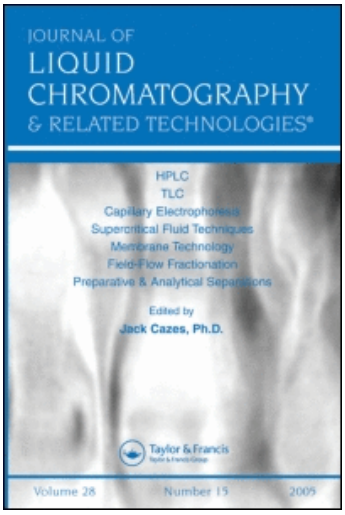


Journal of Liquid Chromatography & Related Technologies

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597273

"PRISMA" Model for Computer-Aided HPLC Mobile Phase Optimization Based on an Automatic Peak Identification Approach

Szabolcs Nyiredy[ab]; Werner Wosniok[c]; Herbert Thiele[d]; Otto Sticher[a]

[a] Department of Pharmacy, Swiss Federal Institute of Technology (ETH) Zurich, Zürich, Switzerland [b] Research Institute of Medicinal Plants, Budakalász, Hungary [c] Department of Statistic, University of Bremen, Bremen 33, Germany [d] Bruker-Franzen Analytik GmbH, Bremen 61, Germany

To cite this Article Nyiredy, Szabolcs , Wosniok, Werner , Thiele, Herbert and Sticher, Otto(1991) '"PRISMA" Model for Computer-Aided HPLC Mobile Phase Optimization Based on an Automatic Peak Identification Approach', Journal of Liquid Chromatography & Related Technologies, 14: 16, 3077 — 3110

To link to this Article: DOI: 10.1080/01483919108049377

URL: http://dx.doi.org/10.1080/01483919108049377

JOURNAL OF LIQUID CHROMATOGRAPHY, 14(16&17), 3077-3110 (1991)

"PRISMA" MODEL FOR COMPUTER-AIDED HPLC MOBILE PHASE OPTIMIZATION BASED ON AN AUTOMATIC PEAK INDENTIFICATION APPROACH

SZABOLCS NYIREDY[1]+, WERNER WOSNIOK[2], HERBERT THIELE[3], AND OTTO STICHER[1]*

[1]*Department of Pharmacy*
Swiss Federal Institute of Technology (ETH) Zurich
CH-8092 Zürich, Switzerland
[2]*Department of Statistic*
University of Bremen
D-2800 Bremen 33, Germany
[3]*Bruker-Franzen Analytik GmbH*
D-2800 Bremen 61, Germany

ABSTRACT

The tripartite "PRISMA" optimization model, as part of the "PRISMA" system, includes all possible solvent combinations between 1-4 solvents, with a possible fifth one as modifier. The solvent composition is characterized by the solvent strength (S_T) and the selectivity points (P_S).

At a constant S_T the correlation between the P_S and the retention data (horizontal function) can be described by a quadratic function. For constant P_S the solvent strengths and retention data correlate (vertical function) with a logarithmic function. These correlations are used to formulate a mathematical model for the dependence of retention times

+Present adress: Research Institute of Medicinal Plants, P.O.Box 11, H-2011 Budakalász, Hungary

(capacity factor) on the mobile phase composition. Unknown compounds are estimated in the mathematical model from a sequence of standard chromatograms after having identified individual peaks by an automatic procedure. Only retention times, relative peak areas, and information about the mobile phase compositions are required as input for the identification approach. The approach involves a combination of statistical methods which exploit both the basic properties of retention data and the mathematical relation between retention data, selectivity points, and solvent strength as derived from the "PRISMA" model. Diagnostic information for checking the identification is generated as a by-product. The mathematical model completed by the estimated constants predicts the expected retention times for each possible mobile phase combination. Peak start and peak end times are predicted in a way similar to the retention times, once the identification has been performed. The most important aspects of a chromatogram can thus be predicted for arbitrary mobile phases.

The separation quality of predicted chromatograms is assessed by the chromatographic response function (CRF). The optimal mobile phase combination is that which theoretically generates the chromatogram with the maximal CRF value. This optimal composition is found by a simple mathematical procedure, which maximises the CRF in dependence upon the mobile phase combination. The optimum found is a local one if the starting set of chromatograms contains no variation of the solvent strength, and a global one if, in the set of starting chromatograms, the solvent strength is varied in a suitable way. Recommendations for the starting position are given.

Twelve measurements are necessary for a local optimum, and 15 for the global one. To increase the accuracy, six measurements at three different solvent strength levels are proposed. Generally the highest and the lowest solvent strength level differ by ±3 (5) % from the middle level.

This strategy is also relevant when modifiers are used in constant amounts. The chromatographic behavior of substances to be separated can be predicted with 1% accuracy from correlations of k' values and selectivity points. Based on these relationships, an automatical mobile phase optimization strategy for isocratic separations is suggested with the "PRISMA" model.

INTRODUCTION

Since the introduction of computer aided HPLC instruments several mobile phase optimization procedures and criteria have been reported by Glajch et al. [1,2], Berridge [3], Schoenmakers et al. [4], Peeters et al. [5], Bridge et al. [6], Fell et al. [7], as well as by Coenegracht [8], Bartha et al. [9], Lankmayer et al. [10], and Bakalyar [11]. The different optimization

criteria and procedures were extensively summarized in 1985 by Berridge [12] and in 1986 by Schoenmakers [13].

Based on the solvent strength prism for the optimization of multisolvent gradient elution due to Kirkland and Glajch [14], the solvent classification of Snyder [15], as well as the seven point optimization method by Glajch, Kirkland, and Snyder [16] the "PRISMA" optimization system [17] was developed for various analytical and preparative planar and column chromatographic methods. The system is based on the "PRISMA" model for the optimization of mobile phase. It was developed first for TLC [18] and HPLC [19], applied later to the various forced-flow planar chromatographic techniques such as OPLC [20-22] and RPC [23, 24]. Application of the "PRISMA" system to planar chromatographic methods was reported in [25]. The transfer possibilities of TLC mobile phase to various preparative column liquid chromatographic techniques are summarized in [26].

The purpose of this paper is to describe a formal method for the task of peak identification and selection of an optimal mobile phase composition. The method is designed for implementation on a personal computer.

The information required is derived from a series of chromatograms generated from the same sample by varying the mobile phase composition according to a standard scheme derived from the "PRISMA" model [27]. It is based on the quadratic correlation between the selectivity of the mobile phase and the capacity factor (k′) and a logarithmic correlation between the solvent strength and retention data [27]. A standard procedure for peak identification is assumed to have been run in order to obtain preliminary retention times and peak areas for each single chromatogram. Those times and areas together with the mobile phase composition information serve as input for the method to be presented. The method output will be

(i) a robust estimate for each peak area based on the totality of all chromatograms, which can be expected to be more accurate than each preliminary value from a single chromatogram.

(ii) a set of equations for the retention surfaces, which relate the

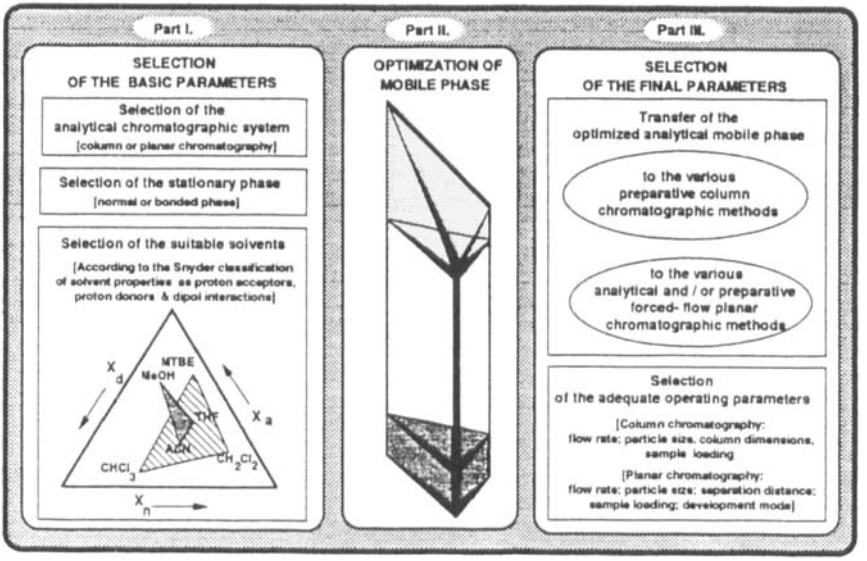


Figure 1 The "PRISMA" system in liquid chromatography.

retention time for each peak with the mobile phase composition which generated it.

(iii) corresponding equations for the start and end of each peak.

(iv) an estimate of the mobile phase composition, which produces a chromatogram with optimally separated peaks.

THEORY

The "PRISMA" System

The "PRISMA" system (Figure 1) consists of three parts. The first part comprises selection of the basic parameters, like the analytical chromatographic system [column (HPLC) or planar (TLC) chromatography], the stationary phase (normal or chemically bonded phase), and the suitable individual solvents. The solvents are selected on the basis of

the Snyder classification [15], according to their properties as proton acceptors and proton donors and their dipole interactions.

In the second part of the system the optimal combination of the selected solvents is established by the "PRISMA" model. The third part of the system concerns the transfer of the optimized mobile phase between the various analytical and preparative column and planar chromatographic techniques. The necessary operating parameters like dimensions, particle size of the stationary phase, flow rate, and, in the case of planar chromatography, the selection of the development mode (circular, linear, anticircular) also belong to the third part of the system.

The "PRISMA" Model for RP-HPLC

The "PRISMA" model, as the pivotal component of the "PRISMA" system, can be visualized as a graphic spatial representation of the solvent strength and the proportions of the components which determine selectivity.

If the S_T values are plotted vertically and if the two dimensional representation of the solvent concentrations, which primarily influence the selectivity, is plotted on the horizontal plane, a prism is obtained with an equilateral triangle as its base. The lengths of the edges of the prism correspond to the solvent strengths of the neat solvents (S_i) in question. As different solvents have differing solvent strengths, the lengths of the edges of the prism are generally unequal and the top plane of the prism will not be parallel and congruent with its base.

If the prism is cut parallel to the base at the height of the lowest edge -determined by the solvent with the lowest solvent strength in the system - the lower part gives a regular prism (Figure 2) where the top and bottom planes are parallel equilateral triangles. This regular part characterizes eluents for reversed phase (RP) HPLC.

The base and top plane of the regular part of the prism are congruent equilateral triangles. The height of this part of the prism corresponds to the solvent strength value of the weakest solvent. Due to the original selection of a decreasing order of solvent strength for tetrahydrofuran ($S_{THF} = 4.5$), acetonitrile ($S_{ACN} = 3.2$), and methanol

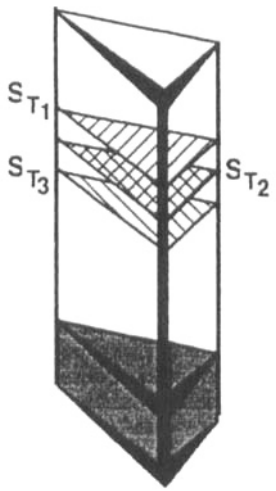


Figure 2 The regular part of the "PRISMA" model for RP-HPLC with the symbol of the modifier.

(S_{MeOH} = 2.6), this is methanol. Eluents characterized by other points on the uppermost plane of the regular part of the model can be obtained by mixing the solvents represented by the apices of the top triangle in the volume proportions corresponding to the point in question. That point of the triangle where the ten-fold values of all three characteristic volume fractions are integers can be defined by a three-digit number. This number - whose digits add up to 10 - can be obtained by multiplying the volume fractions by 10 and arranging them in the order of decreasing solvent strength. The points symbolizing quaternary, ternary, or binary solvent compositions frequently used in optimization (the selectivity points = P_S) are characterized by such three-digit numbers (see Figure 3).

In RP chromatography, the selectivity points on the vertical planes of the regular part of the prism can be obtained by diluting the solvent mixtures with water (S_{HOH} = 0). The solvent strength values decrease from top to bottom; at the base of the prism, the S_T value is zero. If

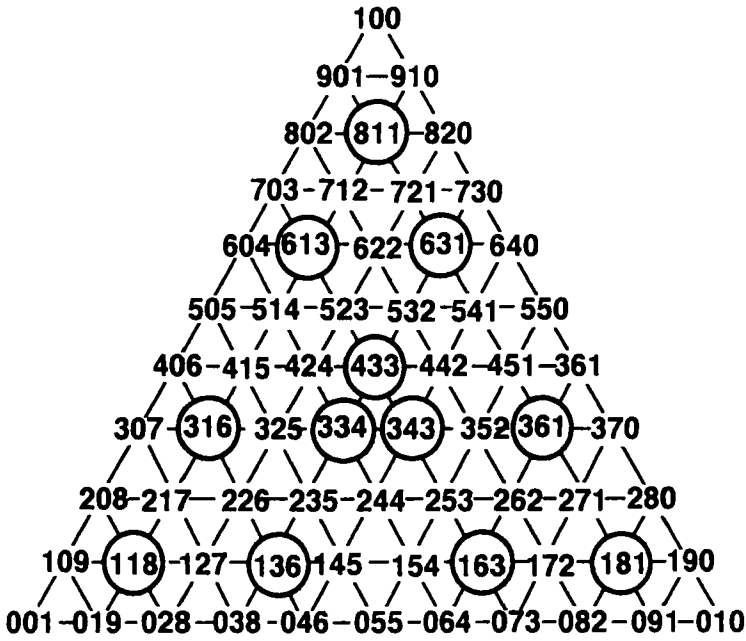


Figure 3 The 12 selectivity points at constant solvent strength for standard sequence of chromatograms for the separation of the local optimum.

sections are taken across the regular prism parallel to the base, triangles of different S_T levels are obtained.

Obviously, all points on one of these triangles represent the same solvent strength, while all points on a vertical straight line correspond to the same selectivity points.

Definitions and Notation

The process of finding an optimal mobile phase combination is based on data of the type produced by standard chromatographic systems. The following notation will be used to describe the procedure:

c_{js} proportion of component s in solvent composition j

$j = 1, 2, \ldots, J, \quad s = 1, 2, \ldots, S. \quad \sum c_{js} = 1$

p_i proportion of the total peak area due to the i-th component in the sample to be separated, i = 1, 2,...,I, $\sum p_i = 1$

m_{jn} mean retention time of th n-th empirical peak, n = 1,2,....,N_j in chromatogram j

a_{jn} relative area of the n-th empirical peak, $\sum a_{jn} = 1$, in chromatogram j

b_{jn} beginning of empirical peak n in chromatogram j

e_{jn} end of empirical peak n in chromatogram j

β_k parameters for the description of individual retention surfaces.

The mobile phase composition will be assumed to consist of four components (S = 4), where the first is the one defining the solvent strength (S_T) while the last three define the selectivity points. The number J corresponds to the number of chromatograms produced. pi denotes the proportions of component i in the sample to be analysed and is unknown a priori. It, as well as the unknown number I of components in the test substance, will be determined in the identification part of the procedure. "Empirical" peaks are those identified in each chromatogram by standard methods. They appear at mean times m_{jn} and have relative areas a_{jn}. They must not be confused with the "individual" peaks, which are the real matter of interest. Each component in the sample to be separated is assumed to generate an individual signal on passing through the column, thus producing its individual peak in the chromatogram. Unfortunately, those individual peaks do not generally appear individually in a real chromatogram, and some of them may lie so close together that, due to limited resolution and random disturbances, they cannot be separated by standard peak identification methods. They are instead identified as only one peak, the "empirical" peak in our terminology.

Different chromatograms recorded for the same sample but with varying proportions c_{js} of the mobile phase composition may result in different clusterings of individual peaks. Retention times belonging to the same individual peak usually vary from one chromatogram to another, if these are generated with different solvent proportions, and even if two

chromatograms are gererated under the same conditions, the results will not be completely identical as a consequence of random disturbances. The relationship between mobile phase composition and mean retention time associated with a component of the sample can be described geometrically by a curved surface in an S-dimensional space whose axes are defined by the first S-1 solvent component and the retention time. The curve is called the retention surface y. Component S can be ignored in this description because of the linear relation $\sum c_{js} = 1$.

Mathematical Model for Retention Surfaces

The relation between retention surface y and solvent composition **c** must be expressed by an equation which generates from a given **c** a prediction for the retention time y. This equation may, of course, contain parameters β_k whose values are unknown in advance and must be estimated from available data. The general form of $y(c,\beta)$ must, however, be specified before applying the methods for identification and optimisation.

The "PRISMA" model has served as a framework for deriving a general mathematical specification for $y(c,\beta)$. Experience [27] has shown that

- the use of different solvent strengths S_T (which corresponds to solvent component c_1) at the same selectivity point changes the logarithm of the retention time linearly:

$$\ln y\,(c_1 \mid c_2, c_3, c_4) = \beta_{10} + \beta_1 c_1, \text{ and that}$$

- the change in the logarithmic retention time generated by the use of different selectivity points (described by c_2, c_3, c_4) and the same solvent strength S_T is well approximated by a quadratic equation [25] of the form

$$\ln y\,(c_2, c_3, c_4 \mid c_1) = \beta_{20} + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_2 c_3 + \beta_5 c_2^2 + \beta_6 c_3^2 .$$

Component c_4 does not appear in the right hand side of the above equation, because it is related to c_2 and c_3 via $c_4 = 1 - c_2 - c_3$ which means that c_4 supplies no further information beyond that contained in c_2 and c_3 and can therfore be omitted here.

Combining the two relations gives a general form for the relation between solvent strength, selectivity point, and retention time:

(RS) $\ln y(c, \beta) = \beta_0 + \beta_1 c_1 + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_2 c_3 + \beta_5 c_2^2 + \beta_6 c_3^2$.
where $c = (c_1, c_2, c_3, c_4)$ and $\beta = (\beta_0, \beta_2, \ldots, \beta_6)$.

The mathematical model for a retention surface defined by (RS) contains a total of 7 unknown parameters, which must be estimated from available data. Estimation of all 7 parameters requires a minimum of $J = 7$ (appropriately choosen) chromatograms which must be produced in order to make the estimation possible. Using only 7 chromatograms, however, implies that there is no possibility of checking the formal assumptions on which (RS) is based nor of generating proper confidence limits for parameters or predicted retention times. Usually it is recommended to produce at least two chromatograms for each of the parameters to be estimated for the purpose of checking.
The decision regarding the compositions c with which to generate the J chromatograms is left largely to the experimenter. The formal condition which must be fulfilled is that the design matrix **X** must have full rank. **X** is defined as having J rows, each row corresponding to the production data of one chromatogram, where the members of row j are

(TF) $x_{j1} = 1, x_{j2} = c_1, x_{j3} = c_2, x_{j4} = c_3, x_{j5} = c_2 c_3, x_{j6} = c_2^2, x_{j7} = c_3^3$.

The requirement that **X** be of full rank is equivalent to demanding that each chromatogram must be produced under conditions which are not yet covered by other chromatograms. A reasonable design for a set of chromatograms which meets this requirement for a local optimum is shown in Figure 3.

For the global optimum of separation a standard sequence is given in Table 1.

The set of standard chromatograms proposed in Table 1 serves to determine the global optimum of the mobile phase composition, because it allows estimation of the parameter β_1, which measures the effect of

Table 1. The 18 selectivity points with variable solvent strength for determination of the global optimum.

Solvent strength level		
1	2	3
811	811	811
361	631	181
181	181	163
118	136	118
316	118	613
433	343	334

solvent strength, as well as determination of β_2,... β_6. The simpler set proposed in Figure 3 involves no variation of solvent strength, and consequently the effect of S_T (c_1) cannot be incorporated in an estimation or optimisation procedure. Only the local optimum of c_2, c_3, and c_4, conditional on the S_T value employed in the standard set, can be found.

The total number of parameters which could reasonably be used to mathematically model the set of all retention surfaces is bounded by the number of independent data points. If the number of observed peak times is $N = \sum N_j$ and J chromatograms are involved, there are only 2N - J data points actually bearing independent information, and consequently no more parameters should be estimated from the data.

The retention time of an individual peak changes according to the value predicted by its retention surface, when chromatograms with different solvent proportions are produced. The sequence of retention times generated in this way is sometimes called the "path" of the peak.

The following method of deriving an optimal solvent composition consists of two parts. In the first part the individual peaks are identified, and an equation describing the relation between mobile phase composition and retention time, the retention surface, is fitted to each set

of peak times. The second part consists in finding an optimal solvent composition, given the previously found retention surfaces and band widths.

The procedures described in the following rely on the general assumption that each substance of the sample generates (apart from random disturbances) the same relative peak area p_i when chromatograms of the same test substance are produced with different proportions c_{js} of the solvent components.

Initial Identification of Individual Peaks and Estimation of Retention Surfaces

Initial identification of individual peaks is performed in a stepwise iterative manner, which is in part inspired by the backfitting algorithm introduced by Friedman and Stuetzle [28].

Step 0: Initialization
Start with I, the unknown number of individual peaks set to 1.

Step 1: Identification of retention surface I and initial estimate of y_I.
The empirical peak areas a_{jn} are sorted in increasing order, regardless of the indices j and n. This ordered set of areas $A = \{a_{(i)}\}$, which has $N = \sum N_j$ members, is used to derive the empirical distribution function F of the peak areas, defined by

$$F(a) = (\text{no. of peak areas} < a) / N.$$

In an ideal situation with all individual peaks separated, the function F would be a step function with jumps of size J / N at the points p_i, i =1,...,I. This cannot be expected to happen in practice, where individual peaks do cluster and peak times are disturbed by noise. An example from a practical situation is illustrated by Fig. X. However, an individual peak will not be clustered with another peak in each chromatogram so that it can be assumed that at least some of the empirically found peaks are in fact individual ones. These should differ only slightly in size from one another, due to random noise, which means that a part of the F(a) curve should be

nearly vertical. This part can be found numerically by looking for that peak area a_{kernel}, which has the smallest mean difference to its 2b nearest neighbours in the set A:

$$a_{kernel} = \min_{b<n\leq N-b} \frac{1}{2b+1} \sum_{i=n-b}^{n-b} a_{(i)}$$

A first recommendation for the value of the span b is b = int (J / 2). Some special treatment not covered by the above formula is necessary for the mean difference assigned to $a_{(n)}$ if $n \leq b$ or $n > N - b$.

The point a_{kernel} is considered as the kernel of a cluster of peak areas, which should contain at least some members that belong to the path of one individual peak. Possible members of such a path are those areas which are not too different from a_{kernel}, collected in the set A kernel:

$$A_{kernel} = \{a_{jn} : | a_{kernel} - a_{jn} | < \delta_0 \; a_{kernel}\}.$$

The set M contains the retention times corresponding to the areas in A kernel:

$$M = \{ m_{jn} : a_{jn} \varepsilon A_{kernel} \}.$$

Retention surfaces are fitted through all the paths z that can be formed from members of M using standard least squares estimation. The mathematical model for retention surface yl as derived from the "PRISMA" model (cf. equation (RS)) is

$$y(c_j, \beta_l) = \beta_{l0} + \beta_{l1} c_{j1} + \beta_{l2} c_{j2} \beta_{l3} c_{j3} + \beta_{l4} c_{j2} c_{j3} + \beta_{l5} c_{j1}^2 + \beta_{l6} c_{j2}^2 .$$

In the above example, the mobile phase compositions, their squares, and an interaction term are used to model the retention surface. This special form, derived from the "PRISMA" model, reflects current knowledge about the relation between retention time and

chromatographic parameters. If further experience shows that other expressions for y are more appropriate, there will be no general difficulties in updateing the procedure accordingly. Of course, an iterative procedure may become necessary if y is modelled by an equation which is nonlinear in its parameters.

Denoting the variables actually used to model y as x_k, k = 1, 2,..., K (cf.(TF)), a linear model like the one above can always be written as

$$\mathbf{y}(\mathbf{X}; \beta) = \mathbf{X}\beta$$

where **y** is a column vector of predicted retention times (length n), **X** is a matrix containing the explaining variables, and β is a column vector of unknown parameters. Observed retention times z and predicted times y of a single individual peak are related via

$$\mathbf{z} = \mathbf{y} + \varepsilon$$

where ε is an error vector with $E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2 1$. The least squares estimate β for β given a vector of observations **z** is obtained from

$$\beta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{z}$$

and the covariance matrix of the estimates is given by

$$\mathrm{Cov}(\beta) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}, \text{ where}$$
$$\sigma^2 = (\mathbf{z} - \mathbf{X}\beta)^T(\mathbf{z} - \mathbf{X}\beta)/(n-K).$$

For details concerning the linear regression approach, see Draper & Smith [29]. Each path and its regression equation is assigned a score, e.g. the residual sum of squares (RSS) or the coefficient of determination r^2. The curvature of y, which is expressed by the quadratic summands in (RS), can also be made part of the score assigned to a path. The members of the path with the "best" score (minimal RSS, maximal r^2, minimal curvature) are considered as being individual peaks generated by the same test substance component. The corresponding regression equation with parameters β_l defines a first estimate for the retention

surface y_l. A first estimate for p_i, the area of individual peak I, is the mean of the a_{jn}'s belonging to the path finally chosen.

Step 2: Completion of the path for peak I and refinement of parameter estimates

The path for peak I obtained so far may be incomplete, because individual peak I might actually have appeared individually only in some of the J chromatograms, and only these appearences have been detected so far.

The estimated retention surface, however, predicts retention times for all of the chromatograms, so that empirical peaks to which peak I contributes and which are not yet identified can be identified.

To this end, a tolerance region is defined for the predicted retention times. The size of the tolerance region is determined using variances of the regression estimate by which the parameter vector β_l was derived. The peaks with retention times within the tolerance region are compared with respect to their areas with the current estimate for p_l. If there are peaks of comparable area, the one next to the predicted time is choosen as a member of the path of individual peak I. If no peaks show comparable size, the peak next to the predicted retention time having sufficient size is chosen as that one, to which peak I contributes. This selection is performed for all chromatograms in which no assignment has yet been made.

Having selected the empirical peaks through which individual peak I passes, a refined estimate for p_l is possible. p_l is now estimated as the mean of all peak areas on the identified path, which are not suspected to be composed of several peaks. A refined estimate for β_l is obtained similarly by fitting the equation for y_l to all J assigned peak times.

Step 3: Check whether all individual peaks have been identified

If the sum of all current estimates p_i , i= 1, 2,...,l, is sufficiently close to 1, the procedure is continued with step 4. Otherwise the value p_l is subtracted from all those empirical peak areas, which have been identified as members of the path of peak I. The number I of individual

peaks is increased by 1 and the procedure is repeated, starting with step 1 and using the set of peak areas just modified instead of the previous one.

Step 4: Joint estimation of all p_i's

Now the number of individual peaks I, the retention surface equations y_i, and the peak areas p_i are determined for all individual peaks. The estimates for the p_i's, however, can still be refined by estimating all p_i's simultaneously using the path information. This can be performed by least squares estimation of **p** in the mathematical model

$$\mathbf{a} = \mathbf{X}\,\mathbf{p},$$

where **a** is the (Nx1) (column) vector of all original empirical peak areas, **p** is a ((I - 1)x1) column vector containing the first I - 1 individual areas to be estimated (p_I is obtained from $p_I = 1 - \sum p_U$), and **X** is a (Nx(I-1)) matrix with entries

$$x_{ni} = \begin{cases} 1 \text{ if peak i contributes to } a_n \\ 0 \text{ otherwise} \end{cases}$$

All row members of a and X are decreased by 1, if peak I contributes to this row, in order to account for the linear relation among the p_i's.

Step 5: Checking the identification

At the end of the procedures, the set of empirical peak areas a_{jn} should have been reduced to a set of zeros, because each time a peak has been identified, its area is removed from all the empirical peaks belonging to his paths. If at the end of the procedure there remain empirical peaks of substantial area, or if there are considerable negative values instead of zeros, an error has occurred in the identification procedure. A second diagnostic tool is available in the sum of the squares of the retention surfaces. The sum of squares belonging to each retention surface should explain a considerable part of the variation in

retention time displayed by the corresponding peak. A warning should be given to the experimenter in either case in order to give him/her the opportunity to make corrections.

The peak identification and area estimation procedure performed so far gives results $\beta_{backfit}$ and $p_{backfit}$ which can normally be expected to be exact enough as a basis for the optimisation that is to follow. It must be noted, however, that retention surfaces and peak areas have not been estimated jointly. This, on one hand, may have resulted in biassed estimates for parameters and standard errors which are both required for the optimisation. On the other hand, a general check of the peak identification is desirable for safety reasons.

These aspects have led us to develop an additional method for the joint estimation of retention surfaces and peak areas, which produces more precise parameter and standard error estimates, and at the same time serves as a diagnostic tool for checking the previous peak assignments.

The method works iteratively. Its success depends largely on the availability of "good" starting values for the parameters. The backfitting-like algorithm already presented serves to derive sufficiently good starting values.

Check of Identification and Final Parameter Estimation: Surface Fitting

The central idea behind this algorithm is to transform the identification problem into an optimization problem, which can be tackled by standard optimization methods. To this end we transform the empirical data m_{jn} and a_{jn} from each chromatogram j into an "idealized" synthetic signal

$$h_j'(t) = \sum_n a_{jn}\, 1_{(m_{jn} > t)} \qquad \text{with } 1_{(x)} = \begin{cases} 1 \text{ if condition x holds} \\ 0 \text{ otherwise} \end{cases}$$

$$t > 0$$

This idealized signal is a function of time t, which contains all of the information provided by the empirical peak. The retention time m_{jn} of a peak is represented by a jump in h', and the peak area a_{jn} supplies the

height of the jump in h′ at time m_{jn}, where the information from one peak is reduced to the time of its occurrence, m_{jn}, which is a point of discontinuity of h′ at m_{jn}. This specific form of collecting all the information of the data makes it possible to perform simultaneous parameter estimation and identification checking.

As the discontinuities of h′are inconvenient to handle in an optimization procedure, we approximate the indicator $1_{(m > t)}$ by a continuous sigmoid function:

$$1_{(m > t)} \approx \frac{1}{1 + \exp[\gamma\ (m_{jn} - t)]},$$

where $\gamma > 0$ is a constant which controls the accuracy of the approximation: the larger the value of γ, the better the approximation; though from a pragmatic viewpoint the value of γ should be kept reasonably low in order to avoid overflow problems. γ values of 10 have been found to provide sufficiently good approximations.

In this way the information from chromatogram j used in the procedure to estimate retention surfaces is compressed to the function

$$h_j(t_{jl}) = \sum_n a_{jn} \frac{1}{1 + \exp[\gamma\ (m_{jn} - t_{jl})]}, \qquad \begin{matrix} j = 1, 2, \ldots, J \\ l = 1, 2, \ldots, L_j \end{matrix}$$

We are free to choose the numbers L_j and positions t_{jl} of time points, at which h_j is computed. From the fact that each of the N_j peaks in chromatogram j contributes only 2 independent pieces of information, it is clear that reasonably $L_j = 2\,N_j$. The positions of t_{jl} are not subject to restrictions. They should be chosen in a way which provides a good description of the height and position of the jumps in h_j. One recommended selection is

$$\begin{aligned} t_{j,\,2(n-1)+1} &= m_{jn} - \gamma / 2 \\ t_{j,\,2n} &= m_{jn} + \gamma / 2 \end{aligned} \qquad n = 1, 2, \ldots, N_j .$$

The retention surface y_i, which predicts the mean time y_i of the occurrence of individual peaks for the substances to be separated is modelled as before as a function of the mobile phase proportions c_{js}, or more generally as a function of explaining variables denoted by x_k, k = 1, 2,...,K. The specific form of y is chosen according to the same considerations as led to equation (RS). No general problems will arise, if, due to new data, a form of y other than the one given by (RS) will be favored in the future, if y is kept differentiable with respect to the parameters β in order to facilitate the estimation of β. The number of parameters in y must as before be kept in a reasonable relation to the number of data points bearing independent information.

The predicted counterpart of the transformed empirical signal h_j is obtained from the set of retention surfaces y_i, i = 1. 2,..., I, by

$$z_j(\mathbf{x}, t, \beta, \mathbf{p}) = \sum_i \frac{pi}{1 + \exp[\gamma\ (y_i(\beta, x) - t)]}$$

For practical purposes of estimating and testing only values of h_j and z_j at selected time points $t_{j1}, t_{j2}, \ldots, t_{jL_j}$ will be considered.

As estimates for the unknown β and **p** we choose those values β and **p**, which minimize the square of the distance between the transformed empirical signals, $h_j(t)$, and the corresponding predictions:

$$S(\beta, \mathbf{p}) = \sum_j \sum_l \left(h_j(t_{jl}) - z_j(x_j, t_{jl}, \beta, \mathbf{p})\right)^2$$

$$\text{(FRS)} \qquad (\beta, \mathbf{p}) = \min_{\substack{\beta \in R^K \\ p \in [0,1]^I,\ \sum p_i = 1}} S(\beta, \mathbf{p})$$

The minimization of S (β, **p**) can be performed by standard iterative numerical optimization methods. A direct solution of (FRS) is not available.

One standard method for obtaining $\theta = (\beta, p)$ is by means of quasi-likelihood estimation [30]. We ignore for the moment that there might be a slight dependence among the hj values caused by the fact that $h_j(t)$ is partially determined by earlier values $h_j(s)$, $s < t$.

The iterative quasi-likelihood estimation procedure is summarized as follows:

1) Start with the set of parameters obtained from the preceding backfitting-like algorithm: $\theta_0 = (\beta_{backfit}, p_{backfit})$. Note that only I - 1 components of p must be incorporated in θ_0, because of $p_I = 1 - \sum p_i$. Let h and $z(\theta_0)$ denote the column vectors of length $L = \sum 2 N_j$, containing the empirical and theoretical values, respectively, of the transformed signal computed for all chromatograms j and all time points t_{jl}, given I and θ_0.

2) Compute a new estimate

$$\theta_{n+1} = \theta_n + (D^T D)^{-1} D^T (h - z(\theta_n)).$$

where $D = (\delta z / \delta \theta_k)$, $k = 1, 2, \ldots, K$ I - 1, computed for $\theta = \theta_n$.

3) If

$$(\theta_{n+1} - \theta_n)^T (\theta_{n+1} - \theta_n) < \varepsilon$$

for an appropriately chosen bound ε the iteration procedure finishes with as estimate

$$\theta = \theta_{n+1}$$

for the unknown true value of θ. If the bound ε is not yet reached, replace n by n + 1 and perform a further iteration beginning with step 2), unless n has become too large. The latter must be interpreted as non-convergence of the procedure.

Non-convergence of the procedure is usually a consequence of bad initial values, where we assume that the procedure has been implemented with the usual safety measures like step halving. Non-convergence due to bed initial values in this context is a hint that the backfitting-like algorithm which supplied the initial values had ended with

an erroneous peak identification, or the parametric form of the y_i's is inappropriate, which in turn may be the reason for bad identification. The same conclusion holds, if the extent of variation explained by the mathematical model is extremely small. If the results of the estimation procedure fail to pass these checks, a further human investigation of the data will become necessary.

Asymptotic variance and covariances for the parameters are, in the case of convergence, obtained as the diagonal and off-diagonal elements, respectively, of

$$(\mathrm{VAR}\theta) \qquad \mathrm{Cov}(\theta) = \sigma^2 (\mathbf{X}^T\mathbf{X})^{-1}$$

where $\mathbf{X} = (x_{jk})$, $j = 1,2,...,J$, $k = 1, 2,...., K+1$ contains the variables used to define the individual retention surfaces y_i, and σ is estimated by

$$(\mathrm{VAR}\sigma) \qquad \sigma^2 = (h - z(\theta))^T (h - z(\theta)) / g$$

with

$$g = L - I(K + 1) + 1$$

From (VARθ) and the fact that each component θ_w of θ has an approximate t_g- distribution, confidence limits for level $1 - \alpha$ are derived as

$$\theta_w - \sigma_w\, t_{g,1-\alpha} \le \theta_w \le \theta_w + \sigma_w\, t_{g,\,1-\alpha}$$

σ_w is the square root of the entry (w, w) of Cov(θ) given in (COVθ). Using the standard errors from (VAR θ) and the specific form of y_i, an estimate $v_i(\mathbf{x})$ for the standard error of a predicted retention time yi can be computed for each value of $\mathbf{X}$.

Selection of Optimal Mobile Phase Combinations

In our frame of reference the optimisation of the mobile phase means that a composition of the mobile phase giving an optimal chromatogram has to be found. Several definitions have been proposed

for the term "optimal" [e.g 3,4]. We choose the chromatographic response function (CRF) given by Berridge [31] and Wright et al. [32].
The general CRF in our notation has the form

$$\text{(CRF)} \qquad CRF = \sum_{i=1}^{l-1} R_{(i)} + l\,\delta_1 - \delta_2\,|\,t_{max} - y_{(l)}|\;\delta_3\,(t_{min} - y_{(1)}).$$

$R_{(i)}$ is the resolution between adjacent pairs of peaks, bounded upwards by 2,

$$\text{(RES)} \qquad R_{(i)} = 2 \min \left\{ (y_{(i+1)} - y_{(i)}) \,/\, (w_{(i+1)} + w_{(i)}),\ 1 \right\}$$

with w_j denoting the peak width at baseline. As before indices in parentheses such as in $y_{(i)}$ denote order, i.e. $y_{(i)}$ is the i-th member in the increasingly ordered sequence of y's. t_{max} is the maximum acceptable time for the last peak, and t_{min} is the desired minimum time for the first peak. The δ's in (CRF) are weights that must be adjusted to user's specific needs. We start here with $\delta_1 = 0$, because the total number of individual peaks cannot vary during optimisation, and $\delta_2 = \delta_3 = 1$, which means that overstepping upper and lower limits for retention times is considered of equal importance.

Having accepted the CRF as a measure of optimality, the search for an optimal chromatogram is equivalent to finding the composition $c^{(m)} = (c_1^{(m)}, c_2^{(m)}, c_3^{(m)}, c_4^{(m)})$ for which the CRF attains its maximum.

The CRF as a function of the mobile phase composition **c**, CRF = CRF(**c**), requires predictions of $y_i(\mathbf{c})$ for arbitrary values of **c** in their domain. These are computed using (RS), (TF) and the β_i-vectors estimated in step I.5 or I.4. Additionally, CRF(**c**) involves the peak width at baseline w_i for each mobile phase composition, w_i is computed as the difference (predicted peak end time - predicted peak start time). Start and end time for each individual peak i are predicted by a peak start surface b_i (**x**) and a peak end surface e_i (**x**) in a way similar to the prediction of the retention time y_i (**x**). The relation between **c** and **x** is again given by (TF).

A least squares approach is used to estimate the parameters $\beta_{ik}^{(b)}$, k = 0, 1, ..., 6; i = 1,2, ..., I in

$$\ln b_i(\mathbf{x}) = \beta_{i0}^{(b)} + \beta_{i1}^{(b)} x_1 + \beta_{i2}^{(b)} x_2 + \beta_{i3}^{(b)} x_3 + \beta_{i4}^{(b)} x_2 x_3 + \beta_{i5}^{(b)} x_2^2 + \beta_{i6}^{(b)} x_3^2.$$

The β_i surface is fitted through the empirical peak start times b_{jn} of those empirical peaks, to which only individual peak i contributes. If the number of chromatograms in which peak i is isolated is not large enough to perform the estimation, only the shift parameter $\beta_{i0}^{(b)}$ is estimated from the empirical peak starts, while the remaining $\beta_{ik}^{(b)}$ for k > 0 are taken from the equation for $y_i(\mathbf{x})$.

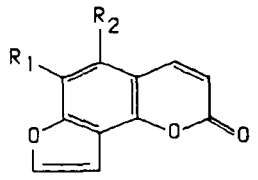
A corresponding procedure is applied for the estimation of the peak and surfaces $e_i(\mathbf{x})$ for i = 1, 2,...,I.
The band width $w_i(\mathbf{x})$ required in (RES) is obtained from

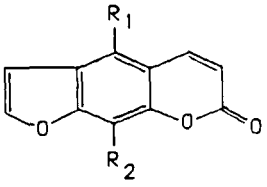
(BW) $w_i(\mathbf{x}) = e_i(\mathbf{x}) - b_i(\mathbf{x})$.

This way of calculating w_i takes into account that the band width may vary depending on the conditions under which the chromatogram is generated, and that in reality peaks frequently do not show a symmetric form.

The CRF value of a chromatogram produced with mobile phase composition **c** can now be predicted for every possible **c**. The complete relation between **c** and CRF value has been established by the equations (TF), (RS), (BW), and (RES).

The search for optimum $\mathbf{c}^{(m)}$ = max CRF(**c**) is performed as a grid search. This means that the corresponding values CRF(**c**) are computed for a grid of **c** values, which covers the whole region of permissible **c** values. The **c** with maximal CRF is taken to be maximum $\mathbf{c}^{(m)}$. The density of the grid should be fixed according to the precision with which components of the mobile phase can be adjusted in the chromatographic apparatus. This rather simple optimisation method is justified by three arguments. The first is that it is sufficient to compute only as many digits of the members in $\mathbf{c}^{(m)}$, as can be used in practice for adjustment of the





R_1	R_2			R_1	R_2		
H	H	Angelicin	(2)	H	H	Psoralen	(3)
H	OCH_3	Iso-bergapten	(1)	H	OCH_3	Xanthotoxin	(7)
OCH_3	H	Sphondin	(6)	OCH_3	H	Bergapten	(4)
OCH_3	OCH_3	Pimpinellin	(5)	OCH_3	OCH_3	Iso-pimpinellin	(8)

Figure 4 Structures of the furocoumarin isomers investigated.

chromatograph. The second argument is that random disturbances in the retention data lead to standard errors greater than zero for all estimated parameters, and consequently also for the predicted $c^{(m)}$. This would make any attempt to compute $c^{(m)}$ with very high precision a somewhat unreasonable undertaking. The final argument for using a grid search is that the computer time required remains within an acceptable limits.

EXPERIMENTAL

Furocoumarin isomers from *Heracleum sphondylium* (for structures see Figure 4) were isolated and identified in the Department of Pharmacy, Swiss Federal Institute of Technology (ETH) Zurich, Switzerland.

The other three furocoumarins were obtained from Roth (Karlsruhe, FRG). Separation of all compounds was performed on a 250 x 4 mm column, filled with 5 μm, 80 Å Eurospher 80/5 C_{18} stationary phase (Eurochrom-Knauer, Berlin, FRG).

All organic solvents [methanol (MeOH), ethanol (EtOH), isopropanol (IPR), tetrahydrofuran (THF), acetonitrile (ACN), and dioxane (DIO)] were of HPLC quality (Romil Chemical, Shepshed, Leics, England). Water was

freshly distilled and filtered through a 0.45 μm membrane filter (Millipore, Bedford, MA, USA). For all experiments, a flow rate of 1ml/min was used. Between the single chromatographic experiments, the column was always equilibrated for 20 min. The temperature of the column was kept at 25° C for all analyses.

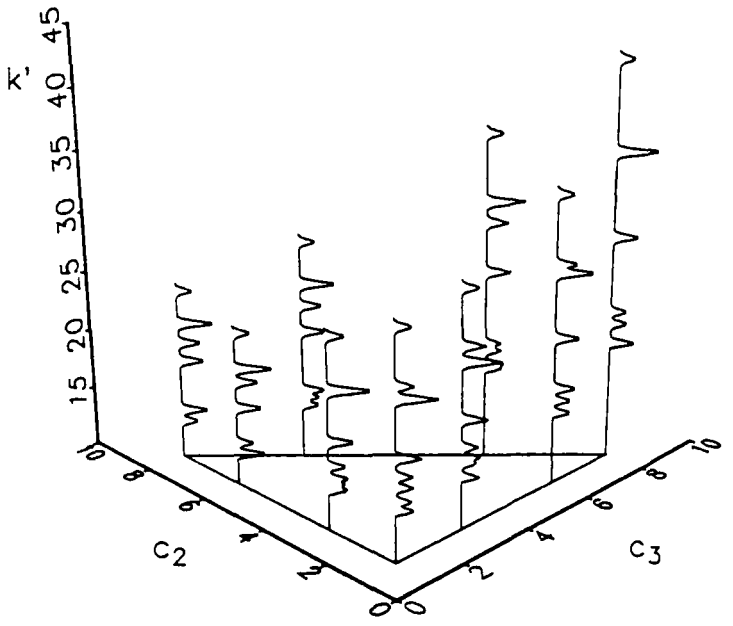
A LC 41 four solvent HPLC instrument with autosampler were used in these experiments. Data acquisition and data processing was performed with the Chromatography Data System ChromStar® (all equipments from Bruker-Franzen, Bremen, FRG). An Epson PC-AX was used as the computer system and a FX-80 printer. A Gilson model 116 UV detector (Gilson Medical Electronics, Middleton, WI, USA) was operated at a wavelength of 313 nm. For the evaluation of the optimum of the mobile phase composition an IBM AT 02 computer was used.

RESULTS AND DISCUSSION

Since C_{18} is the most widely used stationary phase in HPLC analysis, our optimization process always starts with this stationary phase. The strategy for the selection of the suitable solvent is based on the solvent classification by Snyder [15]. After the first experiments with the six neat solvents (MeOH, EtOH, IPR, THF, ACN, and DIO) the solvent strengths were reduced systematically so that the k′ of the at last eluted compound was always between 30 and 35 min. With the help of the calculated CRF values THF, ACN, and MeOH were selected for the further optimization process at a solvent strength levels of 1.1, 1.05, and 1.0.

The optimization steps for the determination of the local optimum at $S_T = 1.05$ are summarized in Fig. 5, 6, and 7b. Nine chromatograms from the twelve measurements (see Fig. 3) are given in Fig. 5.

The capacity factors and the relative peak areas served for initial identification, as shown in Fig. 6a. The finally identified compounds are connected by straight lines, as shown in Fig. 6b. Based on the quadratic correlation of the mobile phase composition and the measured capacity factors, the quadratic retention surfaces could be fitted for all



Fligure 5 HPLC chromatograms of furocoumarin isomers in selected selectivity points at constant solvent strength (S_T=1.05) for the determination of local optimum. (c_1=THF, c_2=ACN, c_3= MeOH, the proportion of c_3 is determined by c_3=10-c_1-c_2)

furocoumarin isomers for the determination of local optimum. The best separation, determined from the twelve measurements, could obtained on solvent strength level 1.05 in P_S = 07-18-75.

In Fig. 7 the CRF values (t_{min}= 3 min, t_{max}= 40 min) are presented at the three tested solvent strength levels (1.1; 1.05; 1.0) with six measurements at each level (at S_T=1.05 with 12 measurements) and the predicted global optimum calculated from 18 measurements. For S_T =1.1 the local optimum was found in P_S= 04-01-95 (Fig. 7a), for S_T =1.05 in P_S= 07-18-75 (Fig. 7b), and for S_T =1.0 in P_S=01-50-49 (Fig. 7c). With the help of the 18 measurements the global optimum (Fig. 7d) was determined in P_S = 97-01-02 at S_T = 0.98. For the CRF value with the

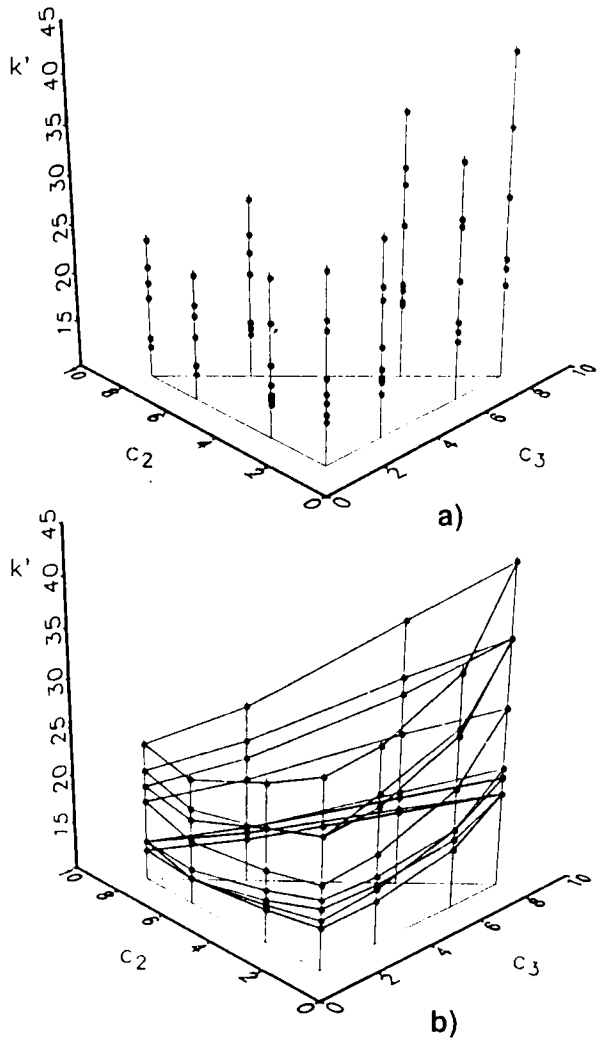


Figure 6 Determination of local optimum of furocoumarin isomers.

a) Base for initial identification; schematic drawing of capacity factors obtained by the chromatographic separation in selected selectivity points at S_T= 1.05.

b) Final identification; capacity factors of the individual peaks are connected by straight lines, as the base for the quadratic retention surfaces (RS).

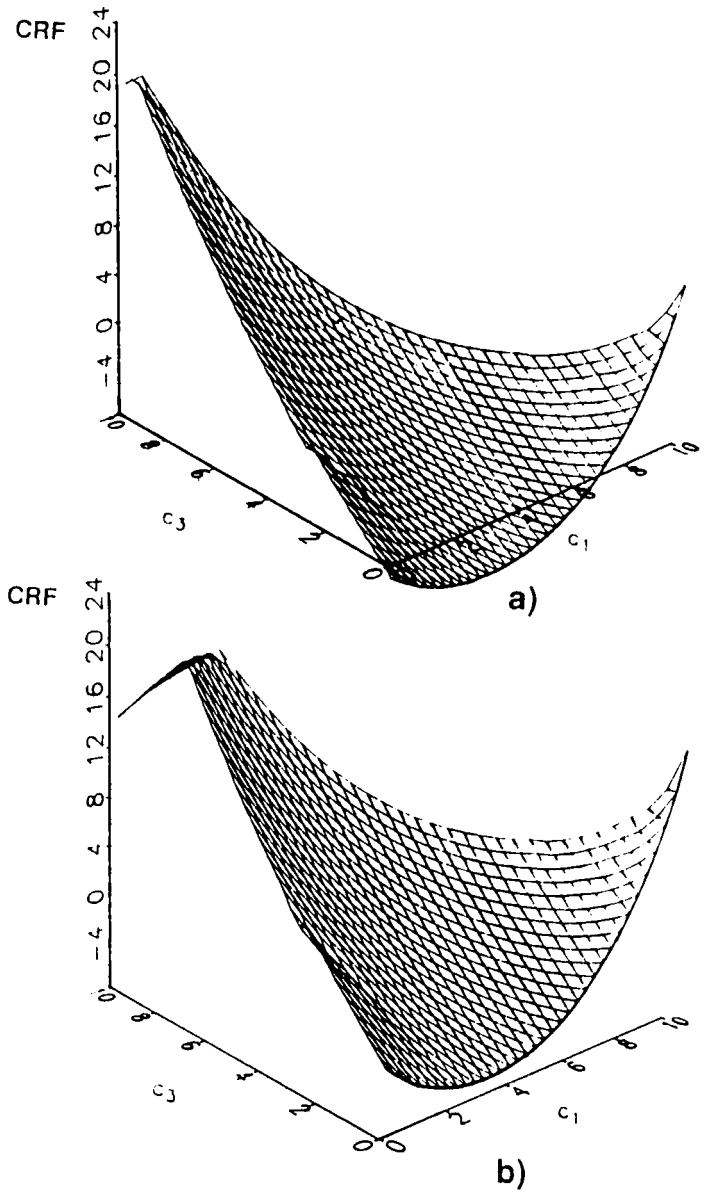


Figure 7 CRF values for the furocoumarin isomers for the determination of local and global optimums.

a) predicted local optimum (P_S=04-01-95) at S_T= 1.1, determined from six measurements.

b) predicted local optimum (P_S=07-18-75) at S_T= 1.05, determined from twelve measurements.

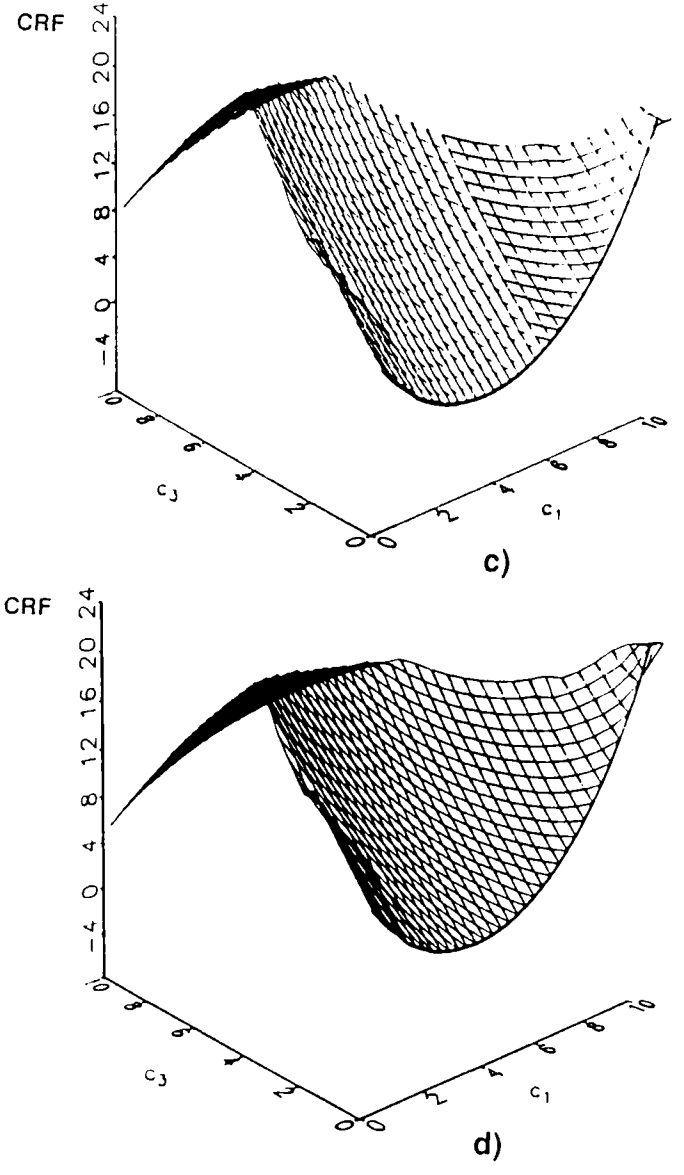


Figure 7 CRF values for the furocoumarin isomers for the determination of local and global optimums.

c) predicted local optimum (P_S=01-50-49) at S_T= 1.0, determined from six measurements.

d) predicted global optimum (S_T= 1.0; P_S=97-01-02) determined from eighteen measurements.

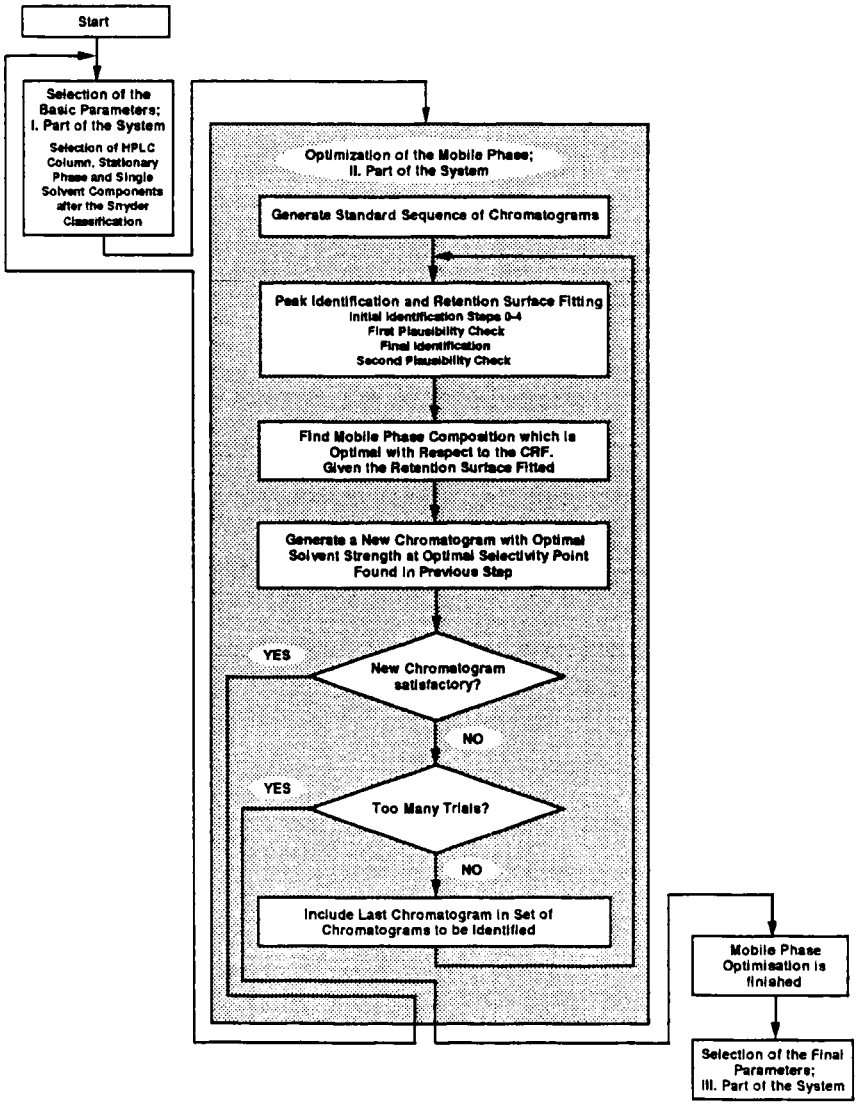


Figure 8 Flow chart of the computer-aided HPLC mobile phase optimization with the "PRISMA" Model.

predicted best mobile phase composition 22.9 was determined, from the measured chromatograms this value was found as 22.7

The predicted and the empirical CRF value for the global optimum both are better than the CRF values for the local optima (CRF=20.0; 20.4; and 20.8 for S_T =1.1; 1.05; and 1.0, respectively), and better than each empirical CRF value. This, together with the good arreement of predicted and observed CRF for the predicted global optimum, means that the optimization procedure has in fact led to an improved separation.

As a conclusion it can be stated, that the quadratic correlation between solvent composition and capacity factor is valid for each S_T level, and that the S_T level acts in a linear way on the logarithm of the capacity factor, as has previously been found [27]. Optimization of the mobile phase can therefore be based on these correlations, using the CRF as a tool to assess optimality. As it is shown in Fig 7, the quadratic correlation are also for the CRF values valid on a certain S_T level. It must be noted, however, that the time limits incorporated in the CRF definition strongly influence the position of local and global optima, as can be seen from Figs. 7a-b.

With this separation example the application and potential of the algorithm is presented for the automatic optimisation of the mobile phase composition in HPLC. The algorithm provides a prediction of an optimal (local or/and global) mobile phase composition. The algorithm is independent of the stationary phase, of the choice of the components of the mobile phase, and of the operating parameters of the instrument.

The position of this algorithm within the complete sequence of actions that finally lead to an optimal chromatogram is shown in by the flow chart in Figure 8.

The algorithm is based on standard retention data, namely empirical capacity factors of peak maximum, peak start and end, the empirical peak areas, as well as the information on the mobile phase compositions involved. This information is usually provided as a standard output from a chromatograph, which means that no additional processing of the detector signal is necessary.

The method has been developed without assumptions concerning peak shapes. Restrictive properties like symmetry or normal distribution

of the peaks are not required. The only important assumption is that each component of the sample to be separated generates a relative peak area which changes from one chromatogram to another only due to random disturbances. This relative area may appear in the form of an isolated peak or as a part of a larger peak, which contains individual relative areas from several compounds of the sample to be separated.

The theoretical background for the mathematical specification of retention surface equations and their relation to the mobile phase composition is derived from the "PRISMA" model. No assumptions are made concerning the crossing behaviour of the retention surfaces. The chromatographic response function (CRF) is used as the optimisation criterion.

For the determination of global optimum the algorithm starts with the identification of individual peaks in a standard sequence of 18 chromatograms, where the allocation of selectivity points and solvent strength is given by Table 1.

The presented automatic peak identification approach, together with the horizontal and vertical correlations between the selectivity points of the "PRISMA" model and the capacity factors allows an automated mobile phase optimization procedure for four solvent HPLC systems.

ACKNOWLEDGEMENT

Financial support from the Swiss KFW Comission is gratefully acknowledged. We would like to thank Dr. K. Gottschall (Eurochrom-Knauer, Berlin, FRG) for his generosity in providing us with several HPLC columns.

REFERENCES

[1] J.L. Glajch, J.J. Kirkland, K.M. Squire and J.M. Minor, *J. Chromatogr.*, 199 (1980) 57.
[2] J.L. Glajch, J.J. Kirkland and J.M. Minor, *J. Liq. Chromatogr.*, 10 (1980) 1727.
[3] J.C. Berridge, *Chemistry in Britain*, 1063 (1987).
[4] P.J. Schoenmakers, N. Dunand, A. Cleland, G. Musch and Th. Blaffert, *Chromatographia*, 26 (1988) 37.

[5] A. Peeters, L. Buydens, D.L. Massart and P.J. Schoenmakers, *Chromatographia* 26 (1988) 101.
[6] T.P. Bridge, M.H. Williams and A.F. Fell, *J. Liquid Chromatogr.*, 12 (1989) 23.
[7] A.F. Fell, H.P. Scott, R. Gill and A.C. Moffatt, *J. Chromatogr.*, 282 (1983) 123.
[8] P.M.J. Coenegracht, A.K. Smilde and A.Knevelman, *J. Liquid Chromatogr.*, 1 (1989) 77.
[9] A. Bartha, H.A.H. Billiet and L. De Galan, *J. Liquid Chromatogr.*, 12 (1989) 173.
[10] E.P.Lankmayr, W. Wegscheider and K.W. Budna, *J. Liquid Chromatogr.*, 12 (1989) 35.
[11] S.R. Bakalyar, in *HPLC Instrumentation & Applications*, International Scientific Communications, Inc. Fairfield, Conn. 1986, pp.73-87.
[12] J.C. Berridge, *Techniques for the Automated Optimisation of HPLC Separation*, Wiley, Chichester, 1985.
[13] P.J. Schoenmakers, *Optimization of Chromatographic Selectivity*, Elsevier, Amsterdam, Oxford, New York, Tokyo, 1986.
[14] J.J. Kirkland, J.L. Glajch, *J. Chromatogr.*, 255 (1983) 27.
[15] L.R. Snyder, *J. Chromatogr. Sci.*, 16 (1978) 223.
[16] J.L. Glajch, J.J. Kirkland and L.R. Snyder, *J. Chromatogr.*, 238 (1982) 269.
[17] Sz. Nyiredy, K. Dallenbach-Toelke, O. Sticher, in *Proceedings of the Fourth International Symposium on Instrumental High Performance Thin-Layer Chromatography* (H. Traitler, H. Studer, R.E. Kaiser, eds) Institute for Chromatography, 1987, pp.289-300.
[18] Sz. Nyiredy, C.A.J. Erdelmeier, B. Meier and O. Sticher,*Planta med.*, (1985) 241.
[19] Sz. Nyiredy, B. Meier, C.A.J. Erdelmeier and O. Sticher, *HRC & CC*, 8 (1985) 186.
[20] Sz. Nyiredy, C.A.J. Erdelmeier and O. Sticher, *in Proc. Intern. Symp. on TLC with Special Emphasis on OPLC* (E.Tyihák, ed), Labor MIM, Budapest (1986) pp.222-231.
21] K. Dallenbach-Toelke, Sz. Nyiredy, B. Meier and O. Sticher, *J. Chromatogr.*, 365 (1986) 63.
[22] Sz. Nyiredy, *Application of the "PRISMA" Model for the Selection of Eluent-systems in Overpressure Layer Chromatography (OPLC)*, Labor MIM, Hungary, 1987.
[23] Sz. Nyiredy, C.A.J. Erdelmeier, B. Meier and O. Sticher, *GIT Suppl. Chromatogr.*, 4/85 (1986) 24.
[24] Sz. Nyiredy, K. Dallenbach-Toelke and O. Sticher, in *Recent Advances in Thin Layer Chromatography* (Ed. I.Wilson), Plenum, London, 1988, pp.45-54.
[25] Sz. Nyiredy, K. Dallenbach-Toelke and O. Sticher, *J. Planar Chromatogr.*, 1 (1988) 336.

[26] Sz. Nyiredy, K. Dallenbach-Toelke, G.C. Zogg and O. Sticher, *J. Chromatogr.*, submitted.
[27] Sz. Nyiredy, K. Dallenbach-Toelke and O.Sticher, *J. Liquid Chromatogr.*, 12 (1989) 95.
[28] J.M. Friedman, W. Stuetzle, *J. Am. Statistical Assoc.*, 76 (1981) 817.
[29] N. Draper, H. Smith, *Applied Regression Analysis*, 2nd Edition, J. Wiley & Sons, New York,1987.
[30] P. McCullagh, F.A. Nelder, *Generalized Linear Models*, Chapman and Hall, London, 1983.
[31] J.C. Berridge, *J. Chromatogr.*, 244 (1982) 1.
[32] A.G. Wright, A.F.Fell, J.C. Berridge, *J. Chromatogr.*, 464 (1989) 27.