*Journal of Chromatography*, 485 (1989) 383–401
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 22 020

# RULE-BASED APPROACH FOR THE DETERMINATION OF SOLUTE TYPES IN UNKNOWN SAMPLE MIXTURES AS A FIRST STEP OF OPTIMIZATION PARAMETER SELECTION IN REVERSED-PHASE ION-PAIR CHROMATOGRAPHY

AKOS BARTHA[a] and GYULA VIGH*
*Chemistry Department, Texas A&M University, College Station, TX 77843-3255 (U.S.A.)*

## SUMMARY

The combination and range of mobile phase variables for selectivity optimization in reversed-phase ion-pair chromatography can be selected rationally by considering the nature (charge type and relative hydrophobicity) of the sample components. An experimental procedure and a rule-based evaluation strategy are described that can be used to determine the charge type and relative hydrophobicity of the components in unknown sample mixtures. Solute-type determination is based on the characteristic retention-shift patterns of charged solutes observed in seven carefully selected methanol–water gradient runs. The gradients are run at three different pH values (2.5, 5 and 7.5) with and without (positively and negatively charged) pairing ions. This retention data set is evaluated by a novel rule-based strategy, which requires neither peak tracking nor other extra-chromatographic information (*e.g.*, spectral data, peak areas). The evaluation rules are based on the combinations of the ideal retention-shift patterns and the experimentally determined retention-shift limits of different solute types. The rule set has been used to develop a computer program, which was tested with a variety of complex samples.

## INTRODUCTION

Optimization of separation selectivity is an important step of chromatographic method development. The success of the optimization process depends greatly on the selection of the mobile phase variables and their ranges (optimization parameter space), irrespective of whether trial-and-error or computer-aided optimization methods are used[1]. Parameter selection is relatively simple in the reversed-phase (RP) chromatographic separation of non-ionic samples (type and concentration of the organic modifier). It becomes a non-trivial task in the RP ion-pair chromatography (IPC) of complex samples which contain non-ionic, ionizable and/or ionic solutes. Here, in addition to the organic modifier, one must also select the pH and ionic strength of the eluent and the type and concentration of the pairing ion used. Owing to the expanded parameter space, optimization becomes much more involved and time consuming, and often yields only local optima.

[a] On leave from the Veszprem University of Chemical Engineering, Veszprem, Hungary.

0021-9673/89/$03.50 

Various approaches such as previous chromatographic experience[2], factorial designs[3–5], statistical mixture designs[6–8] and, recently, expert systems[9–11] have been used to select the type and range of these eluent parameters. However, their usefulness with complex ionic samples, which often contain unknown components, is fairly limited. The efficient use of factorial designs for parameter selection requires the tracking of peaks in a usually large number of sequential chromatograms. Statistical mixture designs often search only a certain part of the optimization parameter space[12]. Expert systems need extensive *a priori* chemical information about the sample components in order to make predictions about the expected retention behavior. The lack of a satisfactory strategy to select the optimization parameter space may be one of the reasons why practising chromatographers find the existing, otherwise powerful, optimization software packages of limited utility.

Recently, Low *et al.*[12] suggested a rational approach for the selection of the primary mobile phase optimization parameters (charge-type of the ion-pairing reagent, pH and/or methanol concentration of the eluent). Based on a study of computer-simulated sample mixtures, preferred combinations of these variables and significantly reduced optimization search areas were found for samples which contain certain solute types. Further studies by Bartha and co-workers extended this approach to the rational selection of the hydrophobicity (chain length) and mobile phase concentration of the pairing ion[13] and the type and concentration of the organic modifiers[14] in RP-IPC systems.

These studies concluded that the mobile phase variables for selectivity optimization of complex sample mixtures can be selected rationally when the nature (charge type and relative retention) of the solutes is known. Bartha *et al.*[15] and Strasters *et al.*[16] succesfully applied this strategy for the practical optimization of sample mixtures containing ionizable and ionic solutes.

However, for the success of this approach, it is imperative to establish the charge type of at least the most retained and the least retained components in the pH 2.5 and 7.5 eluents or, preferably, the charge type and relative hydrophobicity of as many sample components as possible. In fortitious cases this is available as *a priori* information, but in unknown samples it has to be determined prior to parameter selection and selectivity optimization.

Low *et al.*[12] have developed an experimental strategy to derive solute-type information from the retention-time shifts of the sample components that were observed in four successive 0–90% (v/v) methanol–buffer gradients run at pH 2.5 and 7.5. In one run octanesulfonate (at pH 2.5) and in another run tetrabutylammonium (at pH 7.5) pairing ion plugs[17] were injected prior to sample injection. As most complex unknown samples contain components with widely different water solubilities (*i.e.*, different reversed-phase retentions), solvent gradients had to be used in order not to miss the very slightly retained and the very strongly retained components. The gradients with various pH–pairing ion combinations could be easily realized and the chromatographic system rapidly re-equilibrated by the use of the pairing-ion pulse-injection technique[12,17].

Combined with complete tracking of all the peaks to recognize the solute retention shifts in the sequential chromatograms, solute-type assignements could be made by comparing the observed shifts with the ideal retention-shift patterns of charged solutes[12,15,16].

Although the utility of the four-gradient approach in solute-type determinations was successfully demonstrated with a number of samples[12,15], it does have some limitations in the case of complex unknown mixtures. Part of the limitations can be attributed to the use of a single-component pairing-ion plug. As the elution rate of the pairing ion continuously increases during the gradient, the pairing ion cannot sufficiently increase or decrease the retention time of the late-eluting ionic solutes[12]. The other limitation is that peaks must be tracked in all gradient chromatograms in order to recognize their retention shifts. Unfortunately, UV spectra-based peak-tracking methods often cannot be used for mixtures of ionizable compounds owing to the dramatic variation in the spectral properties of the components in eluents of different pH[12]. The injection of standards, as an alternative for peak identification, is time consuming (especially in the gradient mode), and it cannot be applied for unknown mixtures, the very samples for which solute-type determination would be important.

As solute-type determination is only the first step in the process of selectivity optimization, and not a goal in itself, a simple, rapid, easily available and universally applicable method, well within the reach of the average chromatographer, is desired.

In this paper we show that the nature of the solutes in totally unknown aqueous sample mixtures can be determined by combining an extended and improved experimental procedure with a rule-based retention-shift evaluation strategy that is implemented in a computer program. The new approach does not require either peak tracking or the use of any other extra-chromatographic information. The fundamentals and experimental evaluation of this method are discussed using several complex separation examples.

## EXPERIMENTAL

### *Instrumental*

An LC 5500 liquid chromatograph equipped with UV (set at 254 nm, 0.2 a.u.f.s.) and refractive index detectors, a Varian 8085 autosampler and a Model 4270 two-channel integrator (all from Varian Aerograph, Walnut Creek, CA, U.S.A.) were used. A Model 7126 six-port injection valve with a 20-$\mu$l injection loop (Rheodyne, Cotati, CA, U.S.A.) was operated by an electrically controlled pneumatic valve. The chromatographic system was built to allow for delayed and repeated injections independent of the collection of retention data, by programming the injector and the integrator through the external event function of the chromatograph.

A Nova-Pak $C_{18}$ (5 $\mu$m) reversed-phase column (150 × 4.6 mm I.D.) (Waters Assoc., Milford, MA, U.S.A.) was used in the experiments with a flow-rate of 1 ml/min at room temperature (25°C).

### *Chemicals*

High-performance liquid chromatographic (HPLC)-grade methanol was purchased from Mallinckrodt (Paris, KY, U.S.A.). Distilled, deionized water was prepared with a Milli-Q water purification system (Millipore, Milford, MA, U.S.A.). Gold Label quality triethylamine (TEA) and phosphoric acid (85%, w/w) were used as buffer components. Tetraalkylammonium bromides and sodium alkylsulfonates were used as ion-pairing reagents (all from Aldrich, Milwaukee, WI, U.S.A.).

*Mobile phases and gradient sequence*

Triethylamine phosphate buffers (10 m*M*) of pH 2.5 and 7.5 were prepared by directly titrating the organic base with phosphoric acid (10%, w/w). Pairing ion solutions for the "pulsed" injection experiments were prepared in aqueous buffer, and contained mixtures of sodium octyl-, decyl- and dodecylsulfonate or tetrapropyl-, tetrabutyl- and tetrapentylammonium bromide (400, 200 and 100 m*M*), respectively. A 20-$\mu$l aliquot of these cationic or anionic mixtures was injected 5 min prior to the injection of the sample. Samples were injected with a delay, accounting for the pump-to-column dwell volume, at the start of the gradient. The gradient run consisted of four sequences: (i) a linear gradient from 0 to 90% (v/v) methanol concentration at a given pH (2.5, 5 or 7.5) in 15 min, (ii) isocratic elution at a 90% methanol concentration for 5 min, (iii) a reverse linear gradient from 90% methanol to aqueous buffer in 5 min and (iv) re-equilibration of the column with the aqueous buffer for 2 min.

*Computer programming*

The research-prototype rule-based computer program was written using a Turbo Prolog compiler (Borland International, Scotts Valley, CA, U.S.A.). A Powermate II AT compatible personal computer (NEC, Computer Access, College Station, TX, U.S.A.), equipped with an NEC Multisync II color monitor, VGA graphics card, 40 MB harddisk and I/O serial interface card was used for program development and simulation experiments.

## RESULTS AND DISCUSSION

*Retention-shift patterns and experimental retention-shift limits*

Solute-type determination by HPLC is based on the well known retention behavior of charged solutes in RP-IPC. If the retention of an ionic or ionizable solute is measured under different RP-IPC conditions, its charge type can be determined from the observed retention changes. The direction of these retention shifts can be predicted for ideally behaving solutes, leading to a characteristic retention-shift pattern for each solute type[12]. Previously, solute retention shifts measured at four different pH and pairing ion combinations were used for solute-type determination, by simply matching the observed behavior with the ideal retention-shift patterns[12,15,16].

However, this four-gradient design[12] can be used only when all retention shifts can be recognized by tracking the motion of all peaks in all chromatograms. In complex, unknown sample mixtures (containing many components of different charge types), the interchange and/or coelution of the shifting peaks, spectral similarities and changes of the spectral features with pH and solvent composition often make (UV spectra-based) peak tracking difficult and uncertain. As a result, solute-type determination also becomes difficult, if not impossible.

The changes of discrimination between the different solute types can be improved by creating retention shifts in all possible directions. This can be achieved in additional gradient chromatograms with other pH and pairing-ion combinations: at pH 2.5 and 7.5 with and without the addition of a positively and a negatively charged pairing ion. In order to discriminate between non-charged solutes and a mixture of a weak acid and a weak base (all having the same retention time in non-charged form), the inclusion of a medium pH gradient is also needed.

Based on these requirements, we designed an extended set of seven gradients which leads to a full retention-shift pattern (see Fig. 1a–e) for strong acids (SA), strong bases (SB), weak acids (WA), weak bases (WB) and non-ionic (N) solutes. The retention of each solute type in pH 2.5, 5 and 7.5 eluents in the absence of pairing ions is plotted on the three middle bars in Fig. 1a–e. The two bars on the left show how retention changes in pH 2.5 eluents as either a positively or a negatively charged pairing ion is added (in the form of an injected plug) to the eluent. The two bars on the right show how retention changes in pH 7.5 eluents as either a positively charged (symbol +) or a negatively charged (symbol −) pairing ion is added to the eluent. The pH limits (2.5 and 7.5) encompass the p$K$ values of most common weak acids and bases.

These idealized retention-shift patterns can be used for solute-type determination only when: (i) actual solute retention on "real-life" reversed-phase columns is similar to the ideal behaviour and (ii) the experimental retention-shift limits of the various solute types are known.

There are two important experimental factors which have to be carefully adjusted in order to obtain ideal solute retention and significant retention shifts. First, the retention of strong acids and bases (and non-charged compounds) must be invar-

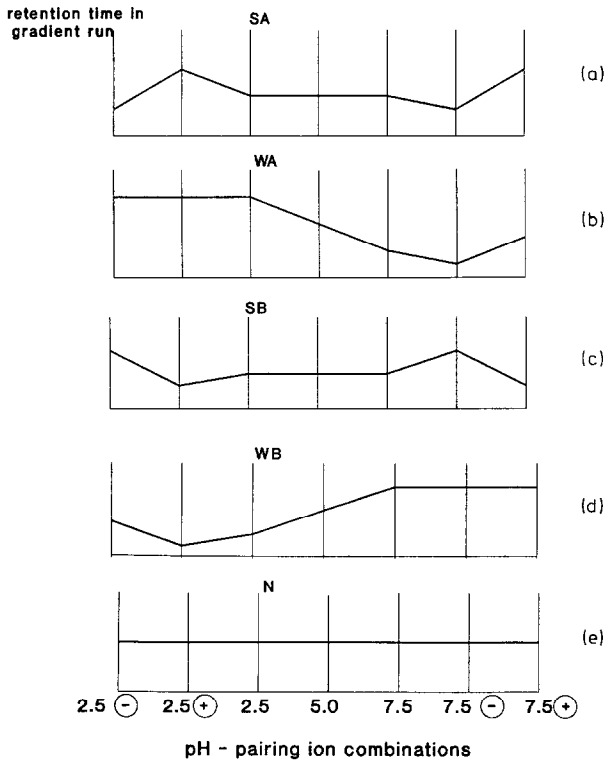


Fig. 1. Idealized retention-shift patterns of (a) strong acid (SA), (b) weak acid (WA), (c) strong base (SB), (d) weak base (WB) and (e) non-charged (N) solutes in the seven-gradient design at different pH and positively charged (symbol +) or negatively charged (symbol −) pairing-ion combinations.

iant for pH and buffer composition changes. This can be achieved by using a triethylamine–phosphate buffer to suppress the silanophilic interactions[6,8,12]. Second, the retention shifts induced by a pairing ion (through Coulombic effects) must be felt over the full duration of the gradient. However, a single-component pairing-ion plug will be eluted only over a short segment of the gradient. This means that the plug of a single pairing ion cannot influence the elution time of the much more retained components. The effects of premature pairing-ion elution cannot be compensated for by simply increasing the amount of the pairing ion which is loaded onto the column[12,17]. Rather, a mixture of pairing ions with increasing adsorption strength must be used. This should provide sufficiently high pairing-ion concentrations along the entire column, during the entire gradient run. Based on our previous adsorption isotherm measurements[18–20], mixtures of tetrapropyl-, tetrabutyl- and tetrapentylammonium ions and octyl-, decyl- and dodecylsulfonate ions were selected. We found that mixtures of these pairing ions lead to larger retention-time shifts (and better defined retention-shift patterns) than single pairing ions do, especially if the relative hydrophobicities of the solutes span a broader range[21].

Experimental retention shifts were determined for a large number of differently charged solutes with mixed pairing-ion plugs on the Novapak $C_{18}$ column, with triethylamine–phosphate buffer–methanol gradients. Typical shift limits derived from this dataset are summarized in Table I. The sign of the retention change (relative to that measured at lower pH or in a gradient without the pairing-ion plug) and the expected extent of the shift (smaller or larger compared with the limit given) are also indicated.

The maximum allowed retention shift of non-charged solutes is related to the reproducibility of the gradients, *i.e.*, the retention time difference for a non-charged solute in any two chromatograms must be less than ±5% relative. Similar shifts are allowed for strong acids and bases when the pH is varied, while at least a 15% change in the retention time is expected for weak acids and bases. Both the repulsion and the attractions between the fully ionized, single-charged solutes and the pairing-ion mixture injected as a plug must produce at least a 10% relative decrease or increase in the retention time. The data in Table I also represent the limits of retention-time windows, which could be drawn around the idealized retention-shift patterns in Fig. 1.

TABLE I

CHARACTERISTIC RETENTION-SHIFT LIMITS (% RELATIVE CHANGE IN RETENTION TIME) OF VARIOUS SOLUTE TYPES IN THE SEVEN-GRADIENT EXPERIMENTAL DESIGN USING PULSE INJECTIONS OF MIXED PAIRING IONS

| *Solute type* | *Relative change (%) in retention time* | | | | |
|---|---|---|---|---|---|
| | *pH variation (pH 2.5, 5, 7.5)* | *Attraction* | | *Repulsion* | |
| | | *pH 2.5* | *pH 7.5* | *pH 2.5* | *pH 7.5* |
| Non-charged (N) | < ±5 | < ±5 | < ±5 | < ±5 | < ±5 |
| Strong acid (SA) | < ±5 | > +10 | > +10 | > −10 | > −10 |
| Weak acid (WA) | > −15 | < ±10 | > +10 | < ±10 | > +10 |
| Strong base (SB) | < ±5 | > +10 | > +10 | > −10 | > −10 |
| Weak base (WB) | > +15 | > +10 | < ±10 | > −10 | < ±10 |

A knowledge of these shift limits is necessary for solute-type classification, irrespective of the method used for the evaluation of the solute retention shifts (complete peak tracking and/or rule-based strategy).

*Principles of the retention-shift rule set*

As peak tracking, which is used to recognize the retention shifts of different solute types in the (gradient) chromatograms of different pH and pairing-ion combinations, is fraught with difficulties, an alternative solute-type identification method was sought. The task is to obtain solute-type information for unknown sample mixtures solely from chromatographic retention-time data, without using any additional (spectal data, peak tracking, peak areas, etc.) information. The factual knowledge available to solve this problem consists of (i) the ideal retention-shift patterns of non-charged solutes, weak/strong acids and bases (Fig. 1), (ii) the experimental retention-shift limits (Table I) and (iii) for each given sample mixture the retention data of all peaks in 2–7 gradient chromatograms.

It must be pointed out that a simple matching of the measured retention times with all possible ideal shift patterns, within the shift limits, does not guarantee the presence of a given solute type. For example, if a given sample mixture does not contain non-charged solute(s), peaks of charged solutes may still elute (owing to their multi-directional shifts) at the same retention time (within $\pm 5\%$) in all chromatograms. Therefore, simply finding a peak within this retention time window in all chromatograms must not be interpreted as proof of the presence of a non-charged solute. The same holds for matching the retention-shift patterns of the other solute types with the measured peak retention data. The combination of retention times and peak areas also will not help, owing to possible coelution of components with varying spectral characteristics.

With this dilemma, the only viable solution is not to verify the presence of a given solute type, but rather to accept that its presence cannot be excluded. In other words, a viable strategy must concentrate on the exclusion of the impossible solute types, rather than on trying to establish which ones are present. The recognition of this simple, but important, fact prompted us to develop our rule-based solute-type evaluation strategy.

Our starting hypothesis is that in any chromatogram any peak can correspond to any solute type(s). In other words, neither the number of the components which can be present in the sample nor the number of the solute-types (weak/strong acid or base, non-charged) which can be assigned to an individual peak are restricted initially. If the sample does not contain certain solute types, their retention shift pattern will not be present. It was found that although the combination of retention-shift patterns becomes intricate, they nevertheless remained unique as the different solute types were combined into mixtures of increasing complexity. As a result, the absence (and/or in very simple instances the presence) of certain retention-shift patterns [*i.e.*, given solute-type(s)] can be recognized from peak retention data alone.

Chromatograms of ideally behaving solute mixtures of increasing complexity were simulated using the individual retention shifts of the various solute types in the seven gradients. The extensive library of these chromatograms was then analyzed and the rules which relate the resulting retention patterns to the different sample compositions were determined. There is a loose hierarchical order in the rule set (and in the

corresponding computer program), which reflects the approach an experienced chromatographer would take in order to minimize the work involved in the solution of the problem.

The retention-shift rule set combines information from the retention-shift windows of Table I and the ideal shift patterns which occur for various mixtures of the different solute-types. The rules can be grouped as follows.

*Rules for extreme cases.* These rules represent cases when the composition of the sample is so simple or fortuitous that the solute type(s) can be determined immediately. For example, if all solutes in the pH 7.5 gradient are eluted before the retention time of the first peak in the pH 2.5 gradient (see Fig. 2a), then all solutes must be weak acids (in both chromatograms). The generalized rule covering this case reads: "If the retention time of the first peak at pH = *X* without/with the addition of a positively/negatively charged pairing ion is shorter/longer than the retention time of the last peak at pH = *Y* without/with the addition of a positively/negatively charged pairing ion, then all solutes in the sample are *Z*". When applied to the previous example the rule reads: "If the retention time of the first peak at pH = 2.5 without a pairing-ion pulse is longer than the retention time of the last peak at pH = 7.5 without a pairing-ion pulse, then all solutes are weak acids in both chromatograms". In such simple cases, two gradient runs may contain enough information to allow classification of the solute mixture (*e.g.*, as in ref. 15).

*Rules to exclude solute types.* The goal of these rules is to exclude the presence of certain solute types by identifying the retention-shift patterns which are missing. The basis of these rules is that peaks belonging to certain solute types must have identical retention times in certain other chromatograms. For example, the same strong acids, strong bases and non-charged components must elute with the same retention times in the pH 2.5, 5 and 7.5 gradients. Therefore, if a corresponding peak is missing in any of the linked chromatograms, the peak cannot belong to a strong acid, strong base or non-charged solute (*e.g.*, see Fig. 2b). The generalized rule is: "If the retention time of a peak at pH = *X* cannot be matched with the retention time of a peak at pH = *Y*, then the peak cannot be *Z*". When applied to the example shown in Fig. 2b, this generalized rule reads: "If the retention time of a peak in the pH = 2.5 chromatogram in the absence of a pairing ion cannot be matched with the retention time of a peak in the pH = 7.5 chromatogram in the absence of a pairing ion, then this peak in the pH = 2.5 chromatogram cannot belong to either a non-charged solute or a strong acid or a strong base". The retention times of two peaks match only if their relative difference falls within the retention shift limit (window) of the given solute type.

*Rules for the first and last peaks.* Special exclusion-type rules are used to compare the first (and last) peaks in certain chromatogram pairs. For example, if the retention time of the first peak in the pH 2.5 gradient with a negative pairing ion is longer than the retention time of the first peak in the pH 2.5 gradient without the pairing ion, then the solute in the first peak of the latter gradient must be a weak and/or strong base (see Fig. 2c). If the retention time of the last peak in the pH 2.5 gradient with a negative pairing ion is shorter than the retention time of the last peak in the pH 2.5 gradient without the pairing ion, then the solute in the last peak of the latter gradient must be a strong acid (see Fig. 2d). The first and last peak rules can also be generalized.

*Conflict-resolving rules.* The conflict-resolving rules follow from our initial as-

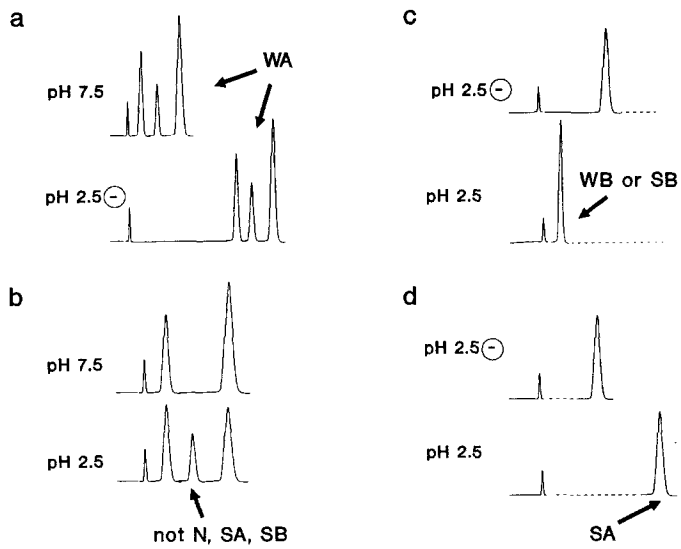


Fig. 2. Typical examples of the different evaluation rules: (a) extreme case or one-solute-type rule, (b) exclusion of solute-type rule, (c) first peak rule and (d) last peak rule. Symbols as in Fig. 1. See text for discussion.

sumption which states that any peak in any chromatogram can belong to any solute type. For example, if a certain solute type cannot be present in any of the peaks in one of the chromatograms (because, *e.g.*, it was eliminated by the use of exclusion rules), it also cannot be present in any peak in any other chromatogram.

*Absence projection rules.* A group of rules permits the projection of the absence of a certain solute type from a given peak to the other linked chromatograms. For example, if the peak at 4 min in the pH 2.5 gradient cannot be a strong acid acid (*e.g.*, owing to a first peak exclusion rule), then the peak(s) at 4 ± 0.2 min (*i.e.*, within the retention-shift window) in the pH 5.0 and 7.5 gradients cannot belong to a strong acid(s) either.

At the moment, the rule set contains 125 different rules which one can use to evaluate the retention time data and classify the solute types. Owing to the large number of rules and, also, the large number of repeated numerical comparisons of the solute retention times, the manual solute-type classification process is laborious and time consuming. Therefore, a computer program was developed to implement this rule-based evaluation strategy.

*Computer program for solute-type evaluation*

In order to test the rule-based solute-type evaluation approach, a research prototype computer program was written using a logic language. The (Turbo) Prolog language was used because it has powerful logics, flexible pattern matching and recursive analysis features, in addition to extensive data input–output capabilities and built-in database options[22].

Different rule representations were used, each suited for a group of rules. Rules for extreme cases were represented as simple "if...then..." rules comparing first and

last peak retention times in certain pairs of chromatograms. Most of the exclusion rules were transformed into recursive algorithms, where the retention time of every single peak is matched with the retention time of all other peaks in the other linked chromatogram(s). The first and last peak rules, owing to their large number although simple form, were transformed into a frame-type representation[23,24]. The rules were used in a forward-chaining manner[25]. Prolog's backtracking features were used only within the recursive algorithms.

The program can be divided into three main parts: In the "Data Input" section the program receives and manages the retention times of all the peaks observed in the gradient chromatograms, together with the respective pH and pairing-ion pulse information. The retention times and chromatograms can be entered in any order.

In the "Type Evaluation" subprogram the retention data of all peaks from all chromatograms are compared (at least two chromatograms are needed to run the program). First, as a starting hypothesis, a list containing all five possible solute types (SA, SB, WA, WB, N) is assigned to all peaks in all the chromatograms. Then, recursively using the retention-shift rule set, the program eliminates the impossible solute-type designations for all the peaks.

In the "List Results" subprogram the solute-type assignments are listed. The user can request the list of rules (and their hierarchical sequence) which were used in the solute-type assignment process, for each individual peak.

The computer program was designed to be fast (typical run times for the type evaluation block are between 10 and 40 s), user friendly (menu-driven, etc.) and flexible. It can be used for the evaluation of the retention time data in a minimum of two to a maximum of seven gradient chromatograms, without any constraints on the order of the measurements and/or data input. Classification of simple solute mixtures may need only 2–4 gradient runs, whereas with complex samples (many different charge types and/or components) all seven gradients may be needed.

The operation of the solute-type determination program is demonstrated by using a simulated sample mixture. The mixture contains a strong acid (SA), a strong base (SB) and two weak bases (WB1, WB2) (see Fig. 3). In this case, four gradients provide enough information to determine all solute types unambiguously. The idealized retention-shift patterns of the above solutes were combined to yield coelution of at least two components in each of the four eluents (at pH 2.5 and 7.5, without and with a negatively and a positively charged pairing ion, respectively), as shown in Fig. 4.

The simulated retention time data in these four eluents (for real samples they must be determined by chromatographic experiments) are listed in Table II. The hold-up time of the column and the peak retention times represent the only actual input for the solute-type determination program, no peak identification being given or used. The chromatograms are identified by their pH (2.5, 5 or 7.5) and the charge type of the ion-pairing (IP) reagent ($-1$, pulsed injection with negatively charged IP; 0, no pairing ion; $+1$, pulsed injection with positively charged IP).

Next, the program is run and the results are listed for each chromatogram. As the selection of the optimization parameters is based on the retention *vs.* pH behavior of the sample components[12–16], solute types found in the pH 2.5 and 7.5 eluents are of primary interest. The results of solute-type determination for these two eluents and an abbreviated version of the explanations of the rules used are given in Tables III and IV, respectively.

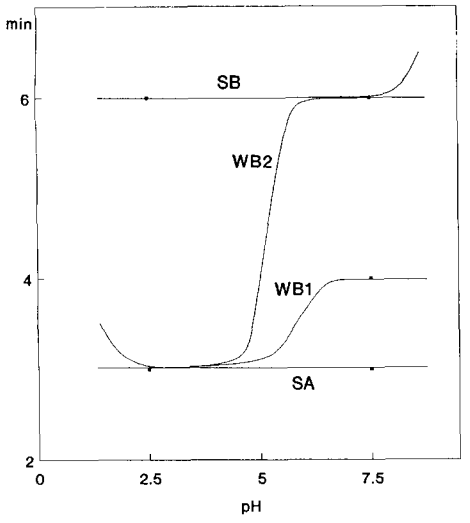


Fig. 3. Idealized retention time *vs.* pH behavior of a four-component solute mixture, containing a strong acid (SA), a strong base (SB) and two weak bases (WB1, WB2).

Clearly, all solute types have been correctly identified by the program (compare Fig. 3 with Tables III and IV). In more complex cases (see the application examples below), all seven gradients may be needed and a single solute type assigned only to the first and last peaks; two or more solute-type assignments can remain for the rest of the peaks. However, even this information is usually sufficient for the rational selection and significant reduction of the optimization parameter space.

*Application examples*

The application of the seven-gradient design and the rule-based evaluation program is demonstrated here using two complex samples. Both samples contain

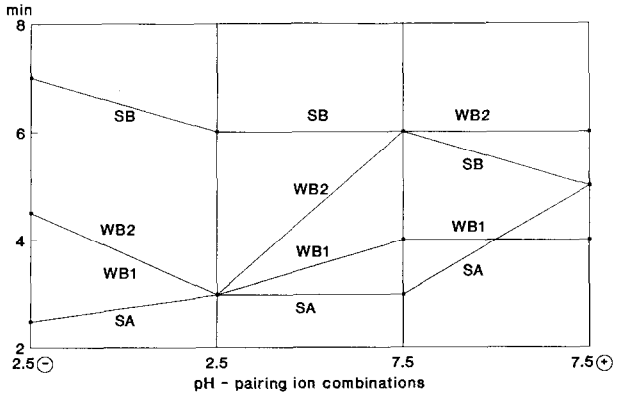


Fig. 4. Simulated retention-time data for the solute mixture in Fig. 3, obtained by combining the ideal retention-shift patterns of the different solute types, at four pH and pairing-ion combinations. Symbols as in Fig. 1.

TABLE II

SIMULATED RETENTION TIME DATA FOR A SAMPLE MIXTURE SHOWN IN FIGS. 3 AND 4, CONTAINING A STRONG ACID, A STRONG BASE AND TWO WEAK BASES, WITH FOUR DIFFERENT pH AND PAIRING ION COMBINATIONS

Column hold-up time: 1 min.

| *Chromatogram* | | *Peak retention times (min)* |
|---|---|---|
| *pH* | *Pairing ion* | |
| 2.5 | No pairing ion (0) | 3, 6 |
| 2.5 | Negative pairing ion (−1) | 2.5, 4.5, 7 |
| 7.5 | No pairing ion (0) | 3, 4, 6 |
| 7.5 | Positive pairing ion (+1) | 4, 5, 6 |

several solutes of different charge types. In order to determine the true charge type of each solute, the retention shifts of all peaks were also determined in all chromatograms by the injection of individual standards. However, it must be stressed that this information was not used by the rule-based computer program: for the program the samples were considered to be completely unknown mixtures.

The first sample is a five-component mixture of dinitroaromatic compounds, containing a non-charged solute (1,4-dinitrobenzene), a strong acid (2,4-dinitrobenzenesulfonic acid), two weak acids (2,4-dinitrophenol and 3,5-dinitrobenzoic acid) and a contaminant, which is also a weak acid. The experimental chromatograms obtained by the seven-gradient design are shown in Fig. 5. All five components are separated in the gradient run at pH 7.5. Retention shifts and peak coelutions occur with both the pH variation and pairing-ion injections, which indicates that the mixture is fairly complex with respect to solute charge types. The corresponding retention shifts for each components are shown in Fig. 6.

TABLE III

RESULTS OBTAINED BY THE RULE-BASED SOLUTE-TYPE IDENTIFICATION PROGRAM FOR THE SOLUTE MIXTURE GIVEN IN FIG. 3 AND IN TABLE II FOR THE pH 2.5 ELUENT

| *pH* | *Pairing ion* | *Retention time (min)* | *Solute types found* |
|---|---|---|---|
| 2.5 | None (0) | 3 | Strong acid, weak base |
| Rules: | | | |
| — Not N, because no matching peaks are found in the (−1) pulsed chromatograms at pH 2.5 and 7.5 (*exclusion rule*) | | | |
| — Not WA, because no matching peak is found in the (−1) pulsed chromatogram at pH 2.5 (*exclusion rule*) | | | |
| — Not SB, because no SB is found in the pH 7.5 eluent at the same retention time (*absence projection rule*) | | | |
| 2.5 | None (0) | 6 | Strong base |
| Rules: | | | |
| — Not N, because... (*same as above*) | | | |
| — Not WA, because... (*same as above*) | | | |
| — Not WB, because no peak of higher retention is found at pH 7.5 (*last peak rule*) | | | |
| — Not SA, because no peak of higher retention is found in the (−1) pulsed chromatogram at pH 7.5 (*last peak rule*) | | | |

TABLE IV

RESULTS OBTAINED BY THE RULE-BASED SOLUTE-TYPE IDENTIFICATION PROGRAM FOR THE SOLUTE MIXTURE GIVEN IN FIG. 3 AND TABLE II FOR THE pH 7.5 ELUENT

| *pH* | *Pairing ion* | *Retention time (min)* | *Solute types found* |
|---|---|---|---|
| 7.5 | None (0) | 3 | Strong acid |
| Rules: | | | |
| — Not N, because no matching peaks are found in the pulsed chromatograms at pH 2.5 and 7.5 (*exclusion rule*) | | | |
| — Not WB, because no matching peak is found in the (+1) pulsed chromatogram at pH 7.5 (*exclusion rule*) | | | |
| — Not SB, because no peak of lower retention is found in the (+1) pulsed chromatogram at pH 7.5 (*first peak rule*) | | | |
| — Not WA, because this solute type has been completely eliminated from the chromatogram at pH 2.5 (*conflict resolution rule*) | | | |
| 7.5 | None (0) | 4 | Weak base |
| Rules: | | | |
| — Not N, because no matching peaks are found in the non-pulsed and (−1) pulsed chromatograms at pH 2.5 (*exclusion rule*) | | | |
| — Not SA or SB, because no matching peak is found at pH 2.5 (*exclusion rule*) | | | |
| — Not WA, because... (*same as above*) | | | |
| 7.5 | None (0) | 6 | Strong base, weak base |
| Rules: | | | |
| — Not N, because no matching peak is found in the (−1) pulsed chromatogram at pH 2.5 (*exclusion rule*) | | | |
| — Not WA, because no peak of higher retention is found at pH 2.5 (*last peak rule*) | | | |
| — Not SA, because no longer retained peak is found in the (+1) pulsed chromatogram at pH 7.5 (*last peak rule*) | | | |

Even a simple comparison of the actual retention shifts in Fig. 6 with the idealized retention-shift patterns in Fig. 1 can lead to solute-type information, as demonstrated previously[12,16]. For example, the retention of the first-eluting 2,4-dinitrobenzenesulfonic acid follows exactly the idealized shift pattern of strong acids. In Fig. 7 the gradient retention times of all five solutes are shown as a function of pH, without pairing-ion addition. The curves represent the idealized retention behavior of the different solute types.

Again, the column hold-up time (1 min), the identification of the chromatograms (pH and pairing ion) and the peak retention times are the only input data used by the solute-type identification program. The possible solute types found by the program for the pH 2.5 and 7.5 eluents are given on the left- and right-hand sides of Fig. 7, respectively. The program correctly determines the charge type of both the first- and last-eluting peaks in the pH 2.5 chromatogram, and totally excludes the presence of strong bases.

The second sample is a seven-component mixture of five anilines (weak bases), 2-cyanopyridine (non-charged) and a contaminant, which is also a weak base. This solute mixture contains only two different charge types (WB and N), but it contains more components than the previous one. The seven gradient chromatograms at different pairing ion and pH combinations are shown in Fig. 8. Again, all components are separated in the pH 7.5 gradients. At high pH, where all solutes are in non-

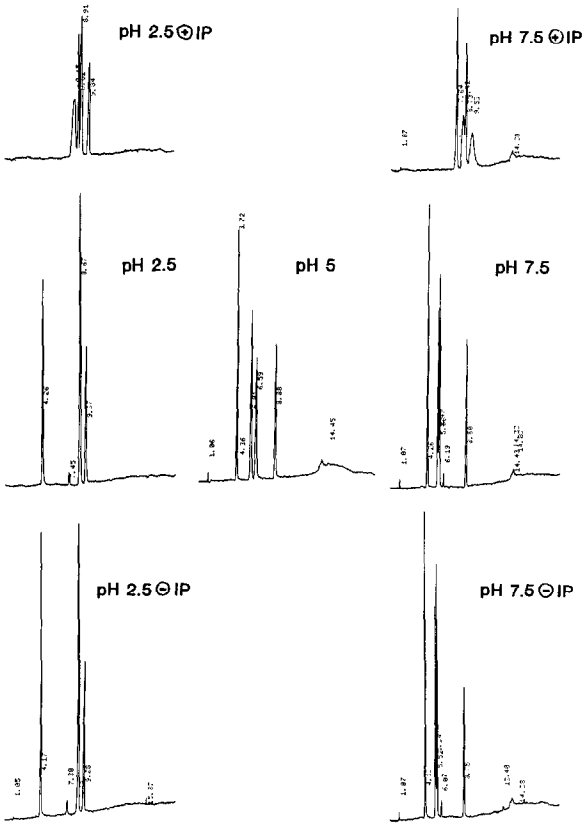


Fig. 5. Experimental gradient chromatograms of a five-component mixture of dinitro-aromatic compounds. Conditions: 15-min linear gradient from 0 to 90% methanol, followed by 5-min isocratic elution at 90% methanol, in 10 m*M* triethylamine–phosphate buffers of pH 2.5, 5 and 7.5, without and with pulsed injection of positively (symbol +) and negatively (symbol −) charged pairing ions, on a Novapak $C_{18}$ column. See Experimental for details.

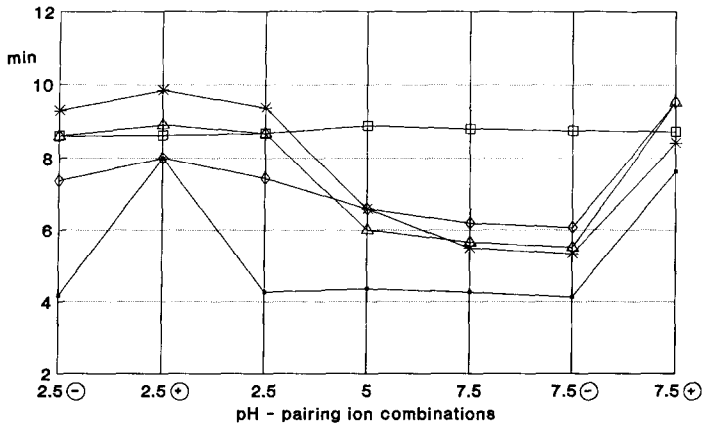


Fig. 6. Experimental retention shifts of a five-component mixture of dinitroaromatic compounds in the seven-gradient design. Solutes: ■ = 2,4-dinitrobenzenesulfonic acid; △ = 2,4-dinitrophenol; ∗ = 3,5-dinitrobenzoic acid; □ = 1,4-dinitrobenzene; ◇ = a contaminant. Conditions and symbols as in Fig. 5.

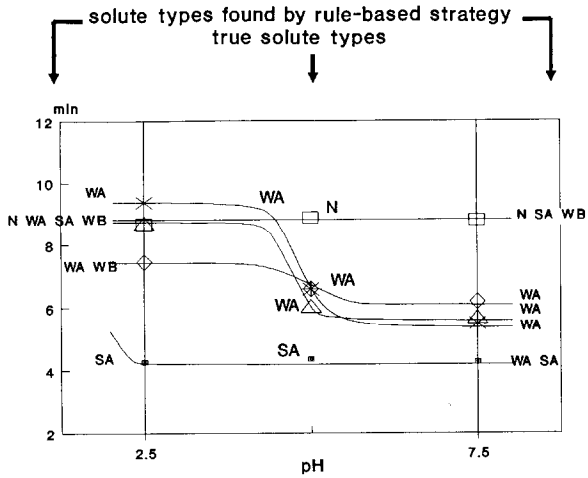


Fig. 7. Gradient retention times of the dinitroaromatic compounds as a function of pH, without pairing-ion addition. The true solute types are represented by the idealized retention curves and indicated in the middle of the plot. The possible solute types, as found by the rule-based computer program for the pH 2.5 and 7.5 eluents, are shown on the left- and right-hand sides of the plot, respectively. Solutes as in Fig. 6.

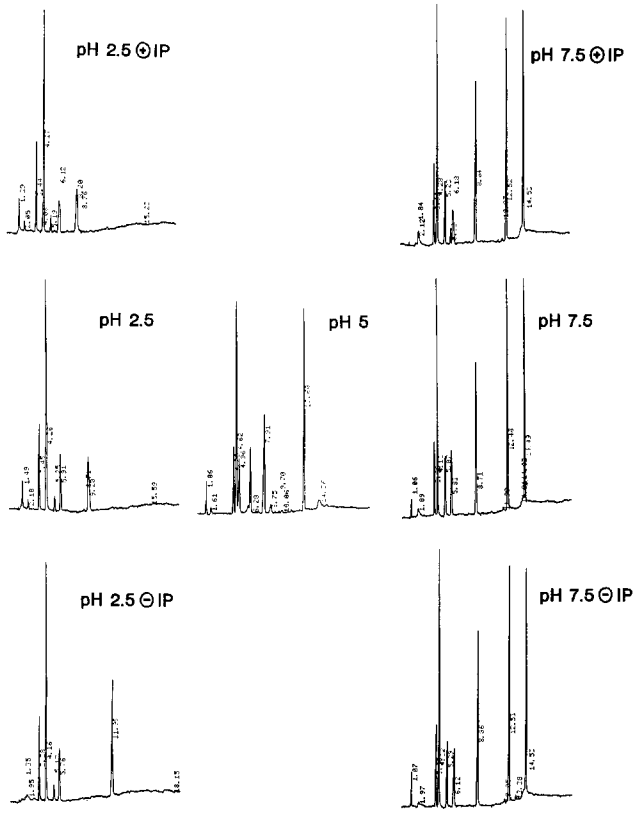


Fig. 8. Experimental gradient chromatograms of a seven-component mixture of anilines and 2-cyanopyridine. Conditions and symbols as in Fig. 5.

charged form, all three chromatograms are very similar. On the other hand, there are large differences between the chromatograms measured in the high and the low pH buffers. Not only the retention times, but also the peak areas change dramatically, indicating that the UV spectra of the weak bases also change significantly with the eluent pH.

The retention shifts for all components are shown in Fig. 9. Again, comparison of this figure with the ideal retention-shift patterns in Fig. 1 allows one to classify the solute types. The real charge type and the idealized retention *vs.* pH behavior of the sample components are shown in Fig. 10.

The results obtained for the pH 2.5 and 7.5 eluents by the rule-based computer program are listed on the left- and right-hand sides of Fig. 10. The program correctly identified the type of the first- and last-eluting peaks in both the pH 2.5 and 7.5 chromatograms, and left two or more (in the case of the non-charged solute peak) possible type assignations for the other peaks.

It is important to point out for both samples that the true solute type(s) present in a given peak is (are) always listed in the possible solute types.

A current limitation of our rule-based approach is that the amphoteric and zwitterionic solutes are not considered. The inclusion of their retention-shift patterns, however, requires further experimental and theoretical work (determination of experimental retention-shift limits and construction of new rules). With very complex samples this fact may lead to incomplete solute-type elimination.

One of the advantages of the rule-based evaluation method is that it does not exclude the use of any additional chemical information about the sample. If certain solute types are known to be absent from the sample mixture, they can be simply omitted from the list of the possible solute types. For example, knowing that our second sample can contain only (weak/strong) bases and non-charged solutes (*i.e.*, that it does not contain acidic components), the stringency of solute-type determina-

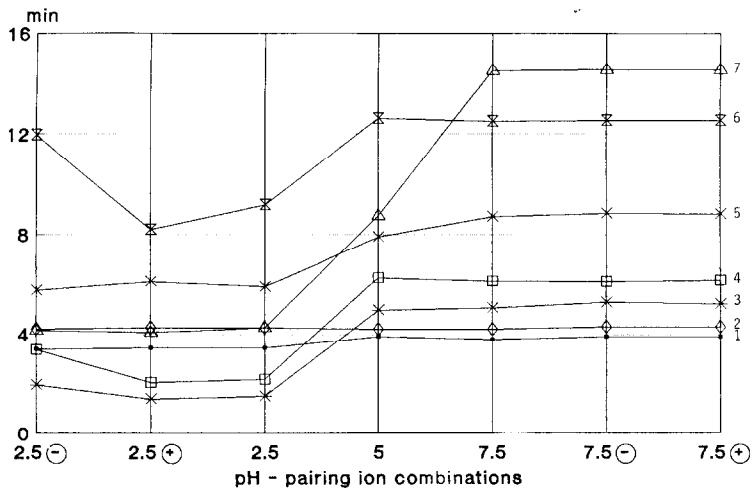


Fig. 9. Experimental retention shifts of a seven-component mixture in the seven-gradient design. Solutes: 1 = a contaminant; 2 = 2-cyanopyridine; 3 = aniline; 4 = 4-nitroaniline; 5 = methylaniline; 6 = 2,6-diethylaniline; 7 = N,N-diethylaniline. Conditions and symbols as in Fig. 5.

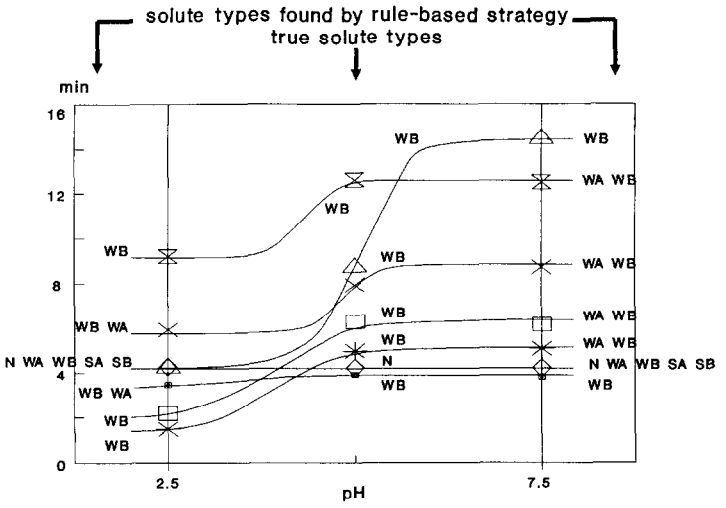


Fig. 10. Gradient retention times of anilines and 2-cyanopyridine as a function of pH, without pairing-ion addition. Solutes as in Fig. 9. Solute types found by the rule-based computer program (at pH 2.5. and 7.5) are shown on the left- and right-hand sides of the plot and the true solute types are shown in the middle of the plot.

tion can be improved, and only one SB assignment is left in the peak of N. An extended version of the Prolog program will allow the input of such *a priori* information and restrict the search only to certain solute types.

The main advantage of the rule-based evaluation method is that it requires only the measurement of retention times in gradient runs, measurements which can be easily automated with a ternary gradient system. The analysis of the retention data set by the rule-based computer program provides solute-type information very rapidly. This information (and the retention data obtained from the gradient experiments) will be used by the next module of a knowledge-based system (currently under development) to select the optimization parameters in RP-IPC rationally, and provide input for existing (and future) optimization software packages.

## CONCLUSIONS

The selection of mobile-phase variables for selectivity optimizaiton in RP-IPC can be rationalized by considering the nature (charge type and relative hydrophobicity) of the sample components. With complex unknown samples it is imperative to establish this knowledge. In this paper we have shown that the nature of the solutes can be determined for totally unknown sample mixtures by a chromatographic experimental procedure in combination with a rule-based evaluation strategy. The new method does not require either peak tracking or the use of other extra-chromatographic information.

In this approach, the unknown sample is analyzed in the reversed-phase mode by running seven linear gradients between 0 and 90% methanol with three different buffers (pH 2.5, 5 and 7.5) and two different pairing-ion combinations. Mixtures of pairing ions of different hydrophobicity are used in pulsed injections to effect signif-

icant retention shifts of all charged solutes throughout the entire gradient run. As soon as a chromatogram is completed (two are needed to start the program), the retention times of all peaks are transferred to a rule-based computer program and evaluated. If further data is not required for unambiguous solute-type identification, there is no need to run the rest of the seven gradients, *i.e.*, the number of experiments depends on the complexity of the sample mixture.

Solute-type identification (without peak tracking) is based on a set of rules, constructed by considering the combinations of the ideal retention-shift patterns and the experimentally determined retention-shift limits of different solute types. First, an assumption is made that any peak can correspond to any solute type (weak/strong, acid/base, non-charged) in any chromatogram. Then, the different (extreme, exclusion, first and last peak, extension, etc.) rules are applied to eliminate sequentially the impossible solute types. The rule set has been implemented in an efficient computer program. When tested with a variety of complex samples, the program correctly identified the type of the first- and last-eluting peaks and excluded the impossible solute-type designations for the rest of the peaks.

ACKNOWLEDGEMENTS

The authors are indebted to M. Huggler and M. J. Roberts (Varian, Sugar Land, TX, U.S.A.) for the loan of the Varian 5500 LC system, and to Dr. J. MacLennan (Waters Assoc., Milford, MA, U.S.A.) for the Novapak $C_{18}$ column and for his interest in this work. Partial financial support by the Advanced Research Program of the Texas Coordination Board of Higher Education, Grant No. 3370, is gratefully acknowledged.

REFERENCES

1 P. J. Schoenmakers and M. Mulholland, *Chromatographia,* 25 (1988) 737.
2 M. T. W. Hearn (Editor), *Ion-Pair Chromatography,* Marcel Dekker, New York, 1985.
3 W. Lindberg, E. Johansson and K. Johansson, *J. Chromatogr.,* 211 (1981) 201.
4 P. Wester, J. Gottfries, K. Johansson, F. Klinteback and B. Windblad, *J. Chromatogr.,* 415 (1987) 261.
5 R. Stromberg, *J. Chromatogr.,* 488 (1988) 1.
6 A. P. Goldberg, E. Nowakowska, P. E. Antle and L. R. Snyder, *J. Chromatogr.,* 316 (1984) 241.
7 H. A. H. Billiet, J. Vuik, J. K. Strasters and L. de Galan, *J. Chromatogr.,* 384 (1987) 153.
8 P. M. J. Coenegracht, N. V. Tuyen, H. J. Metting and P. J. M. Coenegracht-Lamers, *J. Chromatogr.,* 389 (1987) 351.
9 R. Bach, J. Karniczky and S. Abott, in T. H. Pierce and B. A. Hohne (Editors), *Artificial Intelligence Applications in Chemistry*, American Chemical Society, Washington, DC,. 1986, p. 278.
10 H. Gunasingham, B. Srinivasan and A. L. Ananda, *Anal. Chim. Acta,* 182 (1986) 193.
11 Y. Hu, A. Peeters, G. Musch and D. L. Massart, *Anal. Chim. Acta,* 223 (1989) 1.
12 G. K. C. Low, A. Bartha, H. A. H. Billiet and L. de Galan, *J. Chromatogr.,* 478 (1989) 21.
13 A. Bartha, Gy. Vigh and Z. Varga-Puchony, *J. Chromatogr.,* 499 (1990) 423.
14 A. Bartha, Gy. Vigh and J. Ståhlberg, *J. Chromatogr.,* 485 (1989) 403.
15 A. Bartha, H. A. H. Billiet and L. de Galan, *J. Chromatogr.,* 464 (1989) 173.
16 J. K. Strasters, F. Coolsaet, A. Bartha, H. A. H. Billiet and L. de Galan, *J. Chromatogr.,* 499 (1990) 523.
17 V. V. Berry and R. E. Shansky, *J. Chromatogr.,* 284 (1984) 303.
18 A. Bartha and Gy. Vigh, *J. Chromatogr.,* 260 (1983) 337.
19 A. Bartha, Gy. Vigh, H. A. H. Billiet and L. de Galan, *J. Chromatogr.,* 291 (1984) 91.
20 A. Bartha, Gy. Vigh, H. A. H. Billiet and L. de Galan, *Chromatographia,* 20 (1985) 587.

21 A. Bartha and Gy. Vigh, *J. Chromatogr.*, in preparation.
22 K. M. Rich and P. R. Robinson, *Using Turbo Prolog*, McGraw-Hill, Berkeley, CA, 1988.
23 P. Harmon and D. King, *Expert Systems*, Wiley, New York, 1985.
24 C. Townsend, *Advanced Techniques in Turbo Prolog*, Sybex, Alameda, CA, 1987.
25 C. Townsend, *Mastering Expert Systems with Turbo Prolog*, Howard W. Sams, Indianapolis, IN, 1988