Journal of Chromatography, 485 (1989) 219–236
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 21 963

EXPLANATIONS AND ADVICE PROVIDED BY AN EXPERT SYSTEM FOR SYSTEM OPTIMIZATION IN HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY

PETER J. SCHOENMAKERS*
Philips Research, P.O. Box 80000, 5600 JA Eindhoven (The Netherlands)
and
NICK DUNAND
Philips Scientific, York Street, Cambridge CB1 2PX (U.K.)

SUMMARY

An expert system can assist the consulting person (the "user") in making decisions for which the help of an expert would otherwise be desirable. A genuine expert system can also provide an explanation of its recommendations ("explain function") and can offer additional advice or assistance when prompted by the user ("help function").

A particular set of explain and help functions are described that were developed for and implemented in an expert system for system optimization in isocratic high-performance liquid chromatography. Although the expert system has been developed in a typical expert-system tool (Knowledge Craft), it has now been re-implemented in Pascal. The features of the new implementation are also described.

The expert system selects the optimum column from a user-defined column database and suggests the optimum operating conditions (flow-rate, sample size) and instrumentation (detector cell, time constant). The optimum separation is defined as that which yields (i) sufficient separation and (ii) sufficient sensitivity in (iii) the shortest possible time. On a given column, the shortest possible time corresponds to the highest possible flow-rate. The optimization is performed by calculating the higest permissible flow-rates ("ceilings") with regard to seven different parameters and the lowest permissible flow-rates ("floors") with regard to five other parameters. A valid result is obtained if the lowest ceiling is higher than the highest floor.

An explanation of the optimization process can be obtained for each column in combination with the optimum detector (cell) and detection time constant. If only one column is considered, the explanation can be obtained for each column–detector–time constant combination, no matter whether a valid result is obtained or not. The explanation is provided in the form of a bar chart, depicting all the different floors and ceilings. This allows the user to identify rapidly which factor(s) limit the speed of the analysis. Further advice to the user is available for each of the twelve parameters considered during the optimization. This may help the user to reduce the constraints imposed by the limiting factor(s) and hence to reduce the analysis time.

0021-9673/89/$03.50

INTRODUCTION

We have recently described an expert system for optimizing separations in terms of (i) sufficient separation and (ii) sufficient sensitivity in (iii) the shortest possible time[1]. This previous paper discussed an expert system in the conventional sense. It was implemented in a genuine expert-system tool (Knowledge Craft), it featured question-and-answer interplay with the consulting user and it used the inference engine of the sytem to reason towards an optimal result. In this paper we discuss a re-implementation of this expert system in a conventional programming language (Pascal). Because the system performs the same task as the expert system developed in Knowledge Craft and because performing this task without the system would require a genuine expert in chromatography, we continue to refer to it as an expert system.

When an expert system is rewritten in a conventional programming language, such as Pascal, the help and explain facilities of the original tool are no longer available. The same facilities can, however, readily be programmed in Pascal. Explain facilities offered by expert-system development tools can be used to inform the user that because certain conditions were met, a new fact has been established (*i.e.*, a rule has "fired"). The rules that have been applied can be traced back to follow the reasoning of the system in the opposite direction ("How" facility). Also, the user may enquire why certain facts are required by the system ("Why" facility). In our experience, the conventional explain facilities provided in expert-system development tools (*i.e.*, Knowledge Craft and many others) are generally not very helpful. Therefore, instead of mimicing such facilities in our re-implementation in Pascal, we have developed different explain and help facilities, which are unique to the present system. The description of these functions is the purpose of this paper. The expert system itself was described previously[1]. In this paper we provide a brief introduction and describe the differences between the Knowledge Craft and Pascal versions.

As a starting point, the system requires a chromatogram obtained with isoratic high-performance liquid chromatography (HPLC). It will typically be used for a chromatogram in which the peak shape and efficiency are adequate, retention is in the optimum range and selectivity is sufficient to achieve the separation with the required resolution and sensitivity in a reasonable time. As such, the system can be seen

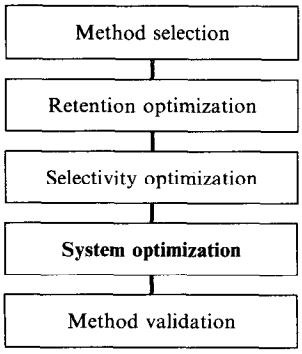

Fig. 1. Outline of a complete expert system for method development in HPLC. The system-optimization module (bold type) is discussed in this paper.

to form part of a large package for method development in HPLC[2]. An overview of this complete system is shown in Fig. 1.

In real life, there are different stages at which an expert chromatographer may consider optimizing the operating conditions. For example, there may be a trade-off between the possible gains of and the time required for selectivity optimization. The user may first want to investigate whether a separation can be realized with a given combination of a mobile and a stationary phase. Based on the outcome of the system optimization (*e.g.*, the analysis time needed to reach the required values for resolution and sensitivity), the user may then decide whether or not the selectivity of the system should be optimized, for example by varying the composition of the mobile phase. Therefore, knowledge on when best to apply the different modules should form part of a complete (integrated) system for method development in HPLC. Building such integrated systems is a major goal within the European co-operation project ESCA ("Expert Systems for Chemical Analysis")[2].

A genuine expert system should offer a number of features to the user: (i) it should offer valuable advice in an area requiring expert knowledge; the problem addressed by the expert system should be both sufficiently important and sufficiently difficult; (ii) it should be able to provide some kind of explanation for its reasoning ("explain function"); and (iii) it should be able to provide additional assistance to the user, both to assist in the consultation of the system and to clarify the advice ("help function").

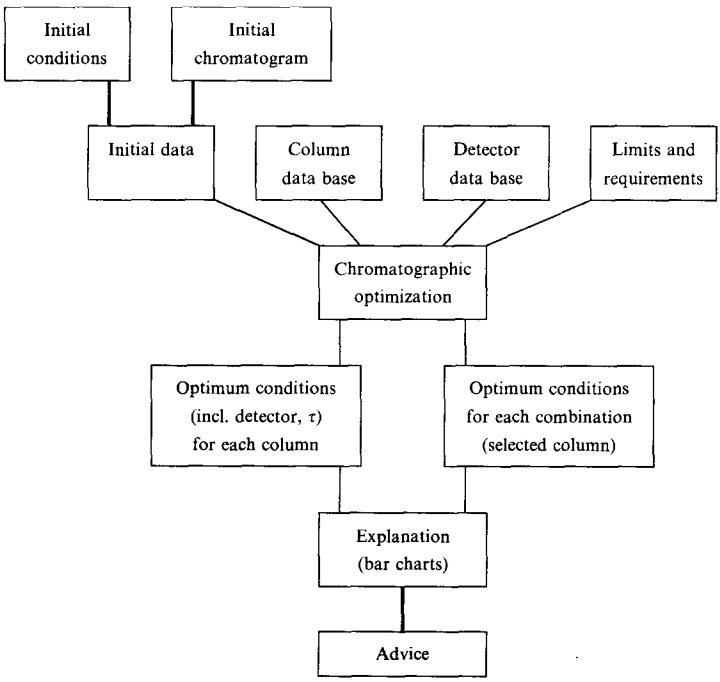

Fig. 2. Outline of an expert system for system optimization in HPLC. Through the user interface, the user can readily alter any of the initial data, entries in the column and detector databases and limits and requirements. Explanation can be asked for any column–detector–time constant combination displayed on the screen and further advice can be asked for each parameter occurring on the explanation chart.

OVERVIEW OF THE SYSTEM

A detailed description of the expert system for system optimization in HPLC has been provided previously[1]. In this section we summarize the system briefly, putting emphasis on differences between the present system and that described previously and on the explain and help facilities, which have not been described before.

Fig. 2 shows an outline of the present system. To run the system, an initial chromatogram is needed, together with the relevant information on how it was recorded. An example of the required input is given under Results. On the second line in Fig. 2 are two databases, one describing the available columns and the other the available detectors (or detector cells). These databases usually need to be created once in a given laboratory, but they can be edited or modified by the user at any time. The column database contains information on the dimensions of the column, its maximum and minimum flow-rates and pressure drops and the maximum amount of sample that can be injected. It also contains empirical factors to describe the actual efficiency and permeability of a column. These parameters allow, for example, that an older column, of which the efficiency has decreased by 20%, can still be considered by the system. Because the entire column database is usually depicted on the screen during a consultation, it is advisable to define different databases for different types of columns (*e.g.*, silica, octadecylsilica, cyanopropylsilica). It may also be desirable to limit a column database to columns from one or a few manufacturers. This can be done either by defining different databases or by selecting a limited number of columns that are relevant for a certain application. The latter can be done very easily and rapidly with the present expert system.

The detector database contains the volume of the detector cells and empirical factors for their sensitivity and noise levels. If all detectors are of the same type, then the sensitivity factors may be independent of the solutes. For example, for UV detectors the sensitivity will be proportional to the optical path length. The system does allow different detectors to be included in the detector database, but in that case the (relative) sensitivities and noise levels must be altered for each application of the system.

The limits and requirements consist of three different sets of values:

(i) *Limits*, these are the overall minima and maxima for the flow-rate and the pressure drop and the maximum volume of sample that can be injected. These values may, for example, be determined by the instrument.

(ii) *Distortion factors*, indicating the maximum permissible extra-column dispersion in the time domain (detection time constant) and the volume domain (detector-cell volume).

(iii) *Requirements*, *i.e.*, the required resolution and signal-to-noise ratio for the analysis.

The chromatographic optimization is outlined in Fig. 3. For each combination of a column, a detector (cell) and a detection time constant, there are seven factors that set an upper limit to the flow-rate. If the flow-rate increases above a certain value, the detection time constant may become so large as to cause unacceptable peak distortion. The acceptable level is defined by the user in the form of a maximum permissible distortion in the time domain. This parameter is defined as

$$\delta_t = \tau/\sigma_{t,0} \tag{1}$$

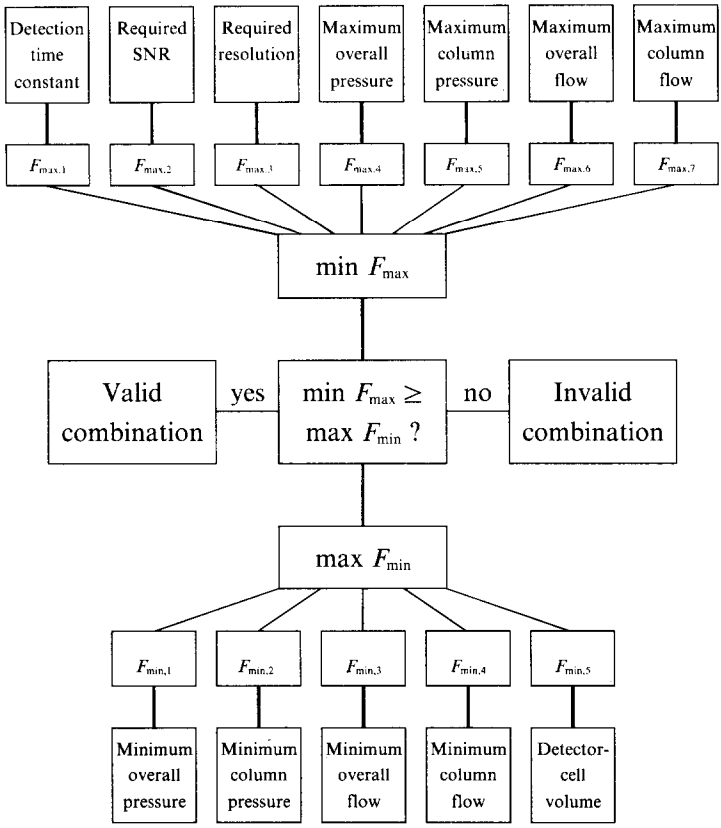

Fig. 3. Strategy followed for optimization.

where τ is the detection time constant and $\sigma_{t,0}$ is the standard deviation of an unretained peak in time units. The default value of δ_t is 0.5, but it may be altered by the user. Higher values than 0.5 may give rise to serious peak distortion and a significant decrease in resolution.

A second factor is the signal-to-noise ratio (SNR). When the flow-rate increases, the column efficiency decreases. When concentration-sensitive detectors are used, as is typical in LC, the peaks become lower and, therefore, the SNR value decreases. The minimum required SNR, as specified by the user, corresponds to a maximum flow-rate. Likewise, the resolution decreases with increasing flow-rate and the specified minimum required resolution corresponds to a maximum flow-rate.

The other four parameters that define an upper limit for the flow-rate are the specified values for the maximum flow-rate and pressure drop for (a) the column and (b) the overall analysis. The column limits can be seen as a protection against abuse. These values are contained within the column database, but they can be altered by the user at any time. The overall values can be selected by the user for a given analysis. Alternatively, default values may be defined for the instrument.

There are five parameters that set lower limits to the required flow-rates. These

are indicated at the bottom of Fig. 3. Four of these correspond to the upper limits, *i.e.*, the lower column and overall limits on the flow-rate and pressure. There may be a minimum pressure for proper operation of the pump, but the value may also be zero. Likewise, the minimum pressure drop over the column is not likely to be a vital parameter. Of more importance is the minimum overall flow-rate. This parameter should preferably be higher than the minimum specified flow-rate of the instrument to guarantee reproducible flow-rates.

The detector-cell volume also imposes a lower limit on the flow-rate. While working above the minimum in the *H vs. u* curve, a reduction in the flow-rates causes an increase in the column efficiency and thus a decrease in the peak width in volume units. When the flow-rate is decreased too far, a given detector cell will no longer be allowed.

The present system is based on a large number of relationships between chromatographic parameters[1]. Many of these are rigorously correct, but some are approximate. All equations are monotonous, which is obviously not correct for the relationship between flow-rate and efficiency (plate count). The minimum flow-rate specified for a column is an essential parameter, because it is used to make sure columns are operated at a flow-rate that is at least as high as that corresponding to the minimum in the Van Deemter curve (*i.e.*, in terms of the linear velocity $u \geqslant u_{opt}$). For small solutes in reversed-phase LC we suggest that the minimum specified flow-rate (in ml/min) should be above

$$F_{min,col} = \gamma d_c^2/d_p \tag{2}$$

where the constant γ has a value of 0.1 (mm^2/min) if d_c is the column diameter in mm and d_p is the particle size in μm. This recommendation is contained within the help facility of the expert system (see below).

The final parameter that sets a lower limit to the flow-rate is the extra-column dispersion, which is characterized in the present system by the detector-cell volume. Other contributions to the extra-column dispersion are not considered. The user is advised (by the system) to adapt the entire instrumentation to the dispersion of the flow cell, for example by using short pieces of narrow-bore capillary tubing in connection with small flow cells. Assembling (and trouble-shooting) the entire instrumentation is beyond the scope of the present system.

The maximum permissible extra-column dispersion can be specified as a distortion factor in the volume domain, which is defined as

$$\delta_V = \frac{(\sigma_{V,0})_{obs} - (\sigma_{V,0})_{col}}{(\sigma_{V,0})_{col}} \tag{3}$$

where $\sigma_{V,0}$ is the standard deviation of an unretained peak in volume units. The relationship between the observed (obs) band broadening and that caused by the column (col) is described by

$$(\sigma_{V,0})^2_{obs} = (\sigma_{V,0})^2_{col} + V^2_{det} \tag{4}$$

The default value for δ_V in the present system is 0.2. Higher values may give rise to serious peak distortion and a significant decrease in resolution.

In the centre of Fig. 3 it is indicated how the actual optimization is performed. Out of the seven upper limits to the flow-rate, the lowest maximum (min F_{max}) is selected. Likewise, there is a highest minimum denoted as max F_{min}. The lowest maximum indicates the highest flow-rate that can be used. At this flow-rate, the optimum result (sufficiently large resolution and SNR in the shortest possible time) can be obtained using a given combination of a column, a detector (cell) and a time constant. This is true, provided that the lowest maximum flow-rate exceeds the highest minimum value. If this is not the case, then there is no flow-rate at which sufficient resolution and sufficient sensitivity can be obtained within the constraints specified by the user. In this case, we speak of an invalid combination of a column, a detector and a time constant.

The optimization illustrated in Fig. 3 is performed for each combination. If ten columns, five detectors and five time constants are available, there are 250 possible combinations. For each column, out of 25 possible combinations, the optimum one is selected based on the following criteria: the shortest analysis time; the highest resolution; the highest SNR; the largest detector-cell volume; and the largest time constant. These criteria are listed in order of priority, *i.e.*, the second criterion only applies if two combinations give rise to equal analysis times, etc. It should be noted that the selection of the optimum combination for a particular column only involves valid combinations, so that the predicted resolution and SNR will always be above the required values specified by the user.

If all 25 possible combinations for a particular column turn out to be invalid, we speak of an invalid column. The output of the system is a list of valid columns with the optimum choice of detector and time constant, the optimum flow-rate, the maximum sample size and the predicted values for the analysis time, resolution, SNR and pressure drop over the column. If ten columns are considered, the results table may consist of anywhere between zero and ten possible columns, listed according to the criteria described above.

If only a single column in the database is considered during the optimization, then all possible combinations are listed on the screen. The valid combinations are listed at the top in order of preference. Invalid combinations are listed, but without predicted values for the flow-rate, resolution, pressure drop and SNR. This is a useful option in several situations, for example to see what the effect would be of using the optimum column with different detectors or time constants or to find out why a column is invalid.

The explain facility of the present system can be used to clarify the optimization process for any combination appearing in the results table, *i.e.*, either the optimum conditions for each valid column if more than one column is considered or all possible combinations involving a single column. In the latter instance, explanations can be obtained both for valid and invalid combinations.

IMPLEMENTATION

The previously described version of this expert system[1] was implemented in the expert-system development tool Knowledge Craft (Carnegie Group, Pittsburgh, PA,

U.S.A.) and ran on a Digital Equipment (Maynard, MA, U.S.A.) Micro VAX workstation. Although we demonstrated the feasibility of building the expert system using this tool, there are several disadvantages, the two most important of which are as follows. (i) Knowledge Craft is an expensive software package, which makes it difficult to use the system at different locations. Also, the possibilities of Knowledge Craft are not fully exploited if it is used to implement a relatively small expert system. (ii) Like all other systems based on the LISP language, but probably more so than most, Knowledge Craft appeared to suffer from a phenomenon called "garbage collection". In practice, using a VAX computer, this meant waiting periods of anywhere between 5 and 20 min at frequent, but irregular intervals.

Since the earlier report, two new versions of the system have been implemented. One of these exploited another expert-system development tool, NExpert Object (Neuron Data, Palo Alto, CA, U.S.A.). This tool is smaller than Knowledge Craft and, therefore, the first of the above problems becomes less serious. Also, NExpert Object is based on the C language rather than on LISP, so that garbage collection is no longer a problem.

Another implementation, which avoids both problems, has been written in a conventional programming language (Pascal). This meant that all the features of the system (including the explain and help functions) had to be built "from scratch". Because an expert system is essentially a computer program, any such system may eventually be written in a conventional computer language.

In our opinion, expert-system development tools play an important role in building new systems. Prototypes can rapidly be built while the computer expert (the "knowledge engineer") can concentrate on understanding the expert's knowledge, rather than on writing a user interface. Expert-system tools allow rapid and flexible prototyping and allow the expert to change his or her mind on (parts of) the domain knowledge during the development of the system.

In the later stages of developing an expert system, the knowledge is likely to change only in detail. When this is the case, an implementation in a conventional programming language may be considered. The program becomes less flexible and it becomes difficult to alter or add substantial pieces of knowledge. On the other hand, the system may become faster and it can more easily be used at different locations.

The results described below were obtained using the Pascal implementation running on a Micro VAX workstation. All parts of the system can be displayed on the screen simultaneously, although a large table of results (more than ten combinations) may push parts of the other windows to the background. All functions of the system can be addressed very quickly by clicking a button on the mouse. The keyboard is needed only to start the system, enter new values or save tables.

The most important performance characteristic of the system is the time required to perform a so-called edit–optimize–edit cycle. This implies that the user changes one (or several) parameters and asks for a renewed optimization and advice. In the present implementation, there is hardly any significant waiting time. Even if ten columns, five detector cells and five time constants are considered, the entire optimization process is a matter of seconds. The process of changing a single parameter, *i.e.*, selecting the relevant window, choosing the parameter to be altered and entering it via the keyboard, takes approximately 10 s. Chainging several parameters within the same window will only take a few seconds more. Therefore, the interaction between the user and the system can be very rapid.

Building column and detector databases requires about 10 min, once the required data (*e.g.*, test chromatograms on all columns) are available. Building a column database is necessary when the system is used for the first time in a laboratory or if a completely new type of column is considered. Consulting the system for a new problem requires the introduction of the initial conditions and a peak table. The latter process may in principle be automated by transferring a peak table from a data station to the expert system. This coupling has not been realized with the present system, but entering the data manually requires only 5–10 min (depending on the number of peaks).

RESULTS

Explain facility

The explain facility provided by typical expert systems is based on retracking the rules that have been applied in a reasoning system. For a system for optimizing LC separations, such an explanation might be as follows. On prompting the system with "Why?", it may reveal to the user that because of rule 173 (stating that the pressure drop should be less than the overall maximum pressure drop) and because rule 154 had been previously applied (to predict a pressure higher than the specified overall maximum), the flow-rate must be lowered. If asked again "Why?" the system would typically retrack one step further and reveal that rule 32 states that the flow-rate is proportional to the pressure drop.

This type of explain facility, based on retracking the reasoning of the system, can be a facility provided by the expert-system development tool. Understanding the reasoning of the system and the foundations for its conclusions may require the retracking of many steps in the process and may therefore be very difficult. For the

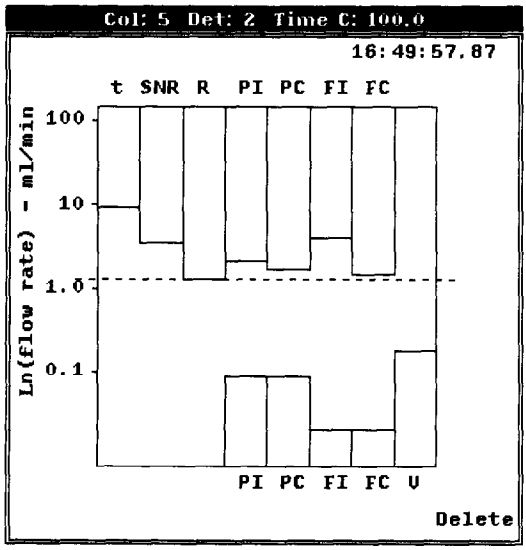

Fig. 4. Example of a bar chart for a valid column–detector–time constant combination. This bar chart corresponds to the strategy outlined in Fig. 3.

present system we have built our own explain facility, which is not based on the features of an expert-system development tool. The explanation is provided in the form of a bar chart, which corresponds closely to the operation of the system illustrated in Fig. 3.

Fig. 4 shows an example of a bar chart obtained for a valid combination of a column, a detector and a time constant. The seven upper limits to the flow-rate are depicted as bars coming down from the top ("ceilings"), while the five lower limits are represented by standing bars ("floors"). The (logarithmic) vertical axis on the bar chart may be changed from coarse ($100 \geqslant F \geqslant 0.001$ ml/min) to fine ($20 \geqslant F \geqslant 0.2$ ml/min) by clicking a button on the mouse.

The horizontal dashed line is the optimum flow-rate, which is chosen to be as high as possible, without cutting through any of the bars. If the lowest ceiling is lower than the highest floor, such a dashed line cannot be drawn and the combination is invalid. The bar chart for an invalid combination, without a dashed line indicating the optimum flow-rate, can be examined by selecting only the column of interest from the column database.

The bar chart illustrates which parameters prohibit a further increase in flow-rate and, hence, a further reduction in the analysis time. If one of the ceilings is much lower than all the others, then a considerable improvement may be possible if the

(a)

HELP: Resolution.

If this is the lowest ceiling, then the factor limiting the speed of analysis is the minimum required resolution between any (relevant) pair of peaks in the chromatogram. The specified minimum resolution is one of the most important factors in the optimization process and its value should be carefully selected. If the mimimum resolution is the limiting factor for reducing the analysis time, then consider whether the specified value can be reduced. However, you better be safe than sorry!

Notes:
1. Verify whether the minimum resolution in the chromatogram does concern two peaks that need to be separated (i.e. at least one relevant peak).
2. Because relative peak heights have been considered in calculating resolution values, verify that the initial chromatogram did not give rise to more extreme peak-height ratios than the (anticipated) samples.

Fig. 5.

(b)

HELP: Detector-cell volume

If this is the highest floor and if the highest floor is higher than the lowest ceiling, then the factor preventing a valid solution is the extra-column band broadening, as characterized by the detector-cell volume.
A cell with a smaller volume will be required to obtain a valid solution for this column.

Notes:
1. It is assumed in the present system that the detector-cell has the largest single contribution to the extra-column band broadening. It is assumed that in installing a small cell all other factors are taken into account simultaneously. It is advisable to use specifically designed low dispersion instrumentation (and connnections) for use in conjunction with small flow cells.
2. You may reduce the constraints on the detector-cell volume by increasing the maximum distortion factor for extra-column dispersion. However, values above 0.2 may lead to serious distortion of the peaks.

Fig. 5. Two examples of the "Help" files that are available for each of the twelve potentially limiting factors considered in Figs. 3 and 4. (a) Resolution (upper limit); (b) detector-cell volume (lower limit).

constraint can be relieved. If several of the ceilings are very close to the dashed line, then a significant reduction in the analysis time cannot easily be achieved on the column being considered. For example, if both resolution and pressure drop are limiting factors, an increase in the maximum permissible pressure drop would not allow an increase in the flow-rate. Also, if the required resolution is decreased, the flow-rate cannot be increased without increasing the pressure ceiling. This is illustrated in Fig. 4, where the ceilings imposed by the required resolution, the maximum pressure drop over the column and its maximum flow-rate are all at similar levels.

The bar charts can therefore be used to establish rapidly which factors are limiting the speed of analysis, which factors cause a particular combination to be invalid and whether or not a major reduction in the analysis time is still feasible.

Help facility

In addition to providing an explanation of the optimization process, the present expert system provides additional advice to the user. For all the twelve possible limiting factors, a "Help" file can be called upon, which contains information on the

(a)

peak table

peak no.	retention time secs	peak height mm	asymmetry (1)	plate count	relevant yes/no
0	45	1	1	20000	no
1	49.5	70	1.3	17200	yes
2	67.5	25	1.1	19200	yes
3	108	30	1.2	21200	yes

(b)

initial values

column	
length (cm)	15
diameter (mm)	4.5
particle (um)	3
efficiency (1)	1
permeability (1)	1
porosity (0.67)	0.63
detector	
sensitivity (1)	0.4
noise (1)	2
time constant (ms)	500
flow rate (ml/min)	2
pressure drop (bar)	150
resolution	6.86
snr	1000
plate count	19200
sample volume (ul)	0.5

(c)

columns

column no.	phase	length cm	dia. mm	part. um	eff (1)	perm (1)	por (0.67)	f_max ml/min	f_min ml/min	dp_max bar	dp_min bar	q_max ul	avail. yes/no
2	ODS	15	4.6	5	1	1	0.61	10	0.3	250	25	10	yes
3	ODS	10	4.6	3	0.9	1.1	0.58	10	0.5	300	50	10	yes
12	ODS	10	4.6	5	1	1	0.61	10	0.3	250	25	10	yes
22	ODS	5	4.6	5	1	1	0.61	10	0.3	250	25	10	yes
13	ODS	5	4.6	3	0.9	1.1	0.58	10	0.5	300	50	10	yes
23	ODS	3	4.6	3	0.9	1.1	0.58	10	0.5	300	50	10	yes

(d)

detectors

detector no.	volume ul	sensitivity	noise
1	8	1	1
2	2.4	0.9	1.5
3	1.2	0.4	2.0
4	0.5	0.2	5
5	0.1	0.03	5

Fig. 6.

(e)

time constants

msecs
20
50
100
500

(f)

distortion factors

time constant dist.	0.5
extra-column disp.	0.2

(g)

limits

max flow rate ml/min	10
min flow rate ml/min	0.02
max press. drop bar	250
min press. drop bar	10
max sample (ul)	0.5

(h)

requirements

resolution	1.5
s-n-r	50

(i)

results

column no.	detector no.	time constant msecs	time secs	flow ml/min	q ul	resolution	snr	pressure bar
23	3	20.0	5.74	7.24	0.500	1.50	227	124
13	2	50.0	7.92	8.75	0.500	1.76	760	250
22	2	50.0	8.61	8.46	0.500	1.50	615	75.3
12	2	50.0	14.6	10.0	0.500	1.95	400	178
2	2	100	23.3	9.37	0.500	2.47	477	250
3	2	100	31.7	4.37	0.500	3.52	1075	250

Fig. 6. Examples of the tables appearing as windows on the screen during a consultation of the expert system. (a) Peak table (for initial chromatogram); (b) initial conditions; (c) column database; (d) detector database; (e) list of possible time constants; (f) maximum permissible distortion factors; (g) overall limits; (h) user requirements; (i) results.

factor in question and on possibilities for improvement. Two examples are shown in Fig. 5.

The required resolution is one of the most important parameters. Within the "Help" file (Fig. 5a) the user is reminded of this, but it is suggested that the required resolution may be lowered to reduce the analysis time. A couple of useful hints are also provided by the system. The peak table representing the initial chromatogram allows peaks to be identified as either relevant or irrelevant. The resolution between two irrelevant peaks will not be considered by the system. Also, the system makes use of complex resolution equations[3], which correct for the effects of (relative) peak heights,

individual peak widths and peak asymmetry on the resolution. These equations yield two values for the resolution between any pair of peaks. If both peaks are relevant, then the lowest value is taken. If only one of the two peaks is relevant, then its value will be used. This implies that separating a relevant small peak from an irrelevant large peak is more difficult than separating a relevant large peak from an irrelevant small peak. The possibility of using better, but complex, resolution equations is a bonus of the present expert system.

If the optimization is performed with a reference sample in which the relative peak heights of a particular pair of successive peaks differ much more (or much less) than in the samples to be analysed, the predictions may be inadequate. Ideally, the user will use a test sample that represents the most difficult situation expected in practice, *i.e.*, peak-height ratios that are as large, but not larger than, those expected in typical samples.

Fig. 5b provides an example of the "Help" files for one of the floors. Of all the factors that set a lower limit on the flow-rate, the extra-column band broadening is probably the most important. The lower limits are relevant only for invalid combinations. Once a valid solution has been obtained, improvement should be sought by relieving the upper limits. This is the first part of the advice given to the user. It is also made clear that the detector-cell volume is singled out as the most obvious exponent of extra-column band broadening, but that it is not the only factor. Finally, the user is told how to loosen the constraints on the extra-column band-broadening by increasing the value of δ_V (eqn. 3), but a warning about the possible negative effects of such an increase is contained in the system.

Example

In this section we discuss a simple example of the use of the present expert system. Fig. 6a (peak table) and 6b (initial values) together represent the initial chromatogram, which was taken from a paper by Freebairn and Knox[4]. In a typical (practical) situation, all the required data are available or can easily be obtained from the chromatogram. For the present example, some data were not available and estimates were used. The asymmetry factors can be measured from a chromatogram, ideally at 13.5% (e^{-2}) of the peak height[3]. For the column porosity, the pressure drop and the SNR value for the lowest relevant peak (peak 2 in the table) estimated values were used. The values for the resolution and the (average) plate count do not need to be entered in the table of initial values, because these data are calculated from the peak table. The detector used in ref. 4 had a volume of 1 μl and a shorter optical path length than a conventional 8-μl cell. The relative sensitivity was assumed to be 40% of that of a conventional cell, while the noise level was assumed to be twice as high (see also Fig. 6d).

Fig. 6c shows the column database used for this study. Because the initial value for the resolution was large, the column database was limited to a number of small columns for rapid analysis. The present example concerns a simple application of the system to a chromatogram with excess separation. However, we have also applied the sytem to complex separations requiring long columns. We have used the system successfully to study the possibilities of packed-capillary[5] and miniaturized packed-capillary columns[6]. The (relative) efficiency and permeability factors[1] have been used to express our experience that columns packed with 3-μm particles may be less efficient

and may yield a higher pressure drop than expected. These factors are a correction for the theoretical predictions. If all values are set equal to unity, the system will compute the plate counts and pressure drops of the different columns from well established theory.

Fig. 6d–h show a number of other windows that appear on the screen. Fig. 6d shows the detector database used for this example. The first three cells may be typical,

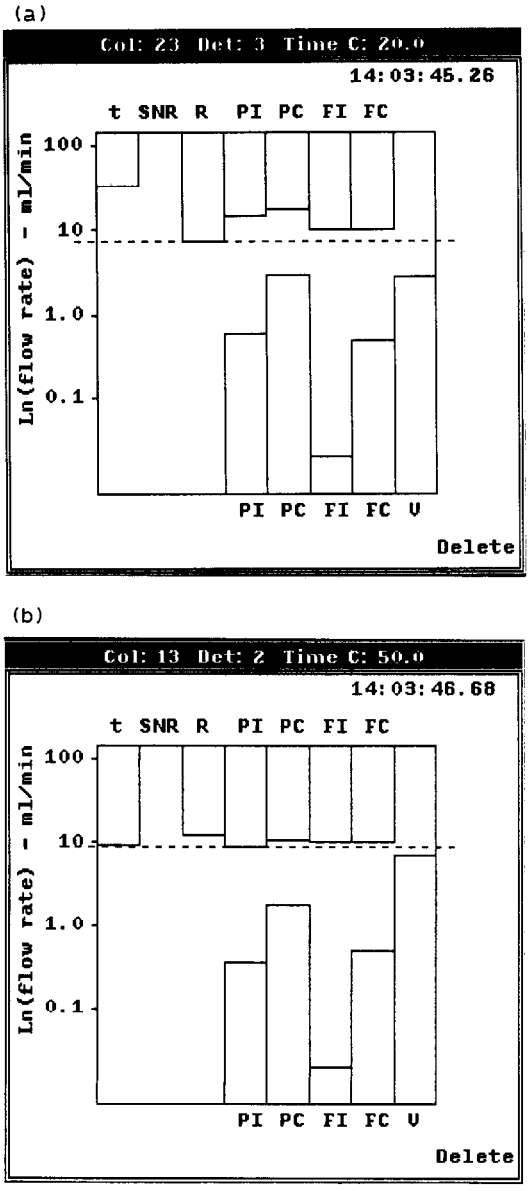

Fig. 7. Examples of bar charts explaining the results of the optimization illustrated in Fig. 6. (a) Column 23 (optimum column); (b) column 13.

commercial ones, whereas the latter two may be research cells. Four different values for the time constant were considered (Fig. 6e) and the default values were used for the maximum permissible distortion factors (Fig. 6f).

Fig. 6g shows the overall limits, which can be specified by the user. The maximum flow-rate may be the instrumental limit, while the maximum pressure drop may have been set at 250 bar to allow reliable analyses to be performed. Together with the specified requirements for resolution and signal-to-noise ratio (Fig. 6h), these limits can be used by the user to express the analytical needs.

Immediately on entering or altering data in the system, a "results window" is displayed such as that shown as Fig. 6i. This table lists all valid columns with the optimum detector cell (see Fig. 6d) and time constant. Column 23 is predicted to yield the best results, but it requires the use of a smaller detector cell (and time constant) than the other columns. Based on the predicted values for the analysis time (retention time of the last peak), flow-rate, resolution, SNR and pressure drop, the user can decide on the optimum choice in practice, which may be column 23 or one of the other columns. However, if a user selects column 2 because of the higher resolution, it is better to ask for the optimum separation with the minimum resolution specified at a higher value. This higher value for the resolution may be obtained in a shorter analysis time with one of the other columns. This is an example of an edit–optimize–edit cycle, which is one of the attractive features of the present system.

Fig. 7 shows the bar charts that can be displayed for the top two columns in Fig. 6i. It is seen that for column 23 the required resolution is the limiting value, whereas for column 13 this is the maximum overall pressure drop. In the latter instance, the detection time constant and the detector-cell volume come close to being limiting factors. If either were to become too large, the system will automatically select better (smaller) values. Because short columns with small particles are used for the analysis, the SNR is not a limiting factor and it forms no ceiling underneath the roof of the bar charts in Fig. 7.

Fig. 8 shows simulated chromatograms, corresponding with the input chromatogram and the predicted optimum. These chromatograms were drawn using an

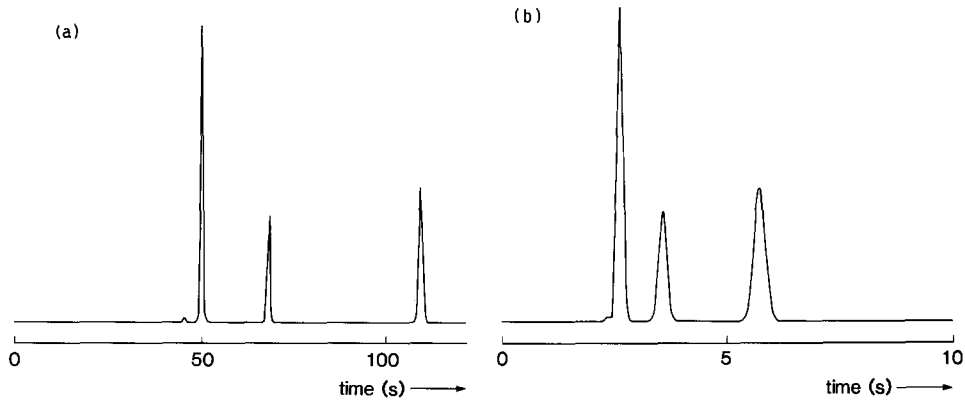

Fig. 8. Simulated chromatograms corresponding to the example described in Fig. 6. (a) Initial chromatogram[4]; (b) predicted optimum chromatogram [corresponding to Fig. 6i (top line) and Fig. 7a].

TABLE I

SUMMARY OF THE ADVANTAGES AND DISADVANTAGES OF CONSULTING THE PRESENT EXPERT SYSTEM AND A GENUINE EXPERT

Expert system	*Genuine expert*
Advantages:	*Advantages:*
Can be made available to many users	Vast areas of knowledge and experience
Always available (day and night)	May indicate the accuracy of the predictions
Patient explanation with every consultation	Should improve with age
Exactly reproducible	
Always considers all possibilities	*Disadvantages:*
Suggests the optimum conditions for each column	Few and far between
Allows the use of complex (resolution) equations	Not available when needed
	Gets tired of explaining
Disadvantages:	Not reproducible (performs best on rainy days)
Knowledge limited to a small domain	Tends to concentrate on one solution
Not aware of the approximations made	Suggests a possible (adequate) set of conditions
This system will never learn	Favours rough and rapid calculations

external Pascal program. At present the simulated chromatogram cannot be made visible directly within the expert system. It can be seen that a much faster chromatogram is predicted. However, because the conditions suggested by the system are within the possibilities and constraints specified by the user, it should be possible to achieve this kind of result in practice. Verifying the applicability of the present system in a laboratory environment, including a comparison of the predicted and experimental chromatograms, is one of the current tasks within our project.

DISCUSSION

In this section we discuss the benefits of the present expert system and compare its performance with that of the conventional method of optimization, *i.e.*, asking a real expert. A summary of the advantages and disadvantages of consulting the present expert system and (what is thought to be) a genuine expert is presented in Table I. We present this table not with the intention of insulting any experts (or systems). All negative characteristics of domain experts can be thought of as applying only to the first author of this paper.

It is not clear *a priori* who should be consulted for the quickest advice. The expert system can be used to address a new problem in about 15 min. Within this time the genuine expert may come up with an educated guess, but not with a computed prediction. It is likely that the expert system will come up with more complete and more consistent advice, having considered all possibilities within its knowledge. The expert may add to the advice much useful information that is not contained in the expert system. While the expert system may outperform the expert in a small area, it appears to be impossible at this stage to include all the knowledge and experience of even a single expert in a computer program.

The expert system can be consulted by many different people at any time, limited only by the availability of computers and software. This is a very attractive feature,

both for the users and for resident experts, who may be disturbed somewhat less. It is our experience that the present system can be of great benefit to a wide variety of users, including the genuine experts themselves. To the many dedicated chromatographers in small laboratories genuine experts may seem to be at endless distances, while expert systems will become increasingly close.

CONCLUSIONS

An expert system has been built for system optimization in isocratic liquid chromatography. The system predicts the optimum column, detector cell, time constant and operating conditions based on an initial chromatogram and limits and requirements set by the user.

The system described is not a conventional expert system but is a re-implementation of a conventional expert system in a programming language (Pascal) that is not normally associated with expert systems. The explain and help facilities described are unconventional and are unique to the present system.

The Pascal version of the system works very rapidly and conveniently in practice, showing a performance superior to the implementations as conventional expert systems in both Knowledge Craft and NExpert Object. Consulting the system for a new chromatogram may typically require 15 min. An edit–optimize–edit cycle in which one or a few parameters are altered can easily be performed within 1 min.

The optimization process is explained to the user in the form of a bar chart, from which the limiting factors can immediately be identified. Bar charts can be obtained for both valid and invalid combinations of a column, a detector (cell) and a time constant.

For each of the potentially limiting factors during the optimization, an additional "Help" file can be called upon from the bar chart. In principle, a similar "Help" function may be added in other parts of the system.

A simple example has been described, but the system has been applied to a number of different chromatograms of varying complexity. The accuracy of the predictions has yet to be evaluated in practice.

ACKNOWLEDGEMENTS

This research was supported by the European Commission as part of ESPRIT Project 1570 (ESCA). We acknowledge the assistance of Anita van Berkel in installing and maintaining the expert system at Philips Research, Eindhoven.

REFERENCES

1 P. J. Schoenmakers, N. Dunand, A. Cleland, G. Musch and T. Blaffert, *Chromatographia*, 26 (1988) 37.
2 D. Goulder, T. Blaffert, A. Blokland, L. Buydens, A. Chhabra, A. Cleland, N. Dunand, H. Hindriks, G. Kateman, H. van Leeuwen, D. Massart, M. Mulholland, G. Musch, P. Naish, A. Peeters, G. Postma, P. J. Schoenmakers, M. de Smet, B. Vandeginste and J. Vink, *Chromatographia*, 26 (1988) 237.
3 P. J. Schoenmakers, J. K. Strasters and A. Bartha, *J. Chromatogr.*, 458 (1988) 355.
4 K. W. Freebairn and J. H. Knox, *Chromatographia*, 19 (1984) 37.
5 M. Verzele, C. Dewaele and M. de Weerdt, *LC · GC Int., Mag. Chromatogr. Sci.*, 2 (1989) 10.
6 T. Takeuchi, D. Ishii and A. Nakanishi *J. Chromatogr.*, 285 (1984) 97.