Journal of Chromatography A, 1218 (2011) 5205–5215

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

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

B. Debrus[a,*,1], P. Lebrun[a,1], J. Mbinze Kindenge[a,b], F. Lecomte[a], A. Ceccato[c], G. Caliaro[d], J. Mavar Tayey Mbay[b], B. Boulanger[e], R.D. Marini[a], E. Rozet[a,2], Ph. Hubert[a]

[a] Laboratory of Analytical Chemistry, CIRM, Department of Pharmacy, University of Liège, Avenue de l'Hôpital 1, B36, B-4000 Liège, Belgium
[b] Service d'Analyse des Médicaments, Département de Galénique et d'Analyse des Médicaments, Université de Kinshasa, B-212 Kinshasa XI, Democratic Republic of the Congo
[c] Odyssea Pharma, Rue du Travail 16, B-4460 Grace-Hollogne, Belgium
[d] Orailac Quality Solutions, Belgium
[e] Arlenda Laboratory Solutions, Avenue de l'Hôpital 1, B-4000 Liège, Belgium

ARTICLE INFO

Article history:
Received 24 March 2011
Received in revised form 30 May 2011
Accepted 30 May 2011
Available online 12 June 2011

Keywords:
Antimalarial drugs
Counterfeit
Design of experiments
Design space
Quality by design
Independent component analysis

ABSTRACT

An innovative methodology based on design of experiments (DoE), independent component analysis (ICA) and design space (DS) was developed in previous works and was tested out with a mixture of 19 antimalarial drugs. This global LC method development methodology (i.e. DoE–ICA–DS) was used to optimize the separation of 19 antimalarial drugs to obtain a screening method. DoE–ICA–DS methodology is fully compliant with the current trend of quality by design. DoE was used to define the set of experiments to model the retention times at the beginning, the apex and the end of each peak. Furthermore, ICA was used to numerically separate coeluting peaks and estimate their unbiased retention times. Gradient time, temperature and pH were selected as the factors of a full factorial design. These retention times were modelled by stepwise multiple linear regressions. A recently introduced critical quality attribute, namely the separation criterion (*S*), was also used to assess the quality of separations rather than using the resolution. Furthermore, the resulting mathematical models were also studied from a chromatographic point of view to understand and investigate the chromatographic behaviour of each compound. Good adequacies were found between the mathematical models and the expected chromatographic behaviours predicted by chromatographic theory. Finally, focusing at quality risk management, the DS was computed as the multidimensional subspace where the probability for the separation criterion to lie in acceptance limits was higher than a defined quality level. The DS was computed propagating the prediction error from the modelled responses to the quality criterion using Monte Carlo simulations. DoE–ICA–DS allowed encountering optimal operating conditions to obtain a robust screening method for the 19 considered antimalarial drugs in the framework of the fight against counterfeit medicines. Moreover and only on the basis of the same data set, a dedicated method for the determination of three antimalarial compounds in a pharmaceutical formulation was optimized to demonstrate both the efficiency and flexibility of the methodology proposed in the present study.

1. Introduction

Malaria remains one of the most extended illnesses worldwide. Over past decades, improper use of antimalarial drugs contributed to widen resistance against malaria parasite to several drugs (e.g. chloroquine) [1,2]. Nevertheless, artemisinin-based combination therapies (ACT) bring new hopes in the fight against malaria [3–6]. Furthermore, in some African countries, up to 80% of medical products are counterfeit [7]. In this context, analytical chemistry and especially chromatographic methods can help to fight against counterfeit medicines.

In this perspective, a HPLC-UV method was developed for the screening of 19 active antimalarial ingredients (AAI): amodiaquine (AQ), arteether (AE), artemether (AM), artemisinin (ART), artesunate (AS), atovaquone (AT), chloroquine (CQ), cinchonine (CC), dihydroartemisinin (DHA), halofantrine (HF), lumefantrine (LF),

* Corresponding author at: Laboratory of Analytical Chemistry, Department of Pharmacy, University of Liège, CHU, B36, Avenue de l'Hôpital 1, B-4000 Liège, Belgium. Tel.: +32 4366 4324; fax: +32 4366 4317.
E-mail address: b.debrus@ulg.ac.be (B. Debrus).
[1] These authors have equally contributed to this article.
[2] Postdoctoral Researcher F.R.S.-FNRS (Belgium).

0021-9673/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2011.05.102

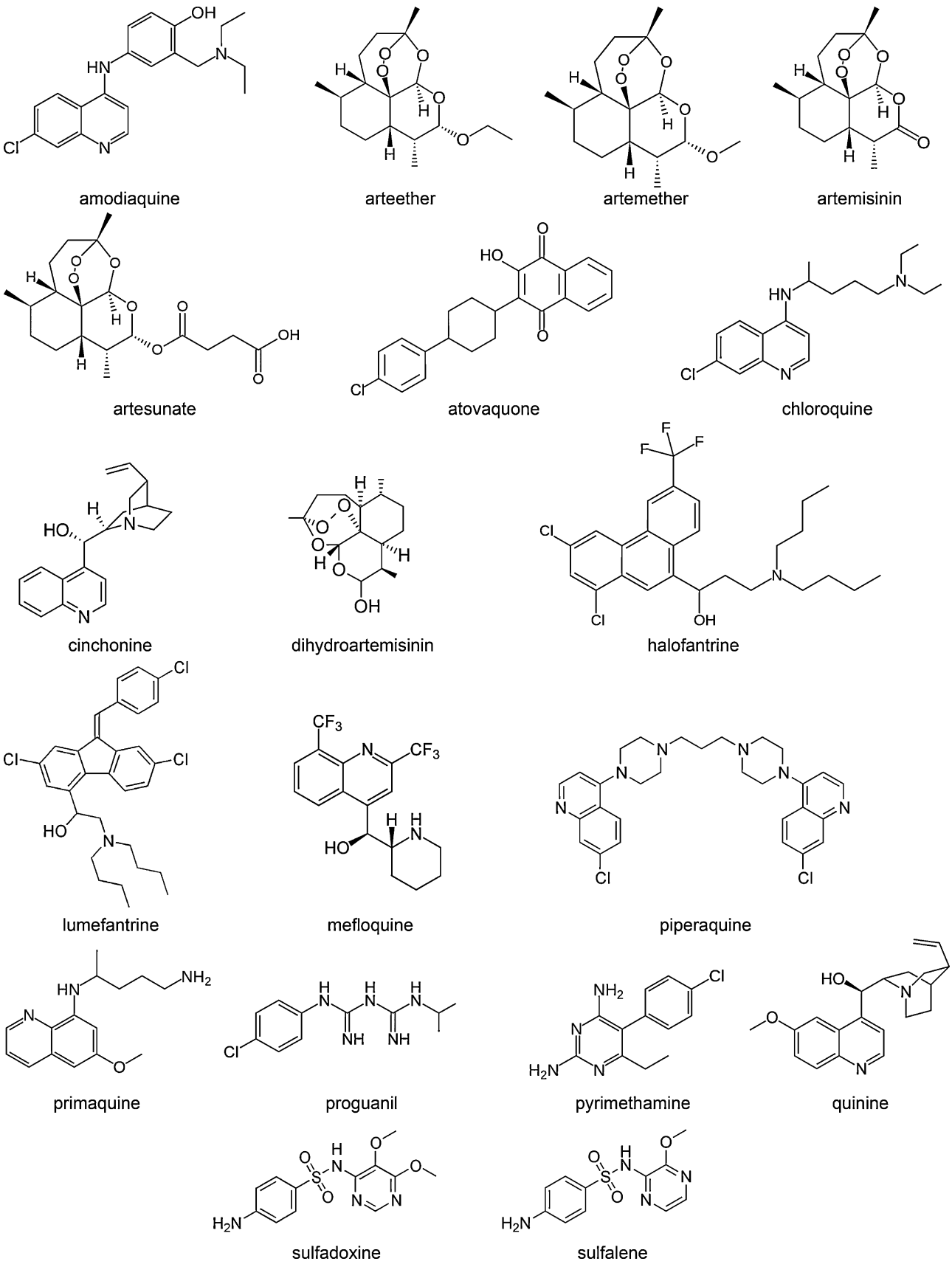

Fig. 1. Chemical structures of the 19 studied antimalarial drugs.

Table 1
pKa and log *P* (at 25 °C) found on SciFinder®, calculated using Advanced Chemistry Development (ACD/Labs©) Software V11.02.

Compounds	pKa		log *P*	Compounds	pKa		log *P*
	Most acidic	Most basic			Most acidic	Most basic	
Amodiaquine	9.43 ± 0.50	5.62 ± 0.50	3.126 ± 0.840	Lumefantrine	13.44 ± 0.20	8.71 ± 0.50	8.671 ± 0.405
Arteether	NA	NA	3.330 ± 0.864	Mefloquine	12.81 ± 0.20	9.24 ± 0.10	2.197 ± 1.149
Artemether	NA	NA	2.820 ± 0.864	Piperaquine	NA	8.92 ± 0.50	6.796 ± 1.413
Artemisinin	NA	NA	2.269 ± 0.680	Primaquine	NA	10.38 ± 0.10	2.740 ± 0.255
Artesunate	4.28 ± 0.17	NA	3.291 ± 0.883	Proguanil	NA	11.15 ± 0.10	2.485 ± 0.263
Atovaquone	5.01 ± 0.10	NA	6.465 ± 0.729	Pyrimethamine	NA	7.18 ± 0.10	2.750 ± 0.328
Chloroquine	NA	10.47 ± 0.25	4.412 ± 0.758	Quinine	12.80 ± 0.20	9.28 ± 0.70	2.823 ± 0.431
Cinchonine	12.98 ± 0.20	9.33 ± 0.70	2.788 ± 0.415	Sulfadoxine	6.16 ± 0.50	2.18 ± 0.10	0.460 ± 0.419
Dihydroartemisinin	12.61 ± 0.70	NA	2.190 ± 0.859	Sulfalene	6.61 ± 0.40	1.46 ± 0.10	0.880 ± 0.456
Halofantrine	13.57 ± 0.20	9.44 ± 0.50	8.902 ± 1.302				

mefloquine (MQ), piperaquine (PPQ), primaquine (PQ), proguanil (PG), pyrimethamine (PM), quinine (QN), sulfadoxine (SD) and sulfalene (SL). Chemical structures are presented in Fig. 1, calculated pKa and log *P* are given in Table 1.

Nowadays, HPLC method development can be achieved using different methodologies. Some have already led to some commercial softwares (e.g. Drylab, ACD/LC simulator, Chromsword, Osiris). These softwares make use of chromatography-based theory such as solvophobic theory and linear solvent strength relationship, to optimize the separation of sample mixtures while maintaining the number of test experiments to a minimum [8–14]. These strategies are generally fast and efficient. Nevertheless, in the current trend of Quality by Design (QbD), these softwares do not provide the ability to advisedly compute or estimate the robustness, also sometimes called ruggedness in some regulatory documents. Consequently, classical robustness tests have to be carried out at the end of the method validation phase to estimate the method ability to remain unaffected by small, but deliberate variations in method parameters [15–17]. In the present work, a distinct and innovative methodology combining design of experiments (DoE), independent component analysis (ICA) and design space (DS) – with regard to ICH Q8(R2) [18] – was used to simultaneously optimize the separation and estimate the method robustness over the whole experimental domain (i.e. the knowledge space) instead of around the optima only.

In the ICH pharmaceutical development guidelines, Q8(R2) [18], 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*". Therefore, the multidimensional combination and interaction of input variable correspond to a subspace, so-called the DS, where assurance of quality has to be proved. Thus, the DS is necessarily encompassed within the experimental domain which is the multidimensional space formed by the factor ranges used during method development. The main concept lying behind the ICH Q8(R2) definition of DS is assurance of quality (also known as quality risk management). Hence, a HPLC method development process which does not take into account the prediction error (i.e. the prediction accuracy [19] estimated from the experimental data) in order to manage the risk cannot be considered as QbD compliant and will not allow to identify or compute the DS. Even if ICH Q8(R2) is the most appropriated guideline when discussing about QbD and DS applied to pharmaceutical sciences, the given DS definition and the examples shown at the end of the document are divergent. It seems that the identification of the multidimensional zone where a quality criterion lies in given acceptance limits or is higher or lower than an advisedly selected threshold do not define the DS which have to take into account assurance of quality. In other terms, in LC, to predict the multidimensional region where the resolution (R_S) is higher than 1.5, does not define a DS as only quality is predicted but not assurance of quality. Indeed, the risk assessment of not being within the acceptance limits is not carried out. On the other hand, using the probability for a given quality criterion to be in the acceptance limits is a way to define DS (e.g. $P(R_S \geq 1.5)$). When computing such a probability, the quality risk management is carried out. Interesting discussions about QbD and DS applied in LC [20] and in pharmaceutical development [21–23] were already published in the scientific literature. Some DoE–DS LC applications were also previously published [24–27].

Therefore, the determination of the DS for a LC method development implies to consider also the error on the studied responses and criteria. The variability of the retention times has to be taken into account during the development phase. These considerations hold completely with the QbD definition [18]. Furthermore, ICH Q8(R2) guideline states that "*working within the design space is not considered as a change. Movement out of the design space is considered to be a change and would normally initiate a regulatory post approval change process*". Consequently, the DS defines a zone of robustness as no significant changes in terms of separation quality should be observed on the resulting chromatograms.

Therefore, in this work, an innovative methodology using DoE and ICA was used to optimize the separation and identify the DS for the above mentioned AAIs. The present study is a useful and relevant application of complementary strategies previously published [24,28]. The overall objective of the present work was both to demonstrate the ability of the DoE–ICA–DS methodology to provide optimal and robust HPLC method and to apply it as a case study to the development of a useful method for the screening of 19 antimalarial drugs in the framework of the fight against counterfeit medicines.

2. Experimental

2.1. Chemicals and reagents

Methanol (HPLC gradient grade), hydrochloric acid, ammonium hydroxide and ammonium hydrogen carbonate (99%) were purchased from Merck (Darmstadt, Germany). Ammonium formate (99%) was provided by Alfa Aesar (Karlsruhe, Germany). Ultrapure water was obtained from a Milli-Q Plus 185 water purification system from Millipore (Billerica, MA, USA).

Artesunate (99.8%), arteether (99.5%), artemisinin (99.6%), artemether (99.8%) and dihydroartemisinin (100.0%) were purchased from Apoteket AB (Stockholm, Sweden). Lumefantrine was gratefully provided by Novartis International Pharmaceutical Ltd. (Villevorde, Belgium). Amodiaquine hydrochloride (99%), chloroquine diphosphate (98%), cinchonine (99.9%), mefloquine hydrochloride (99%), quinine hydrochloride (96.1%) and sulfadoxine (99.9%) were purchased from Sigma Aldrich (St. Louis, MO, USA). Atovaquone (98.9%), proguanil hydrochloride (99.9%) and halofantrine were gratefully provided by GlaxoSmithKline (Genval, Belgium). Piperaquine was extracted from the P-Gvaxin formu-

lation from Bliss GVS Pharma (Mumbai, India). Sulfalene was purchased from Fagron NV/SA (Waregem, Belgium). Finally, Arteplus® pharmaceutical formulation was obtained from KIM Pharma Ets. (Kinshasa, D.R. Congo).

2.2. Standard samples preparation

A mixture of the 19 AAIs was used and was prepared as follow. In a first step, 10 mg of AQ, AT, CC, CQ, HF, LF, QN, MQ, PG, PM, PPQ, PQ, SD and SL, were dissolved in a 100.0 mL volumetric flask with methanol. This solution was annotated S1. In a second step, 10 mg of AE, AM, ART, AS and DHA were dissolved in a 2.0 mL volumetric flask with a mixture S1-methanol (50:50, v/v). An aliquot of this solution was filtered with 0.20 μm PTFE syringe filtration disks into a vial for injection in the HPLC system.

Other groups of compounds were prepared in the same way. Group 1 (see Section 3.3) contained AE, AM, AQ, ART, AS, AT, CC, CQ, HF, LF, QN, MQ, PG, PM, PPQ, PQ, SD and SL. Group 2 (see Section 3.3) contained AE, AM, AQ, AS, AT, CC, CQ, DHA, HF, LF, QN, MQ, PG, PM, PPQ, PQ, SD and SL. One sub-mixture was also used (see Section 3.5) and contained AS, PM and SL only.

As the present screening method was developed for medicine control, concentrations of analytes were set as close as possible of classical pharmaceutical formulation contents.

2.3. Experiments

Chromatographic experiments were performed on a Waters 2695 separation module coupled to a Waters selector valve 7678 and a Waters 996 Photodiode array detector (Waters, Milford, MA, USA). pH measurements were performed with a SevenEasy S20 pH meter (Mettler Toledo, Columbus, OH, USA). Buffers consisted in 10 mM pH 2.5 formic acid, pH 4.0 ammonium formate, pH 6.0 ammonium formate, pH 8.0 ammonium hydrogen carbonate and pH 10.0 ammonium hydrogen carbonate. The pH was adjusted to the desired value with concentrated formic acid or ammonia 35% aqueous solution.

Chromatograms were recorded at wavelengths ranging from 210 nm to 400 nm with a step of 1.2 nm and with an acquisition time of 500 ms. Separations were carried out on an XBridge C18 250 mm × 4.6 mm, 5 μm, combined with an XBridge guard column C18 20 mm × 4.6 mm, 5 μm, both provided by Waters. Experiments were carried out at a flow rate of 1.0 mL/min and injection volume was 10 μL. Peaks were integrated at 230 nm.

The choice of the chromatographic column is justified by two main reasons. First, the chemistry of the column (i.e. trifunctionally bonded C18) allows a relatively good retention for most of the organic compounds. Second, the silica matrix-ethylene bridges and the octadecyl triple bonding strengthen the silica stability in very acidic and alkaline media. Therefore, this column was selected to obtain a generic methodology (i.e. usable for most mixtures in pharmaceutical analytical sciences).

2.4. Design of experiments

In order to model the chromatographic behaviour of each peak, a full factorial design was selected. It comprised three factors: pH of the aqueous part of the mobile phase (pH), gradient time (t_G) to linearly modify to proportion of methanol from 5% to 95%, and column temperature (T). Factors and their respective levels are summarized in Table 2. A total of 45 experimental conditions (5×3^2) were defined by this full factorial design. In the present case, a full factorial design was used to simultaneously optimize the method, estimate its robustness and evaluate the adequacy between chromatographic behaviours predicted by the theory of liquid chromatography and those obtained by the mathematical models. On the other hand, if method optimization is the unique objective, lighter DoE can be envisaged (e.g. fractional factorial or central composite designs). As shown in Table 2, temperature range was quite narrow. This factor was investigated to test out how the method robustness depended on this factor rather than optimizing the separation while varying T. As the resulting screening HPLC method should be used in different African countries with various or more frequently no column oven, it is of first importance to evaluate the effect of this factor on the studied responses. Moreover, the temperature could not be higher than 35 °C because some molecules were found to be relatively unstable at higher temperature. Gradient time and pH ranges were expanded as much as possible in order to widen the experimental domain and to minimize the risk of not finding any good separation within it. XBridge C18 columns can sustain pH from 1 to 12. pH range was slightly narrowed from 2.5 to 10 in order to maintain suitable column lifetime. Gradient time range was also wide (from 20 to 60 min). These factors were selected to test out their effects on the selectivity (a priori, mainly by modifying pH) and to shorten the time of analysis (a priori, mainly by decreasing t_G while preserving acceptable separation).

Table 2
Factors and levels of the full factorial design.

Factors	Levels				
pH	2.5	4.0	6.0	8.0	10.0
Gradient time (t_G, min)	20	40	60		
Temperature (T, °C)	25	30	35		

The experiments at a same pH were carried out in row for evident practical reasons. Then, the within pH blocks experiments were conducted in a random order. It is preferable to carry out the experiments in a totally random order to avoid experimental biases. Nevertheless, the column equilibration and conditioning times when constantly changing mobile phase pH drastically increase the time devoted to achieve the DoE results. Furthermore, the pH measurement error can be assumed to be equal to 0.1%. Other error sources (i.e. mobile phase composition during gradient, temperature, etc.) generated higher response errors and the pH blocking did not lead to poor predictive errors. pH blocking effect was negligible in the present study.

The central point (i.e. pH 6.0, t_G = 40 min, T = 30 °C) was independently repeated twice (i.e. carried out thrice) with the preparation of new buffer and fresh mobile phase. The central points for lower and higher temperatures (i.e. pH 6.0, t_G = 40 min, T = 35 °C and pH = 6.0, t_G = 40 min, T = 25 °C, respectively) were also repeated once (i.e. carried out twice).

2.5. Independent component analysis

ICA is a statistical method allowing the numerical separation of sources maximizing the independence between them based on non-Gaussianity [29]. In chromatography, ICA was already used to numerically separate coeluting peaks in order to estimate their integration limits (i.e. the times at the beginning and end of a peak) [27,28]. Indeed, for coeluting peaks, when using a drop-line valley separator, the estimation of integration limits is highly biased. Then, the modelling of responses based on these biased times could lead to poor prediction accuracy. In order to avoid this situation, ICA was used to numerically separate coeluting peaks of antimalarial drugs. The integration limits of non-coeluting peaks were obtained manually on the chromatograms.

2.6. Modelling and optimization methodology

2.6.1. Retention times modelling

Dewé et al. [30] and Lebrun et al. [24] provided a new approach for retention times modelling and DS computation. In these works, they showed that the modelling of the resolution ($R_{S,crit} = 2(t_{R,2} - t_{R,1})/(w_{b,1} + w_{b,2})$; with $t_{R,2} > t_{R,1}$ and $w_{b,1}$, $w_{b,2}$ are the baseline peaks widths; $t_{R,1}$ and $t_{R,2}$ being the retention times of the critical pair peaks (i.e. the two most proximate peaks)) can lead to poor prediction caused by its non-linear and non-continuous behaviour when selectivity drastically changes.

Hence, the retention times at the beginning, the apex and the end of each peak (respectively t_B, t_R and t_E) were measured. The studied responses were the logarithm of the retention factor (i.e. $\log(k_{tR})$ with $k_{tR} = (t_R - t_0)/t_0$, t_0 being the column dead time) and the logarithm of both half-widths (i.e. $\log(w_l)$ and $\log(w_r)$; with $w_l = k_{tR} - k_{tB}$ and $w_r = k_{tE} - k_{tR}$). These responses were modelled by an identical polynomial equation (see Eq. (1)) using a multiple linear stepwise regression maximizing the adjusted coefficient of determination (R^2_{adj}).

$$\log(k_{tR}) = \beta_0 + \beta_1 \cdot \mathrm{pH} + \beta_2 \cdot \mathrm{pH}^2 + \beta_3 \cdot \mathrm{pH}^3 + \beta_4 \cdot t_G + \beta_5 \cdot t_G^2 + \beta_6 \cdot T + \beta_7 \cdot T^2 + \beta_8 \cdot \mathrm{pH} \cdot t_G + \beta_9 \cdot \mathrm{pH} \cdot T + \beta_{10} \cdot t_G \cdot T + \beta_{11} \cdot \mathrm{pH} \cdot t_G \cdot T + \varepsilon \quad (1)$$

where $\beta_0 \ldots \beta_{11}$ are the estimated parameters of the model and ε is the estimated error of the model.

2.6.2. Quality criterion, error propagation and design space computation

The retention times were predicted using Eq. (1). The prediction error was considered as a Gaussian distribution centred on 0 and with a standard deviation equal to the standard deviation of the residuals distribution. Thus, for each operating condition, the predicted responses were normally distributed and the predicted retention time for a compound followed a log-Normal distribution.

A quality criterion, also known as critical quality attribute, was selected to assess the quality of a separation. Lebrun et al. [24] proposed to use the separation criterion (S) defined as the difference between the beginning of the second peak and the end of the first peak of the critical pair ($S_{crit} = t_{B,2} - t_{E,1}$; with $t_{B,2} > t_{E,1}$; $t_{E,1}$ and $t_{B,2}$ being the end time and the beginning time of the critical pair peaks, respectively). The separation criterion is clearly easier to compute and to interpret. If $S_{crit} \geq 0$, critical pair peaks are baseline-resolved.

Then, the prediction errors obtained for the retention times are propagated to the quality criterion. Practically, Monte Carlo simulations were used to obtain, for a given operating condition, the distribution of S from the distributions of t_R, t_B and t_E of the critical pair peaks. Finally, using the S distribution, the probability for S to be higher than a selected threshold was used to determine the DS. In mathematical terms, the DS can be defined by Eq. (2).

$$\mathrm{DS} = \{\mathbf{x}_0 \in \chi | E_{\hat{\theta}}[P(S > \lambda)|\hat{\theta}] \geq \pi\} \quad (2)$$

where $\mathbf{x}_0$ is a point of the experimental domain (χ). λ is the acceptance limit for criterion S, π is the quality level and $\hat{\theta}$ is the set of estimated parameters of the model. P and E respectively correspond to the estimators of probability and mathematical expectation. In the following sections, λ was set at 0 or 0.1 min.

In other words, the DS defines a subspace where the probability to obtain baseline-resolved peaks (i.e. $S > 0$ min) is higher than a predefined quality threshold (e.g. $\pi = 65\%$). In practice, π was selected to be 85% of the optimal probability to have $S > 0$. No regulatory document yet provides guidelines on how to compute or estimate the design space quality level [27]. Nevertheless, a high $P(S > 0)$ induces the achievement of robust chromatographic methods.

2.6.3. Prediction of optimal separation

The experimental domain was investigated with a grid search method. A multidimensional grid was defined over the experimental domain. Then, the quality criterion value and its associated prediction error (the distribution of S obtained propagating the error using Monte Carlo simulations) were computed for each of the experimental condition defined by the grid. The optimum was selected as the point of the grid giving the highest probability value. In practice, the number of points was set as high as possible while keeping the total computing time lower than 12 h (i.e. one night computation).

2.7. Software

An in-house computer program was developed to perform the retention times modelling with stepwise multiple linear regressions, the error propagation and the grid search method. The coding was carried out with R 2.11.1 [31]. ICA-based numerical separations were performed using FastICA algorithm [32].

3. Results and discussion

3.1. Peak detection and peak matching

As the compounds from the artemisinin group (i.e. AE, AM, AS and DHA) present very similar and non-specific UV spectra, these five molecules were injected individually to identify and match them. Then, for the rest of the compounds, in case of coelution, ICA was used to determine the times at the beginning (t_B), the apex (t_R, the retention time) and the end (t_E) of each peak [28]. Finally, for the non-coeluting peaks, these times were manually read on the chromatograms. Examples of coeluted peaks numerical separation using ICA can be found in previous works [27,28].

Furthermore, at alkaline pH and high temperature (pH 10, $T = 35\,^{\circ}C$), DHA peak was split in two coeluting peaks and several unidentified peaks were also observed. These results suggested that DHA degraded in alkaline conditions at 35 °C. Therefore, pH 10 experiments were not used for responses modelling.

3.2. Retention times modelling

Retention times modelling were achieved by stepwise multiple linear regressions which selected the terms of Eq. (1) to maximize R^2_{adj}. As three times (t_B, t_R and t_E) of 19 peaks were modelled, 54 models were obtained. The estimated model parameters ($\beta_0 \ldots \beta_{11}$) are presented in Table 3 and the R^2_{adj} are summarized in Table 4.

In the present study, the terms were considered highly significant when their P-value was smaller than 0.001 ($\alpha = 0.1\%$). In Table 3, some P-values can seem high but the corresponding terms were selected as they increased R^2_{adj} which was selected as the stepwise regression criterion [33].

The model parameters presented in Table 3 represent the chromatographic behaviour of the studied compounds. Therefore, the model parameters can be used to corroborate the chromatographic behaviours thanks to some physico-chemical properties such as pKa values. These interpretations are listed in following Sections 3.2.1 and 3.2.3). The objective of these considerations is not to highlight some quantitative structure retention relationships but to understand the resulting models thanks to some chromatographic behaviour.

Table 3
Estimated model parameters ($\beta_0 \ldots \beta_{11}$, see Eq. (1)) and *P*-values (Pr(>|*t*|)) of their significance tests for log(k_{tR}) modelling of the studied compounds: amodiaquine (AQ), arteether (AE), artemether (AM), artemisinin (ART), artesunate (AS), atovaquone (AT), chloroquine (CQ), cinchonine (CC), dihydroartemisinin (DHA), halofantrine (HF), lumefantrine (LF), mefloquine (MQ), piperaquine (PPQ), primaquine (PQ), proguanil (PG), pyrimethamine (PM), quinine (QN), sulfadoxine (SD) and sulfalene (SL).

Compounds	AQ		AE		AM		ART		AS		AT		CQ	
Term	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)
Intercept	4.557	**<0.001**	5.323	**<0.001**	5.282	**<0.001**	5.107	**<0.001**	5.081	**<0.001**	5.358	**<0.001**	4.323	**<0.001**
pH	0.594	**<0.001**	−0.004	0.468	0.000	0.338	0.001	0.558	−0.120	**<0.001**	−0.050	**<0.001**	0.212	**<0.001**
pH^2	0.173	**<0.001**	NS	NS	0.001	0.112	0.001	0.300	−0.023	**<0.001**	−0.077	**<0.001**	0.276	**<0.001**
pH^3	−0.195	**<0.001**	0.006	0.327	NS	NS	−0.001	0.712	0.039	**<0.001**	−0.041	**<0.001**	0.114	**<0.001**
t_G	0.336	**<0.001**	0.439	**<0.001**	0.428	**<0.001**	0.406	**<0.001**	0.406	**<0.001**	0.445	**<0.001**	0.309	**<0.001**
t_G^2	−0.081	**<0.001**	−0.103	**<0.001**	−0.103	**<0.001**	−0.096	**<0.001**	−0.096	**<0.001**	−0.106	**<0.001**	−0.073	**<0.001**
T	−0.016	0.004	−0.006	**<0.001**	−0.011	**<0.001**	−0.013	**<0.001**	−0.016	**<0.001**	−0.010	**<0.001**	−0.021	**<0.001**
T^2	NS	NS	−0.005	0.052	−0.002	0.006	−0.003	0.006	−0.003	0.207	−0.001	0.457	NS	NS
pH × t_G	0.053	**<0.001**	NS	NS	0.000	0.610	0.001	0.401	−0.010	**<0.001**	−0.007	**<0.001**	0.043	**<0.001**
pH × T	0.019	0.010	NS	NS	0.001	0.296	NS	NS	−0.003	0.103	NS	NS	0.015	0.036
t_G × T	NS	NS	−0.004	0.042	−0.002	**<0.001**	−0.002	0.004	−0.003	0.115	−0.001	0.497	NS	NS
pH × t_G × T	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS

Compounds	CC		DHA		HF		LF		MQ		PPQ	
Term	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)
Intercept	4.673	**<0.001**	5.112	**<0.001**	5.329	**<0.001**	5.450	**<0.001**	5.077	**<0.001**	5.197	**<0.001**
pH	0.039	0.010	0.006	0.018	0.101	**<0.001**	0.253	**<0.001**	0.023	**<0.001**	0.428	**<0.001**
pH^2	−0.058	**<0.001**	0.013	**<0.001**	0.074	**<0.001**	0.043	**<0.001**	0.052	**<0.001**	−0.576	**<0.001**
pH^3	0.352	**<0.001**	0.004	0.079	0.011	0.178	−0.071	**<0.001**	0.032	**<0.001**	0.244	**<0.001**
t_G	0.337	**<0.001**	0.405	**<0.001**	0.435	**<0.001**	0.408	**<0.001**	0.422	**<0.001**	0.375	**<0.001**
t_G^2	−0.077	**<0.001**	−0.097	**<0.001**	−0.100	**<0.001**	−0.088	**<0.001**	−0.100	**<0.001**	−0.085	**<0.001**
T	−0.012	0.003	−0.013	**<0.001**	−0.008	**<0.001**	−0.013	**<0.001**	−0.011	**<0.001**	NS	NS
T^2	−0.007	0.310	−0.003	0.012	−0.004	0.262	−0.005	0.314	−0.004	**<0.001**	NS	NS
pH × t_G	0.056	**<0.001**	0.001	0.191	−0.022	**<0.001**	−0.057	**<0.001**	0.002	0.008	0.102	**<0.001**
pH × T	NS	NS	0.001	0.267	NS	NS	−0.012	0.003	0.004	**<0.001**	NS	NS
t_G × T	−0.005	0.299	−0.002	0.030	NS	NS	0.003	0.404	−0.001	0.109	NS	NS
pH × t_G × T	NS	NS	NS	NS	NS	NS	0.006	0.208	NS	NS	NS	NS

Compounds	PQ		PG		PM		QN		SD		SL	
Term	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)	Estim.	Pr(>|*t*|)
Intercept	4.897	**<0.001**	4.926	**<0.001**	4.853	**<0.001**	4.763	**<0.001**	4.394	**<0.001**	4.325	**<0.001**
pH	−0.005	0.372	0.011	**<0.001**	0.232	**<0.001**	0.056	**<0.001**	−0.177	**<0.001**	−0.133	**<0.001**
pH^2	−0.047	**<0.001**	0.002	0.116	0.048	**<0.001**	−0.053	**<0.001**	−0.241	**<0.001**	−0.331	**<0.001**
pH^3	0.070	**<0.001**	−0.001	0.778	−0.098	**<0.001**	0.306	**<0.001**	−0.102	**<0.001**	−0.206	**<0.001**
t_G	0.379	**<0.001**	0.385	**<0.001**	0.370	**<0.001**	0.356	**<0.001**	0.260	**<0.001**	0.226	**<0.001**
t_G^2	−0.090	**<0.001**	−0.092	**<0.001**	−0.086	**<0.001**	−0.080	**<0.001**	−0.063	**<0.001**	−0.058	**<0.001**
T	−0.015	**<0.001**	−0.019	**<0.001**	−0.016	**<0.001**	−0.015	**<0.001**	−0.026	**<0.001**	−0.027	**<0.001**
T^2	−0.006	0.027	−0.007	**<0.001**	−0.007	0.029	NS	NS	NS	NS	NS	NS
pH × t_G	0.007	0.003	−0.001	0.204	0.024	**<0.001**	0.046	**<0.001**	−0.053	**<0.001**	−0.071	**<0.001**
pH × T	−0.002	0.343	NS	NS	0.007	0.006	NS	NS	NS	NS	NS	NS
t_G × T	−0.005	0.025	−0.003	0.005	NS	NS	−0.006	0.168	NS	NS	NS	NS
pH × t_G × T	0.004	0.167	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS

NS: non-selected during stepwise regression.

3.2.1. pH effect

The retention time of acidic or basic compounds should respectively decrease or increase with respect to pH following a sigmoidal curve. To ease the interpretation of the model parameters given in Table 3, three cases should be envisaged.

First, if pKa is inside to the modelling pH range (MpHR; from 2.5 to 8), the sigmoidal curve inflexion point should be within MpHR.

Table 4
Adjusted coefficients of determination (R^2_{adj}) of the stepwise multiple linear regressions for the 19 AAIs retention factors logarithm (log(k_{tR})), left half-width logarithm (log(w_l)) and right half-width logarithm (log(w_r)).

Compounds	R^2_{adj}			Compounds	R^2_{adj}		
	log(k_{tR})	log(w_l)	log(w_r)		log(k_{tR})	log(w_l)	log(w_r)
Piperaquine	0.997	0.813	0.491	Mefloquine	1.000	0.651	0.391
Cinchonine	0.997	0.656	0.456	Artemisinin	1.000	0.571	0.896
Chloroquine	0.995	0.483	0.667	Dihydroartemisinine	1.000	0.932	0.622
Sulfalene	0.991	0.701	0.373	Artesunate	1.000	0.963	0.904
Amodiaquine	0.996	0.484	0.433	Artemether	1.000	0.299	0.726
Quinine	0.998	0.586	0.762	Halofantrine	0.999	0.842	0.643
Sulfadoxine	0.995	0.870	0.618	Arteether	0.999	0.558	0.635
Pyrimethamine	0.999	0.534	0.053	Lumefantrine	0.998	0.975	0.889
Primaquine	0.999	0.737	0.360	Atovaquone	1.000	0.411	0.284
Proguanil	1.000	0.636	0.561	**Mean**	**0.998**	**0.669**	**0.567**

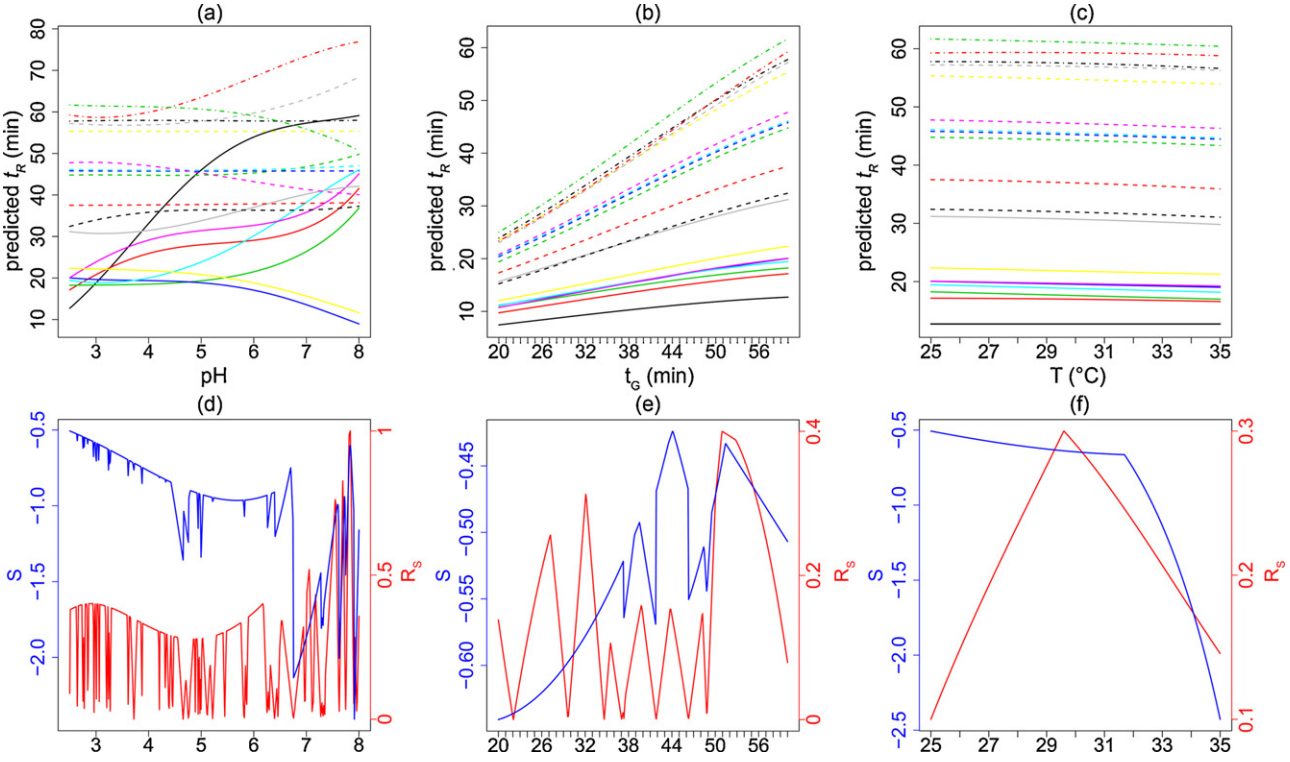

Fig. 2. Predicted retention times (t_R) with respect to DoE's factors. (a) Predicted t_R (min) vs. pH – with t_G = 60 min and T = 25 °C, elution order at pH 8: 1: SL, 2: SD, 3: PM, 4: PQ, 5: PG, 6: AS, 7: CC, 8: PM, 9: QN, 10: ART, 11: AQ, 12: DHA, 13: MQ, 14: LF, 15: AM, 16: AE, 17: PPQ, 18: HF, 19: LF. (b) Predicted t_R (min) vs. t_G (min) – with pH 2.5 and T = 25 °C, elution order at t_G = 60 min: 1: PPQ, 2: CC, 3: CQ, 4: AQ, 5: SL, 6: QN, 7: SD, 8: PM, 9: PQ, 10: PG, 11: MQ, 12: ART, 13: DHA, 14: AS, 15: AM, 16: HF, 17: AE, 18: LF, 19: AT. (c) Predicted t_R (min) vs. T (°C) – with pH 2.5 and t_G = 60 min, elution order at T = 35 °C: 1: PPQ, 2: CC, 3: CQ, 4: AQ, 5: SL, 6: QN, 7: SD, 8: PM, 9: PQ, 10: PG, 11: MQ, 12: ART, 13: DHA, 14: AS, 15: AM, 16: HF, 17: AE, 18: LF, 19: AT. Compound assignation: (black line) PPQ, (red line) CC, (green line) CQ, (blue line) SL, (cyan line) AQ, (magenta line) QN, (yellow line) SD, (grey line) PM, (dashed black line) PQ, (dashed red line) PG, (dashed green line) MQ, (dashed blue line) ART, (dashed cyan line) DHA, (dashed magenta line) AS, (dashed yellow line) AM, (dashed grey line) HF, (dotted-dashed black line) AE, (dotted-dashed red line) LF and (dotted-dashed green line) AT. (d–f) Corresponding predicted S (blue line, left axis) and R_S (red line, right axis). (For interpretation of the references to color in text, the reader is referred to the web version of the article.)

In this case, the pH term (β_1 is positive for bases and negative for acids; it can be close to 0) and mainly the pH^3 term (β_3 is positive for acids and negative for bases) should be highly expressed to mimic the theoretical sigmoidal variation of t_R with respect to pH. AQ, CC, LF, PM, PPQ and QN (basic compounds) and AS (acidic compound) presented t_R variations corresponding to this first case (see Fig. 2a).

Second, if pKa is outside and lower than MpHR, the truncated sigmoidal curve can be thus considered as a decreasing convex quadratic variation for acids and an increasing concave quadratic function for bases. A function is convex if and only if the region above its graph is a convex set. The pH term (β_1 is positive for bases and negative for acids) and principally the pH^2 term (β_2 is positive for acids and negative for bases) should be mainly expressed to fit the quadratic variation of t_R with respect to pH. PQ, a basic compound, presented a t_R variation which corresponded to both first and second case. The pKa corresponding to the quinoline function (pKa ∼ 4.9) was inside MpHR (corresponding to case 1) but the second pKa (∼10.4) was higher than MpHR (corresponding to case 2).

Third, if pKa is outside and higher than MpHR, the truncated sigmoidal curve can be thus considered as a decreasing concave quadratic variation for acids and an increasing convex quadratic function for bases. The pH term (β_1 is positive for bases and negative for acids) and principally the pH^2 term (β_2 is negative for acids and positive for bases) should be mainly expressed to fit the quadratic variation of t_R with respect to pH. AT, SD and SL (acidic compounds) and CQ, HF and CQ (basic compounds) displayed t_R variations corresponding to this third case.

The rest of the AAIs (AE, AM, ART and DHA) are neutral compounds, which did not present t_R variation with respect to pH. The above mentioned interpretations were in accordance with the predicted t_R variations with respect to pH presented in Fig. 2a.

It should be noted that the given pKa are estimated or measured for aqueous media. Therefore, they merely represent an estimation of the pKa in hydro-organic phase.

3.2.2. Gradient time and temperature effects

The general shape of t_R variation with respect to t_G is an increasing concave function which can easily be fitted by a second order polynomial (see Fig. 2b). All β_4 (corresponding to t_G) were positive and all β_5 (corresponding to t_G^2) were negative due to the concave curvature. Van't Hoff law defines the variation of an equilibrium constant with respect to the temperature. In HPLC, a linear relationship generally links $\log(k_{tR})$ with $1/T$. A decreasing curved function should be observed for $\log(k_{tR})$ vs. T. Besides, the selected temperature range was narrow and statistically significant t_R variations were not observed premising method robustness against temperature variation as expected (see Fig. 2c).

3.2.3. Models adequacy

In order to visualize the models adequacy, the appropriateness between predicted and experimental data as well as the corresponding residuals were displayed in Fig. 3. As observed in Fig. 3b, the residuals are distributed between −2 and 2 min. The residuals standard deviations were equals to 0.36, 0.41 and 0.36 min for t_R, t_E and t_B, respectively. Given these results, it is thus reasonable to assume that the retention time predictive error will be around 0.8 min (i.e. 2 × standard deviation) over the whole experimental domain.

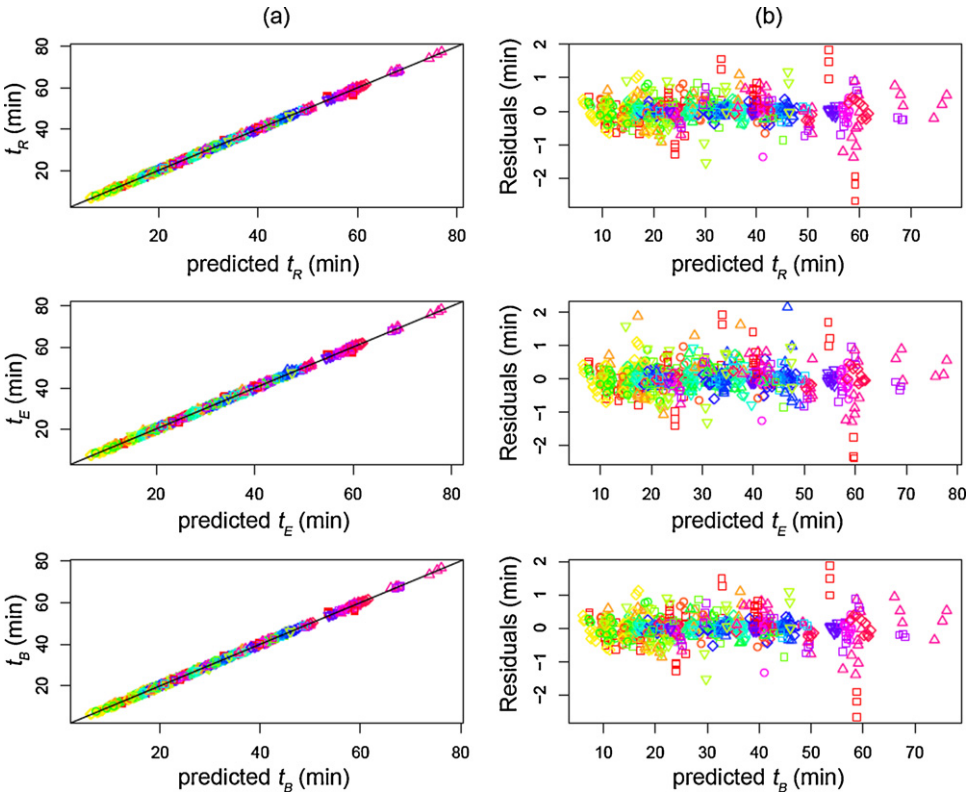

Fig. 3. Modelling results and corresponding residuals. (a) Predicted vs. experimental values for t_R, t_E and t_B. (b) Corresponding residuals plots. Compound assignation: (red square) PPQ, (red circle) CC, (orange head up triangle) CQ, (yellow diamond) SL, (green head down triangle) AQ, (green square) QN, (green circle) SD, (green head up triangle) PM, (green diamond) PQ, (blue head down triangle) PG, (blue square) MQ, (blue circle) ART, (blue head up triangle) DHA, (blue diamond) AS, (blue head down triangle) AM, (purple square) HF, (purple circle) AE, (purple head up triangle) LF, (purple diamond) AT. (For interpretation of the references to color in text, the reader is referred to the web version of the article.)

Furthermore, Shapiro–Wilk Normality tests were carried out on the modelled responses residuals. The *P*-values were all higher than 0.05 which meant that the residuals were actually normally distributed.

From Table 4, one can observe that the R^2_{adj} were higher for t_R than for t_E and t_B. This can be explained by two main reasons. First, the times at the peaks beginning and end could be biased by a poor estimation of integration limits. Even when ICA was applied, the independent component baseline can be distorted at the side of the coelution [27,28] leading to an overestimation of t_B or an underestimation of t_E. Second, w_l and w_r are computed from 2 measurements. For instance, the computation of w_l involved t_B and t_R whereas the computation of k_{tR} only involved t_R thus explaining the lower observed R^2_{adj}.

The smallest R^2_{adj} value is equal to 0.053 for the right half-width of PM peak. This small value did not reflect a poor data adjustment. It only reflected the fact that PM peak right half-width was poorly influenced by the experimental factors and that the response variability was higher than its average variation. As observed in Fig. 3, no aberrant results were observed for this compound or for others (i.e. no significant outliers observed on residuals plots).

Each model was used to predict t_R according to DoE factors. As expected and presented in Fig. 2a, pH is the factor that had the most significant effect on selectivity. The identification of neutral, acidic or basic compounds is also easy. Neutral compounds show no t_R variation with respect to pH. Acidic and basic compounds have a respective decreasing or increasing variation with respect to pH. The huge t_R increasing variation of PPQ can be explained by the four basic nitrogen on the piperazine cycles lying in the PPQ structures.

The temperature has the lowest effect (see Fig. 2c). In other terms, these low t_R variations with respect to T underline the method robustness while changing T as expected routine use in Africa.

Fig. 2d–f shows the behaviour of the predicted quality criteria S and R_S.

One can also observe from Fig. 2d–f that predicted S and R_S share some similarities. The criteria curves change-points occur when critical pair peaks change and discontinuities are resulting from distinct peak asymmetries. These figures also clearly show why R_S or S should not be selected as multiple linear model responses but only as critical quality attributes.

3.3. Quality criterion and design space computation

The separation quality criterion (S) was computed over the whole experimental domain. A grid of 42,875 points (i.e. $35 \times 35 \times 35$) was defined and S was computed for each operating condition. The residuals distribution was used to generate a Gaussian predictive error (i.e. the Gaussian distribution standard deviation was set equal to the residuals standard deviation, see Section 3.2.3), which was used to propagate the predictive error to S using Monte-Carlo simulations [27,34,35]. 2500 Monte-Carlo simulations were thus carried out for each of the 42,875 points. Then, the results were presented as probability surfaces (i.e. $P(S > 0)$) rather than response surfaces.

The very similar chromatographic behaviour between ART and DHA prevented obtaining a separation of the 19 AAIs. Nevertheless, at some operating conditions, they were the only two coeluting peaks. Thus, two groups were formed. The first contained 17 AAIs and ART (group 1) and the second contained the same 17 AAIs and DHA (group 2). These two groups are justified from a therapeutic point of view because ART and DHA are never present in the same

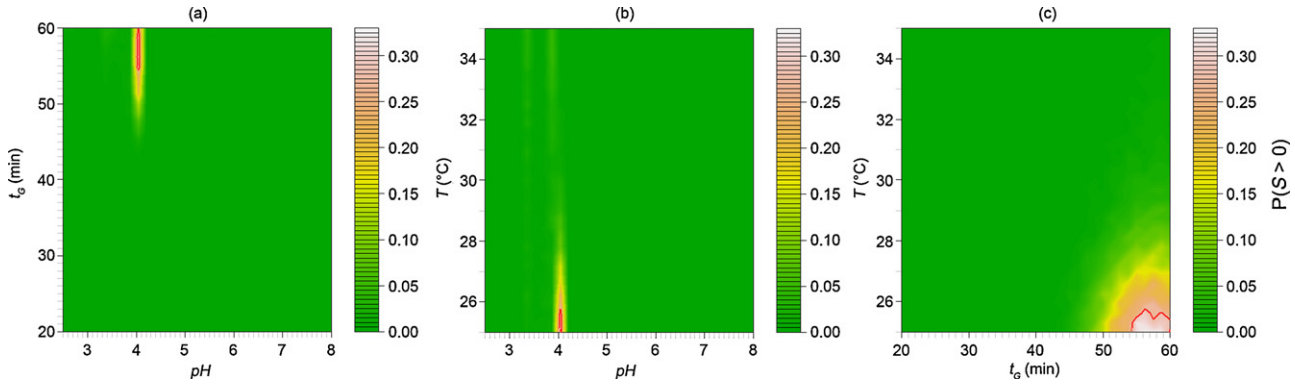

Fig. 4. Probability surfaces (i.e. $P(S>0\,\text{min})$) for group 1 separation. (a) Gradient time (min) vs. pH, (b) temperature (°C) vs. pH and (c) temperature (°C) vs. gradient time (min). The DS ($\pi = 28\%$) is encircled by a red line. (For interpretation of the references to color in text, the reader is referred to the web version of the article.)

pharmaceutical formulation. The optimization process was then repeated for each group independently. The probability surfaces for $P(S>0)$ for group 1 and group 2 are presented in Figs. 4 and 5, respectively.

For a given π, the DS shape is directly linked to the method robustness. Therefore, the DS shapes in Figs. 4 and 5 allowed concluding that the resulting screening methods are relatively highly robust with respect to modifications in t_G (from 54 to 60 min for group 1 and from 58 to 60 min for group 2) and T (from 25 to 25.8 °C for group 1 and from 25 to 26 °C for group 2). Conversely, these screening methods are far less robust with respect to pH (from 4 to 4.1 for both groups). However, the pH measurement variability can be estimated to 0.1% (see Section 2.4). Consequently, the relatively poor method robustness with respect to pH should therefore not be problematic when care is taken during the buffers preparation and pH measurements.

3.4. Prediction of optimal separation

As shown in Figs. 4 and 5, the probability surface and the corresponding DS are very similar for the two groups of investigated compounds. The main difference is that the quality levels are different: $\pi = 28\%$ for group 1 and $\pi = 8\%$ for group 2. This difference between quality levels for group 1 and 2 came from the shorter DHA retention times compared to ART. DHA slightly coeluted with MQ (Peak 12 in Fig. 6) explaining the much lower π for group 2.

Finally, only one operating condition was selected allowing the separation of the 18 AAIs of both groups independently. The optimal operating condition giving the highest probability to have $S>0$ ($P = 33\%$ for group 1 and $P = 9.6\%$ for group 2) was pH 4.05, $t_G = 56.2$ min and $T = 25$ °C. This probability could seem low. But it is important to keep in mind that, at this optimal condition, S depended on $t_{E,2}$ and $t_{B,3}$ of the two most proximate peaks or critical pair (Peak 2 and 3 in Fig. 6, SL and AQ, respectively). As mentioned in Section 3.2, the predictive error was estimated at 0.8 min for both $t_{E,2}$ and $t_{B,3}$. As it can be seen from Fig. 6b, the predicted difference between $t_{E,2}$ and $t_{B,3}$ is smaller than the estimated predictive error. The chromatogram recorded at the optimal condition is depicted in Fig. 6a for group 1 compounds.

Chromatograms recorded for group 2 compounds at the optimal operating condition are very close to those displayed in Fig. 6a for group 1. Despite the inability to separate ART and DHA, it was easy to identify them even if coeluting because their retention times were different enough allowing their distinction (i.e. $t_R = 44.5$ min for ART and $t_R = 44.9$ min for DHA).

3.5. Sub-mixture

One can observe that the screening method (Fig. 6) has a quite long analysis time. It can be also observed that the compounds eluted between 15 and 60 min. The first 15 min are not "used" to separate the compounds in a shorter time. It mainly resulted from the fact that the methanol proportion at the start of the gradient ($\%MeOH_{ini}$) was low (i.e. $\%MeOH_{ini} = 5\%$) and was not introduced as a DoE factor. $\%MeOH_{ini}$ was not added as a DoE factor to keep the number of experiments to an acceptable value (i.e. 45 experiments). The introduction of $\%MeOH_{ini}$ would have increased the number of experiments to 135 (i.e. 3×45 to estimate the quadratic effect of $\%MeOH_{ini}$) for a full factorial design.

Nevertheless, the DoE–DS methodology can also be used to develop methods aiming at reducing the analysis time while optimizing the separation of specific sub-mixtures. Some compounds

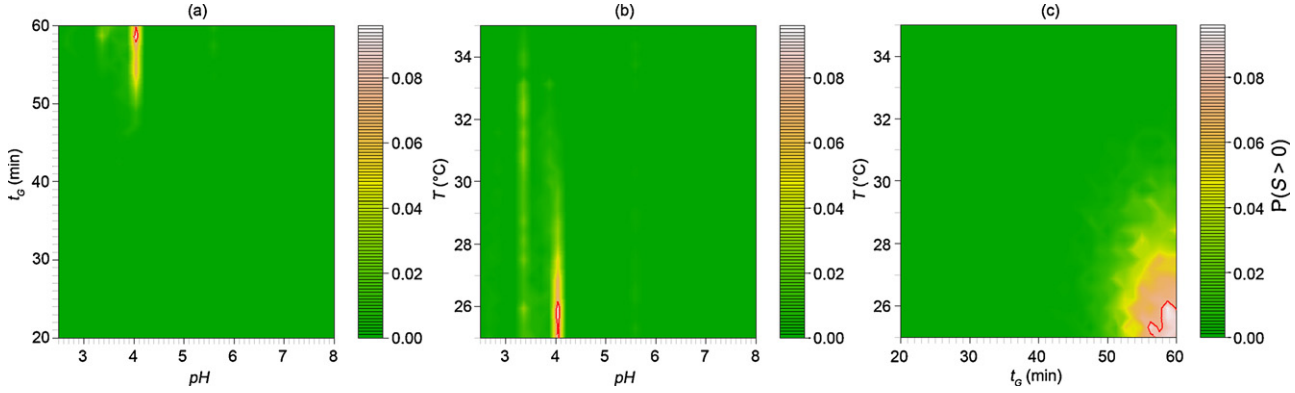

Fig. 5. Probability surfaces (i.e. $P(S>0\,\text{min})$) for group 2 separation. (a) Gradient time (min) vs. pH, (b) temperature (°C) vs. pH and (c) temperature (°C) vs. gradient time (min). The DS ($\pi = 8\%$) is encircled by a red line. (For interpretation of the references to color in text, the reader is referred to the web version of the article.)

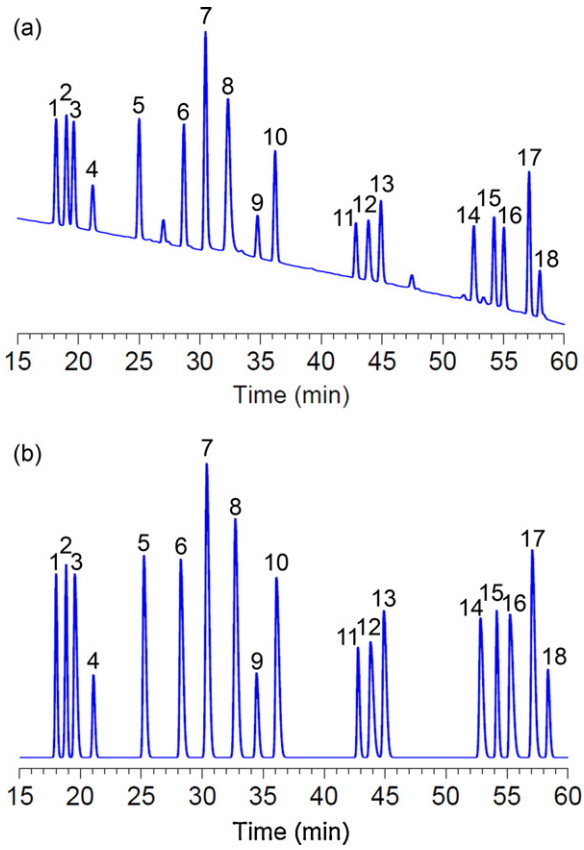

Fig. 6. (a) Experimental chromatogram recorded at pH 4.05, t_G = 56.2 min and T = 25 °C with group 1 compounds. (b) Respective predicted chromatogram. Peak numbering: 1 = CQ, 2 = SL, 3 = AQ, 4 = SD, 5 = CC, 6 = QN, 7 = PM, 8 = PP, 9 = PQ, 10 = PG, 11 = MQ, 12 = ART, 13 = AS, 14 = AM, 15 = HF, 16 = AE, 17 = LM and 18 = AT.

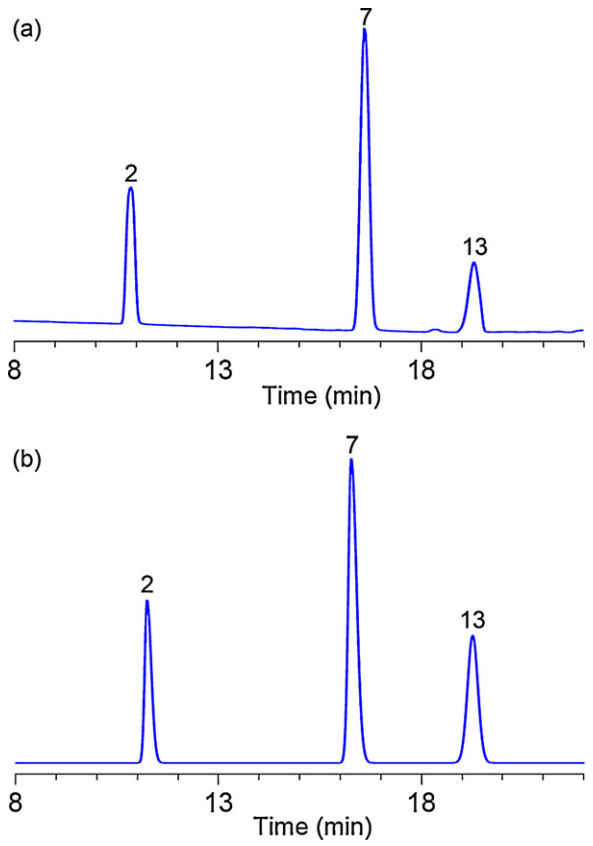

Fig. 8. (a) Experimental chromatogram recorded at pH 5.4, t_G = 20 min and T = 35 °C with sub-mixture compounds. (b) Corresponding predicted chromatogram. Peak numbering: 2 = SL, 7 = PM, 13 = AS.

were therefore selected to test out this opportunity without performing any additional experiments. The selected sub-mixture contained AS, PM and SL. This combination is representative of a pharmaceutical formulation present on the Democratic Republic of the Congo market.

In order to minimize the time of analysis while simultaneously optimizing the separation of these 3 compounds, a multi-criteria optimization was carried out. The criteria were the separation criterion S and the time of analysis (i.e. the retention time at the end of the last peak). The acceptance limits were placed at 0.1 min for S and at 20 min for a more convenient analysis time. The DS thus defined the subspace where the probability to obtain S > 0.1 min concurrently with an analysis time < 20 min was higher than 97.5% as shown in Fig. 7.

One can observe that the DS depicted in Fig. 7a and c is really small. The reason behind this observation is the closeness of the last peak with the analysis time threshold (i.e. <20 min). However, due to the high π, the specific method is very robust. The optimal operating condition is pH 5.4, t_G = 20 min and T = 35 °C. The recorded chromatogram at the optimal condition is displayed in

Fig. 8b presents the predicted chromatogram. The adequacy between predicted and experimental t_R was very good. It is quite obvious that a method development (using already available commercial softwares or a 2^3 full factorial design with t_G and $\%MeOH_{ini}$ as factors) for the separation of these three compounds would

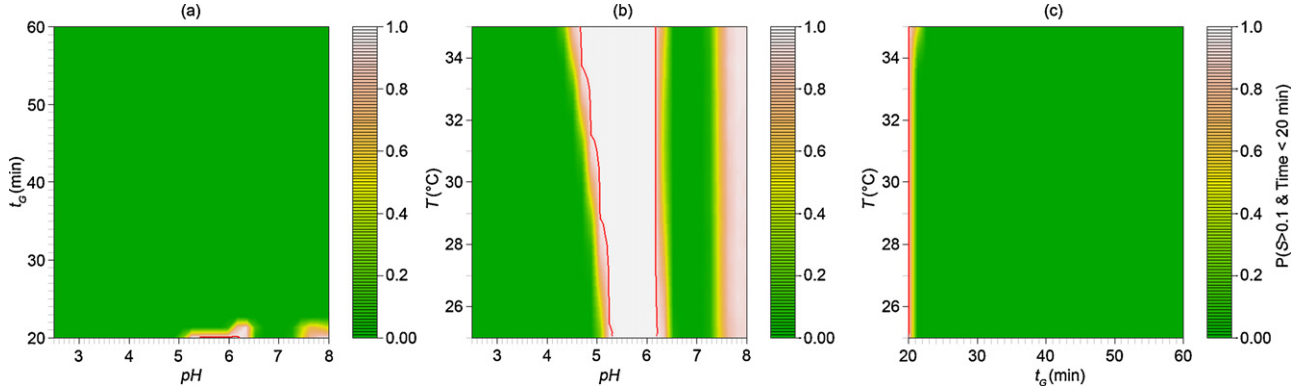

Fig. 7. Probability surfaces (i.e. $P(S > 0.1$ min and analysis time < 20 min)) for sub-mixture separation. (a) Gradient time (min) vs. pH, (b) temperature (°C) vs. pH and (c) temperature (°C) vs. gradient time (min). The DS (π = 97.5%) is encircled by a red line. (For interpretation of the references to color in text, the reader is referred to the web version of the article.)

be able to find operating condition giving a good separation in a shorter analysis time. However, this present method optimization dedicated to the pharmaceutical formulation (Arte-Plus®) was carried out from the same data that those used for the optimization of the screening method. It underlined the fact that the DoE–DS methodology is generic. Here an innovative methodology was used to optimize the separation while simultaneously estimating the robustness of either a general screening method or specific methods.

4. Conclusions

In the current trend of being QbD compliant, it is of first importance to develop methodologies that provide robust optimal separations defined by a DS. With regards to this objective, DoE–ICA–DS allowed obtaining optimal screening methods for the separation of 19 AAIs. The methods robustness was evaluated thanks to the DS quality level, DS shape and the assessment of the factors effects. It resulted that the obtained screening methods were very robust against temperature modification. This result is very important when one of the final aims is a method transfer to African laboratories where column ovens are not always available. Theses screening methods can be considered as a step forward for the fight against counterfeit medicines. The present work also allowed demonstrating the ability of the DoE–ICA–DS methodology to encounter optimal separation for complex mixtures (i.e. containing compounds with very similar structures and physico-chemicals properties). The present study also demonstrated the ability of fitted mathematical models to be used to identify and corroborate theoretical chromatographic behaviours of studied compounds. It highlighted the fact that DoE strategy can be a very useful tool for chromatographists in order to develop or refine the understanding of some chromatographic behaviour. Furthermore, the separation of a 3 compounds mixture was also carried out without performing any additional experiments. The resulting method was also very robust to temperature changes. Finally, the results presented in this manuscript strengthen the fact that DoE–ICA–DS can be considered as a generic QbD compliant methodology for the optimization and the robustness assessment of new chromatographic methods for the analysis of pharmaceutical formulations or more complex matrices such as plant or biological materials.

Nomenclature

E	estimator of mathematical expectation
k	retention factor
P	estimator of probability
R^2_{adj}	adjusted coefficient of determination
R_S	resolution
S	separation quality criterion
T	temperature (°C)
t_B	retention time at the beginning of a peak (min)
t_E	retention time at the end of a peak (min)
t_G	gradient time (min)
t_R	retention time (min)
t_0	column dead-time (min)
w_l	peak left half-width
w_r	peak right half-width
$\mathbf{x}_0$	points of the experimental domain

Greek letters

β	estimated parameters of the model in Eq. (1)
ε	estimated error of the model in Eq. (1)
$\hat{\theta}$	set of estimated parameters of the model
λ	acceptance limit for selected criterion (see Eq. (2))
π	quality level (see Eq. (2))
χ	experimental domain

Acknowledgements

The authors would like to thank the Walloon Region of Belgium for the PPP convention funds N° 917007 and Arlenda SA for partial funding of Optimal-DS project. A research grant from the Belgium National Fund for Scientific Research (F.R.S-FNRS) to E. Rozet is also gratefully acknowledged. Research grants from the Belgian Coopération Universitaire au Développement (CUD) allocated to J. Mbinze Kindenge and J. Mavar Tayey Mbay are thankfully acknowledged.

References

[1] M.K. Laufer, C.V. Plowe, Drug Resist. Updat. 7 (2004) 279.
[2] A.F. Cowman, S.J. Foote, Int. J. Parasitol. 95 (1990) 503.
[3] T.K. Mutabingwa, Acta Trop. 95 (2005) 305.
[4] C.O. Obonyo, E.A. Juma, B.R. Ogutu, J.M. Vulule, J. Lau, Trans. R. Soc. Trop. Med. Hyg. 101 (2007) 117.
[5] N. Valecha, H. Joshi, P.K. Mallick, S.K. Sharma, A. Kumar, P.K. Tyagi, B. Shahi, M.K. Das, B.N. Nagpal, A.P. Dash, Acta Trop. 111 (2009) 21.
[6] N. Singh, M.M. Shukla, G. Chand, P.K. Bharti, M.P. Singh, M.K. Shukla, R.K. Mehra, R.K. Sharma, A.P. Dash, T. Roy. Soc. Trop. Med. H. 105 (2011) 133.
[7] R.D. Marini, J. Mbinze Kindenge, M.L.A. Montes, B. Debrus, P. Lebrun, J. Mantanus, E. Ziemons, S. Rudaz, Ph. Hubert, Chim. Oggi 28 (2010) 10, http://hdl.handle.net/2268/62379.
[8] C. Horváth, W. Melander, I. Molnár, J. Chromatogr. 125 (1976) 129.
[9] P.W. Carr, J. Li, A.J. Dallas, D.I. Eikens, L. Choo Tan, J. Chromatogr. A 656 (1993) 113.
[10] D. Nagrath, F. Xia, S.M. Cramer, J. Chromatogr. A 1218 (2011) 1219.
[11] L.R. Snyder, J.W. Dolan, J.R. Gant, J. Chromatogr. 165 (1979) 31.
[12] L.R. Snyder, J.W. Dolan, D.C. Lammen, J. Chromatogr. 485 (1989) 65.
[13] P. Nikitas, A. Pappa-Louisi, J. Chromatogr. A 1216 (2009) 1737.
[14] G. Vivó-Truyols, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, J. Chromatogr. A 1018 (2003) 169.
[15] B. Dejaegher, Y. Vander Heyden, J. Chromatogr. A 1158 (2007) 138.
[16] R. Ragonese, M. Mulholland, J. Kalman, J. Chromatogr. A 870 (2000) 45.
[17] E. Hund, Y. Vander Heyden, M. Haustein, D.L. Massart, J. Smeyers-Verbeke, Anal. Chim. Acta 404 (2000) 257.
[18] Guidelines of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), Pharmaceutical Development, Q8(R2) Step 4, Geneva, 2009, p. 1.
[19] E. Rozet, A. Ceccato, C. Hubert, E. Ziemons, R. Oprean, S. Rudaz, B. Boulanger, Ph. Hubert, J. Chromatogr. A 1158 (2007) 111.
[20] J.J. Peterson, J. Qual. Technol. 36 (2004) 139.
[21] J.J. Peterson, R.D. Snee, P.R. McAllister, T.L. Schofield, A.J. Carella, J. Qual. Technol. 41 (2009) 111.
[22] J.J. Peterson, J. Biopharm. Stat. 18 (2008) 959.
[23] G.W. Stockdale, A. Cheng, Qual. Technol. Quant. Manag. 6 (2009) 391.
[24] P. Lebrun, B. Govaerts, B. Debrus, A. Ceccato, G. Caliaro, Ph. Hubert, B. Boulanger, Chemom. Intell. Lab. Syst. 91 (2008) 4, http://hdl.handle.net/2268/1640.
[25] B. De Backer, B. Debrus, P. Lebrun, L. Theunis, N. Dubois, L. Decock, A. Verstraete, Ph. Hubert, C. Charlier, J. Chromatogr. B 877 (2009) 4115, http://hdl.handle.net/2268/4442.
[26] F. Krier, M. Brion, B. Debrus, P. Lebrun, A. Driesen, E. Ziemons, B. Evrard, Ph. Hubert, J. Pharm. Biomed. Anal. 54 (2011) 694, http://hdl.handle.net/2268/75222.
[27] B. Debrus, P. Lebrun, A. Ceccato, G. Caliaro, E. Rozet, I. Nistor, R. Oprean, F.J. Rupérez, C. Barbas, B. Boulanger, Ph. Hubert, Anal. Chim. Acta 691 (2011) 33, http://hdl.handle.net/2268/88084.
[28] B. Debrus, P. Lebrun, A. Ceccato, G. Caliaro, B. Govaerts, B.A. Olsen, E. Rozet, B. Boulanger, Ph. Hubert, Talanta 79 (2009) 77, http://hdl.handle.net/2268/13236.
[29] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New York, 2001.
[30] W. Dewé, R.D. Marini, P. Chiap, Ph. Hubert, J. Crommen, B. Boulanger, Chemom. Intell. Lab. Syst. 74 (2004) 263, http://hdl.handle.net/2268/6036.
[31] The R Project for Statistical Computing. http://www.r-project.org/.
[32] A. Hyvarinen, E. Oja, Neural Networks 13 (2000) 411.
[33] N.R. Draper, H. Smith, Applied Regression Analysis, 3rd ed., Wiley, New York, 1998.
[34] M.A. Herrador, A.G. Asuero, A.G. Gonzalez, Chemom. Intell. Lab. Syst. 79 (2005) 115.
[35] G. Vivó-Truyols, V. Concha-Herrera, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, J. Chromatogr. A 1096 (2005) 123.