Journal of Chromatography A, 1229 (2012) 180–189

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography A

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

Approaches to find complementary separation conditions for resolving complex mixtures by high-performance liquid chromatography

C. Ortiz-Bolsico, J.R. Torres-Lapasió*, M.C. García-Álvarez-Coque

Departament de Química Analítica, Universitat de València, c/Dr. Moliner 50, 46100 Burjassot, Spain

ARTICLE INFO

Article history:
Received 24 November 2011
Received in revised form 12 January 2012
Accepted 13 January 2012
Available online 20 January 2012

Keywords:
Complex mixtures
Chromatographic optimisation
Complementary separation conditions
Comprehensive search
Genetic Algorithms
Peak count: Local search

ABSTRACT

Chromatographic problems are usually addressed trying to find out a single experimental condition aimed to resolve all compounds in the sample. However, very often, the chromatographic system is not able to provide full resolution. When a separation fails, the usual choice is introducing a drastic change in the chromatographic system (e.g. column, solvent, pH). There are, however, other possibilities that take advantage of the gathered information in the failed separation, without the need of new experiments, based on the concept of complementary separations (e.g. isocratic mobile phases, gradients, columns, chromatographic modes). One separation condition will focus on the resolution of some compounds in the sample, while the other compounds will be resolved using a second (or subsequent) condition(s). Complementary separations, being a simple and attractive idea, present, however, challenges in terms of computation volume and complexity of the required algorithms. This work describes in detail different approaches that have been developed up-to-date for this purpose, and introduces a new approach based on the peak count concept that is benefited of the best features of the previous approaches: high reliability in finding the solution, accessibility to analysts without specialised programming skills and short computation time.

1. Introduction

Chromatographic analyses demand finding experimental conditions to separate the compounds of interest. This task is often arduous and discouraging, and involves several objectives that are opposed each other (e.g. high resolution, short analysis time, and low economic and environmental costs). When the analyst is faced with a new sample, he/she ignores the full potential of the separation system. The optimised conditions may not be also the best that the system can offer. Fortunately, nowadays, a rigorous evaluation of the system potential is possible using numerical methods, which efficiently and reliably explore the separation performance of a chromatographic system. In spite of the literature available in this field [1–15], new challenges are continuously arising.

The success of the separation is determined by the chromatographic system: the instrumentation features, and the combination of column, modifier(s), experimental conditions, and incidentally, pre-conditioning steps used to change the nature of column or analytes. All these elements must be combined properly to reach an acceptable separation performance. In chromatography, more than in other fields, the experimental factors that can be modified to change the analytical behaviour are numerous, and the quality of the separations may vary drastically when the factors are changed.

The aim of the analysis should be clearly defined before starting the optimisation process. This may vary considerably depending on the problem. In most cases, the analyst is interested in the separation of all compounds in the sample. In others, the aim is less ambitious, focusing the attention on only a few compounds, or even, on a single compound [16]. The problem is most usually addressed trying to find a single experimental condition able to get the resolution of all compounds in the sample. However, very often, the chromatographic system will not be able to achieve full resolution. When a separation fails, the usual choice is introducing a drastic change in the chromatographic system (e.g. column, solvent, pH). There are, however, other possibilities that take advantage of the gathered information in the failed separation, without the need of new experiments.

In 2000, an optimisation strategy was proposed to achieve the chromatographic separation of complex samples to get full resolution, based on the concept of complementary situations [17]. One separation condition (e.g. an isocratic mobile phase, gradient, column) would allow the resolution of some compounds in the sample, while the other compounds would be resolved using a second (or subsequent) condition(s). The idea is attractive, but demanding in computation terms. A substantial reduction in the calculation effort was achieved through the application of natural

* Corresponding author. Tel.: +34 96354 4014; fax: +34 96354 4436.
E-mail address: jrtorres@uv.es (J.R. Torres-Lapasió).

0021-9673/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2012.01.034

computation [17,18]. Recently, a simple approach more accessible to analysts was proposed to find complementary separations conditions (CSCs) [19,20]. This work describes in detail the different approaches that have been developed up-to-date for this purpose. A new approach, based on the peak count concept and assisted by local search, is also reported. The approach is benefited of the best features of the previous approaches.

2. Theory

2.1. Total, partial and specific optimisation strategies

Once the retention and peak profile behaviour for each compound in the target mixture has been appropriately modelled, three optimisation levels can be defined, which have been called total, partial and specific strategies, according to the number of compounds to be resolved in a sample [16]. The way of structuring the calculation is next described in detail for each type of strategy. In all cases, the reduction of the chromatographic information related to the resolution is performed in two consecutive steps. In the first step, a descriptor that measures the success in the separation of each pair of compounds (pair resolution), or of an individual compound from the remaining in the sample (elementary resolution), is obtained. We will only refer here to the descriptors that measure the elementary resolution associated with each compound. In the second step, the elementary resolutions are combined in a descriptor that measures the global resolution in the chromatogram.

The calculations are best outlined in matrix terms. The elementary resolutions are thus arranged in a matrix, **P**. Each column in **P** corresponds to a given compound, and each row is associated with a certain separation condition from a set of hypothetical conditions, whose performance is being investigated (namely, the conditions grid) (Fig. 1a). For each condition in the grid, simulated chromatograms are calculated and from them the elementary resolutions. The global resolution associated with each experimental condition is obtained by multiplying the elements in the corresponding row in **P** (Fig. 1b). If the process is extended to all conditions in the matrix, a global resolution vector is obtained. The element with a maximal value points out the optimal separation condition. On the other hand, the maximal value in each matrix column represents the maximal separation that can be expected for each compound, which is called the "limiting resolution" (Fig. 1c). The limiting resolution values for the different compounds in a sample are of great interest to establish the operative limits of the chromatographic system.

Instead of solving the *n* compounds in the sample using a single separation condition, *n* separation problems can be outlined, each aimed at solving a different compound. This approach is possible provided that the selected resolution function allows an independent evaluation of the contributions of each compound. Similarly to the calculation of the global resolution for a given experimental condition through the product of elementary values, for each matrix row (Fig. 1b), the elementary limiting resolutions can be multiplied as well, resulting in a combined resolution, the "global limiting resolution" (Fig. 1c). This indicates the maximal system performance and can be used to calculate the degree of success by dividing the global resolution associated with the selected conditions (those offering maximal resolution, or any other selected as satisfactory) by the global limiting resolution. The expectancies in getting a substantial improvement in the results, or the magnitude of the expected improvement, can be also evaluated.

The analyst interest is focused, sometimes, to resolve only some compounds in the sample. For this purpose, partial optimisation strategies should be applied, which classify the eluted compounds in two categories: the analytes (whose elementary resolution is optimised) and the interferences (whose resolution is not optimised, but should be taken into account). In this case, only the columns in the elementary resolution matrix that include the compounds of interest are considered, which gives rise to the calculation of partial resolutions (Fig. 1d).

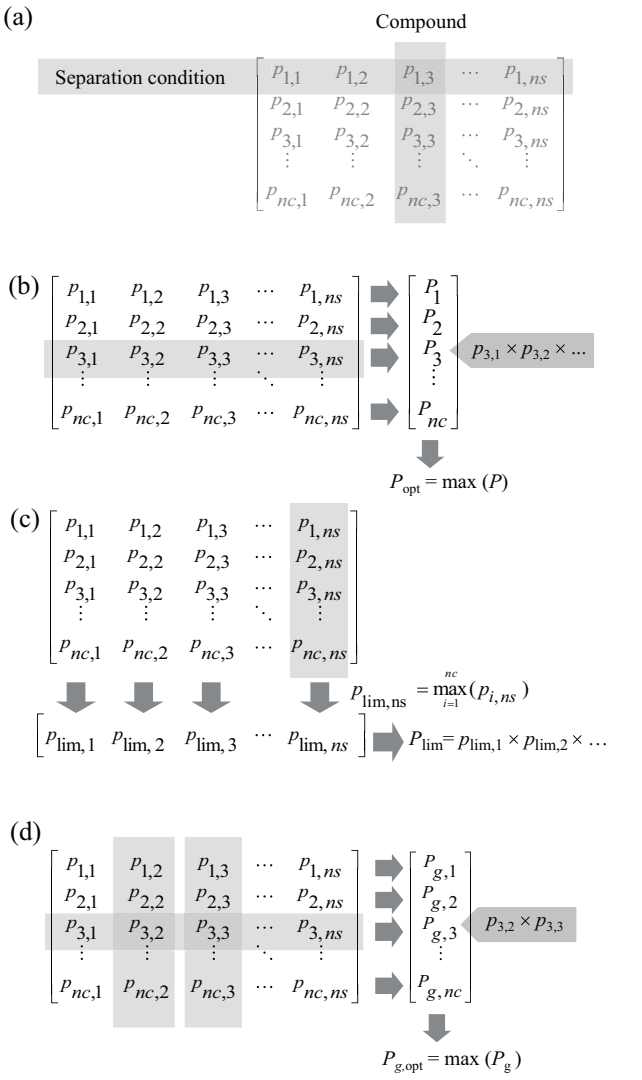

Fig. 1. Matrix of elementary peak purities: (a) each column corresponds to a compound and each row is associated with a certain separation condition; (b) Calculation of the global resolution associated with each separation condition and selection of the optimal value; (c) Calculation of the limiting elementary peak purities and global limiting purity; (d) Optimisation of the resolution of a group of two compounds.

2.2. A mixed strategy: complementary separation conditions

As the complexity of the sample increases, a single experimental condition is unlikely to provide an acceptable separation of all compounds in a sample. A possible solution is the use of a combination of two or more experimental conditions with complementary behaviour. The simplest and most frequent case, which will be taken as example in this work, is the optimisation of complementary mobile phases, using the same solvent system. It should be noted that the concept of complementarity goes beyond the optimisation of isocratic mobile phases: it is possible to optimise gradients, combinations of solvents, chromatographic columns, separation techniques and so on. Hence, the term CSCs used throughout this work.

A CSCs optimisation selects two or more separation conditions, so that each one is dedicated to resolving some compounds, leaving the peaks of other compounds unattended. As a result, the separation space increases, and so the chances of success. The selection is done in such a way that when the results of the optimal CSCs are considered altogether, all compounds are maximally resolved. With this strategy, full (or almost full) resolution may be possible, although the sample should be chromatographed using two or more different separation conditions. Often two conditions are enough to achieve a significant improvement in the resolution.

The CSCs optimisation is a mixed strategy that gathers features of the total and partial strategies: it is a partial strategy because it focuses on the separation of subsets of compounds, but also global, since the final aim is resolving all compounds in a sample. The CSCs optimisation can be considered as an intermediate case between a classical optimisation (where a single condition is searched for resolving maximally the *n* compounds in a sample) and an individual optimisation (where the best condition to separate each compound from the others is searched). Searching CSCs may be closer to one or another strategy, depending on the number of selected conditions. Regarding the individual optimisation, CSCs try to resolve all compounds in a sample with less experimental effort. A reasonable compromise is reached between the system resolution performance (which is not fully exploited, but up to a reasonable level) and the experimental effort (which without being as economical as a single condition, is still acceptable under a practical point of view).

2.3. Search spaces

The calculation of CSCs can be addressed using two perspectives: the formation of groups of compounds and the formation of groups of separation conditions. In both search spaces, the number of CSCs should be previously selected by the analyst.

2.3.1. Formation of groups of compounds

The diagram in Fig. 2a (left) illustrates the process of formation of groups of compounds for a set of 10 compounds (a mixture of diuretics and β-blockers eluted isocratically with mobile phases containing sodium dodecyl sulphate and 1-propanol [17]). This approach is based on the examination of all possible distributions of *ns* compounds in *ng* groups, and the search of the experimental condition where each group is resolved. In the example, 10 compounds are being separated using two CSCs, selected from the conditions grid, where an index is associated to each condition. The compounds are split into two groups: compounds 1, 2, 3, 8 and 10 are assigned to the first group and the other compounds to the second group. The optimisation consists of determining which condition resolves each group the best. With this aim, the resolution of the compounds within each group (i.e. the partial resolution) is calculated, by multiplying the elementary resolutions for each compound within the group, $p(g,s)$. The process is extended to all possible ways of dividing the compounds into two groups. The combination of CSCs with the highest resolution is selected as the optimal. The quality of the global separation is obtained by multiplying the partial resolutions for the optimal CSCs:

$$P = \prod_{g=1}^{ng} \underset{\text{phase}=1}{\overset{np}{\text{MAX}}} \left(\prod_{s=1}^{ns(g)} p(g,s) \right) \quad (1)$$

As observed, the number of elements that are multiplied to get the global resolution matches the number of compounds. Therefore, the results for situations including a different number of CSCs are comparable, and can be related to the global limiting resolution. In some cases, there will be compounds overlapping under one condition and fully resolved with another, but there will be also compounds well resolved in both conditions. In the latter case, the compounds are assigned to the CSC where their elementary resolution is the highest.

Table 1
Resolution (measured as peak purity) achieved with a single mobile phase and with two and three mobile phases containing sodium dodecyl sulphate and 1-propanol, in the separation of a mixture of 10 diuretics and β-blockers [17].

Partial peak purity					
ng	Combination	CSC1	CSC2	CSC3	Global purity
1 separation condition	1111111111				0.751
2 CSCs*	1212211121	0.9543	0.9261		0.884
3 CSCs	1312311121	0.9543	0.9338	0.9985	0.890
Limiting global purity					0.892

*A B C D E F G H I J
1 2 1 2 2 1 1 1 2 1

2.3.2. Formation of groups of separation conditions

Another option is examining the combinations of separation conditions that can be drawn from those included in the conditions grid. In the example in Fig. 2a (right), the group including the separation conditions with indexes 157 and 331 in the **P** matrix is examined. Each compound in the sample is assigned to the separation condition where its elementary resolution is higher. The combined resolution is calculated from Eq. (1).

Table 1 shows the results obtained with the mixture of 10 compounds, using a single mobile phase, two and three CSCs. The encoded figures that appear in the column "Combination" identify the CSC (1, 2 or 3) that best resolves each compound, whereas the order in the list points to the compound (compounds A–J). The partial resolutions are indicated for the optimal CSCs. The last column lists the global resolution assigned to each combination. As noted, the resolution increases with the number of CSCs, tending to reach the global limiting resolution, $P_{lim} = 0.892$.

2.4. Comprehensive search of CSCs

The simplest and most immediate way of finding optimal CSCs is examining one by one all possible combinations that can be established, namely, making a comprehensive search in the conditions grid. Since a finite number of combinations is examined, and each solution is a set of integers, the problem can be classified as a combinatorial optimisation.

The comprehensive search is performed using integer arithmetic: binary for two groups, ternary for three and so on. A systematic sum of the coded groups is carried out up to cover all possible combinations (Fig. 3a and b). However, in order to reduce the computation volume, it should be noted that there are combinations representing the same distribution. For example, in Fig. 3c, after decoding the compounds assignment, two identical sets are obtained, which represent exactly the same distribution, since "group 1" and "group 2" tags are meaningless and can be interchanged. Therefore, only one of these combinations should be examined. Note that the saving in computation time is exponential: with two groups, only half of the combinations should be examined, with three only one-sixth, etc.

There are situations where defining groups of compounds will involve the examination of a smaller number of combinations, and others, where outlining the problem by forming groups of separation conditions will be more advantageous. This depends on the number of compounds to be separated, number of separation conditions in the grid and number of target groups. Fig. 4 can be helpful for deciding which strategy should be adopted. The line

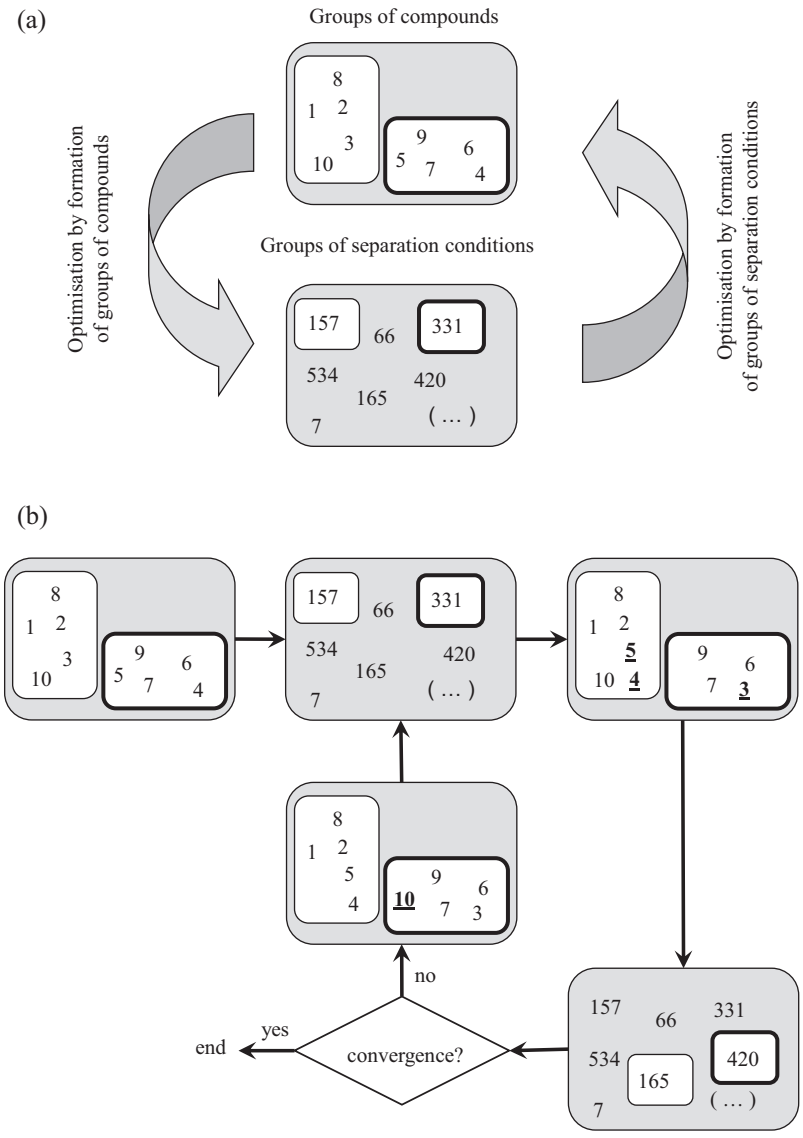

Fig. 2. (a) Search spaces by formation of groups of compounds (left) and formation of groups of separation conditions (right). (b) Local method: the search space of groups of compounds and groups of separation conditions are interchanged. More details are given in the text.

indicates when the selection of the approach is indifferent. Above the diagonal, the problem is best outlined by formation of groups of separation conditions, and below, by formation of groups of compounds.

2.5. Approaches to reduce the computation volume in the search of complementary conditions

The comprehensive exploration of all possible combinations is relatively simple, but it is only practical when the number of combinations to be examined is small. For a high number of combinations, faster approaches are mandatory to get feasible computation times. Some of such approaches are described below.

2.5.1. Genetic algorithms

Genetic algorithms (GAs) facilitate the search of optimal CSCs, using the two perspectives: formation of groups of compounds or formation of groups of separation conditions. GAs shorten the computation time, since the exploration is progressively focused on the most promising regions in the search space. GAs usually start with a random population of encoded solutions, consisting of 20–500 individuals typically, which is called the initial population. This population is made to evolve using rules that simulate natural selection processes (Fig. 5). The first step consists of evaluating the quality of each individual in the population (in our problem, the global resolution in each CSC), which is called fitness. With the fitness information, the initial population is made to undergo some mathematical operations mimicking genetic processes: reproduction, mating, crossover and mutation. These simulate the way in which the best individuals survive and thrive in Nature, crossing their genetic information to give rise to better new individuals. Each cycle is called generation. The process stops when no improvement is observed after a reasonable number of generations, that is, when one or more solutions dominate the population (i.e. the genetic diversity falls below a certain threshold).

The more complex the situation, the more advantageous the application of GAs. GAs will be competitive if the number of function evaluations needed to reach convergence (i.e. number of generations multiplied by the population size) is significantly smaller with regard to the comprehensive search.

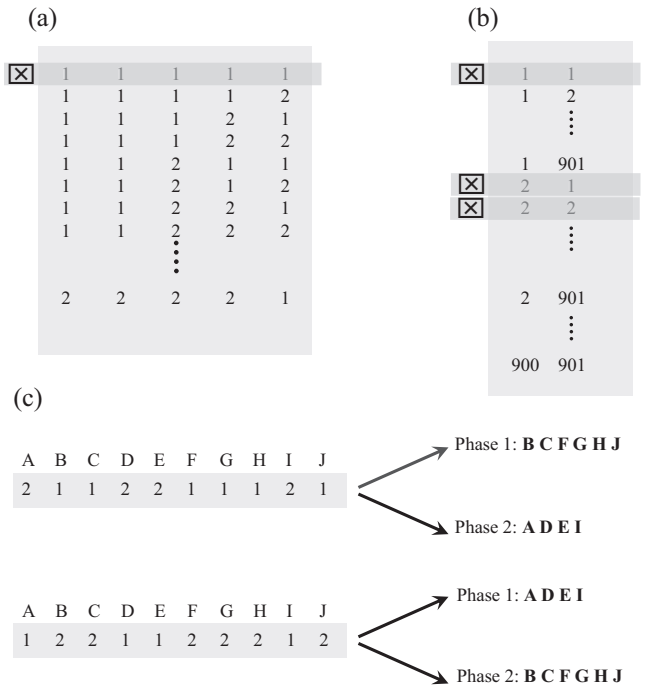

Fig. 3. Systematic combinatorial search: (a) examination of groups of compounds ($ns = 5$ and $ng = 2$), (b) examination of groups of separation conditions ($np = 901$ and $ng = 2$) and (c) combinations that represent the same distribution. The conditions [1 1] and [2 2] in (b) should not be examined because it imply only one condition, and condition [2 1] was already examined as [1 2].

2.5.2. Locally optimised genetic algorithm (LOGA)

In order to find optimal CSCs in complex problems, a more powerful approach was developed. This approach includes an interchange of the search spaces of compounds and separation conditions, looping the interconversions to create an iterative optimisation method (Fig. 2b). After several iterations, no further improvement in resolution will be observed, and the distributions of compounds and separation conditions will not change anymore. Convergence will be reached when the combination of separation conditions becomes the optimal. The interchange of spaces improves the features of the initial population within each generation.

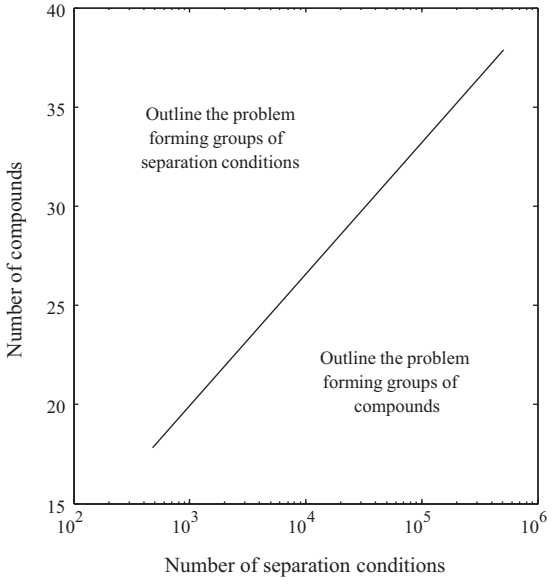

Fig. 4. Selection of the best search space to find out the optimal CSCs, based on the number of compounds and separation conditions.

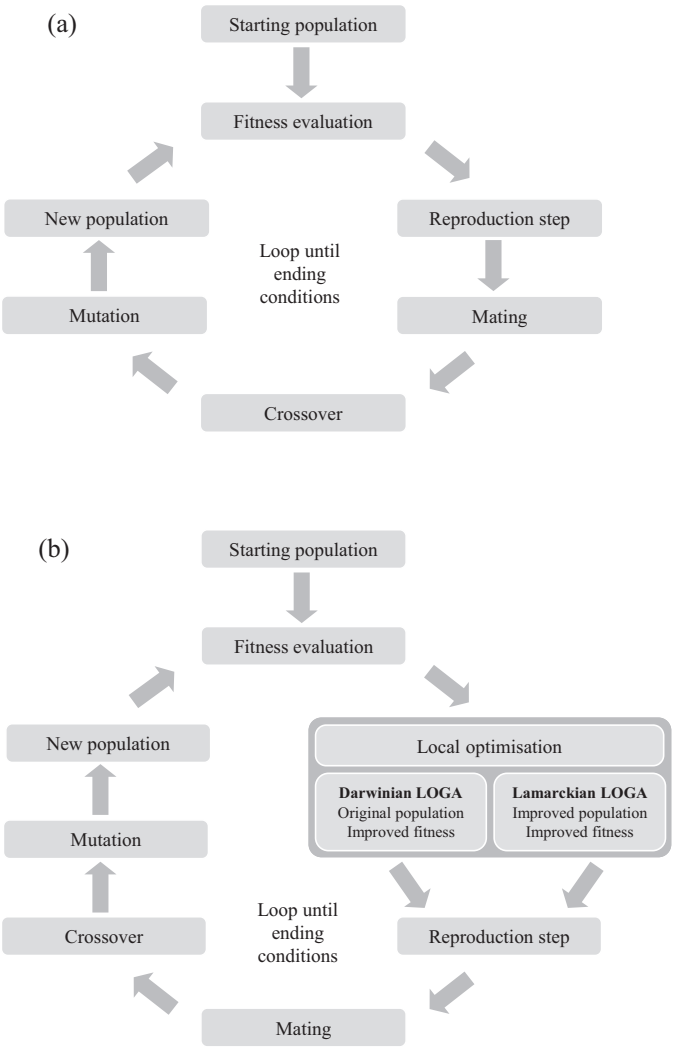

Fig. 5. Genetic algorithms used in the fast computation of CSCs: (a) conventional GA and (b) LOGA. Two kinds of LOGA can be defined according to the selected strategy for reproducing the population: Darwinian and Lamarckian LOGA.

The problem of 10 compounds can be used to illustrate how the local search operates (Fig. 2b). An initial distribution of compounds resolved with the separation conditions with indexes 157 and 331 will be assumed again. The method starts reassigning each compound to the condition offering the highest resolution (157 or 331). A new distribution of compounds is obtained (observe that compounds 3, 4 and 5 have changed to another group). The process follows finding the conditions resolving the new distribution the best (conditions 165 and 420), and is repeated up to convergence, which occurs after a small number of iterations (usually less than five).

The local search method has the disadvantage that, although it improves a given distribution of compounds, it is unable to find the global solution of the problem. However, the solution found is highly precise, due to the stepwise refinement. In contrast, conventional GAs are able to explore the search space globally, obtaining a solution close to the optimal, although with low precision. There is no 100% guarantee that the solution found by the GAs will be the real optimal one, although it is expected to be very close.

The implementation of the local search method as an internal step of a GA algorithm gives rise to a hybrid GA, which was

called "locally optimised genetic algorithm" (LOGA), which takes advantage of both approaches (Fig. 5b) [18]. The local search alters the reproduction pattern or the individuals in the initial population along each generation. Coupling a GA with a local search can be performed in two ways: using the initial population with the reproduction probabilities for the optimised population (Darwinian reproduction mechanism), or the improved population with their own reproduction probabilities (Lamarckian reproduction mechanism). The Darwinian approach leads to a slower loss of diversity, and consequently, it retains the search ability longer, so there will be more chances to succeed with complex problems. This was the option used in this work.

2.5.3. Sequential search of complementary situations based on the peak count concept

A different approach is based on the optimisation of the peak count, PC (i.e. number of peaks that exceed an established resolution threshold, measured as elementary resolution) [20,21]. In this approach, among the conditions that resolve the largest number of compounds, that one showing the highest resolution is selected. This is done based on the fractional peak count:

$$\mathrm{fPC} = \mathrm{PC} + \mathrm{f} \quad (2)$$

where f quantifies the global resolution of the peaks that exceed the threshold.

In this approach, the CSCs are found one after another. For this reason, it was called "sequential" or "stepwise" search [20]. The first selected CSC is the one that resolves the maximal number of compounds in the conditions grid. The next CSC is focused on the compounds unresolved by the first one, again aimed to resolve a maximal number of compounds. The process is repeated with the remaining compounds in the sample, finding additional CSCs, until all compounds are resolved above a certain threshold denoting a satisfactory resolution level that the analyst arbitrarily sets. Ideally, when all CSCs are considered altogether, all compounds in the sample will be resolved above the threshold. However, it may happen that one or more compounds remain unresolved (i.e. below the threshold) under any experimental condition.

Note that this outline is opposed to the search strategy followed in Sections 2.4, 2.5.1 and 2.5.2, where the performance (global resolution) of a number of combinations of CSCs (previously defined) is examined. In the sequential search, the number of CSCs needed to resolve the mixture grows as the required threshold becomes more demanding. Therefore, the user has no direct control on the number of CSCs resulting from the search. Also, subsequent CSCs will have associated a decreasing number of compounds, since the process leaves out those formerly selected.

2.5.4. Locally optimised sequential search (LOSS)

The sequential search is more complex than the comprehensive search, but still rather simple. It has the advantage of a fast computation, being competitive against the comprehensive search and conventional GAs, but it finds the actual optimal solution only for simple problems. Often, the sequential search yields combinations with a global resolution smaller than that found by LOGA. Note that the sequential search just gives a solution fulfilling the resolution requirements that the analyst arbitrarily has set (i.e. all peaks should exceed a certain threshold). The origin of this pitfall is that along the process, the CSCs are found keeping out those compounds selected by a former CSC, and therefore, they are excluded for further improvements. Consequently, the resolution potential is not totally exploited. Overcoming this limitation would need the implementation of a feedback mechanism inside the algorithm structure, which allowed the excluded compounds to participate in further steps of the search. We propose here the local search used in the LOGA algorithm.

As indicated, the sequential search does not allow the control of the number of CSCs. It may happen that it yields, for a given threshold, more CSCs than the target number (e.g. two or three CSCs), if the resolution requirement is too high. In that case, those CSCs exceeding the required number are discarded. The opposite may also happen: the threshold established in the sequential search can yield a number of CSCs below the target number, meaning that there is no need of such a high number of CSCs to reach satisfactory resolution. In other words, the solution found is more economical than the requirements. The implementation of the local method, once the sequential search is completed, allows adapting the yielded number of CSCs to that desired by the analyst, with significant improvements.

3. Experimental

The probe compounds were seven acids (2-, 3- and 4-nitrobenzoic acid, benzoic acid, resorcinol, phenol and *m*-cresol), seven bases (*N*,*N*-dimethylbenzylamine, 2.6-dimethylaniline, 2,4,6-trimethylpyridine, 4-chloroaniline, aniline, *p*-toluidine and pyridine) and one amphoteric compound (3-aminophenol). The experimental design consisted of 36 mobile phases: 3 organic modifier levels (20, 40 and 60% acetonitrile, v/v) and 11 pH levels, covering the 2–13 range [19]. The scan of the conditions grid was designed so that the distance between consecutive conditions corresponded approximately to the experimental uncertainties: 0.1 units in both pH and organic solvent percentage.

A chromatograph equipped with a dual pump and a UV–visible detector was used. The flow-rate was 1 ml min^{-1} for the mobile phases containing 40 and 60% acetonitrile, and 3 ml min^{-1} for 20% acetonitrile. The separation was carried out with a 15 cm × 4.6 mm i.d. polymeric C18 column with 15–20 μm particle size from Polymer Labs (Model PLRP-S 100 Å). All measurements were performed at 25 °C. The pH was measured with a Crison potentiometer (Model MicropH 2002, Barcelona, Spain), with a precision of ±0.002 pH units, using a Ross electrode (Orion Model 8102, a combination of a glass electrode and a reference electrode with 3.0 M KCl aqueous solution as salt bridge) [19].

The routines to compute the CSCs approaches were developed in MATLAB 2010b (The MathWorks Inc., Natick, MA, USA).

4. Results and discussion

4.1. Separation capability of the chromatographic system

The retention of the probe compounds was described using a model based on the normalised solvent polarity, which includes the changes with pH [19]. The performance of the chromatographic column was measured using the peak purity [15], which measures the resolution associated with each peak. The peak purity is a normalised measurement that takes into account the chromatographic peak shape (position, size and elution profile). Its meaning is straightforward: for a given peak, it indicates its non-overlapped fraction. Table 2 shows the limiting elementary peak purities and the optimal values obtained for the mixture of 15 ionisable compounds, using a single experimental condition and two CSCs. The global peak purities are indicated at the bottom of the table.

In this example, the limiting peak purities are close to 1.00 (P_{lim} = 0.994), indicating that almost baseline resolution can be reached for all compounds. The lowest peak purity value corresponds to 3-aminophenol (P_{lim} = 0.993). However, reaching the limiting values would require a specific mobile phase composition for each compound. Naturally, the use of 15 different experimental conditions is unfeasible. In practice, an analyst would select a single experimental condition to resolve the mixture. However, the

Table 2
Elementary peak purities (limiting values and optimal values for a single separation condition and two CSCs), and corresponding partial and global resolutions for the sample containing the 15 ionisable probe compounds.

Probe compound	Limiting resolution and needed conditions		Optimal single separation condition: 20% acetonitrile/pH 6.5	Optimal CSC1: 20% acetonitrile/pH 3.3	Optimal CSC2: 22.4% acetonitrile/pH 10.8
	Peak purity	Acetonitrile/pH		Peak purity	
2-Nitrobenzoic acid	1.000	29.4%/2.8	0.863	**1.000**	0.294
3-Nitrobenzoic acid	1.000	20.0%/2.4	0.434	**0.999**	0.316
4-Nitrobenzoic acid	1.000	20.8%/3.0	0.491	**0.999**	0.019
Benzoic acid	1.000	20.0%/4.8	0.807	**1.000**	0.321
Resorcinol	1.000	20.0%/2.1	0.863	**0.996**	0.501
Phenol	1.000	28.2%/5.3	1.000	1.000	**1.000**
3-Aminophenol	0.993	20.0%/4.5	0.861	0.156	**0.993**
m-Cresol	1.000	27.2%/2.0	0.999	0.490	**1.000**
N,*N*-Dimethylbenzylamine	1.000	20.0%/7.6	0.998	0.977	**1.000**
2,6-Dimethylaniline	1.000	38.4%/4.1	0.993	0.805	**0.999**
2,4,6-Trimethylpyridine	1.000	20.0%/6.6	1.000	0.019	**1.000**
4-Chloroaniline	1.000	38.4%/4.1	0.993	0.682	**0.999**
Aniline	1.000	28.2%/5.3	1.000	0.267	**1.000**
p-Toluidine	1.000	20.0%/4.9	0.998	0.264	**1.000**
Pyridine	1.000	20.0%/11.5	0.999	0.782	**1.000**
Partial peak purity				0.994	0.991
Global peak purity	0.994		0.108	0.986	

presence of compounds with very similar chromatographic behaviours will make the separation to fail at any mobile phase composition (at least inside the conditions grid), using a single mobile phase. This is the case of the problem under study (Table 2).

As can be seen, using the optimal single experimental condition, many compounds in the sample did not reach the maximal system resolution performance (compare the optimal elementary purities with the limiting peak purities). This is the case of 3-nitrobenzoic acid ($p = 0.434$), 4-nitrobenzoic acid ($p = 0.491$), benzoic acid ($p = 0.807$), 2-nitrobenzoic acid ($p = 0.863$), resorcinol ($p = 0.863$) and 3-aminophenol ($p = 0.861$). In fact, only 10.9% of the system resolution performance is reached when a single isocratic mobile phase is optimised. Fig. 6a depicts the chromatogram for the mixture of 15 ionisable compounds, using the optimal composition. The analysis time was 75 min, with the less retained compounds showing severe overlapping.

4.2. CSCs search

4.2.1. Comprehensive search, conventional GA and LOGA

A single mobile phase did not allow the separation of the 15 ionisable compounds. However, since the resolution expectancies for the system were high (i.e. all elementary limiting purities were close to 1.00), the search for optimal CSCs was expected to provide significant improvements. It should be remarked that the same information used to perform a conventional optimisation is all what is needed to find the optimal CSCs (i.e. no additional experiment is required). As described in Section 2, the search of CSCs can be carried out using several approaches, which were applied to the mixture of ionisable compounds.

The comprehensive search (Section 2.4) inspected the global resolution of all possible combinations of mobile phases inside the conditions grid, which contained 101 (% acetonitrile) × 111 (pH) = 11211 mobile phases. According to Fig. 4, the search for a solution through the formation of groups of compounds (see Section 2.3.1) was preferable. The number of combinations to examine (computation time in a personal computer equipped with a Core i7 with 4 GB RAM and Windows 7 64-bit Enterprise is given in parenthesis) was 1.64×10^4 (0.28 min) for two CSCs and 2.38×10^6 (40.6 min) for three CSCs. If the search were outlined by forming groups of mobile phases (see Section 2.3.2), the number of combinations would be 6.28×10^7 (17.9 days) for two CSCs and 2.35×10^{11} (7.63 years) for three CSCs. We selected a non-excessively

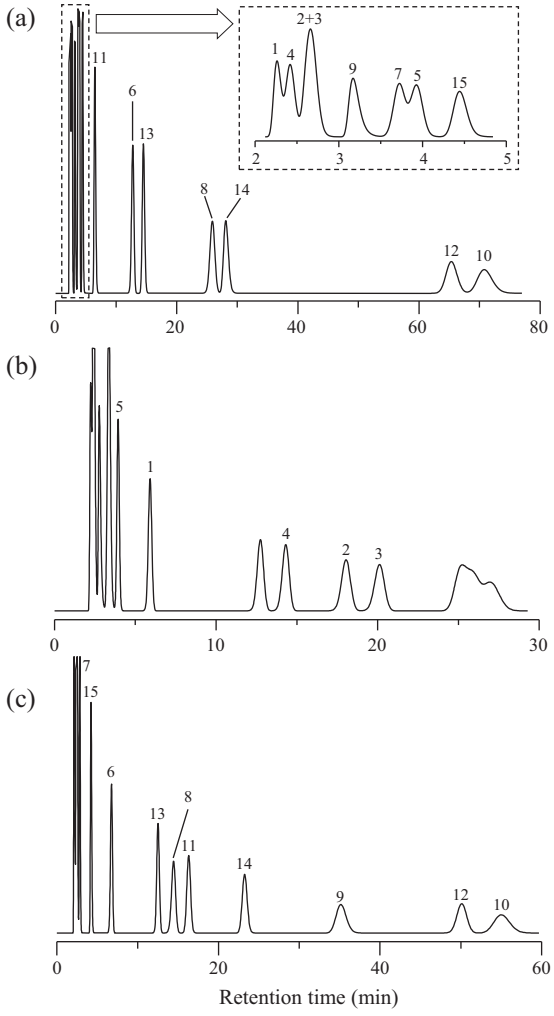

Fig. 6. Simulated chromatograms corresponding to the optimal conditions: (a) single mobile phase (20% acetonitrile at pH 6.5) and (b, c) best combination of two CSCs: CSC1 (20% acetonitrile at pH 3.3) and CSC2 (22.4% acetonitrile at pH 10.8), respectively. In (b) and (c), only the compounds assigned to each CSCs are indicated. Elementary and global resolutions are given in Table 2.

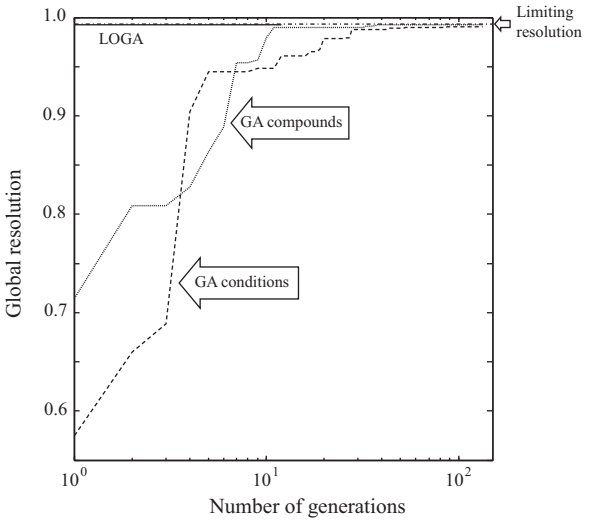

Fig. 7. Evolution of genetic algorithms in the search of three CSCs: solution found versus number of generations for a conventional GA forming groups of compounds or groups of separation conditions, and for LOGA.

complex problem to be able to determine the true optimal composition through a comprehensive search, within a reasonable computation time. In most usual cases (less than 20 compounds and grids with up to 1×10^4 conditions), the outline by formation of groups of compounds is more advantageous.

Fig. 6b and c shows the optimal chromatograms for two CSCs. Table 2 lists the elementary peak purities for each optimal CSC (the assigned compounds to each CSC are marked in bold). CSC1 resolved 5 compounds (all acidic) and CSC2, the 10 remaining compounds. Almost full resolution was achieved with two CSCs, with a global peak purity of 0.986. This represents an increase in the degree of success from 10.9% for a single experimental condition to 99.2% for two CSCs.

For three CSCs, the compositions were 21.8% acetonitrile at pH 2.9 (5 assigned compounds), 29.4% acetonitrile at pH 5.2 (4 compounds) and 23.2% acetonitrile at pH 10.9 (6 compounds). For this sample, the use of three CSCs instead of two would not yield a significant improvement in the separation (the global peak purity changed from 0.986 for two CSCs to 0.999 for three CSCs). Hence, the increased experimental effort is not worth. Therefore, the problem is solved by preparing only an additional mobile phase with respect to the conventional optimisation.

For our discussion, it is interesting to examine the capabilities of a conventional GA and LOGA to find the optimal solution. Using a conventional GA, the population consisted of 30 candidate solutions. With more complex problems, the population should be larger. Due to the random nature of GAs, giving a general indication of the computation time is difficult. However, in all cases examined, it amounted only a few seconds.

Fig. 7 shows, for a conventional GA, the evolution of the optimisation with an increasing number of generations, for the separation of the 15 ionisable compounds using three CSCs and the search spaces of compounds and separation conditions. In this example, the conventional GA converged on the correct solution in a relatively small number of generations. In general, the algorithm should be adapted to the nature and complexity of the problem to achieve convergence in a short time.

Fig. 7 also shows the evolution of LOGA to find the solution. The search efficiency is significantly enhanced: by repeatedly applying the algorithm, usually 0–3 generations were required for both two and three CSCs, to find the right solution using much shorter computation times than the conventional GA (zero generations means that the local method applied to the initial random population provided the right solution without any genetic operation). The magnitude of the evolution of LOGA is so small that it results imperceptible at the figure scale.

To understand the effectiveness of LOGA, it should be explained that conventional GAs progressively concentrate the search effort in the most promising solutions, gradually abandoning other regions in the factor space with smaller success. Also, in order to maintain a certain diversity and sufficient exploration capability, some random individuals are incidentally included in the population. LOGA is able to explore and take benefit of the potential adaptation of each individual in the population, along all generations. This gives rise to a larger accuracy in the search and a higher convergence speed. As indicated, in some instances, it is possible to converge without performing any genetic operation. In the cases where the comprehensive search cannot be performed due to the high computation time, the result provided by LOGA can be considered as a very good approximation to the optimal value. Evidence supporting this idea is that when LOGA is started with different random populations, it finds the same solution in very complex problems.

4.2.2. Sequential search of CSCs based on the peak count concept

The sequential search of CSCs (see Section 2.5.3) requires the establishment of a threshold of elementary peak purities (i.e. minimal value that each compound should reach). As the threshold demand is increased, one can expect that the required number of CSCs will be constant or will increase. The analyst ignores a priori the required threshold to obtain a certain number of CSCs. A tool that can assist in this aim is a graph relating the resolution threshold with the number of CSCs (scan of thresholds). The solution of the sequential search will be given by the highest threshold yielding the desired number of CSCs.

Fig. 8a illustrates such a graph for the problem of 15 ionisable compounds. It should be reminded that in the problem being solved, the limiting elementary purities are close to 1.00 (Table 2), which means that all compounds can be resolved. For problems with lower limiting values, a too demanding threshold will be unattainable for some compounds. The irregular behaviour at high thresholds observed in Fig. 8a can be explained by considering that the approach performs a sequential search of CSCs, where each assignation of compounds to a CSC affects the next assignation to a subsequent CSC. This is translated in different solutions depending on the adopted threshold (i.e. for neighbouring thresholds, the assignation of at least one compound to a given CSC can change).

As discussed in Section 2.5.3, the sequential search may not find the real optimum, although the solution can be satisfactory and even similar to that found by the comprehensive search. For the selected problem, the compositions found by the sequential search were 21.6% acetonitrile at pH 10.7 (10 assigned compounds) and 20.0% acetonitrile at pH 3.3 (5 compounds), for the first and second CSCs, respectively. The partial resolutions were 0.970 and 0.994, respectively, and the global resolution: 0.964.

The result obtained by the sequential search differed from the other approaches. The resolution with two CSCs was poorer. It is, however, interesting to note that the computation times with this approach are always very short, since the CSCs are straightforwardly found from the examination of the elementary peak purities matrix, calculating the fractional peak count (Eq. (2)) for each mobile phase in the conditions grid (11211 mobile phases for this problem). Because a full inspection of the combinations of compounds or separation conditions is not carried out, the number of evaluations is significantly smaller with regard to the comprehensive search. Note, finally, that the sequential search yields simultaneously, as a result, groups of compounds and the associated CSCs.

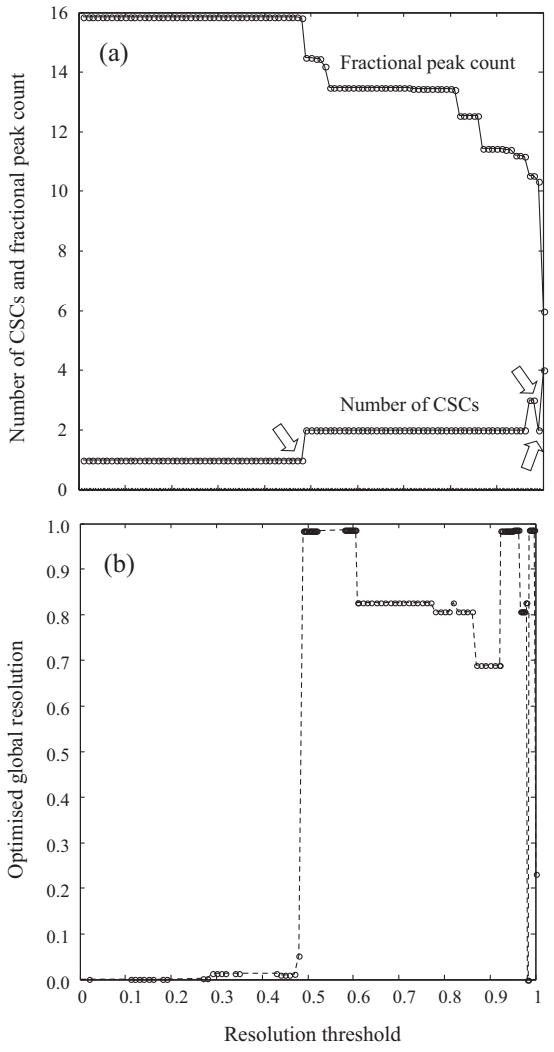

Fig. 8. Implementation of the approaches based on the peak count concept: (a) number of CSCs required by the sequential search against the resolution threshold (the fractional peak count is also plotted; note that for a given number of CSCs, the compositions found by the sequential method are those associated to the largest threshold, which is marked with an arrow); (b) Optimisation of two CSCs with LOSS.

4.2.3. Locally optimised sequential search (LOSS)

This approach combines the unsupervised sequential search of CSCs, based on the peak count concept, with the iterative local method that interchanges the search spaces of compounds and separation conditions (see Section 2.5.2). For short, we have called this approach LOSS from "locally optimised sequential search". In contrast to the sequential search, this approach requires that the analyst decides a priori a target number of CSCs. The method starts with a candidate solution consisting of the CSCs obtained from the sequential search. The local search is applied to these CSCs after adapting them to the target number, without taking into account the previous solute assignation. Each compound is reassigned to that CSC where it reaches the highest elementary resolution. Then, the mobile phase in the conditions grid resolving optimally each group of compounds obtained in this way is selected, which gives rise to a new combination of CSCs. Owing to the internal optimisation, this new combination will have a global resolution equal or larger than that yielded by the sequential search. Here, a new iteration begins: the compounds are again reassigned among the new CSCs, and the process is repeated up to convergence, that is, up to the global resolution cannot be improved anymore, or up to obtain two consecutive equal combinations of CSCs.

After running the new method, there is however no guarantee that another combination with an even better global resolution exists. This drawback was solved by restarting the process with different thresholds (i.e. performing a scan of thresholds). Each threshold gives rise to a different combination of CSCs (a different candidate solution), making the process similar to starting a GA with different initial populations. The only decision the analyst should take is the threshold range to scan and the step size (e.g. 0.65–1.0 in steps of 0.01). Similarly to simulated annealing [22], in further steps, the scanning effort is concentrated in those promising regions where the global resolution is higher, using a smaller step size. LOSS eliminates some pitfalls found in the sequential search approach: those compounds excluded in the sequential search are optimised later in the local search, and if an extra CSC is not needed, it will be left aside in the search space of compounds.

Fig. 8b depicts the global resolutions calculated in the search of two optimal CSCs for the sample of 15 ionisable compounds, as a function of the threshold value. The best CSCs are found in three regions in the threshold plot (around a resolution threshold of 0.55, 0.95 and 0.99). The plot was obtained by making three successive scans of deeper detail centred in those threshold regions found in the previous scan, which yielded the best resolution after applying the local method (i.e. focusing the scanning effort in the best regions). The solution provided by LOSS agreed with LOGA. We have found that LOSS is able to succeed even in highly complex problems. Occasionally, it may even improve the solution given by LOGA, with a similar computation time.

5. Conclusions

In the analysis of complex samples, an unsatisfactory separation is very often obtained using a single experimental condition (e.g. a single isocratic mobile phase, a single gradient, a single solvent system, column or separation technique). This makes the development of new methodologies able to increase the separation space or enhance the selectivity mandatory, which can be achieved by applying chemical or physico-chemical modifications (e.g. changing the type of stationary phase or introducing a secondary equilibrium). Such a change means discarding all the work done and developing an optimisation to find out another single separation condition from a new experimental design, which may succeed or not.

In this work, we discuss a way to increase the separation space by selecting two or more CSCs (e.g. two or more isocratic mobile phases, gradients, combination of columns or even separation techniques, etc.). This can be done by taking advantage of the experimental work already available, using the same data collected to carry out the conventional optimisation that failed (no more work is needed). The increase in the separation space gives a chance of improving the resolution and eventually the optimisation robustness.

The idea of obtaining complementary separation conditions is straightforward and apparently simple, but the computation volume and complexity can be a challenge. This work describes and discusses in detail several approaches that have been developed to find out optimal CSCs. This task requires the implementation of sophisticated algorithms to reduce the computation time for complex problems. The sequential search of CSCs, developed in earlier work [20], simplifies the computation. However, for complex problems, the solution reached is not the optimal (although, it may be close enough to it and sufficiently satisfactory).

The hybrid method (LOSS) developed for this work, based on the peak count concept and assisted by a local search, solves the drawbacks of the sequential search, picking out the best features of previous approaches: the simplicity of the sequential search, the

global nature of genetic algorithms through the scan of thresholds, the accuracy of the local optimisation and the fast computation of LOGA. Therefore, LOSS has many advantages: high reliability in finding the solution, reduced complexity in the implementation of the algorithm, which makes it more accessible to analysts without specialised programming skills, and short computation time.

Finally, it is interesting to compare LOSS with LOGA. LOSS improves each solution independently from the others (i.e. there is no improvement through a cooperation with other solutions), while Darwinian LOGA is able to evolve solutions allowing to combine and change the information over a few generations. Nevertheless, the possibilities of success in LOSS are very close to those in LOGA. To understand the reason for the success of LOSS, it should be considered that LOSS is similar to a generation in LOGA with Lamarckian evolution, where the initial individuals are replaced by those optimised. The difference is that in the Lamarckian LOGA, the initial population is randomly generated, while in LOSS it is generated through a systematic scan of thresholds, which greatly increases the chances of success, without the need to evolve through generations.

Acknowledgements

This work was supported by Project CTQ2010–16010/BQU (Ministerio de Ciencia e Innovación, MICINN of Spain) and FEDER funds. C. Ortiz-Bolsico thanks a FPI grant from MICINN.

References

[1] P.J. Schoenmakers, Optimisation of Chromatographic Selectivity, Elsevier, Amsterdam, 1986.
[2] J.C. Berridge, Techniques for the Automated Optimisation of HPLC Separation, Wiley, New York, 1988.
[3] S. Ahuja, Selectivity, Detectability and Optimisation in HPLC, Wiley, New York, 1989.
[4] J.W. Dolan, D.C. Lommen, L.R. Snyder, J. Chromatogr. 485 (1989) 65.
[5] L.R. Snyder, J.W. Dolan, D.C. Lommen, J. Chromatogr. 485 (1989) 91.
[6] R.M. Smith, Retention and Selectivity in Liquid Chromatography, Elsevier, Amsterdam, 1995.
[7] R. Cela, M. Lores, Comput. Chem. 20 (1996) 175.
[8] R. Cela, M. Lores, Comput. Chem. 20 (1996) 193.
[9] S. Heinisch, E. Lesellier, C. Podevin, J.L. Rocca, A. Tchapla, Chromatographia 44 (1997) 529.
[10] L.R. Snyder, Practical HPLC Method Development, 2nd ed., Wiley, New York, 1997.
[11] J.R. Torres-Lapasió, M.C. García-Álvarez-Coque, J.J. Baeza-Baeza, Anal. Chim. Acta 348 (1997) 187.
[12] A.M. Siouffi, R. Phan-Tan-Luu, J. Chromatogr. A 892 (2000) 75.
[13] W.D. Beinert, R. Jack, V. Eckert, S. Galushko, V. Tanchuck, I. Shishkina, Int. Lab. 31 (2001) 16.
[14] I. Molnár, J. Chromatogr. A 965 (2002) 175.
[15] M.C. García-Álvarez-Coque, J.R. Torres-Lapasió, J.J. Baeza-Baeza, Anal. Chim. Acta 579 (2006) 125.
[16] J.R. Torres-Lapasió, M.C. García-Álvarez-Coque, J. Chromatogr. A 1120 (2006) 308.
[17] G. Vivó-Truyols, J.R. Torres-Lapasió, M.C. García-Álvarez-Coque, J. Chromatogr. A 876 (2000) 17.
[18] G. Vivó-Truyols, J.R. Torres-Lapasió, M.C. García-Álvarez-Coque, Chemom. Intell. Lab. Syst. 59 (2001) 89.
[19] J.R. Torres-Lapasió, M.C. García-Álvarez-Coque, E. Bosch, M. Rosés, J. Chromatogr. A 1089 (2005) 170.
[20] A. Ortín-Sebastián, J.R. Torres-Lapasió, M.C. García-Álvarez-Coque, J. Chromatogr. A 1218 (2011) 5829.
[21] A. Ortín, J.R. Torres-Lapasió, M.C. García-Álvarez-Coque, J. Chromatogr. A 1218 (2011) 2240.
[22] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C++: The Art of Scientific Computing, 2nd ed., Cambridge University Press, New York, 1992.