 5, (blue square) compound 6, (green diamond) compound 7, (purple triangle) compound 8 and (pink diamond) compound 9.

2.3.1.3. Optimal separation

One optimal experimental condition was identified at pH 3.1 and with TG=30 min (fig. 5, point 1). Fig. 6b shows the chromatogram recorded at this optimal condition. The predicted S was 0.34 min (i.e. predicted R_S = 2.06) and the observed one was -0.09 min (i.e. observed R_S = 1.23). Some reasons are behind the difference between predicted and experimental values. First, among the modelled responses, the logarithms of peak half-width were those that offer the lowest accuracy in prediction (see Table II). The modelled half-widths allowed calculating the beginning and end of the peak with the value of retention time, which itself was affected by its error. Thus, the observed errors for t_B and t_E have a double source. It depends on the error of the retention factor and the specific half-widths errors. Moreover, S was directly calculated

from the beginning and end times of two peaks, its value is also tarnished (though less than R_S) with a double error coming from the modelling of t_R and w_l or w_r. Secondly, as observed in fig. 3, independent component coming from coeluted peak can show a local negative baseline on the side where coelution was initially observed. This can be clearly observed on the independent components in fig. 3a. This deformation generates a relatively small underestimation of peak widths, resulting in a relatively small overestimation of predicted separation.

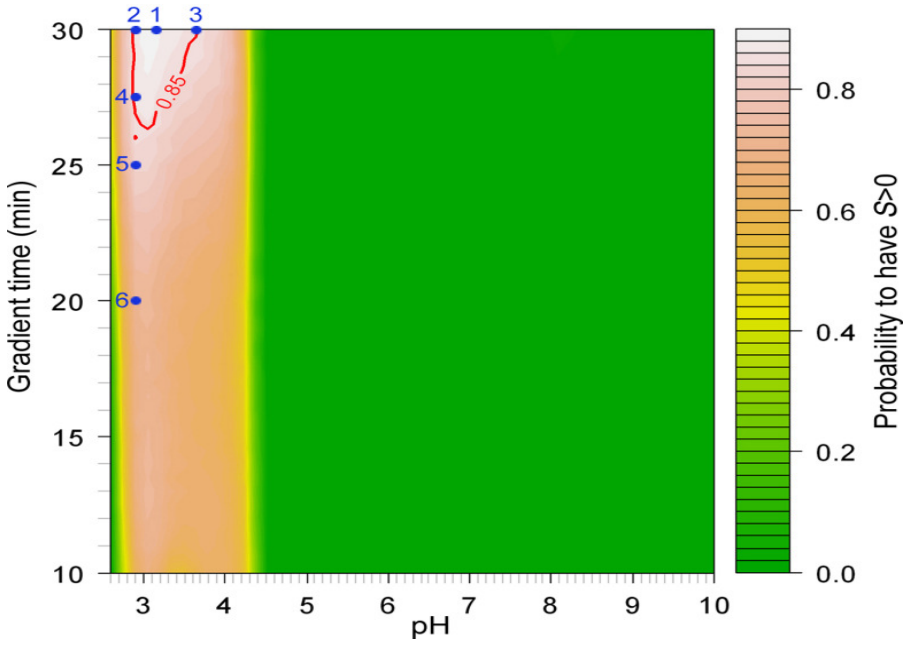


Fig. 5. Probability surface for the separation criterion (acceptance limit λ=0 min) with the design space hedged-in the black line (π=85%). Blue dots depict additional tested experimental conditions.

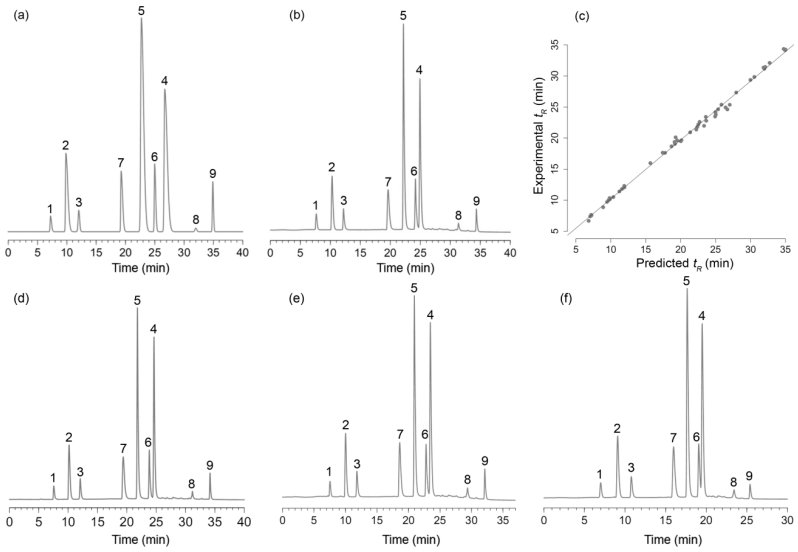


Fig. 6. (a) Predicted chromatogram at the optimal condition 1 (pH 3.1 and T_G = 30.0 min), (b) chromatogram recorded at the optima condition 1 (pH 3.1 and T_G = 30 min). (c) Relationship between experimental retention times versus predicted ones for the nine peaks and for the six tested experimental conditions. The black line is a simple line with 0 intercept and slope of 1. (d) Chromatogram recorded at condition 3 (pH 3.7 and T_G = 30.0 min), (e) chromatogram recorded at condition 4 (pH 2.9 and T_G = 27.5 min) and (f) chromatogram recorded at condition 6 (pH 2.9 and T_G = 20.0 min).

Work is underway to overcome this problem. Nevertheless, the predictions are very accurate on average. The correlation coefficient between the predicted retention times (i.e. t_R) and observed ones is 0.998 as reported in fig. 6c. The chromatogram recorded at the optimal operating condition (see fig. 6b) also confirms that the slight coelution existing between compound 4 and 6 do not prevent the quantitative determination of these two compounds.

2.3.1.4. Design space robustness assessment

As it can be seen in fig. 5 (points 2-6), five experimental conditions were selected to verify models ability to predict and to assess the robustness. The experimental separation criteria, S, measured in experimental conditions 2, 3 and 4 were not significantly different from the separation measured at the optimal condition (point 1) chromatograms. Even if the corresponding chromatogram present slight coelution between compound 4 and 6 (S ~ 0 min), the observed separations did not

change significantly within DS. On the other hand, chromatograms were recorded at experimental condition numbered 5 and 6 (i.e. outside the DS, see fig. 5) in order to confirm robustness in DS. In both of them, the coelution of compound 4 and 6 was so important that their accurate quantification would be impossible (see fig. 6f).

Compounds of sample 1 were first matched using their UV spectrum (i.e. each peak was labelled with a number). Secondly, Eli Lilly & Co provided peak identification with structures and names of labelled peaks. Originally numbered from 1 to 9, the compounds were respectively, atenolol, pindolol, a licensed compound, warfarin, indoprofen, naproxen, propranolol, an impurity of retinoic acid and retinoic acid.

2.3.2. Sample 2 – pharmaceutical formulation

As sodium cyclamate, mannitol and orange flavor do not absorb in the usable UV range (above 200 nm), they were not taken into account in the present optimization process. Also, sunset yellow FCF does not appear in the following sections as it did no absorb UV at acidic pH (at pH 2.6). Peak integration was carried out by the same operator using ICA in order to avoid additional error sources. For each chromatogram, the peaks were integrated and identified using their UV spectra. Then, the measured values of t_B, t_R and t_E were stored in a file for the further computational process.

2.3.2.1. Automatic reading of chromatograms

When coeluted peaks were observed, ICA was used to numerically separate the peaks to avoid a poor estimation of integration limits. Then, the integrations were carried out on the respective components resulting from the ICA numerical separation. Compounds were manually identified thanks to the comparison between their UV spectra and reference UV spectra [59]. These identifications were also confirmed by the injections of individual solution of each compound.

2.3.2.2. Responses modelling and criterion computation

On one hand, the adjusted coefficients of determination (R^2_{adj}) were computed in order to assess the polynomial fit quality. As shown in Table III, the logarithm of the retention factors were well modelled since the adjusted R^2_{adj} were higher than 0.98. R^2_{adj} were computed from the models (see Eq. 11).

Table III. Adjusted coefficient of determination (R^2_{adj}) and residuals standard deviation (σ) for the logarithm of the retention factors ($\log(k_{tR})$), for the left half-width ($\log(w_l)$) and for the right half width ($\log(w_r)$) for sample 2. Peak numbering: 1=maleate, 2=phenylephrine, 3=saccharin, 4=acetaminophen, 6= chlorpheniramine, 7=PVP-K30.

Compound N°		1	2	3	4	6	7	Mean
R^2_{adj}	$\log(k_{tR})$	0.993	0.998	0.982	0.999	0.999	0.999	0.995
	$\log(w_l)$	0.524	0.914	0.444	0.841	0.906	0.995	0.771
	$\log(w_r)$	0.912	0.947	0.982	0.845	0.486	0.961	0.856
σ	$\log(k_{tR})$	0.027	0.016	0.013	0.013	0.010	0.002	0.014
	$\log(w_l)$	0.047	0.068	0.110	0.148	0.063	0.035	0.079
	$\log(w_r)$	0.079	0.089	0.044	0.094	0.268	0.082	0.109

On the other hand, fig. 7 illustrates the relationship between the experimental and the predicted responses as well as the corresponding residuals. The Gaussian distributions of the residuals also corroborated the quality of the multiple linear regressions and allowing accurate predictions of retention times. As previously, a Shapiro-Wilk test was performed on the residuals distribution to verify the response normality. The *p*-values were equal to 0.056, 0.062 and 0.101 for $\log(k_{tR})$, $\log(w_l)$ and $\log(w_r)$, respectively. The residuals distributions were thus considered normal (*p*-values > 0.05).

2.3.2.3. Optimal separation

Two optimal separation conditions were encountered. The first one (optimum n°1 at pH 4.3 and with T_G = 29.5 min and the second one (optimum n°2) at pH 10.0 and with TG = 10.0 min. When an acceptance limit λ = 0.1 min for S was selected, the quality level had to be adjusted to 40% to find DS. At this value, two DS were found around each optimum as seen on fig. 8a and 8b.

The chromatograms recorded at optimal conditions 1 and 2 are presented in fig. 9a and 9b, respectively. To test the precision of these operating conditions, three chromatograms were independently recorded with new buffer solutions at each optimal experimental condition. The repeatabilities of retention times were very good with CV (%) values of 0.6% and 0.8% for optimum 1 and 2, respectively. Fig. 9a shows that saccharin (compound 3) and acetaminophen (compound 4) were slightly coeluted.

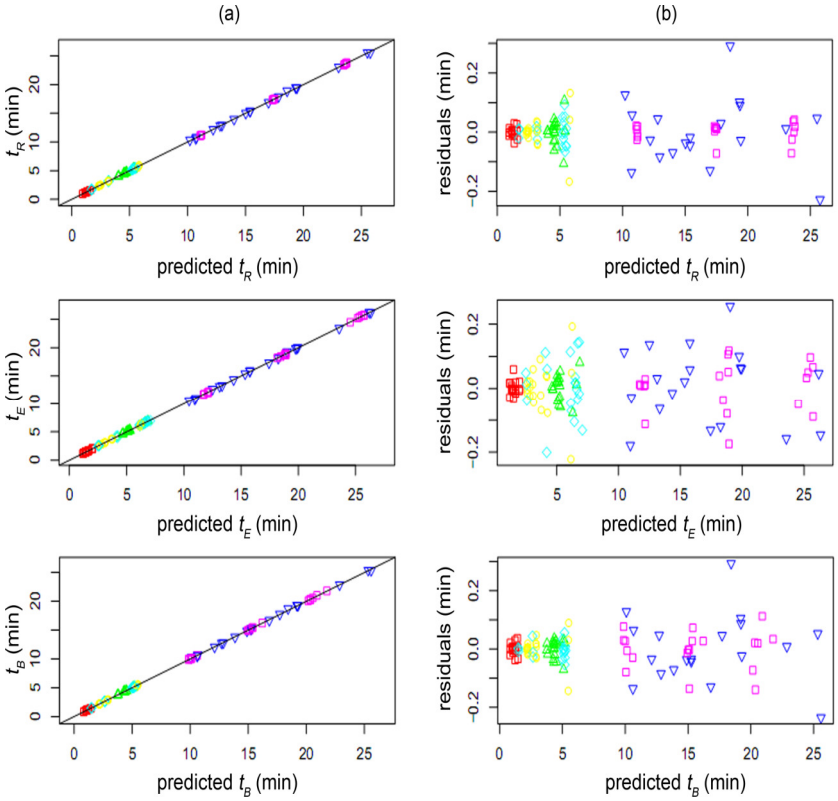


Fig. 7. Modelling results for sample 2. (a) Predicted versus experimental values for t_R, t_E and t_B. (b) Corresponding residuals plots. Compound assignation: (red square) maleate, (yellow circle) phenylephrine, (green triangle) saccharine, (blue diamond) acetaminophen, (blue triangle) chlorpheniramine and (pink square) PVP-K30.

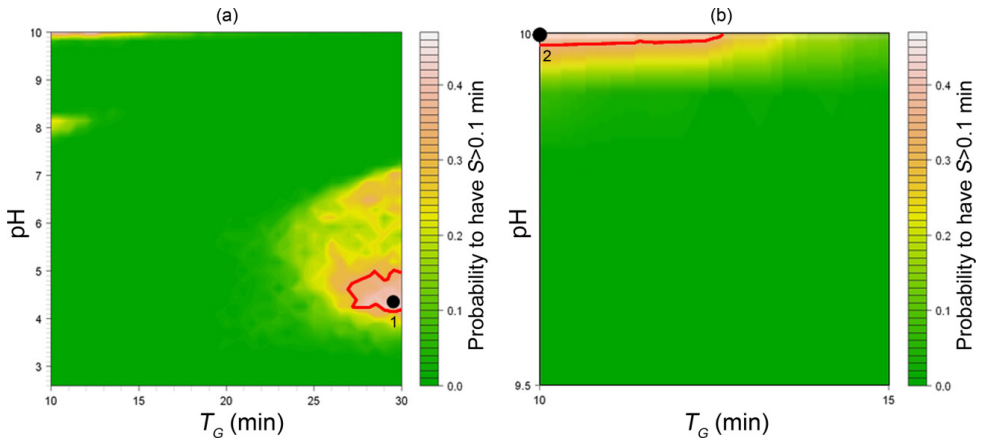


Fig. 8. (a) Probability surface for the separation criterion (acceptance limit λ = 0.1 min) with the design space 1 (π = 40%) hedged-in the red line; optimum 1 is depicted by the black circle (pH 4.3 – T_G = 28.0 min). (b) Magnification of the upper left zone of the probability surface with the design space 2 (π = 40%) hedged-in the red line, optimum 2 is depicted by the black circle (pH 10 – T_G = 10 min).

As the probability of the prediction to get S > 0.1 minute was only of 40%, this small coelution (S < 0) was not surprising. On the contrary, the chromatograms recorded at optimum 2 (fig. 9b), the separation S of the critical pair was higher than 0.1 minute and seemed to be robust as a significant variation of S was not observed during the independent repetitions.

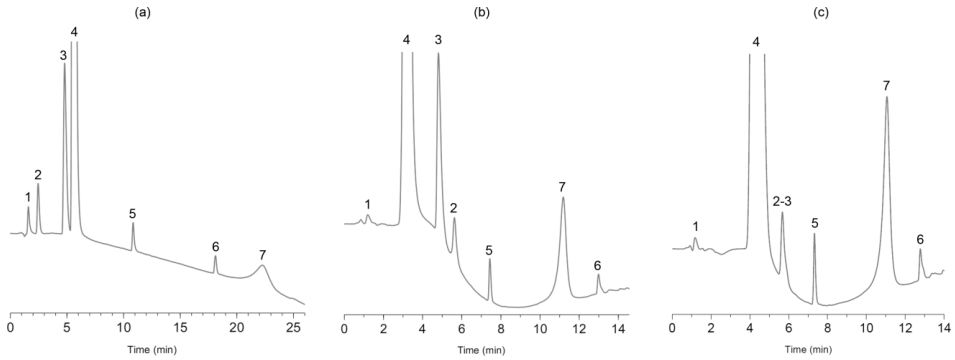


Fig. 9 (a) Chromatogram recorded at optimum 1 (pH 4.3 – T_G = 28.0 min). (b) Chromatogram recorded at optimum 2 (pH 10.0 – T_G = 10 min) (c) Chromatogram recorded outside the DS as a counterexample (pH 9.5 – T_G = 10 min). Peak numbering: 1 =maleate, 2 = phenylephrine, 3 = saccharin, 4 = acetaminophen, 5 = sunset yellow FCF, 6 = chlorpheniramine, 7 = PVP-K30.

2.3.2.4. Design space robustness assessment

The computed DS were large enough to evaluate the separation robustness. In the DS 1 (corresponding to optimum 1), three experimental conditions were tested, at pH 4.3 with T_G = 28.0 min and at pH 4.8 with T_G = 28.0 and 29.5 min. The measured separations, S, on the corresponding chromatograms were not significantly different from the initial measured separation (optimum 1). As for optimum 1, the separation criterion threshold (i.e. λ=0.1 min) was not reached. However, these separations were barely acceptable (S < 0 min but > −0.1 min) as the coelution between peak 3 and 4 was very slight. Moreover, the observed separations did not change significantly within DS involving an expected robustness. The shape of DS 2 (corresponding to optimum 2) was thinner, so only one experimental condition was added, at pH 10.0 and with T_G = 12.0 min, to evaluate the DS. In this chromatogram, a minor decrease of S was noticed but it was still higher than the selected acceptance limit (S > 0.1 min). Therefore, this DS defined a zone of acceptable robust separation. As previously carried out with sample 1, a counterexample was performed outside of the DS, at pH 9.5 and with T_G = 10.0 min (see fig. 9c). At this experimental condition, a large coelution was observed between saccharin (compound 3) and phenylephrine (compound 2). In addition, a decrease in saccharin peak size between fig. 9b and 9c was observed. It can be explained by the modification of the molar absorption coefficient with pH change.

2.4. Conclusions

The automated optimization of chromatographic separation is the first critical step in the framework of the automated development of chromatographic method. In this paper, DoE, ICA, multiple linear regression, error propagation and DS methodologies were successfully applied to separate nine compounds of an unknown sample mixture in less than 40 min and the seven compounds of a pharmaceutical formulation. Furthermore, for this latter sample, the analysis time was shortened to less than 14 min. This global methodology is also very flexible as the choice of each criterion and their respective acceptance limit are made by the analyst. In addition, an evaluation of DS robustness was carried out during the present study. The separation criterion S clearly demonstrated its robustness capability within the identified DS. It strengthens the fact that DS defines a space wherein the separations are complete and the method is robust and even more robust if the DS is large. Nevertheless, as the size of DS depends on the value of π, further works are still required to define possible adjustment strategy for this parameter. Eventually, if one compound is discarded from one of the tested mixtures, such as an active ingredient of the pharmaceutical formulation, the present DS would obviously have defined optimal and robust space for the separation of the subsequent mixture without any additional experiments.

Acknowledgements

The author would like to thank the Walloon Region of Belgium for the FIRST-DEI convention funds N°516130, the Eli Lilly and Company for partial funding of ADAM project and the fund Léon Fredericq for the grant allocated to his working period in Madrid. A research grant from the Belgium National Fund for Scientific Research (F.R.S-FNRS) to E. Rozet is also gratefully acknowledged. Bernard A. Olsen and Brian Fahie from Eli Lilly & Co are also gratefully acknowledged for their contribution to this work.

3. Application of a new optimization strategy for the separation of tertiary alkaloids extracted from *Strychnos usambarensis* leaves

The HPLC separation of six alkaloids extracted from *Strychnos usambarensis* leaves has been developed and optimized by means of a powerful methodology for modelling chromatographic responses, based on three steps, i.e. DoE, ICA and DS. This study was the first application of a new optimization strategy to a complex natural matrix. The compounds separated are the isomers isostrychnopentamine and strychnopentamine, 10-hydroxyusambarine and 11-hydroxyusambarine, also strychnophylline and strychnofoline.

Three LC parameters have been optimized using a multifactorial design comprising 29 experiments that includes 2 center point replicates. The parameters were the percentage of organic modifiers used at the beginning of a gradient profile which consisted in different proportions of MeOH and MeCN, the gradient time to reach 70 % of organic modifiers starting from the initial percentage and the percentage of MeCN found in the mobile phase. Subsequent to the experimental design application, predictive multilinear models were developed and used in order to provide optimal analytical conditions.

The optimum assay conditions were: methanol/acetonitrile-sodium pentane sulfonate (pH 2.2; 7.5 mM) (33.4:66.6, v/v) at a mobile phase flow rate of 1mL/min during a 40.6 minutes gradient time. The initial organic phase contained 3.7 % MeCN and 96.3 % MeOH.

The method showed good agreement between the experimental data and predictive value throughout the studied parameters space. Improvement of the analysis time and optimized separation for the compounds of interest was possible due to the original and powerful tools applied.

Finally, this study permitted the acquisition of isomers profiles allowing the identification of the optimal collecting period of *Strychnos usambarensis*.

3.1. Introduction

Malaria still remains the major parasitic infection, affecting hundreds of countries around the world, especially tropical regions in Africa, Asia and Latin America. The impact of the disease in terms of morbidity and mortality, as well as the ever increasing emergence of resistance to currently used treatments enhance the importance of developing new drugs.

The plant world represents a huge, varied, promising and a countless source of new potential therapeutic substances.

3.1.1. Alkaloids from *Strychnos usambarensis* leaves

Our study concerns Strychnos usambarensis Gilg (Loganiaceae). This small tree growing widely in Eastern Africa, particularly in Rwanda, is used in traditional pharmacopoeias. It is well known for its use as arrow poison due to quaternary curarizing alkaloids contained in the roots [60,61] but also for its tertiary alkaloids isolated from leaves because some of them possess antiplasmodial [62] and cytotoxic properties [63-66].

These activities could be related to their original structure: they are all asymmetrical monoterpenoid bisindolic usambarane-type alkaloids. The major compounds are isostrychnopentamine (fig. 10a) and its epimer strychnopentamine (fig. 10b), 10-hydroxyusambarine (fig. 10c) and its position isomer 11-hydroxyusambrine (fig. 10d), strychnophylline (fig. 10e), strychnofoline (fig. 10f) and its epimer isostrychnofoline (fig. 10g).

From a therapeutically point of view, isostrychnopentamine (ISP), and to a lesser extent, its isomer strychnopentamine (SP) seem to be the most promising substances [67,68]. ISP proved to be active in vitro and in vivo against Plasmodium falciparum with under μM concentrations.

In the recent years, ISP also showed a potential anti-tumor activity against apoptosis-resistant cancer cells [69,70].

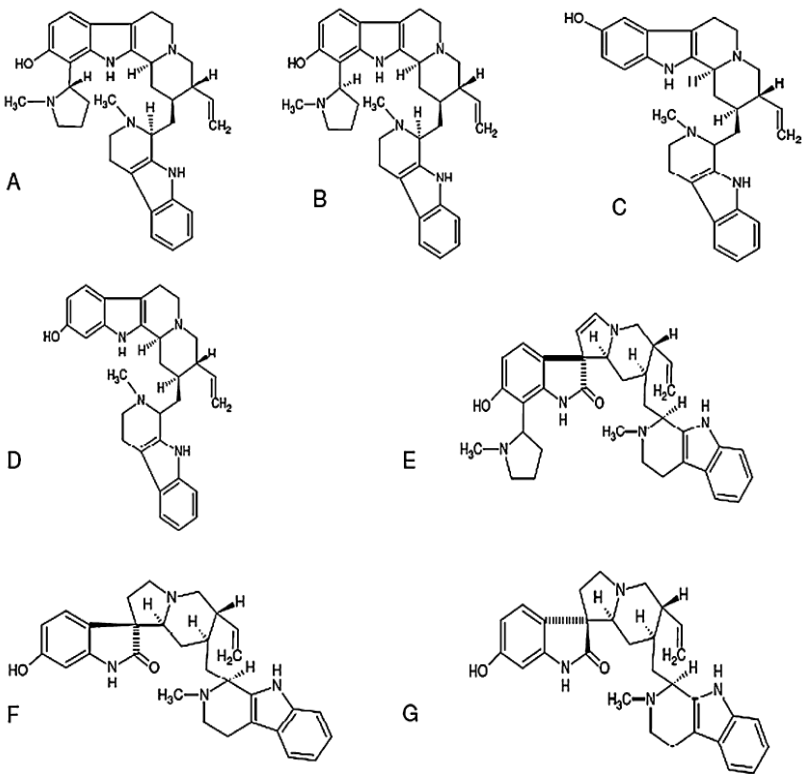


Fig. 10. Chemical structures of (a) isostrychnopentamine, (b) strychnopentamine, (c) 10-hydroxyusambarine, (d) 11-hydroxyusambarine, (e) strychnophylline, (f) strychnofoline and (g) isostrychnofoline.

3.1.2. Analytical and biostatistical methods

To our knowledge, this work is one of the first studies dedicated to the chromatographic separation of these tertiary alkaloids extracted from *Strychnos usambarensis* leaves, representing also the first application of a new optimization strategy to a complex natural matrix. Considering the complexity of the plant material, the difficulty lies in finding a method that could separate and identify all the structurally similar compounds (epimers, isomers).

Two LC procedures have been described in the 90s about alkaloids separation of *Strychnos* species. The first one exposed the isolation of monomeric indole alkaloids in *Strychnos nux-vomica* and *S. ignatii* seeds [71], and the second one, the separation of

bisindolic alkaloids in *S. usambarensis* roots [72]. Both were assessed by reversed-phase chromatography using MeCN, an ion-pairing reagent (sodium acetate and sodium salt of heptanesulfonic acid, respectively) and phosphate buffer (pH ~ 3.00).

The goals of the present study were to develop and optimize HPLC conditions for the separation of six target alkaloids from *Strychnos usambarensis* leaves that are strychnopentamine, isostrychnopentamine, 10–hydroxyusambarine, 11–hydroxyusambarine, strychnophylline and strychnofoline, respectively. To achieve this, three methodologies were combined.

Firstly, DoE was implemented to gather experimental data in order to achieve statistical modelling. The major advantage of using design of experiments to develop this method is that it allows all potential factors to be evaluated concurrently, systematically and quickly [32,34,39].

Secondly, a DS was built over the design of experiment domain to simultaneously optimize the chromatographic method separation and assess the robustness of its future use [26,73,74].

Finally, ICA methodology was used to facilitate peaks detection and identification even for co-eluted peaks [42]. Another aim was to reduce analysis cost by using MeOH as organic modifier rather than MeCN because at the time of the study, the market presented itself with a paucity regarding the last one.

Ion-pairing chromatography was used in this study as it is an efficient strategy to control the retention of protonised bases by reversed phase liquid chromatography (RP-LC) [75-78]. Additionally this new method showed to be precise and suitable to qualitatively identify different samples of *S. usambarensis* leaves collected over a period of three years in various seasons.

3.2. Materials and methods

3.2.1. Chemicals

Orthophosphoric acid 85 % (pro analysis), sodium carbonate, dichloromethane, methanol and acetonitrile of HPLC-grade were obtained from Merck (Darmstadt, Germany). The 1-pentanesulfonic acid (sodium salt monohydrate 99 %) was purchased from Acros Organics (Geel, Belgium). The ultra pure water was obtained with a Milli-Q system (Millipore Plus 185, Billerica, MA, USA).

3.2.2. Plant material

All five *Strychnos usambarensis* leaves batches were collected by two of the authors (M.F and L.A.) in Akagera National Park, Rwanda at different periods of time (June 2007, November 2007, August 2008 and 2009, February 2010). They were quickly air-dried in Rwanda before storage in Belgium at a temperature of 15°C. A voucher specimen of the plant was deposited in the herbarium of the Pharmaceutical Institute, at the University of Liege.

3.2.3. Extraction of the tertiary alkaloids from *S. usambarensis* leaves

For each batch of leaves, powdered dried material was macerated in sodium carbonate and then percolated with distilled ethyl acetate until total extraction of the alkaloids. The extract was acidified by acetic acid 1 % and washed by distilled dichloromethane to remove chlorophyll and pigments. The resulting acidic solution was basified to pH 8.0 using sodium carbonate and repeatedly extracted with distilled dichloromethane. The organic solvent was dried over sodium sulfate and concentrated to yield the alkaloid extract.

3.2.4. HPLC alkaloids solutions

For each sample, approximately 10 mg of the alkaloid extract was precisely weighed and dissolved in 1 ml of mobile phase, specifically a mixture of MeOH/ MeCN (96.3:3.7, v/v) and aqueous solution of pentane sulfonate (7.5 mM) (33.4:66.6, v/v) adjusted to pH 2.2 with orthophosphoric acid (0.5 %). These solutions were then filtered through a 0.45 µm HVLP filtration membrane into vials for injection in the HPLC system.

3.2.5. Instruments and conditions

The HPLC system used was a Waters 2695 separation module coupled to a Waters 996 Photodiode array detector. The chromatographic separation was performed on a LiChrospher® 60 RP-select B column (250 x 4 mm i.d.; particle size 5 µm) from Merck. The UV detector wavelength was set at 273 nm for the entire study. Temperatures of the column and of the samples were set at 25°C and 10°C, respectively. Injections of 5 µL of the alkaloids extracts were accomplished. pH measurements were performed with a SevenEasy Mettler-Toledo pH meter (Schwerzenbach, Switzerland) on aqueous eluent component before addition of the organic modifier.

In the optimized procedure, the mobile phase used for the separation was a 33.4:66.6 (v/v) mixture of an organic phase and an aqueous solution of sodium pentane sulfonate (7.5 mM) adjusted with orthophosphoric acid (0.5 %) to pH 2.2. The

organic phase, containing 3.7 % MeCN and 96.3 % MeOH was delivered at a flow rate of 1.0 mL/ min during a 40.6 minutes gradient time using gradient profile.

3.2.6. Design of Experiments, Independent Component Analysis and Design Space

To define the DoE only the factors that have a significant effect over the separation had to be selected as well as establishing their minimal and maximal limits. The DoE to be selected had to allow the modelling of response surfaces of individual analytes from a complex mixture applying a limited number of experiments in order to find the optimal analytical conditions [79].

Among the classes of experimental designs that are appropriate for obtaining data that will permit the estimation of the coefficients, of all main effects and interactions are the full factorial designs. A full factorial design was selected for this study given the complexity of the mixture and the lack of a priori data on the compounds chromatographic behaviour. It was also employed in order to enable a better understanding of the chromatographic separation and to allow estimation of interactions and quadratic effects.

The three factors selected as well as their three levels are given in Table IV. They are:

- The percentage of organic modifiers (OM) mixture used at the beginning of the gradient profile consisting of different proportions of MeOH and MeCN,
- The gradient time to reach 70 % of organic modifier starting from the initial percentage,
- The percentage of MeCN found in the mobile phase.

Table IV. Factors and levels selected for the full factorial design. % $OM_{initial}$: initial percentage of organic modifiers mixture of the mobile phase; T_G: gradient time; % MeCN: percentage of acetonitrile added into the organic modifiers mixture of the mobile phase.

Factors	Levels		
O.M. $_{initial}$ (%)	25	35	45
T_G (min)	25	35	45
MeCN (%)	2.5	7.5	12.5

This full factorial design leads to 29 experiments including 2 additional independent repetitions at the center of the experimental domain (leading therefore to a total of 3 center points).

The chromatograms obtained from the realisation of the 33 full factorial design, allowed peaks to be detected and indexed at their beginnings (t_B), apexes (t_A) and ends (t_E) which were then transformed in retention factors in order to predict the chromatographic behaviour of the compounds following the methodology proposed by Lebrun et al [26,80]. In order to model the logarithm of the retention factors, multiple linear equations have been used which enable the computation and optimization of the separation factor S [26,42,73,74]. The optimization criterion considered in this study is the separation factor S, firstly because it is a natural criterion and secondly, its value and its associated error are easily computable [26,81,82].

The separation, S, represents the difference between the retention time at the end (t_E) of pair the first peak and the retention time at the beginning (t_B) of the second peak of the critical. ICA is a blind source separation method that could be used to automate the way the peaks are indexed and identified [58,83].

DS can be defined as the robust zone where the quality of the method separation is kept in spite of small environmental variations and this region for method robustness can also be envisioned as the space within the response surface where the critical separation S is at least 0 minutes.

The DS is defined by the set of all factors where the predicted probability that the separation S is higher than 0 minutes must be higher than a certain quality level (π), given the uncertainty of the estimation of model parameters θ [26,42,73,74]. In other words, the DS defines a region of experimental conditions where $S \geq 0$ min is obtained at least π times out of one hundred (e.g. 80 times out of 100).

3.2.7. Software

JMP v 8.0.2 is the statistical software that provided the technologies and algorithms for building and deploying the DoE method applied in this study.

The statistical analysis of the data was performed by means of an in-house computer program. The coding used was based on R 2.12 statistical language [84] for Windows which is open access and available at http://www.r-project.org.

3.3. Results and discussion

3.3.1. Experimental design

The full factorial design employed to optimize the separation, also assisted the development of a better understanding of the interaction of the selected chromatographic factors on separation quality.

The three HPLC factors chosen to be optimized were selected based on their known strong influence on selectivity, empirical preliminary experiments and certain instrumental limitations. For instance, given the fact that work on *Strychnos usambarensis* was a joint study, there was need for a column available in both laboratories that could additionally authorize a preparative liquid chromatography transfer, which in this case was represented by LiChrospher® 60 RP-select B. Also the experiments were conducted at a temperature of 25°C and a pH 2.2 which assured compounds stability. Instead of heptane sulfonate [71], sodium pentane sulfonate was chosen as the ion-pairing agent because of its short alkyl chain that forms an ion-pair complex which is less hydrophobic resulting to a smaller retention time. From preliminary experiments, a small percentage of MeCN was added to the mobile phase because its presence improved the shape of peaks, thus helped reduce the costs of analysis, especially due to the actual shortage and high price of MeCN, in a potential preparative context.

Hence, the three parameters selected for the optimization process were gradient time, percentage of OM and percentage of MeCN in the mobile phase composition.

Table IV shows the established ranges over which the chromatographic factors were to be varied, as follows: percentage of initial organic modifiers mixture (25 %-45 %), gradient time (25 min-45 min) and percentage of MeCN in the mobile phase (2.5 %-12.5 %).

The randomized sequence, in which the 29 experiments (including the 2 additional independent repetitions at the center of the experimental domain) were performed, helps minimize the effect of environmental variables that can introduce a bias on the measurements. Replicates of the central point, that were carried out using freshly prepared buffer, offer the possibility of calculating the independent estimate of the experimental error so that lack-of-fit and the statistical significance of the factors effects could be tested.

Additionally, these 3 independent repetitions allowed for an overall test of quadratic effects.

3.3.2. Optimization analysis – ICA

Given the similarity of the UV-spectra for all 6 compounds of interest shown in fig. 11, their almost identical chromatographic behaviour as well as the great number of potential endogenous compounds interfering with the 6 peaks of interest, the identification becomes very delicate.

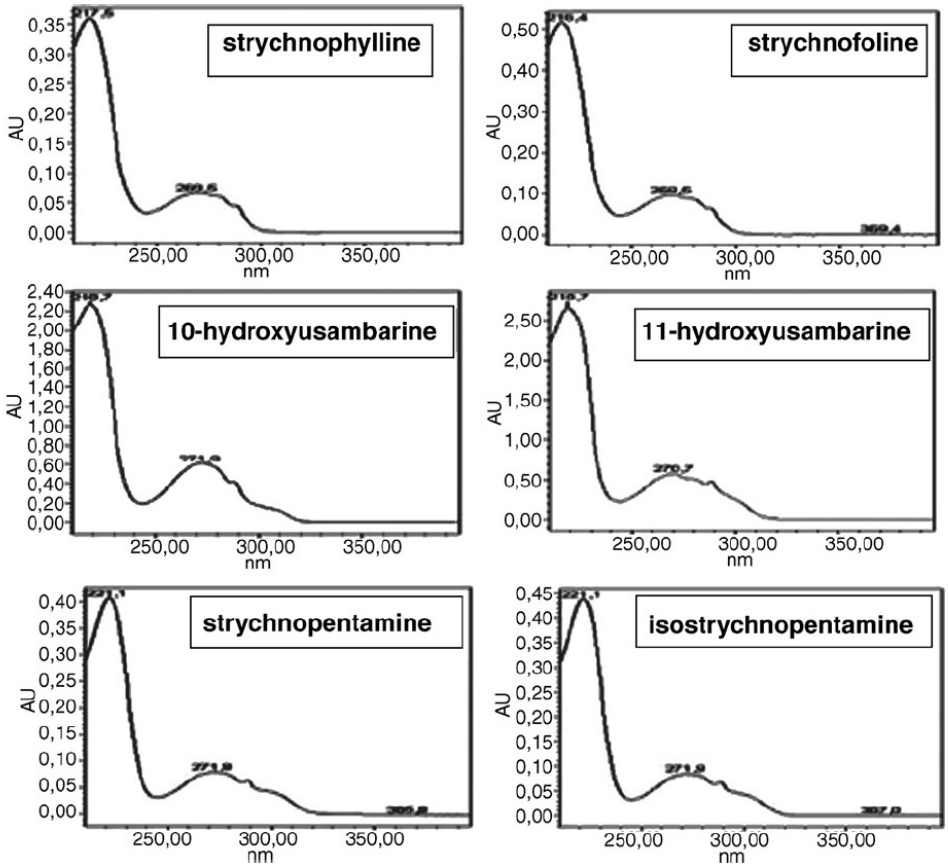


Fig. 11. UV spectra of studied alkaloids

At this stage ICA is usually an interesting tool because it allows a numerical separation for the co-eluted peaks and refines the estimation of the peak retention time at the beginning, apex and end. However in this case, the contribution of ICA is not

very significant mostly because of the similarity of UV-spectra of the compounds of interest. Nonetheless, ICA was useful for the unknown endogenous compounds.

Once all the chromatograms were obtained from the operational chromatographic conditions, the peaks were manually detected and indexed at the beginning, apex and end of each retention time. The logarithm of the retention factor was modelled employing multiple linear equations in order to enable the computation, the optimization of the separation and to create multivariate responses surface model [42].

The importance of checking the fit quality of a model is stressed here because of its strong correlation with the poor accuracy of the prediction of the retention times. In order to assess the quality of the modelling, the observed values and the predicted ones for the beginning, apex and end were compared. A good correlation between these values was observed meaning that the data is well fitted (all $R^2 > 0.998$). To illustrate this, fig. 12 presents the scatter plots of the observed and predicted retention times of peaks.

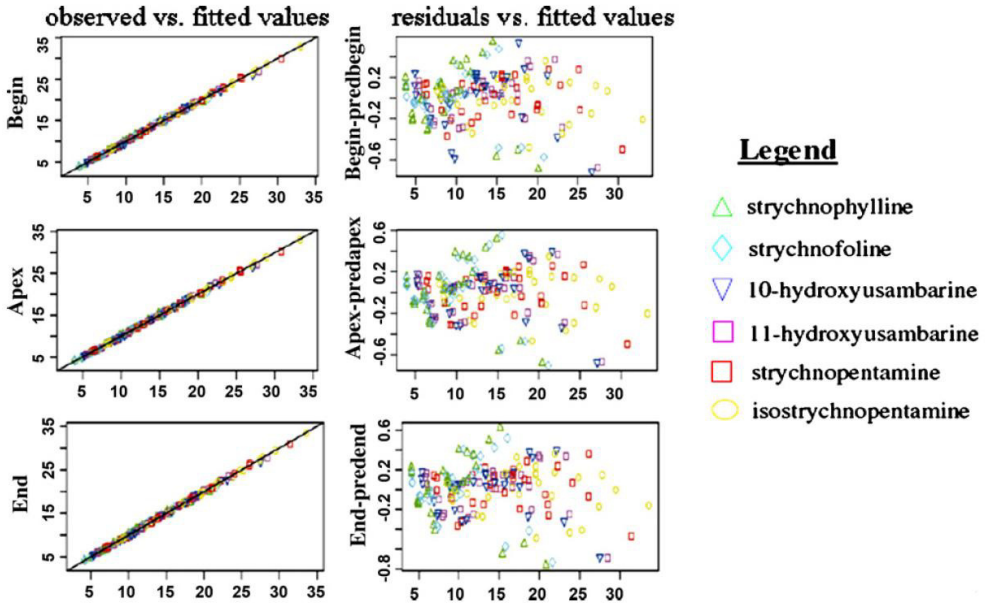


Fig. 12. Actual original responses of retention times (begin, apex and end) versus predicted one (predapex). On the right are the residuals.

The residuals which represent the differences between the predicted output from the model and the measured output from the data set are in a range of [-0.8; 0.6].

And also a normal distribution of residuals was noted as observed in fig. 12 which shows that the model chosen provided a good explanation of the data. The optimization criterion considered in this study was the separation factor S which is defined as the difference between the beginning of a peak and the end of the precedent peak. The propagation of the predictive error through the criterion was analyzed to offer confidence for the optima as proposed by Lebrun et al [26]. In order to assess the robustness, the probability to attain a separation superior or equal to 0 was calculated by means of Monte Carlo simulations and the response surfaces obtained were then replaced by probability surfaces in order to find a region of $S \geq 0$ [73].

3.3.3. Design space computation

Considering the aforementioned conditions, the design space corresponds to the region where this probability is superior to a preset quality level (π), in this case π = 30%. The expected probability to have well-separated peaks at the optimal condition is π = 45 % as illustrated in fig. 13. This DS also corresponds to a robustness zone in accordance with ICH Q8 guidelines [20]. Working within this robust zone defined in the experimental domain is convenient because it is not considered as a modification in the analytical method.

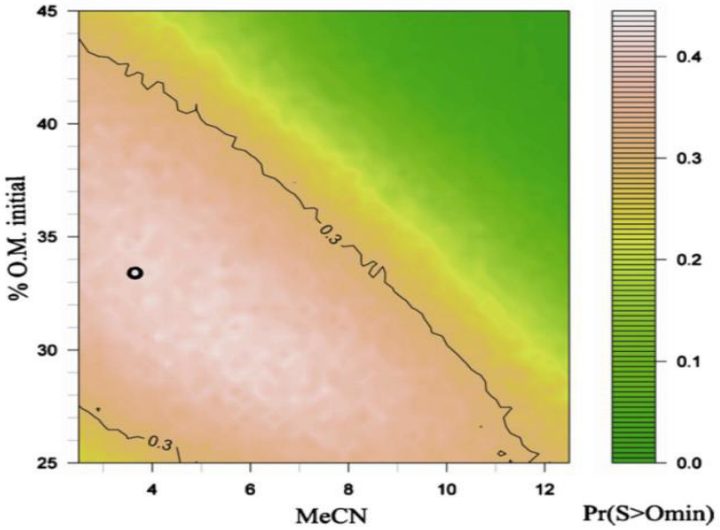


Fig. 13. Representation of the design space of the method on the experimental domain. Probability surface for the separation criterion S > 0 Pr(S > 0 min) with the design space (π = 0.3 or 30 %) included inside the black lines where the expected probability to have well-separated peaks at the optimal condition is π = 45 %. The black circle represents the optimal condition: gradient time of 40.6 min, proportion of MeCN in the organic modifiers (% $OM_{initial}$) mixture of 3.7 % and proportion of the organic modifiers mixture in the mobile phase of 33.4 %.

Despite of the relatively poor DS probability (30 %), a good agreement between the predicted chromatogram and the experimental chromatogram was observed. Indeed, a comparison between the predicted and the real processed chromatogram at the optimum conditions is illustrated in fig. 14. The chromatogram recorded at the optimum conditions is obtained by employing a 33.4:66.6 (v/v) mixture of organic phase and an aqueous solution of sodium pentane sulfonate (7.5 mM) adjusted with orthophosphoric acid (0.5 %) to pH 2.2. The initial organic phase was a mixture (3.7 % MeCN and 96.3 %) delivered at a flow rate of 1.0 mL/ min during a 40.6 minutes gradient time using a gradient profile.

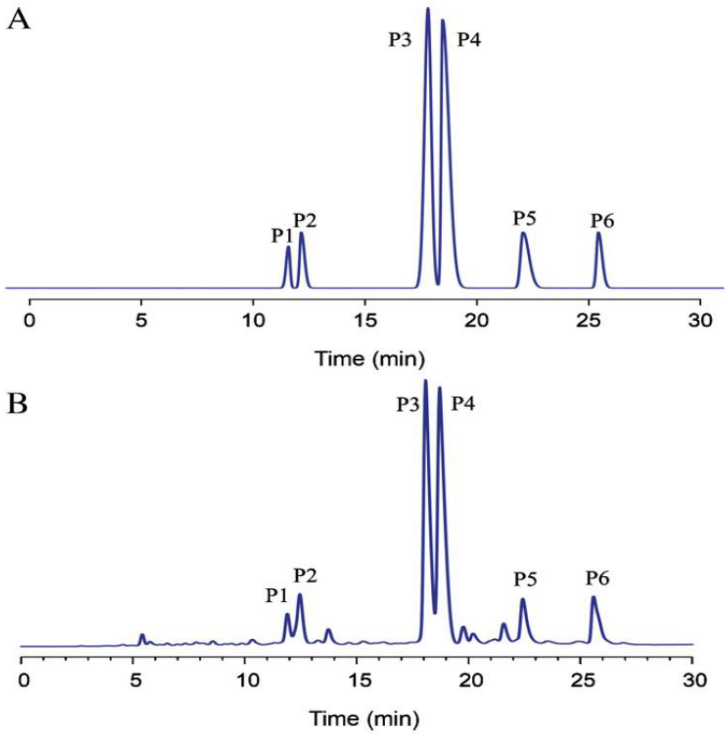


Fig. 14. Comparison between the predicted chromatogram (a) and the observed chromatogram (b) at the optimal condition obtained by employing a 33.4:66.6 (v/v) mixture of organic phase and an aqueous solution of sodium pentane sulfonate (7.5 mM) adjusted with orthophosphoric acid (0.5 %) to pH 2.2; the initial organic phase was a mixture (3.7 % MeCN and 96.3 %) delivered at a flow rate of 1.0 mL/ min during a 40.6 min gradient time using a gradient profile. P1: strychnophylline, P2: strychnofoline, P3: 10-hydroxyusambarine, P4: 11-hydroxyusambarine, P5: strychnopentamine and P6: isostrychnopentamine.

The chromatogram obtained at the optimum conditions illustrated on fig. 15a was also compared to the initial chromatogram shown on fig. 15b obtained based on a previous publication [71] and recorded before the application of experimental design and DS approach. The chromatogram illustrated on fig. 15b was assessed by reversed-phase chromatography using MeCN, an ion-pairing reagent (sodium salt of heptanesulfonic acid) (1 g in 420 mL) and fitted at pH 3.2 with phosphoric acid (0.5 %). Firstly, improvement of separation for the isomers of interest (strychnopentamine and isostrychnopentamine), of the analysis time and of the peak shape quality were noted. Secondly, an increase in the separation is also noticed regarding the other two pairs of isomers that are 10-hydroxyusambarine - 11-hydroxyusambarine (R_S=1.5) and strychnophylline - strychnofoline (R_S=4.3).

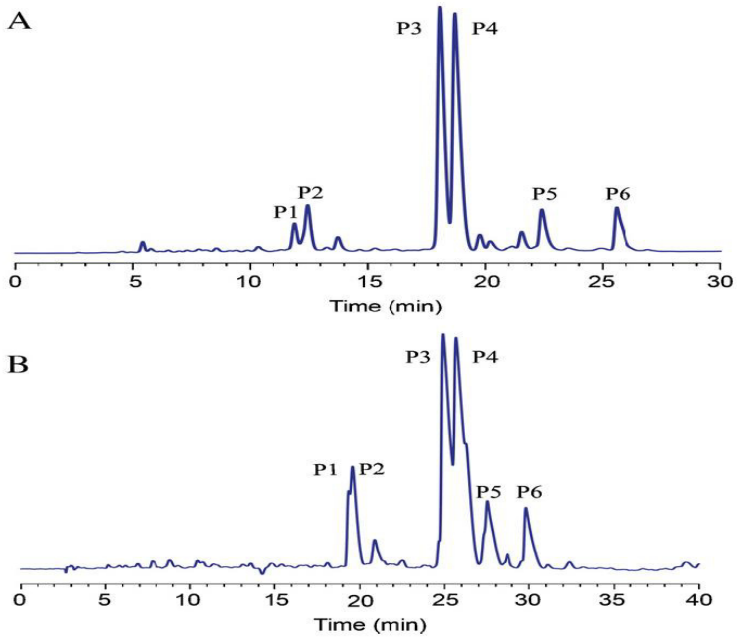


Fig. 15. Comparison between the observed chromatogram at the optimal condition obtained by employing a 33.4:66.6 (v/v) mixture of organic phase and an aqueous solution of sodium pentane sulfonate (7.5 mM) adjusted with orthophosphoric acid (0.5 %) to pH 2.2; the initial organic phase was a mixture (3.7 % MeCN and 96.3 %) delivered at a flow rate of 1.0 mL/ min during a 40.6 min gradient time using a gradient profile (a) and the one at the initial conditions based on acetonitrile (MeCN), aqueous solution of sodium salt of heptanesulfonic acid (1 g in 420 mL) and fitted at pH 3.2 with phosphoric acid (0.5 %), delivered at 1.0 mL/ min (b). P1: strychnophylline, P2: strychnofoline, P3: 10-hydroxyusambarine, P4: 11-hydroxyusambarine, P5: strychnopentamine and P6: isostrychnopentamine.

3.3.4. Qualitative analysis of phytochemical content

The optimized HPLC-UV method was then used to compare the potential differences of its phytochemical constituents over the collection periods of plant samples. As showed in Table V where the batch of June 2007 was arbitrarily chosen as percentage reference, *Strychnos usambarensis* displayed differences in the concentration of its phytochemical constituents according to the period it was collected. Chromatograms showed important differences between the analyzed samples with an emphasis on isostrychnopentamine and strychnopentamine as illustrated in fig. 16a-e.

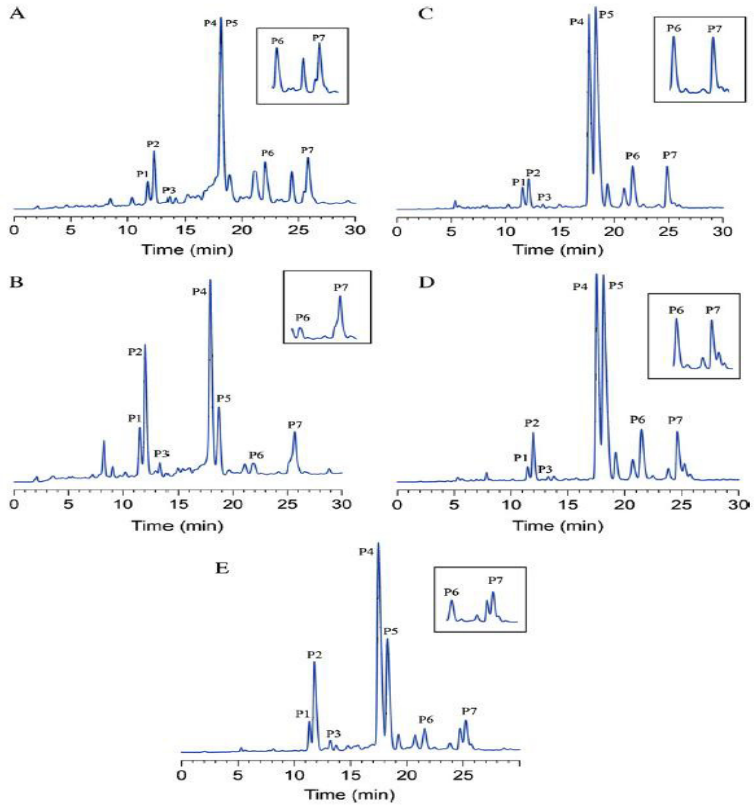


Fig. 16. (a) Chromatogram obtained for the period of June 2007. P1: isostrychnofoline, P2: strychnofoline, P3: strychnophylline, P4: 10-hydroxyusambarine, P5: 11-hydroxyusambarine, P6: isostrychnopentamine and P7: strychnopentamine. (b) Chromatogram obtained for the period of November 2007. (c) Chromatogram obtained for the period of August 2008. (d) Chromatogram obtained for the period of August 2009 (e). Chromatogram obtained for the period of February 2010.

Table V. Variation of the extracts concentration (expressed as percentage) of the 6 compounds of interest during a 3 years period. June 2007 content was arbitrarily chosen as reference. SF: strychnofoline, SPHY: strychnophylline, 10-OH: 10—hydroxyusambarine, 11-OH: 11-hydroxyusambarine, SP: strychnopentamine, ISP: isostrychnopentamine.

Percentage / Period	Strychno foline	Strychno phylline	10-hydroxy usambarine	11-hydroxy usambarine	Strychno pentamine	Isostrychno pentamine
June 07	100	100	100	100	100	100
November 07	202.6	136.3	86.5	223.3	23.3	78.5
August 08	103.7	101.7	235.6	2337.7	188.2	121.8
August 09	166.1	105.9	241.4	2121.2	225.2	120.5
February 10	389.1	274.7	283.7	1086.8	103.8	42.1

The *Strychnos usambarensis* samples collected in August and in February showed the highest content of hydroxylated derivatives: 10- and 11-hydroxyusambarine. Comparing the various periods under study, the highest concentration in isostrychnopentamine was obtained for the month of August, which is also equal to the concentration of its isomer, strychnopentamine SP. However, for the month of November, there was a marked difference in concentration between ISP and SP, as well as between 10-hydroxyusambarine and its 11-hydroxy isomer. Consequently, it is reasonable to conclude that during the dry season the species is a much richer source of alkaloids of interest than in the rainy season.

3.4. Conclusions

A HPLC method developed using a new optimization strategy was used for the separation of tertiary alkaloids extracted from *Strychnos usambarensis* leaves. The methodology involves three steps: DoE, ICA and DS which allowed finding the robust analytical conditions where the separation meet the desired requirements. This was the first time that this strategy was applied to a complex matrix.

The partial replacement of MeCN by MeOH was successful, allowing reducing costs and maintaining a high quality of results in respect with the ICH Q8 guidelines and the DS approach.

An optimal separation was achieved especially for strychnopentamine, isostrychnopentamine, 10–hydroxyusambarine and 11-hydroxyusambarine. A good separation was also obtained for another pair of isomers that is represented by strychnophylline and strychnofoline, succeeding so to acquire for the first time, to our knowledge, a LC-UV profile for the compounds of interest.

In addition, the method resulting from the strategy of simultaneous multifactorial optimization reduced overall assay development time and provided information regarding separation and sensitivity due to the detection of new compounds in the analyzed mixture of plant origin, compounds whose separation and identification are still delicate.

This study has also allowed the identification of the highest concentration period for alkaloids which present the most interesting pharmacological activity.

Acknowledgements

Special thanks are due to WBI (Wallonie Bruxelles International) for a research grant provided to Iolanda Nistor. Our sincere thanks go to National University of Rwanda (Prof. Charles Karangwa and M. R. Muganga. This study was partly sponsored by the CUD (Cooperation Universitaire au Developpement) and by the Belgian National Fund for Scientific Research (FNRS) – grant 3.4533.10. M. Cao is a Research Fellow from the FNRS. A research grant from the Belgium National Fund for Scientific Research (FRS-FNRS) to E. Rozet is also gratefully acknowledged. This study was partly sponsored by the CUD (Cooperation Universitaire au Developpement) - Project PIC ULg-UCL-Rwanda.

4. Implementation of a Design Space Approach for Enatiomeric Separations in Polar Organic Solvent Chromatography

The implementation of a Design Space approach and the CPPs to consider when applying the QbD concepts outlined in ICH Q8(R2), Q9 and Q10 to analytical method development and optimization for three chiral compounds where the focus of this work. In this sense, an HPLC method using a polysaccharide-based stationary phase containing a cellulose tris (4-chloro-3-methylphenylcarbamate) selector in polar organic solvent chromatography mode was considered. The effects of trifluoroacetic acid (TFA) and n-hexane concentration in a MeCN mobile phase were investigated under a wide rage of column temperatures.

Good adequacies were found between the observed data obtained after using a central composite design and the expected chromatographic behaviours predicted by applying the DoE-DS methodology. The critical quality attribute represented here by the separation criterion (S_{crit}) allowed assessing the quality of the enantioseparation. Baseline separation for the compounds of interest in an analysis time of less than 20 minutes was possible due to the original and powerful tools applied which facilitated an enhanced method comprehension.

Finally, the advantage of the DoE-DS approach resides in granting the possibility to concurrently assess robustness and indentify the optimal conditions which are compound dependent.

4.1. Introduction

Food and Drug Administration (FDA) has been particularly active in fostering the concept of QbD which entails building quality into the process and the product in a systematic, science- and risk-based manner during development rather than attempting to test quality into a product retrospectively [20].

Several guidelines have been issued by the ICH to harmonize and facilitate the implementation of QbD framework.

According to ICH Q8(R2) [20] (pharmaceutical development), the goal of QbD is to establish a design space of input variables and operating parameters without negatively affecting the identified CQAs. It is expected that working within the design space will result in an operating space where most process robustness will be found and in a product meeting a predefined quality.

A second guideline of interest is ICH Q9 [21] which intends to link process parameters to CQAs and to reduce quality-influencing risks.

This paper focuses on the factors to consider when applying the QbD concepts outlines in ICH Q8(R2) and Q9 to analytical method development and optimization.

Two of the major elements recommended by the ICH Q8 and Q9 guidelines as a minimum expected quality reside firstly in identifying CQAs of the drug product, so that those product characteristics having an impact on product quality can be studied and controlled and secondly determining the quality attributes of the overall achievement of the analytical method.

Also, QbD requires an understanding of the manner in which process variables influence product quality [85]. Translated to analytical chemistry, one of the objectives is to gain a better understanding of the critical process parameters influence on the critical quality attribute.

These recent years, the importance of chiral separations has been increasingly recognized in pharmaceutical research given the different structural characteristics of drugs with multiple chiral centers which can lead to a potential biological or pharmacological therapeutic activity which is mostly restricted to one of the enantiomers. This is one of the reasons why during drug development a thorough assessment of the potential chiral drug candidates is required by the FDA or EMEA (European Medicines Evaluation Agency).

Different analytical separation techniques can be used in order to obtain chiral separations but the most frequently used in pharmaceutical analysis is high-performance liquid chromatography. Current chiral HPLC methods are either direct, which consist in the formation of transient diastereoisomeric complexes between the analyte and the chiral stationary phase, or indirect which involves derivation of samples [86,87]. The industry leans towards direct chiral separations using CSPs (chiral stationary phases) which are more widely used and more predictable [88].

Numerous studies have been reported in the literature where the most popular and communly used CSPs are polysaccharides, cyclodextrins, macrocyclic glycopeptide antibiotics, Pyrkle types, proteins, ligand exchangers, and crown ether based.

Chromatography using CSPs has proven to be the most useful technique to achieve chiral separations [89].

Polysaccharide-based phases have been identified as very versatile and useful chiral selectors for the separation of a large variety of enantiomeric compounds in different chromatographic modes [86]. The polar organic mode became increasingly popular for various CSPs during last years [90]. This mode combines the advantages of reversed-phase chromatography, i.e. increased solubility, ionic and polar interactions, with those of SFC and NPLC in which solvents are easier evaporated.

Chiral method development is a difficult process as a great variety of chiral selectors for the CSP are commercially available, and the interaction of the mobile phase with both the analyte and CSP has to be considered, because the enantioselectivity of a given CSP for a given compound is a priori unknown and thus could not offer a guarantee for a successful enantioseparation.

The analyst is confronted with the selection of various factors affecting chiral resolution (the nature of chiral compound and the chiral selector, mobile phase conditions, column efficiency and temperature) which can easily transform the process into a time-consuming, expensive and sometimes ineffective one [91,92].

Given the significant pharmacological interest of the chiral molecules examined and the complexity of chiral separations, the aim of the present work was the development of an efficient and robust analytical method that is able to resolve these molecules effectively using a polysaccharide-based stationary phase containing a cellulose tris (4-chloro-3-methylphenylcarbamate) selector [93-96].

To achieve this purpose, a DoE-DS methodology is employed to identify the CPPs and to establish a comprehensive DS which can be further analyzed as to determine optimum analytical conditions [26,39,73,74].

DoE is a tool used to gain maximum understanding of effects and interactions between the most CPPs in order to provide robustness and an optimum analytical method. DS for analytical methods is considered to be a multidimensional combination of parameters that affect method performance which will allow the impact of any deliberate changes within the knowledge space of the particular method to be scientifically assessed.

4.2. Materials and methods

4.2.1. Chemicals and reagents

HPLC grade acetonitrile was obtained from Biosolve (The Netherlands). Analytical grade TFA and diethylamine (DEA) were purchased from Fisher Scientific (Leicestershire, UK), respectively VWR International (Fontenay sous bois, France). HPLC grade n-hexane was provided by Merck (Darmstadt, Germany).

Solutions of sample compounds were prepared by dissolution in MeCN (i.e. the predominant organic modifier in the mobile phase) to produce a concentration of 1 mg/ml.

4.2.2. Instruments and conditions

The chromatographic system used consisted of a Waters 2695 separation module coupled to a Waters 996 Photodiode array detector.

In the present study, the optimization strategy was examined on a polysaccharide-based stationary phase containing a cellulose tris (4-chloro-3-methylphenylcarbamate) selector commercialized as Lux® Cellulose-4 column (250 x 4.6 i.d.; particle size 5μm) from Phenomenex.

The chromatographic experiments were carried out under polar organic solvent chromatography mode. The mobile phase delivered using isocratic conditions, at a flow rate of 0.8 ml min-1 was composed of a mixture of MeCN, n-hexane, DEA and TFA.

The UV detector wavelength was set at 225 nm for the entire study.

The temperature of the samples was set at 10°C. Injections of 10 μl of the solutions were accomplished.

The determinations of the circular dichroism spectra for each compound were performed at 20°C with a Jobin Yvon CD6 (Lonjumeau, France) circular dichrograph. A quartz cell (Hellma Inc.) of 10 mm path length was used to obtain spectra at 0.2 nm intervals from 220 to 240 nm. Spectra result from the averaging of three scans. The spectra were smoothed using Gaussian convolution smoothing.

4.2.3. Design of experiments

This study focuses on implementing a DOE-DS (design of experiments – design space) approach which is compliant with the QbD initiative encouraged these recent years by the FDA which promotes the idea that quality should be built in by design. This means that the analytical process should cope with disturbances and operation within the design space should assure that the quality of the product fulfils the predefined requirements.

In the interest of modelling the process robustness, one must start with a clear definition of what is to be modelled and which process parameters affect the quality of the process.

A DoE is a strategy for planning efficient, systematic and simultaneous collection of responses from the factors evaluated with the purpose of acquiring data on the parameters influencing the quality of a product [42,97].

DoE is also considered a tool used to identify the critical process parameters and determine a response surface design which can be analyzed as to establish the design space which will later provide an optimum analytical condition [26].

In this study, this method is considered because it will help gain a maximum understanding about the process leading to its robustness and yielding an optimum analytical condition. A rotatable central composite design was selected for this study because of the column equilibration time demanded by each chromatographic experiment [98,99]. The design involved 15 experimental points with 3 replicates of the centre point of the domain resulting in 18 randomized runs.

4.2.4. Modelling and DS computation

In the acquired chromatograms obtained from the Central Composite Design that was chosen to explore the experimental domain, the detected peaks were indexed at the beginning, apex and end of retention time [97,100,101]. Because, from the statistical perspective, retention times do not present themselves with a Gaussian distribution, the logarithms of the retention factors (log(k)) were selected as responses and modelled by a stepwise multi-linear regression which used the adjusted coefficient of determination (R^2_{adj}) as to avoid overfitting.

The linear regression model helped define the relationship between the responses and the critical process parameters. At this point in time, a residual analysis will be appropriate in order to check model adequacy.

4.2.4.1. Critical quality attributes

Process robustness implies starting with a clear definition of what is to be modelled and which process parameters affect the performance and quality of the process [102].

ICH Q8 defines CQA as 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 [20].

As such, a quantitative measure of the process quality, the CQA, must be determined, e.g., resolution or separation criterion [38].

A recently introduced critical quality attribute by Lebrun et al. [26], namely the separation factor (S_{crit}), was considered in this study to asses the quality of a separation rather than selecting the resolution.

Considering the definition of the separation criterion as the difference of retention time between the beginning of the second peak and the end of the first peak ($Scrit = t_{B2}-t_{E1}$), its value and associated error are easier computable.

If the separation criterion (S_{crit}) between the least well resolved peak pair (critical pair) is above the threshold value of 0 ($S_{crit}>0$), baseline separation is achieved.

4.2.4.2. Critical process parameters

Employing DoE leads to an assessment of the factors which most strongly influence the CQA, i.e. separation criterion and therefore to a determination of the CPPs. A better understanding of the CPPs will help develop a robust design space [26,85,103].

Among various factors identified for polar organic solvent chromatography as affecting chiral recognition which is usually represented as arising from three points of interaction, are the percentage of mobile phase additives, column temperature and the percentage of n-hexane [99].

These three factors presented in Table VI were studied across the experimental domain mapped out by the DoE.

Table VI. Factors and levels selected for the central composite design. %TFA: percentage of acidic additive added to the mobile phase; temp (°C): column temperature; %n-hexane: proportion of n-hexane added into the mobile phase represented by acetonitrile.

Factors	*Levels*				
TFA (%)	0.05	0.07	0.1	0.13	0.15
temp (°C)	10	13	17.5	22	25
n-hexane (%)	1	3	5.5	8	10

Additives are typically incorporated into chiral mobile phases to minimize peak distortion arising from unwanted interactions between polar solutes and the stationary phase and thus to sharpen tailing peaks and to increase resolution of chiral compounds [88,89,95].

DEA and TFA concentration ranges were expanded as much as possible in order to widen the experimental domain and to minimize the possibility of not finding any good separation within. The concentration range between 0.05% and 0.15% of the mobile phase acidic additive was also chosen considering column limitations and keeping in mind a possible memory effect that can be presented in time.

Because of the difficulty encountered in the screening step with regards to the long equilibrium time of the CSP when both DEA and TFA were varied, the selection of only TFA as a factor for the DoE was considered [98,99,104]. For all experiments, the basic additive concentration was established at 0.1%.

The effect of temperature which can be described as a function of the balance between enthalphic and enthropic contributions to the retention times of the chiral compounds presented here was assessed for the critical quality attribute (i.e. the separation criterion).

In this paper, the effect of temperature was studied in the range 10°C-25°C. The limitation at 25°C was set by the stability of the chiral molecules studied. In addition to these two factors, a third one represented by the percentage of n-hexane was added. The values for the lower and upper limit for this factor were set between 1% and 10%.

4.2.4.3. Design Space

The methods and tools such as high-end statistic (i.e. multivariate analysis) and DoE were used to help understand the most CPPs and map out the DS, an operating space where most process robustness will be found, which will be subsequently employed to determine the optimal analytical condition.

The QbD-oriented guise, i.e. ICH Q8(R2) is incoherent with regards to the examples given at the end of the document and the DS definition presented.
In other words, the DS is described by the functional relationship between process inputs (critical quality parameters) and their influence on CQA (represented here by the separation criterion) in order to define a zone where not only quality is predicted, but also assurance of quality (also known as quality risk management) is established.

For an analytical method, the first step towards predicting quality can be described by the separation criterion being above its threshold value.

The second one, where quality risk management is considered, is to quantify the reliability or level of assurance by computing a probability distribution that can provide over time assurance of quality and detect process drift [105]. With this intent, Monte Carlo simulations were performed in order to have a better understanding of the impact of the prediction error propagation from the responses on the quality criterion chosen. Thus, the design space was consolidated as the multidimensional subspace where the probability for the separation criterion to be within acceptance limits was higher than a defined quality level [73].

Taking this into account, the mathematical medium by which a broader variation understanding and robustness are attained may be expressed as Eq. 18:

$$DS = \left\{ x_o \in \chi \middle| E_\theta \left[P(CQA > \lambda) \middle| x_o, \theta \right] \geq \pi \right\} \qquad \text{Eq. 18}$$

Where xo is a point of the experimental domain χ and θ is the set of estimated parameters of the model. CQA corresponds to the separation criterion which is supposed to be higher than a selected threshold of acceptance λ. P and E represent the estimators of probability and mathematical expectation and π is the preselected reliability level for the design space to provide assurance of quality.

Process validation can be envisioned as creating a design space, a multidimensional space that defines how different variables interact. The design space for an analytical method encompasses the acceptable ranges for the critical process

parameters demonstrated to provide quality assurance for the delivered product, and defines a knowledge space in which process adjustments and refinements may be undertaken without regulatory involvement.

4.2.4.4. Optimum analytical condition

In order to obtain an optimum analytical condition, a grid search was carried out across the experimental domain mapped out by the DoE. The impact of the model prediction error on the CQA using Monte Carlo simulations was studied for each point of the grid in order to compute the probability of reaching the separation time of 0 minutes. The point of the grid having the highest probability value is considered to be the optimum analytical condition.

4.2.5. Software

In order to build and deploy the DoE methodology applied in this study, JMP v 8.0.2, which is a statistical software providing the necessary technologies and algorithms, was used.

The statistical analysis of the data was performed by means of an in-house computer program. The coding used was based on R 2.12 statistical language [84] that is open access and also available on-line at http://www.r-project.org.

4.3. Results and discussion

4.3.1. Peak detection

All experiments for the three molecules studied (fig. 17) were carried out with an UV-DAD detector which lacks the possibility to determine the optical activity of compounds analyzed. Thus, in order to assess chiral activity, the structures of the molecules were characterized by circular dichroism (CD) spectroscopy. For each molecule, the three corresponding HPLC peaks were named after their elution order as follows: P1 (first peak), P2 (second peak) and P3 (third peak).

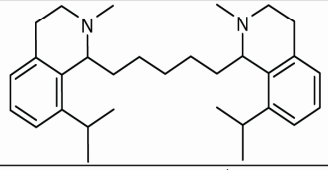
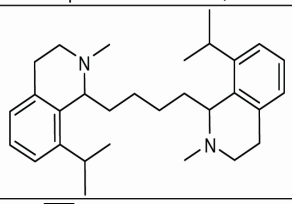
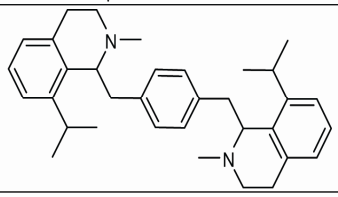
Reference	Structure
BEL097	
BEL174	
BEL169	

Fig. 17. Chemical structures of the three compounds studied: BEL097, BEL174 and BEL169.

The results for the BEL097 molecule displayed opposite CD spectral patterns (fig. 18) for P1 and P3. P1 exhibited positive ΔA values in the 220 – 228nm region, while P2 presented itself with optical inactivity, along with the ratio (1:2:1) obtained in the HPLC analysis, possible indicating this second peak as the meso form of the structure of BEL097.

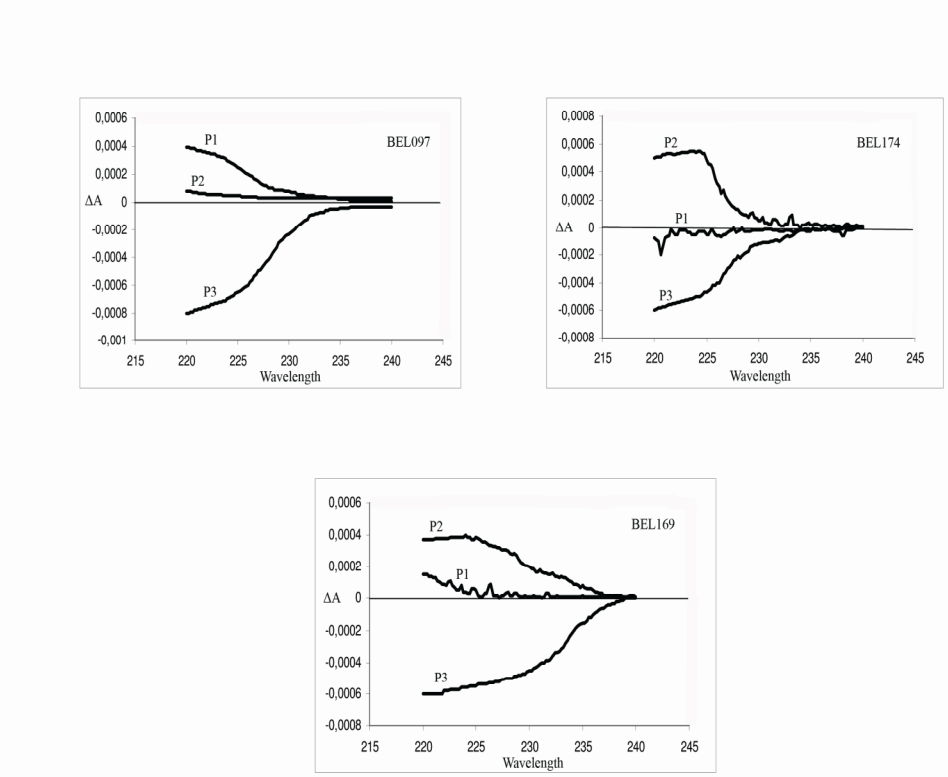


Fig. 18. Circular dichroism (CD) spectra of BEL097, BEL174 and BEL169 solvated in MeOH. Spectra were obtained after isolation of HPLC eluates.

As for BEL174, P2 has positive bands in the wavelength region of 220 – 230nm, while P3 shows the converse behaviour (fig. 18). The opposite effects indicate that the two compounds are probably a pair of enantiomers. On the other hand, P1 has no optical activity and keeping in mind the same reasoning as for BEL097, this is a possible indication of the meso form of the BEL174 structure.

The spectra for the P2 and P3 of BEL169 were mirror images of each other, as shown in fig. 18, confirming their enantiomeric nature. The same meso form indication as for P2 of BEL097 and P1 of BEL174 is possible for P1 of BEL169.

4.3.2. Retention times modelling

To define the relationship between the responses and the factors, a linear regression model was applied on the basis of a multiple linear regression presented in Eq. 19.

$$Y = \beta_0 + \beta_1 \cdot TFA + \beta_2 \cdot temp + \beta_3 \cdot n-hexane + \beta_4 \cdot temp \cdot n-hexane + \beta_5 \cdot temp \cdot TFA + \beta_6 \cdot TFA \cdot n-hexane \quad \text{Eq. 19}$$

Firstly, the experiments were used to model the chromatographic behaviour of each peak and concurrently evaluate the method robustness.

Secondly, this model which describes process responses enables an enhanced approach to how process understanding and predictability can be attained with a few well designed experiments alongside the model.

The objective of these considerations is not to highlight some quantitative structure retention relationships but to understand the resulting model thanks to the chromatographic behaviour.

Slight structural changes in the analogue molecules presented here can lead to large differences in chiral recognition which confirms the difficulty in elucidating chiral recognition mechanisms and predicting the separation in any given changes.

4.3.3. Effects analysis

Usually, hydrogen-bonding interactions and π-π interactions are essential to chiral recognition with polysaccharide-based chiral stationary phases, but will not suffice for chiral discrimination in polar organic mode.

The chiral recognition mechanism for the polysaccharide-based stationary phase containing a cellulose tris (4-chloro-3-methylphenylcarbamate) selector studied here is influenced by many factors, including the structure of the chiral solute, the mobile phase composition and temperature.

For example, the acidic additive can be envisioned as attenuating or even suppressing the ionizable functional groups of the analytes leading to enhanced H-bonding interactions which are configuration dependent between their neutral form and the CSP. In addition to that, temperature can play a role in influencing the ionization degree of a compound and then onto the retention mechanism.

This mechanism can also be significantly influenced by the recognition ability of the polysaccharide stationary phase which is improved by the substituents introduced to the phenyl group of the carbohydrate polymer: the electron-donating methyl group situated in the meta-position of the phenyl ring and an electron-withdrawing halogen (i.e. chlorine) at the para-position of the phenyl ring. These interactions are illustrated below by the effect of the TFA percentage, temperature and n-hexane proportion exert on the separation.

4.3.3.1. TFA effect

Increasing the additive concentration has a high impact on the separation, but the analysis time remains constant in the 20 minutes range thanks to the n-hexane factor.

At a concentration higher than 0.06% TFA, partial separation is observed for BEL174, while BEL097 and BEL169 present baseline separation. Also TFA concentration affects retention in the sense that peak shift is observed for all the molecules studied starting from a concentration of the acidic additive higher than 0.06%.

For BEL097, TFA enhances the initial separation characteristics until it reaches a concentration of 0.1% where even though peak shape improves significantly, the separation is influenced by a decreasing trend suggesting a certain saturation of this effect at low levels of the acidic additive. The same responses are noted for BEL174 and BEL169, but at different TFA concentrations 0.14%, respectively 0.12%.

4.3.3.2. Temperature and n-hexane effects

To reduce analysis time without decreasing the resolution, the influence of temperature and the percentage of n-hexane on a polysaccharide-based stationary phase containing a cellulose tris (4-chloro-3-methylphenylcarbamate) selector were studied.

In the domain range of 10°C – 25°C, the column temperature factor presented a curbature effect on enantioresolution because of the changes in the ionization degree of the compounds analysed over the domain studied, but sub-ambient temperatures as presented in section 4.3.6. lead to an improvement in selectivity and resolution.

Increasing the n-hexane content in the mobile phase had a direct effect on retention times shortening analysis times to under 20 minutes.

4.3.3.3. Statistical effects analysis

The effects analysis of the three CPPs considered in this study is presented for each peak at the beginning, apex and end of retention time.

By taking a brief look at fig. 19 (BEL097), the factor with the strongest influence on the overall separation is represented by the TFA (the acidic additive) which is impacting in a positive manner the retention times of the P1, P2 and P3 compounds.

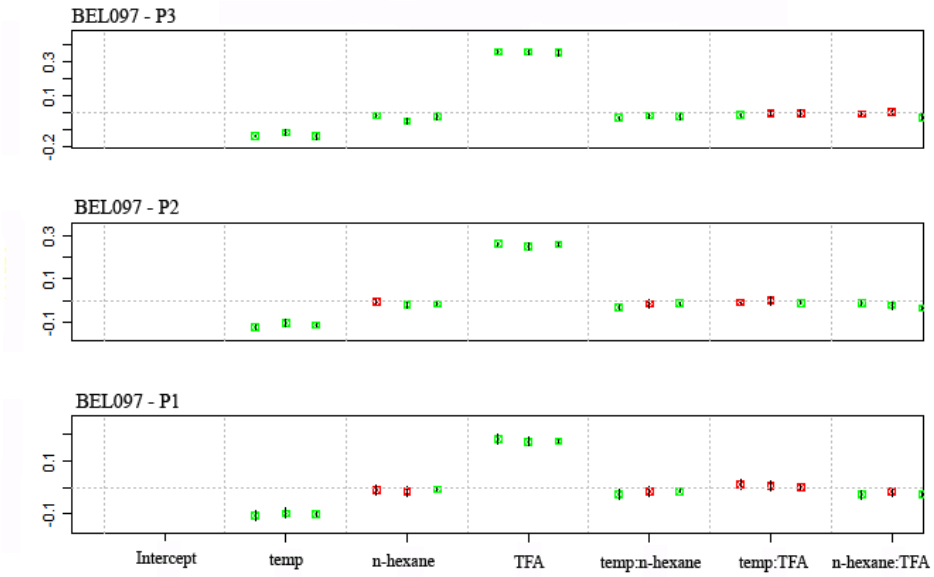


Fig. 19. The significance (non nullity) of regression coefficients at 0.95% for BEL097. The parameters for which the 95% credible intervals contain 0 are said to be nonsignificant (red), while significantly ones differ from 0 (green).

Closely follows the temperature which with a slightly smaller, but nonetheless important influence has a negative effect on the beginning, apex and end of retention times for the 3 peaks.

On the other hand, the factor which has the least impact on the separation is represented by the n-hexane as seen from fig. 19 (BEL 097).

Even though there is a narrow difference between the structure of BEL097 and BEL174 as they are analogue molecules, the attention turns to the TFA presented in fig. 20 (BEL174) which is impacting in a positive manner the retention times of the possible enantiomers (P2 and P3) and has a negative effect on the retention time of the

P1 compound. This may also indicate that the impact this CPP has on the separation is compound dependent.

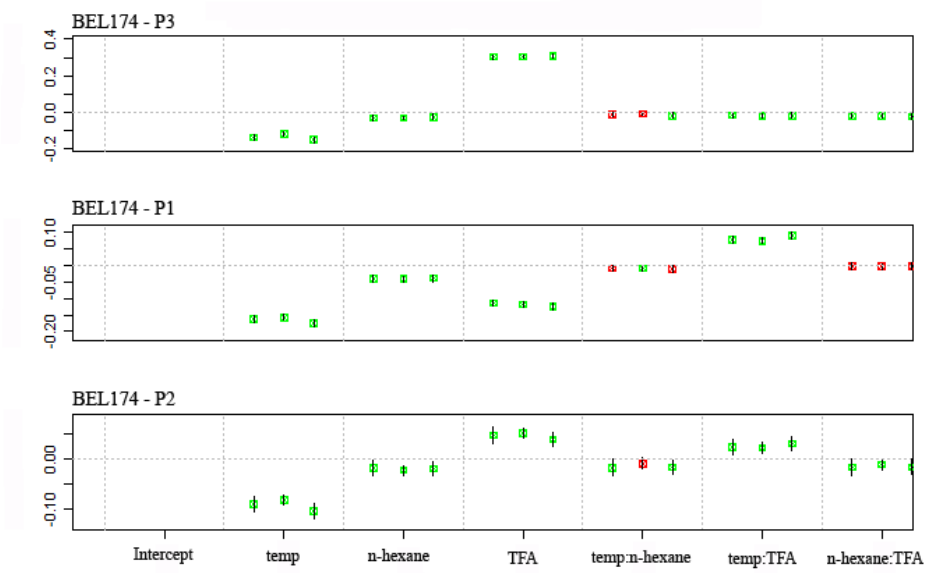


Fig. 20. The significance (non nullity) of regression coefficients at 0.95% for BEL174. The parameters for which the 95% credible intervals contain 0 are said to be nonsignificant (red), while significantly ones differ from 0 (green).

At a closer look at fig. 20 (BEL174), the temperature exerts a negative effect on the 3 peaks, but has a slightly stronger impact than the one presented for the BEL097 compound in fig. 19.

The n-hexane parameter exerts the same slightly negative effect on the separation as for the BEL097.

The trend indicating the acidic additive as having the strongest impact on the critical quality attribute is sustained by the third molecule investigated as seen in fig. 21 (BEL169).

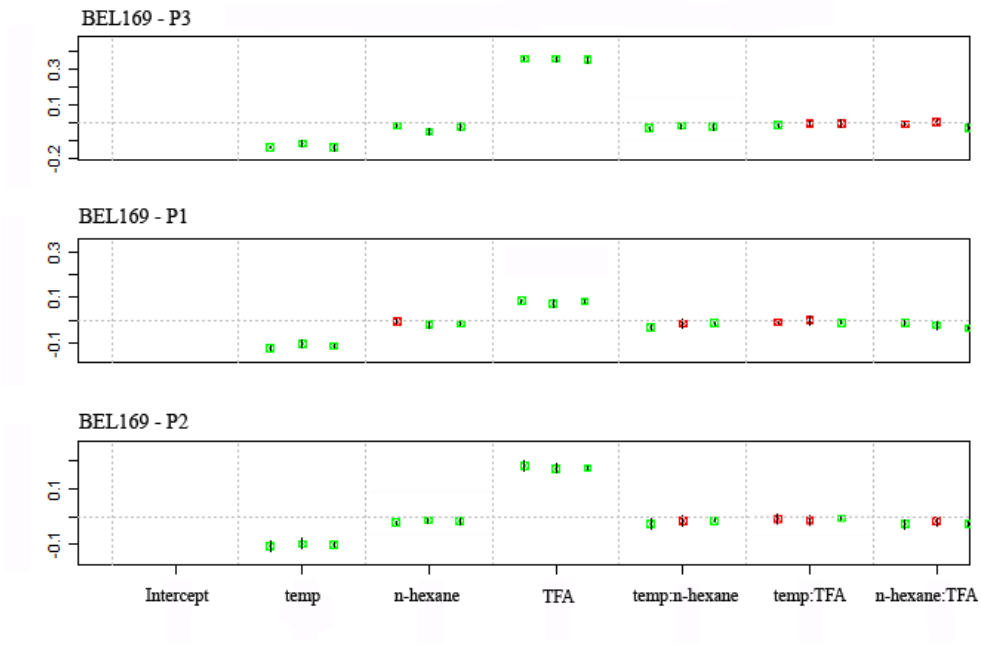


Fig. 21 The significance (non nullity) of regression coefficients at 0.95% for BEL169. The parameters for which the 95% credible intervals contain 0 are said to be nonsignificant (red), while significantly ones differ from 0 (green).

A similar negative effect of the temperature as for the other 2 molecules (i.e. BEL097 and BEL174) is enforced by the results presented in fig. 21(BEL169).

Even though n-hexane does not have a very strongly felt effect on the separation, it influences the retention times given that all analysis were carried out within 20 minutes.

4.3.4. Model adequacy

It is recommended to check the suitability in order to enhance the understanding of the capabilities of the model.

A plot of the observed data against the predicted responses along with the corresponding residuals is depicted in fig. 22 – 24 for each of the three molecules studied.

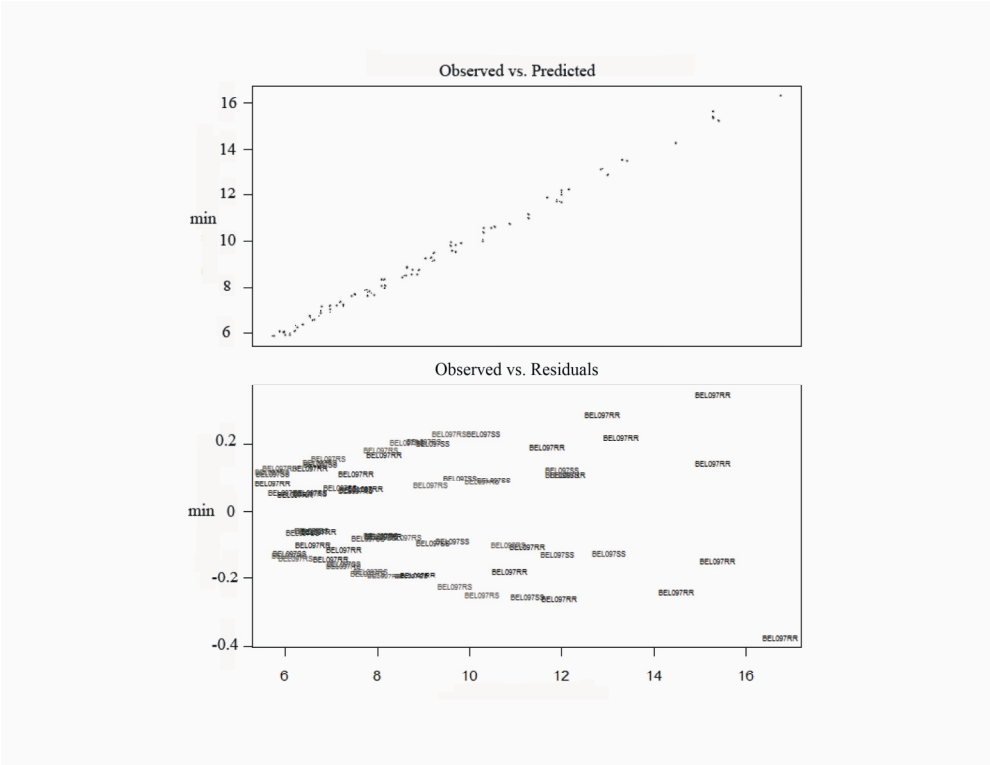


Fig. 22. A good model adequacy for BEL097 is depicted. This is expressed by the residuals (bottom) obtained after comparing observed responses of retention times at beginning, apex and end versus predicted ones (top).

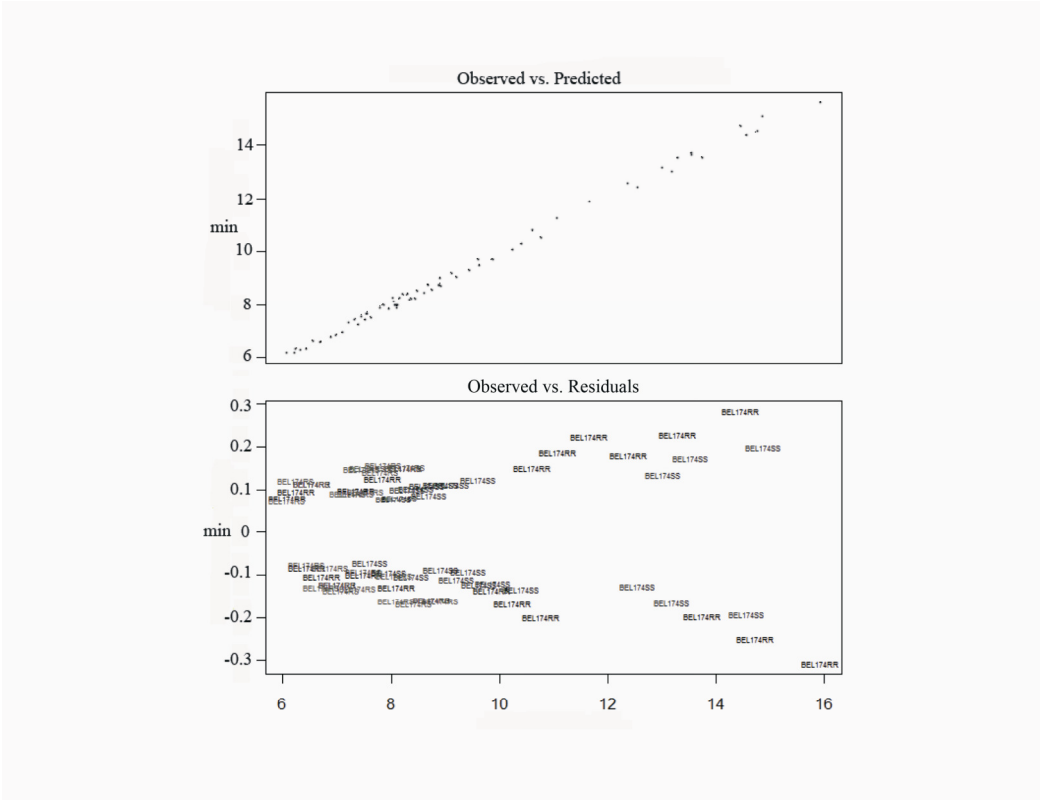


Fig. 23. A good model adequacy for BEL174 is depicted. This is expressed by the residuals (bottom) obtained after comparing observed responses of retention times at beginning, apex and end versus predicted ones (top).

As observed in fig. 22 – 24, the residuals are distributed between -0.4 and 0.4 minutes and no aberrant responses were observed for either of the compounds. The normal distribution of the residuals is an indicator that the data is generally well explained by the model chosen.

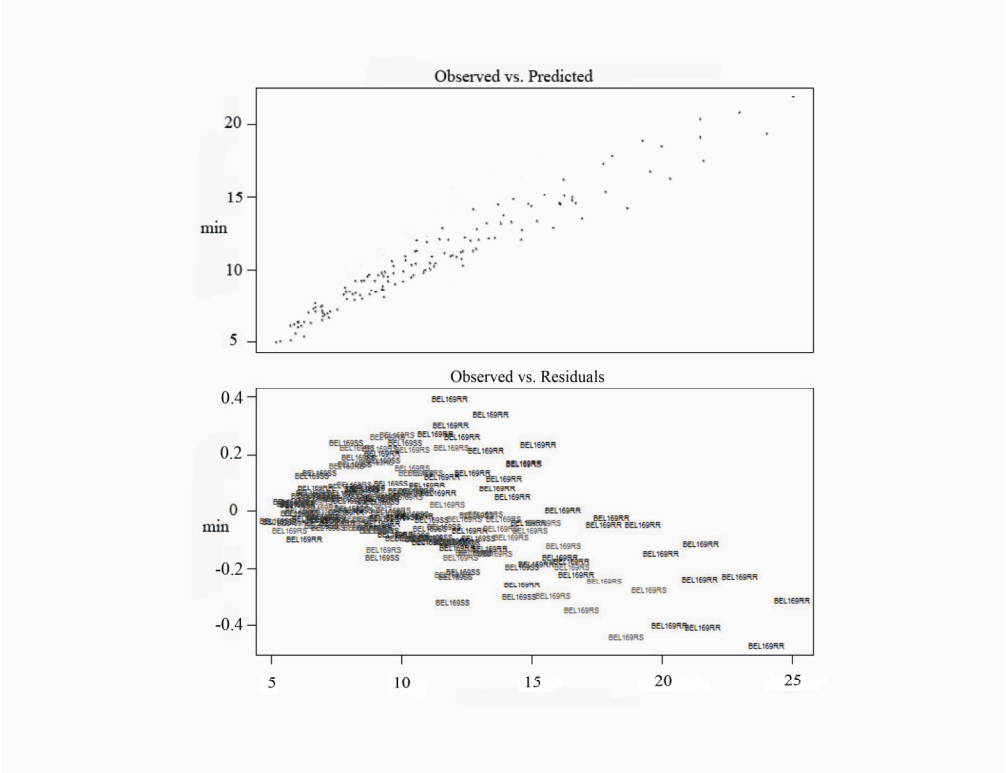


Fig. 24. A good model adequacy for BEL169 is depicted. This is expressed by the residuals (bottom) obtained after comparing observed responses of retention times at beginning, apex and end versus predicted ones (top).

4.3.5. Quality criterion and DS computation

The criterion selected in order to achieve the desired objective of finding the best separation was the minimal separation (Scrit > 0min). Monte-Carlo simulations were carried out in order to propagate the predictive error from the responses obtained by using Eq. 19 to the critical quality attribute so that the probability of reaching the desired objective is computed. After a grid search was applied on every point of the experimental domain, probability surfaces were recorded for each of the molecules studied as follows:

For BEL097 a design space (DS) was found (fig. 25), for a specified quality level of π =0.413, chosen as 95% of the quality level of the optimal chromatographic solution.

Given that the probability surfaces recorded are an indication of the joint expected probability to obtain a separation within specifications, an expected probability of 0.41 for BEL097 means that within the DS, there is a risk of 0.59 not to be within specifications. In order to have a better understanding of results, a look at the marginal predictive distributions of the predicted optimal point presented in fig. 28 could be considered.

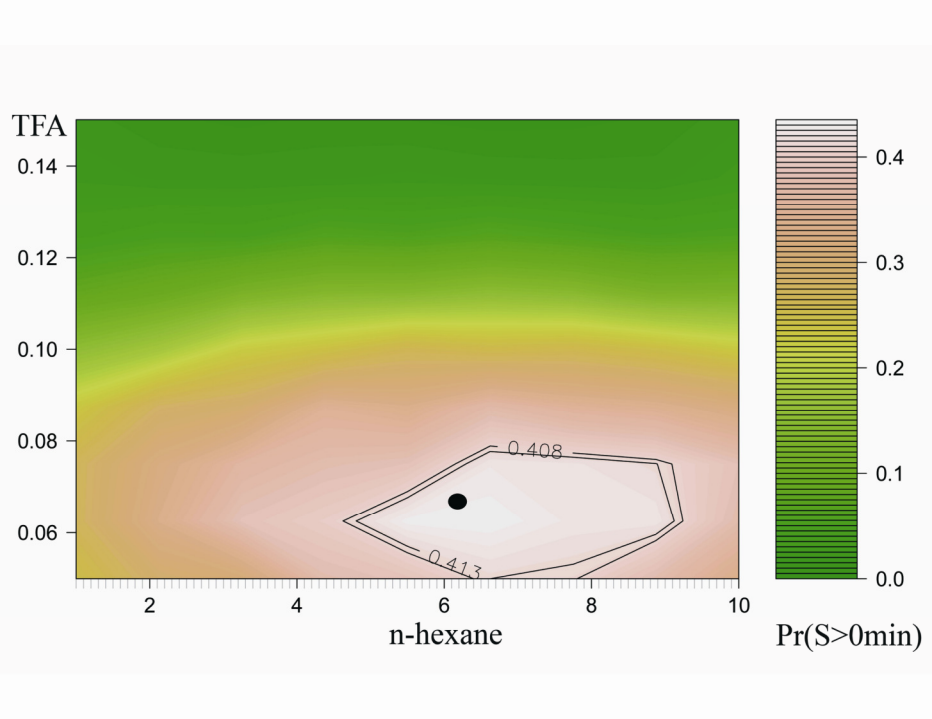


Fig. 25. Representation of the Design Space for BEL097. Probability surface for the separation criterion S>0 Pr(S>0min) with the design space (π=0.413 or 41.3%) included inside the black lines where the expected probability to have well-separated peaks at the optimal condition is π =45%. The black circle represents the optimal condition that is depicted in Fig 28b: TFA concentration of 0.07%, column temperature of 13°C and proportion of n-hexane in the mobile phase of 6%.

As shown in fig. 26, a DS was found for BEL174 at a specified quality level of π =0.616, while an even higher quality level (i.e. π =0.655) presented in fig. 27 was found for BEL169.

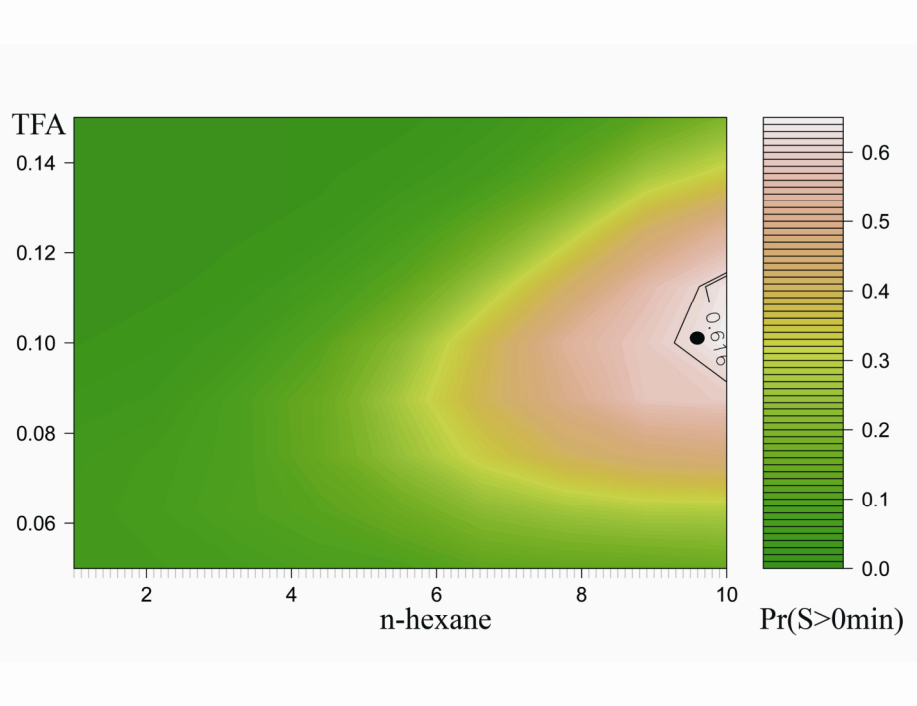


Fig. 26. Representation of the Design Space for BEL174. Probability surface for the separation criterion S>0 Pr(S>0min) with the design space (π=0.616 or 61.6%) included inside the black lines where the expected probability to have well-separated peaks at the optimal condition is π =65%. The black circle represents the optimal condition that is depicted in fig. 29b: TFA concentration of 0.1%, column temperature of 10°C and proportion of n-hexane in the mobile phase of 10%.

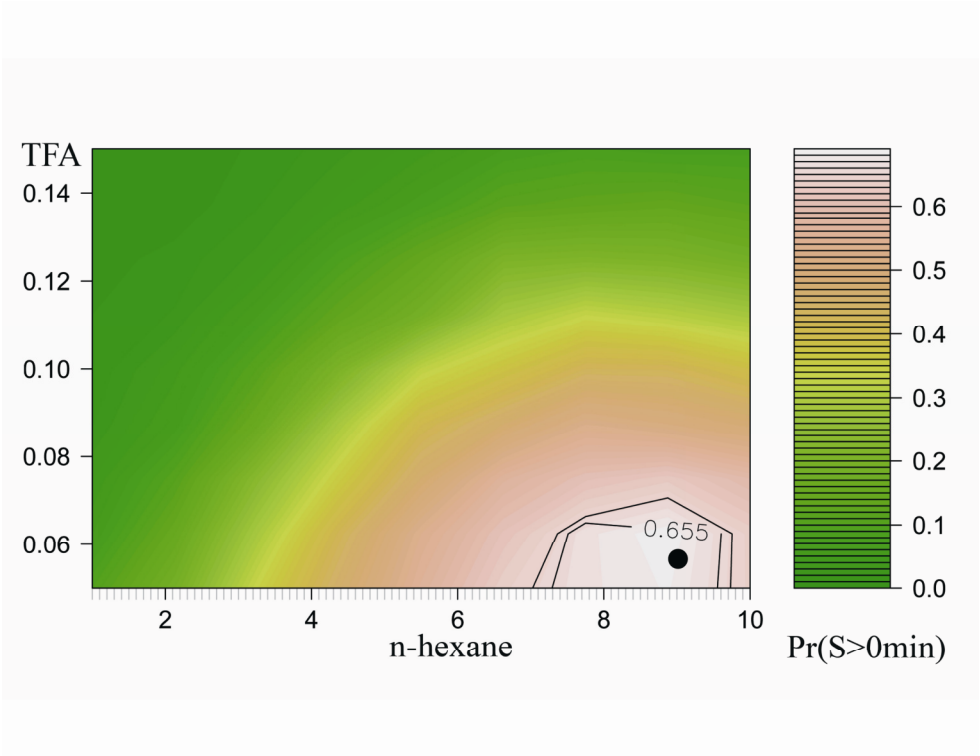


Fig. 27. Representation of the Design Space for BEL169. Probability surface for the separation criterion S>0 Pr(S>0min) with the design space (π=0.655 or 65.5%) included inside the black lines where the expected probability to have well-separated peaks at the optimal condition is π =69%. The black circle represents the optimal condition that is depicted in fig. 30b: TFA concentration of 0.06%, column temperature of 13°C and proportion of n-hexane in the mobile phase of 9%.

4.3.6. Optimal separation

For BEL097 the optimal separation condition was identified at 0.07% TFA, 13°C and 6% n-hexane as shown in fig. 28 where a comparison between the predicted chromatogram and the recorded one for this point within the design space presented earlier is depicted. A good correlation is found between the two chromatograms.

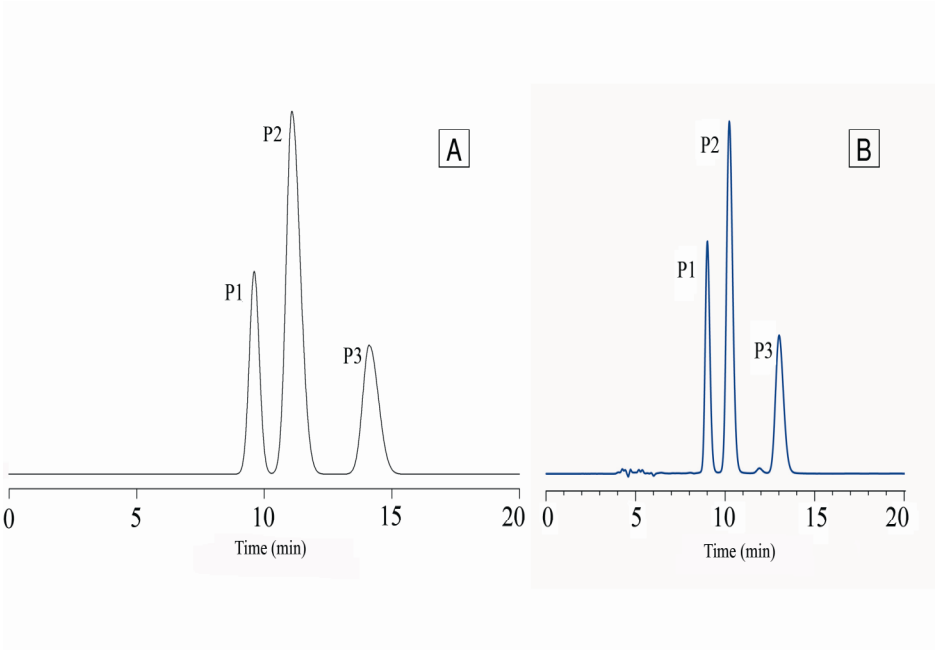


Fig. 28. Comparison for BEL097 between the predicted chromatogram (a) and the observed chromatogram (b) at the optimal condition obtained by employing a TFA concentration of 0.07% and a proportion of 6% n-hexane in the mobile phase (MeCN) at a column temperature of 13°C, delivered at a flow rate of 0.8 mL/min. The three peaks obtained for BEL097 were named after their elution order as follows: P1 (first peak), P2 (second peak) and P3 (third peak).

As presented in fig. 29, chromatograms predicted and observed for BEL174 at the optimal separation condition that was identified at 0.1% TFA, 10°C and 10% n-hexane were tested. Even though the value for the quality level was 0.616, similarity is manifested between the two.

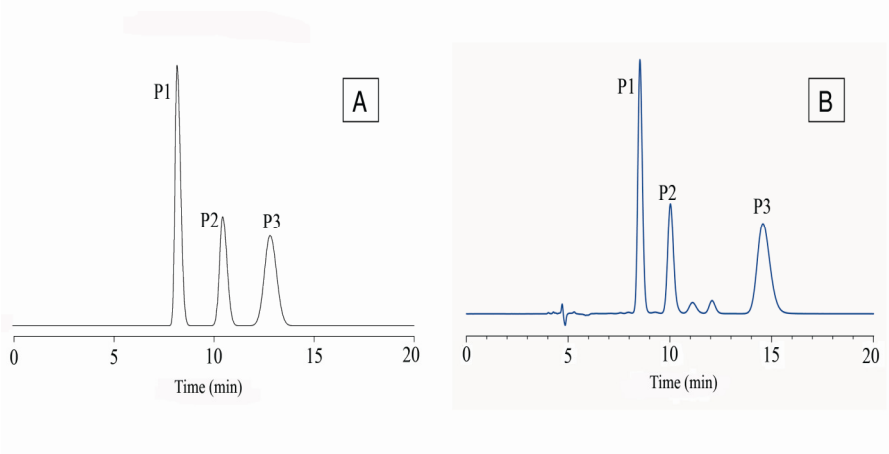


Fig. 29. Comparison for BEL174 between the predicted chromatogram (a) and the observed chromatogram (b) at the optimal condition obtained by employing a TFA concentration of 0.1% and a proportion of 10% n-hexane in the mobile phase (MeCN) at a column temperature of 10°C, delivered at a flow rate of 0.8 mL/min. The three peaks obtained for BEL174 were named after their elution order as follows: P1 (first peak), P2 (second peak) and P3 (third peak).

Finally, the chromatograms obtained for BEL169 at the optimal separation condition of 0.06% TFA, 13°C and 9% n-hexane reveal good agreement as displayed in fig. 30.

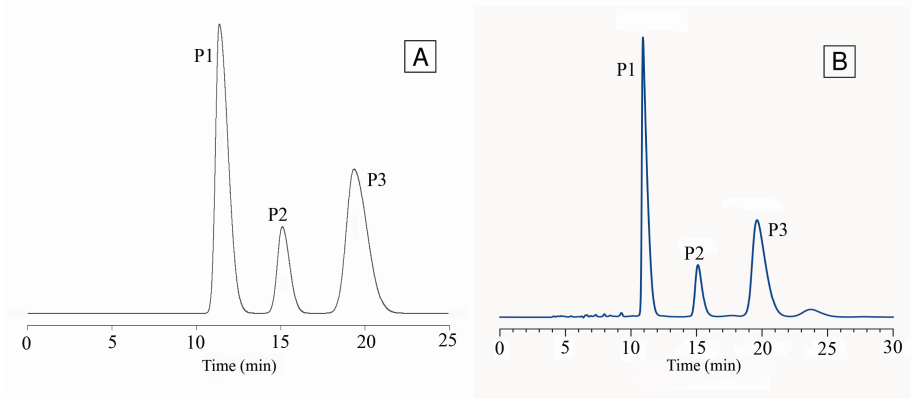


Fig. 30. Comparison for BEL169 between the predicted chromatogram (a) and the observed chromatogram (b) at the optimal condition obtained by employing a TFA concentration of 0.06% and a proportion of 9% n-hexane in the mobile phase (MeCN) at a column temperature of 13°C, delivered at a flow rate of 0.8 mL/min. The three peaks obtained for BEL169 were named after their elution order as follows: P1 (first peak), P2 (second peak) and P3 (third peak).

4.4. Conclusions

The set objective of reaching baseline enatioseparation was achieved given that the three compounds studied were enantiomerically separated using a QbD compliant DoE-DS methodology in the polar organic solvent chromatography mode. Preliminary results with regards to the nature of the compounds were obtained using the HPLC method combined with circular dichroism and UV-VIS spectrometry.

A good understanding of the CPPs effects on a polysaccharide-based stationary phase containing a cellulose tris (4-chloro-3-methylphenylcarbamate) selector represents a step forward to apprehend the chromatographic behaviour of the three chiral compounds. In addition, the CQA chosen provided acceptable separation in an overall short analysis time.

Moreover, this promising methodology allowed for the optimal condition and robustness to be concurrently determined which along with its effectiveness strongly indicates a powerful strategy that can be applied in the analysis of pharmaceutical compounds.

The results presented here demonstrated the applicability of a DoE-DS methodology to identify the DS on an enantioselective case.

Acknowledgements

Special thanks are due to WBI (Wallonie Bruxelles International) for a research grant provided to Iolanda Nistor. A research grant from the Belgium National Fund for Scientific Research (FRS-FNRS) to E. Rozet is also gratefully acknowledged. Additional thanks goes to Dr. F.Rosu for the use of the circular dichroism detector.

2.5. General conclusions

The main objective of the present work was to test an innovative methodology for the development of chromatographic methods. This methodology was created in order to be compliant with the concepts of quality by design (QbD) and design space (DS) recommended by the International Conference on Harmonisation (ICH). This latter methodology is based on the modelling of the retention times at the beginning, the apex and the end of each peak corresponding to the compounds of a mixture by using the separation criterion (S) rather than the resolution (R_S) as a critical quality attribute. Stepwise multiple linear regressions were used to create the models.

Translated to method development, DS can be defined as the subspace, necessarily encompassed in the experimental domain (i.e. the knowledge space), within which the probability for the criterion to be higher than an advisedly selected threshold is higher than a minimum quality level. From a chromatographic point of view, the design space thus defines a region where, on average, with respect to a given quality level, peaks do not coelute.

Concurrently to these works, the usefulness of independent component analysis (ICA) to automatically read chromatograms was demonstrated. This statistical technique allowed to numerically separate coeluting compounds in order to estimate the retention times at their respective beginnings and ends.

In order to validate this global methodology, a mixture of unknown compounds and a pharmaceutical formulation were first used. This study confirmed its ability to significantly speed up the development process and to obtain optimal separations. The optimization based on this new ICA-DoE-DS methodology paved the way towards the automated development of chromatographic methods.

This was the first time that this strategy was successfully applied to a complex matrix represented by the extract from the *Strychnos usambarensis* leaves.

Also, the partial replacement of acetonitrile by methanol was successful, allowing reducing costs and maintaining a high quality of results in respect with the ICH Q8 guidelines and the design space approach.

In addition, the method resulting from the strategy of simultaneous multifactorial optimization reduced overall assay development time and provided information regarding separation and sensitivity due to the detection of new compounds in the analyzed mixture of plant origin, compounds whose separation and identification are still delicate.

A second study was conducted on chiral molecules in order to demonstrate the powerfull methodology presented. Nowadays, enatioseparation of diastereoisomers is a subject of high interest for the pharmaceutical domain and presents a high degree of difficulty. The strategy used offered a good understanding of the CPPs effects which further allowed for the optimal condition and robustness to be concurrently determined in a Quality by Design environment.

In conclusion, the global methodology developed during this work, demonstrated its abilities to automatically read chromatograms and to numerically separate coeluting peaks. It also confirmed its aptitudes to optimize the separation of numerous mixtures and to identify the DS corresponding to a zone of robustness. The proposed methodology finally proved its efficiency as well as its flexibility.

2.5. Originality of the thesis

The novelty of this work resides in the fact that by bringing together HPLC notions, regulatory support for biopharmaceutical development along with chemometrics new possibilities for the future of analytical method development are being offered.

The advantages offered by the DoE-DS methodology presented here consisting of costs reduction, higher robustness, speeding up the procedure, achieving a wider process understanding and a better knowledge of the CPPS chosen represent an important step that leads towards automated chromatographic methods.

The implementation of this approach to a complex natural matrix such as the extract of *Strychnos usambarensis* as to identify the best collecting periods and thus isolate a greater concentration of tertiary alkaloids that possess antiplasmodial and cytotoxic properties brings an added value and originality.

The work conducted on the three chiral compounds presented in the last part of this thesis is a subject of high interest especially for the pharmaceutical domain. The successful enantioseparations achieved in spite of the degree of difficulty that is attributed to this type of separation, enhance the value of this highly efficient and flexible methodology.

References

[1] Research and Markets. *HPLC systems and accessories: A global strategic business report*. [Online]. Available from: http://www.strategyr.com/HPLC_Systems_and_ Accessories_ Market_Report.asp. [Accessed 19th May 2012].
[2] MarketsandMarkets. *Life science and chemical instrumentation market (2011 - 2016) global trends, end user and competitive analysis*. [Online]. Available from: http://marketpublishers.com/report/medicine_pharmaceuticals_biotechnology/healthcare_equipment_services/life_science_chemical_instrumentation_market_2011_2016_global_trends_end_user_competitive_analysis.html. [Accessed 19th May 2012].
[3] Tswett M. *Ber Dtsch Chem Ges*, 1906;24:316-323.
[4] Martin AJP, Synge RLM. A new form of chromatogram employing two liquid phases: A theory of chromatography. 2. Application to the micro-determination of the higher monoamino-acids in proteins. *Biochem J*, 1941;35(12):1358-1368.
[5] Ahuja S, Rasmussen H. *HPLC method development for pharmaceuticals*. New York: Academic Press Inc;2007.
[6] Ajuja S, Scypinski S. *Handbook of modern pharmaceutical analysis*. New York:Academic Press Inc;2010.
[7] Meyer VR. *Practical high-performance liquid chromatography*. New York:John Wiley & Sons;2010.
[8] Scott RPW. *Liquid chromatography column theory*. New York:John Wiley and Sons;1992.
[9] Dong MW. *Modern HPLC for practicing scientists*. New York:John Wiley & Sons;2006.
[10] Grushka E, Grinberg N. *Advances in chromatography*. Florida:CRC Press Inc.;2009.
[11] Neue UD. *HPLC columns: theory, technology, and practice*. New York:John Wiley & Sons;1997.
[12] Snyder LR, Kirkland JJ, Glajch JL. *Practical HPLC method development*. New York:Wiley-Interscience; 2010.
[13] Guillarme D, Ruta J, Rudaz S, Veuthey J-L. New trends in fast and high-resolution liquid chromatography: a critical comparison of existing approaches. *Anal Bioanal Chem* 2010;397:1069-1082.
[14] Rostagno M. *Monolithic columns for fast high-performance liquid chromatography separations, SciTopics.* [Online]. Available from: http://www.scitopics.com/Monolithic_columns_for_fast_high_performance_liquid_chromatography_separations.html. [Accessed 19th May 2012].
[15] Dejaegher B, Heyden YV, Experimental designs and their recent advances in set-up, data interpretation, and analytical applications. *J Pharm Biomed Anal*, 2011;56:141-158.

[16] [Online]. Available from: http://www.molnar-institute.com/.[Accessed 20th May 2012].
[17] [Online]. Available from: http://www.chromsword.de/.[Accessed 20th May 2012].
[18] [Online]. Available from: http://acdlabs.eu/products/com_iden/meth_dev/lc_sim/.[Accessed 20th May 2012].
[19] [Online]. Available from: http://www.datalys.net/index.htm.[Accessed 20th May 2012].
[20] European Medicines Agency. *ICH Q8, Pharmaceutical development*;2009.
[21] European Medicines Agency. *ICH Q9, Quality risk management*;2011.
[22] European Medicines Agency. *ICH Q10, Pharmaceutical quality systems*;2011.
[23] Potter C, PQLI application of science- and risk-based approaches (ICH Q8, Q9 and Q10) to existing products. *J Pharm Innov* 2009;4:4-23.
[24] Rathore A S, Winkle H. Quality by design for biopharmaceuticals. *Nat Biotechnol* 2009;27:26-34.
[25] Rosas J G, Blanco M, Gonzalez J M, Alcala M. Quality by design approach of a pharmaceutical gel manufacturing process, part 1: Determination of the design space. *J Pharm Sci* 2011;10:4432-41.
[26] Lebrun P, Govaerts B, Debrus B, Ceccato A, Caliaro G, Hubert Ph, Boulanger B. Development of a new predictive modelling technique to find with confidence equivalence zone and design space of chromatographic analytical methods. *Chemom Intell Lab Syst* 2008;91:4.
[27] Bondi Jr RW, Drennen III JK. Quality by design and the importance of PAT in QbD. *Separ Sci Technol* 2011;10:195-224.
[28] Massart DL, Vandeginste BGM, Buydens LMC, De Jong S, Lewi P, Smeyers-Verbeke J. *Handbook of chemometrics and qualimetrics: Part A. Volume 20A in Data handling in science and technology*. Amsterdam: Elsevier;1997.
[29] Einax JW. Chemometrics in analytical chemistry. *Anal Bioanal Chem* 2008;390:1225-1226.
[30] Martens H, Martens M. *Experimental Designs*. New York:John Wiley & Sons;2001.
[31] Cochran W G, Cox GM. *Experimental Designs*. New York:Wiley;1962.
[32] Box GEP, Hunter JS, Hunter WG. *Statistics for experimenters: design, innovation, and discovery*. New York:Wiley;2005.
[33] Montgomery DC. *Design and analysis of experiments.* New York:John Wiley & Sons;2005.
[34] Anthony J. *Design of experiments for engineers and scientists*. Oxford:Butterworth–Heinemann;2003.
[35] Oehlert GW, Freeman WH and Company. *A first course in the design and analysis of experiments*. New York:Freeman;2000.
[36] Hubert Ph, Nguyen-Huu J-J, Boulanger B, Chapuzet E, Cohen N, Compagnon P-A, Dewe W, Feinberg M, Laurentie M, Mercier N, Muzard G, Valat L, Rozet E.

Harmonization of strategies for the validation of quantitative analytical procedures. A SFSTP proposal – Part III. *J Pharm Biomed Anal* 2007;45:82-96.
[37] Rozet E, Hubert C, Ceccato A, Dewe W, Ziemons E, Moonen F, Michail K, Wintersteiger R, Streel B, Boulanger B, Hubert Ph. Using tolerance intervals in pre-study validation of analytical methods to predict in-study results. The fit-for-future-purpose concept. *J Chromatogr A* 2007;1158:126-137.
[38] Schoenmakers P, Abbott RW. *Anal Chim Acta* 1988;208:357.
[39] Dewe W, Marini R, Chiap P, Hubert Ph, Crommen J, Boulanger B. Development of response models for optimizing HPLC methods. *Chemometr Intell Lab Syst* 2004;74:263-268.
[40] Peterson JJ. *J Qual Tech* 2004;36:139.
[41] De Backer B, Debrus B, Lebrun P, Theunis L, Dubois N, Decock L, Verstraete A, Hubert Ph, Charlier C. *J Chromatogr B* 2009;877:4115.
[42] Debrus B, Lebrun P, Ceccato A, Caliaro G, Govaerts B, Olsen BA, Rozet E, Boulanger B, Hubert Ph. *Talanta* 2009;79:77.
[43] Biswas KM, Castle BC, Olsen BA, Risley DS, Skibic MJ, Wright PB. *J Pharm Biomed Anal* 2009;49:692.
[44] García A, Rupérez FJ, Marín A, de la Maza A, Barbas C. *J Chromatogr B* 2003;785:237.
[45] Olmo B, García A, Marín A, Barbas C. *J Chromatogr B* 2005;817:159.
[46] Marín A, Barbas C. *J Pharm Biomed Anal* 2006;40:262.
[47] Marín A, Barbas C. *J Pharm Biomed Anal* 2004;35:769.
[48] Marín A, García E, García A, Barbas C. *J Pharm Biomed Anal* 2002;29:701.
[49] Vander Heyden Y, Nijhuis A, Smeyers-Verbeke J, Vandeginste BGM, Massart DL. *J Pharm Biomed Anal* 2001;24:723.
[50] Dejaegher B, Vander Heyden Y. *J Chromatogr A* 2007;1158:138.
[51] Dejaegher B, Dumarey M, Capron X, Bloomfield MS, Vander Heyden Y. *Anal Chim Acta* 2007;595:59.
[52] Dejaegher B, Capron X, Smeyers-Verbeke J, Vander Heyden Y. *Anal Chim Acta* 2006;564:184.
[53] Dejaegher B, Smeyers-Verbeke J, Vander Heyden Y. *Anal Chim Acta* 2005;544:268.
[54] Atkinson AC, Donev AN. *Optimum Experimental Designs*. Oxford:Clarendon Press;1992.
[55] International Conference on Harmonization (ICH) of Technical Requirements for registration of Pharmaceuticals for Human Use. *Topic Q2(R1): Validation of Analytical Procedures: Text and Methodology*. Geneva:ICH;1996.
[56] Herrador MA, Asuero AG, Gonzalez AG. *Chemom Intell Lab Syst* 2005;79:115.
[57] Hyvärinen A, Karhunen J, Oja E. *Independent Component Analysis*. New York:Wiley;2001.
[58] Hyvärinen A, Oja E. *Neural Networks* 2000;13:411.

[59] Moffat AC, Osselton MD, Widdop B. Clarke's Analysis of Drugs and Poisons. Electronic version:Pharmaceutical Press;2004.
[60] Angenot L. De l'existence en Afrique centrale d'un poison de flèche curarisant, issu du Strychnos usambarensis. *Ann Pharm Fr* 1971;29:353–364.
[61] Angenot L, Dubois M, Ginion C, van Dorsser W, Dresse A. Chemical structure and pharmacological (curarizing) properties of various indole alkaloids extracted from an African Strychnos. *Arch Int Pharmacodyn Ther* 1975;215:246–258.
[62] Wright CW, Bray DH, O'Neill MJ, Warhurst DC, Phillipson JD, Quetin-Leclercq J, Angenot L. Antiamoebic and antiplasmodial activities of alkaloids isolated from Strychnos usambarensis. *Planta Med*, 1991;57:337–340.
[63] Bassleer R, Depauw-Gillet MC, Massart B, Marnette JM, Wiliquet P, Caprasse M, Angenot L. Effets de trois alcaloïdes extraits du Strychnos usambarensis sur des cellules cancéreuses en culture. *Planta Med* 1982;45:123–126.
[64] Tits M, Desaive C, Marnette JM, Bassleer R, Angenot L. Antimitotic activity of strychnopentamine, a bisindolic alkaloid. *J Ethnopharmacol* 1984;12:287–292.
[65] Leclercq J, Gillet MC, Bassleer R, Angenot L. Screening of cytotoxic activities of Strychnos alkaloids (methods and results). *J Ethnopharmacol*, 1986;15:305-316.
[66] Bonjean KA, Gillet MC, Quentin-Leclercq J, Angenot L, Bassleer R. In vitro cytotoxic activity of two potential anticancer drugs isolated from Strychnos: strychnopentamine and usambarensine. *J Anticancer Res*, 1996;16:1129-1137.
[67] Quetin-Leclercq J, Bouzahzah B, Pons A, Greimers R, Angenot L, Bassleer R., Barhason H. Strychnopentamine, a potential anticancer agent. *Planta Med* 1993;59:59-62.
[68] Frédérich M, Tits M, Goffin E, Philippe G, Grellier P, De Mol P, Hayette MP, Angenot L. In vitro and in vivo antimalarial properties of isostrychnopentamine, an indolomonoterpenic alkaloid from Strychnos usambarensis. *Planta Medica* 2004;70:520-525.
[69] Frédérich M, Bentires-Alj M, Tits M, Angenot L, Greimers R, Gielen J, Bours V, Merville MP. Isostrychnopentamine, an indolomonoterpenic alkaloid from Strychnos usambarensis, induces cell cycle arrest and apoptosis in human colon cancer cells. *J Pharmacol Exp Ther* 2003;304:1103-1110.
[70] Balde EHS, Mégalizzi V, Cao M, Angenot L, Kiss R, Van Damme M, Frédérich M. Isostrychnopentamine, an indolomonoterpenic alkaloid from Strychnos usambarensis, with potential anti-tumor activity apoptosis-resistant cancer cells. *Int J of Oncol* 2010;36:961-965.
[71] Gadi Biala R, Tits M, Wauters JN, Angenot L. A new HPLC method for the assay of alkaloids in Strychnos nux-vomica and Strychnos ignatii. *Fitoterapia* 1996;62:163-165.
[72] Frédérich M, Tits M, Angenot L. Qualitative and quantitative evaluation of bisindole usambarane alkaloids in Strychnos usambarensis roots by high performance liquid chromatography-diode-array. *Phytochem Anal*, 1998;9:63–66.

[73] Debrus B, Lebrun P, Ceccato A, Caliaro G, Rozet E, Nistor I, Oprean R, Rupérez FJ, Barbas C, Boulanger B, Hubert P. Application of new methodologies based on design of experiments, independent component analysis and design space for robust optimization in liquid chromatography. *Anal Chim Acta* 2011;691:33-42.

[74] Debrus B, Lebrun P, Rozet E, Nistor I, Ceccato A, Caliaro G, Oprean R, Boulanger B, Hubert P. Nouvelle méthodologie pour le développement automatisé de méthodes analytiques en chromatographie liquide pour l'analyse de mélanges de composés inconnus. *Spectra Analyse* 2009;268:28-33.

[75] LoBrutto R, Jones A, Kazakevich YV, McNair HM. Effect of the eluent pH and acidic modifiers in high-performance liquid chromatography retention of basic analytes. *J Chromatogr A* 2001;913:173-187.

[76] Kazakevich YV, LoBrutto R, Vivilecchia R. Reversed-phase high-performance liquid chromatography behaviour of chaotropic counteranions. *J Chromatogr A* 2005;1064:9-18.

[77] Flieger J. The effect of chaotropic mobile phase additives on the separation of selected alkaloids in reversed-phase high-performance liquid chromatography *J Chromatogr A* 2006;1113:37-44.

[78] Flieger J. Effect of mobile phase composition on the retention of selected alkaloids in reversed-phase liquid chromatography with chaotropic salts. *J Chromatogr A*, 2007;1175:207-216.

[79] Myers RH, Montgomery DC, Anderson-Cook CM. *Response surface methodology: process and product optimization using designed experiments*. New York:Wiley;2009.

[80] Lebrun P, Boulanger B, Lambert P. Use of Bayesian multivariate prediction models to optimize chromatographic methods. [Oral communication] in Bayes 2010, Braine l'Alleud, Belgium.

[81] Lebrun P, Boulanger B. Expected Design Space: a Bayesian perspective based on modelling, prediction and multi-criteria decision method. [Oral communication] in NCB 09, Boston, USA.

[82] Vanbel PF, Tilquin BL, Schoenmakers PJ. Criteria for optimizing the separation of target analytes in complex chromatograms. *Chemometr Intell Lab*, 1996;35:67-86.

[83] Wang G, Ding Q, Hou Z. Independent component analysis and its applications in signal processing for analytical chemistry. *Trends Analyt Chem* 2008;27:368-376.

[84] R Development Core Team. *R: A Language and Environment for Statistical Computing*; 2010.

[85] Rathore AS, Mhatre R. *Quality by Design for biopharmaceuticals: Principles and case studies.* New York:Wiley, John&Sons;2009.

[86] Morante-Zarcero S, del Hierro I, Fajardo M, Sierra I. HPLC with polysaccharide chiral stationary phase in polar-organic phase mode: application to the asymmetric epoxidation of allylic alcohols. *J Sep Sci* 2009;32:3055-63.

[87] He BL, Shi Y, Kleintop B, Raglione T. Direct and indirect separations of five isomers of Brivanib Alaninate using chiral high-performance liquid chromatography. *J Chromatogr B* 2008;875:122-135.
[88] Ates H, Mangelings D, Heyden YV. Fast generic chiral separation strategies using electrophoretic and liquid chromatographic techniques. *J Pharm Biomed Anal* 2008;48:288-294.
[89] Ye YK, Stringham R, Wirth MJ. Origin of enhanced chiral selectivity by acidic additives for a polysaccharide-based stationary phase. *J Chromatogr A* 2004;1057:75-82.
[90] Lynam KG, Stringham RW. Chiral separations on polysaccharide stationary phases using polar organic mobile phases. *Chirality* 2006;18:1-9.
[91] Younes AA, Mangelings D, Heyden YV. Chiral separations in normal-phase liquid chromatography: Enantioselectivity of recently commercialized polysaccharide-based selectors. Part II. Optimization of enantioselectivity. *J Pharm Biomed Anal* 2011;56:521-537.
[92] Peng L, Jayapalan S, Chankvetadze B, Farkas T. Reversed-phase chiral HPLC and LC/MS analysis with tris(chloromethylphenylcarbamate) derivates of cellulose and amylose as chiral stationary phases. *J Chromatogr A* 2010;1217:6942-55.
[93] Chankvetadze B, Yashima E, Okamoto Y. Chloro-methyl-phenylcarbamate derivatives of cellulose as chiral stationary phases for high performance liquid chromatography. *J Chromatogr A* 1994;670:39-49.
[94] Chankvetadze B, Kartozia I, Yamamoto C, Okamoto Y. Comparative enantioseparation of selected chiral drugs on four different polysaccharide-type chiral stationary phases using polar organic mobile phases. *J Pharm Biomed Anal* 2002;27:467-478.
[95] Dossou KSS, Chiap P, Chankvetadze B, Servais A-C, Fillet M, Crommen J. Enantioresolution of basic pharmaceuticals using cellulose tris(4-chloro-3-methylphenylcarbamate) as chiral stationary phase and polar organic mobile phases. *J Chromatogr A* 2009;1216:7450-7455.
[96] Ates H, Mangelings D, Heyden YV. Chiral separations in polar organic solvent chromatography: Updating a screening strategy with new chlorine-containing polysaccharide-based selectors. *J Chromatogr B* 2008;875:57-64.
[97] Nistor I, Cao M, Debrus B, Lebrun P, Lecomte F, Rozet E, Angenot L, Frederich M, Oprean R, Hubert P. Application of a new optimization strategy for the separation of tertiary alkaloids extracted from Strychnos usambarensis leaves. *J Pharm Biomed Anal* 2011;56:30-37.
[98] Caccamese S, Bianca S, Carter GT. Direct High-Performance Liquid Chromatographic separation of the enantiomersof an aromatic amine and four aminoalcohols using polysaccharide chiral stationary phases and acidic additive. *Chirality* 2007;19:647-653.

[99] Matthijs N, Maftouh M, Heyden YV. Screening approach for chiral separation of pharmaceuticals IV. Polar organic solvent chromatography. *J Chromatogr A* 2006;1111:48-61.
[100] Lebrun P, Krier F, Mantanus J, Grohganz H, Yang M, Rozet E, Boulanger B, Evrard B, Rantanen J, Hubert Ph. Design space approach in the optimization of the spray-drying process. *Eur J Pharm Biopharm* 2012;80:226-234.
[101] Debrus B, Lebrun P, Mbinze Kindenge J, Lecomte F, Ceccato A, Caliaro G, Mavar Tayey Mbay J, Boulanger B, Marini RD, Rozet E, Hubert Ph. Innovative high-performance liquid chromatography method development for the screening of 19 antimalarial drugs based on a generic approach, using design of experiments, independent component analysis and design space. *J Chromatogr A* 2011;1218:5205-5215.
[102] Vogt FG, Kord AS. Development of Quality-by-Design analytical methods. *J Pharm Sci* 2011;100:797-812.
[103] Degerman M, Westerberg K, Nilsson B. Determining critical process parameters and process robustness in preparative chromatography – A model-based approach. *Chem Eng Technol* 2009;32:903-911.
[104] Dossou KSS, Farcas E, Servais A-C, Chiap P, Chankvetadze B, Crommen J, Fillet M. Optimization of the liquid chromatography enantioseparation of chiral acidic compounds using cellulose tris(3-chloro-4methylphenylcarbamate) as chiral selector and polar organic mobile phases. *J Chromatogr A* 2012;1234:56-63.
[105] Peterson JJ, Kenett RS. Modeling opportunities for statisticians supporting Quality by Design efforts for pharmaceutical development and manufacturing. *Biopharm Rep* 2011;6-13.