Journal of Chromatography, 485 (1989) 267–281
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 22 029

EXPERT SYSTEM PROGRAM FOR ASSISTANCE IN HIGH-PERFORMANCE LIQUID CHROMATOGRAPHIC METHOD DEVELOPMENT

STEPHEN S. WILLIAMS*, JOSEPH F. KARNICKY, JEAN-LOUIS EXCOFFIER and SETH R. ABBOTT
Varian Associates, 2700 Mitchell Drive, Walnut Creek, CA 94598 (U.S.A.)

SUMMARY

A series of rule-based expert system programs which advise chromatographers during high-performance liquid chromatographic method development is described. The system is based on the VIS inference engine, and operates on both IBM-PC/AT compatible and VAX computers. The relationship of the programs to each other and to the chemist developing methods is described, and a coordinated example of program use is presented.

INTRODUCTION

The process of method development in liquid chromatography is often difficult and time consuming. This is due primarily to the large number of interdependent parameters which exist in the practice of high-performance liquid chromatography (HPLC) and the consequent requirement to study these parameters during method development through multiple chromatographic runs. Many advances in HPLC have been aimed towards loosening the method development bottleneck; examples are automated sampling[1] and data collection[2] systems, and decreases in time per analysis through faster chromatography[3]. These address the method development problem through increased efficiency, rather than by a change in method development processes. Computer programs may also be employed to aid the chromatographer by helping to extrapolate results through application of algorithmic models[4–7], thus reducing the number of experiments required. In contrast to these approaches, experienced chromatographers are able to produce working methods in a shorter time by making judgements which eliminate many paths toward unsuccessful results. Unfortunately, such expertise is often in short supply, while the range of expertise required to develop methods for todays complex samples is increasing.

A potential remedy to this problem is through the use of expert system programming techniques[8–20]. Expert system programs contain the general knowledge of experts over a limited domain, and have the capability of inferring factual information from a specific instance related to the domain. This is typically conceived as an inference engine (a general purpose logic processor), and a knowledge base (the expert knowledge in a formal framework)[8]. Expert systems are thus fundamentally

0021-9673/89/$03.50

different than procedural (conventional) computer programs, in which the knowledge and the procedures which operate on the knowledge are intermixed[9]. Expert system technology is applicable to diverse problems within the field of analytical chemistry[10,11], and in particular provides a way to capture conveniently the heuristic knowledge that expert chromatographers use in developing HPLC methods[12]. Expert systems have recently been applied to portions of the HPLC method development process, including mobile phase selection[13,14], detector selection[15,16] and optimization of selectivity[17,18]. Additionally, the advent of expert system programs has been associated with systematic reviews of method development processes[19,20] leading to increased formalization and understanding of method development.

ECAT (Expert Chromatographic Assistance Team) is a collection of expert system programs intended to assist the inexperienced chromatographer in HPLC method development much in the way that an expert chromatographer would. The user interacts with the program via a question and answer session. Inferencing takes place between each set of questions to formulate both conclusions and follow-up questions. The recommendations of the program are intended to lead the user logically through the entire method development process towards a workable HPLC separation.

Many changes have occurred to ECAT since an earlier version of the program was presented[12], and these are the main focus of this paper.

KNOWLEDGE ABSTRACTION RELATED TO HPLC METHOD DEVELOPMENT

Knowledge in general, and especially within the context of expert systems, implies that information within a domain is organized. As a domain grows, the level of organization within the domain must increase. Rule-based expert systems usually accomplish increased organization of knowledge through contexts attached to rules or sets of rules, and by breaking up the knowledge into smaller sub-domains.

The complexity of the method development process in HPLC requires that the knowledge in an HPLC expert system be very organized. ECAT models knowledge in HPLC by following both the logical grouping of subjects and the order within subjects that is used by chromatographic experts. This structure (described in greater detail below) was determined through informal interviewing of experts, and by accumulation of ideas and experience from chromatographers through the literature. As a consequence, the designers of the knowledge base structure were required to be knowledgeable in both chromatography and expert systems techniques, an idea somewhat at odds with reported descriptions of the "ideal" expert system project[9].

Most chromatographic experts start the method development process by collecting information about general sample chracteristics. From this, they use more specific information to develop a "separation", which is usually a consistent set of specifications for a mode (reversed-phase, normal-phase, ion-exchange, etc.), a column, a mobile phase and a detector. This is followed by experiments which refine or optimize the method. Concurrent with this is a consideration of sample pretreatment, which may be required to remove interferences or to increase the detectability of a particular analyte. A body of fundamental chemical information is always available to experts, and must consequently be available to an expert system.

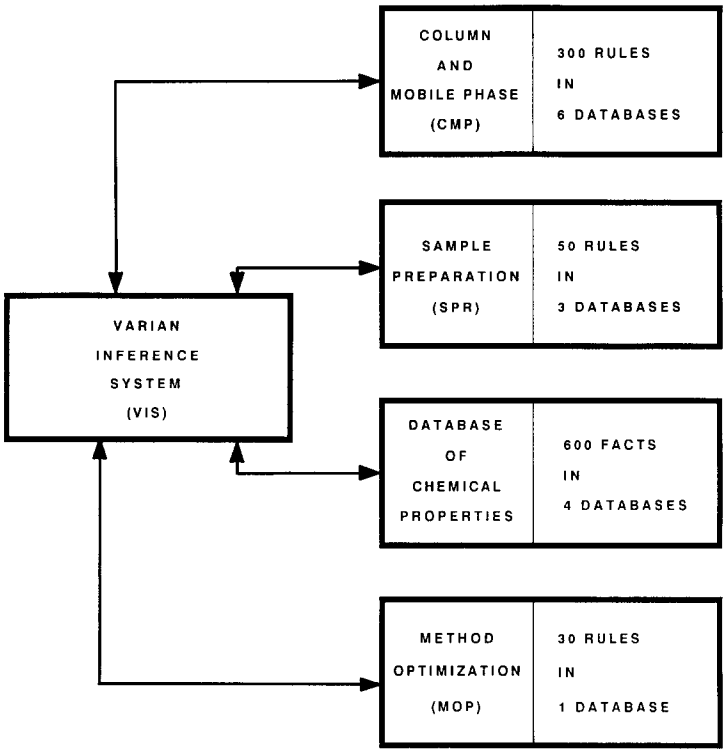

Fig. 1. Block diagram of ECAT expert system.

ECAT COMPONENTS

ECAT (Fig. 1) consists of four modules connected to an inference engine. Column and mobile phase (CMP) receives inputs about analyte characteristics, and makes recommendations for the column packing, column geometry, mobile phase liquids and mobile phase modifiers. Method optimization (MOP) leads the user through a series of experiments in which analyte mobility is determined and the separation is optimized in both total time and acceptable resolution. Sample preparation (SPR) helps the user to determine whether or not the sample requires pretreatment in order to enhance some quality of the separation. Available to the other modules is a module containing factual information about specific chemicals and classes of chemicals. Each of these modules is described in more detail in the sections below.

All of the modules include basic rules which describe the principles of chromatography within the domain. Rules which define and order the development process (procedural rules) are also added, allowing the expert system to proceed through the process as an expert would. Rules are grouped into sets related by contexts, and these sets are activated and deactivated under control of other rules. These features provides a powerful environment for abstracting the complex interactions in HPLC method development.

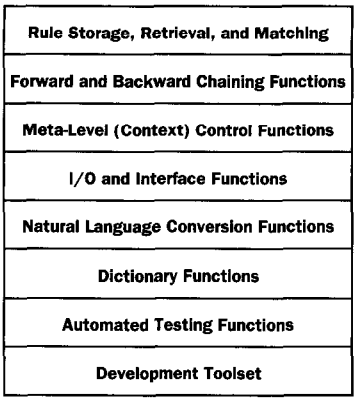

Fig. 2. VIS inference engine.

ECAT is written in Common LISP[a] and operates on VAX computers running DEC Common LISP under the VMS operating system and on IBM-PC/AT and similar computers using Gold Hill Common LISP.

Inference engine

The inference engine used in ECAT is the Varian Inference System, or VIS. Fig. 2 shows the major functional blocks in VIS.

VIS works with knowledge encoded as rules and facts. Facts are represented as propositions in predicate logic[b]. Rules have the form:

IF <LHS> THEN <RHS>,

where <LHS> and <RHS> are propositions combined by logical connectives, such as "AND", "OR" or "NOT". Elements of propositions (other than predicates) may contain variables instead of specific values in order to generalize the meaning of propositions in rules and facts; inferencing in such a system is merely the process of assigning specific values (not necessarily numerical) to variables in a chaining fashion through a symbol matching process.

Inferencing in VIS occurs by forward chaining (from data to conclusions) or

---

[a] LISP (list processor) is a programming language which used parenthesized lists for both procedural operations and storage of data, and makes heavy use of symbolic, as opposed to numerical, operations. These features make LISP the preferred languages for expert systems and other artificial intelligence applications.

[b] A proposition is a statement whose truth is assumed to be determinable through the application of propositional calculus. Predicate logic is a form for stating propositions; in VIS, the specific form is:

(PREDICATE SUBJECT [PROPERTY-1] [PROPERTY-2] . . .)

For example,
(father George-Washington our-country)
(has-pK_a benzoic-acid 4.2)

reverse chaining (from facts towards hypotheses which support the facts). Rules and facts in VIS are assigned to contexts, which may be enabled and disabled to allow higher structuring of the knowledge. A facility is also included for triggering of questions from the rules, so that an "intelligent" question and answer session is possible. Finally, provision is made for procedural attachment to predicates, which allows conventionally programmed routines to be used to resolve propositions. This allows a mixing of heuristic and procedural techniques.

While the previous paragraph described what is properly considered the "inference engine", VIS also contains additional functions to support expert system development and operation. Functions are included for general-purpose input and output between the user and the inferencing functions, a facility for translation of internal LISP propositions to English phrases and tools for maintenance of the knowledge base.

Column and mobile phase selection

CMP was the first ECAT module built, and is the largest, with approximately 300 rules. CMP is in many respects the key module in ECAT, as it is the starting point for development of a method. The chemical properties database was originally a part of CMP; although now a separate module, it is still used extensively by CMP.

CMP was developed through the use of specific probes, or trial separations for which valid separation methods were known. The first two probes (phenols and opium alkaloids, representing moderately polar–weakly acidic compounds and polar–heterocyclic basic compounds, respectively) lead to approximately 25 rules which describe the column and mobile phase selection process for these two compound classes. The third probe (an acid extract of urine) was not successfully predicted using these 25 rules; three additional rules (to differentiate between methods for strong and weak acids) were required to produce a correct CMP recommendation. In this way, each additional probe revealed inconsistencies in the rule base which were corrected by adding new rules or changing old ones. The automatic testing capabilities of VIS were used with each change in the rule base to re-evaluate all probes until every probe gave the "correct" answer. If a recommendation from CMP changed, but was thought to be reasonable, then the new recommendations were verified chromatographically. This probe–modify–test procedure builds a validation process into the rule-base development cycle.

Approximately 50 probes have been used in developing the 300 rules in the current rule base. These probes include separations which allow CMP to deal with many different kinds of LC separations; some examples are normal-phase, reversed-phase, ion-exchange and hydrophobic interaction modes; ionic, ion-pair and ion-suppressed submodes; and peptide and protein separations. The rules have been generalized as much as possible while being kept strictly correct, so that redundancy in the information contained within the rules is minimized.

CMP begins by requesting information about the analyte class (or classes) in the sample. VIS forward chains from this initial factual input in an attempt to develop a more general identity for the class, or to conclude the existence of properties associated with the class. If the compound class is known to a rule, or is recognized as being part of the database, then inferences may be drawn. As an example, myoglobin is known (from the data base) to be a protein (CMP differentiates proteins and peptides

TABLE I

USER INPUTS TO CMP

Chemical class
Functional class
Analyte amount
Molecular weight information
Analyte names
Acid–base characteristics
Isoelectric points (for proteins)
Polarity
Hydrophobicity

at 5000 daltons), proteins are (from a rule) separated using the reversed-phase mode, proteins require (from another rule) asking the user if maintenance of bioactivity is desired, etc. These newly inferenced facts will be available for subsequent inferencing throughout the CMP session. If the compound is unknown to CMP or the database, then its fundamental properties are determined by asking the user to provide the information directly. These properties forward-chain with other rules in the knowledge base to cause additional conclusions to be asserted, or to force additional queries of the user.

Table I lists typical questions asked in the course of a CMP session. If the user is unable to provide a definitive answer to any question (*i.e.*, answers "unknown"), then the recommendations are broadened to support the maximum possible number of the potential responses. When the chain of inferencing is exhausted, then the session is searched for the existence of key facts, which are then translated from internal LISP forms to produce a list of CMP recommendations.

While it is not the intent of this paper to describe the rule bases for the modules in detail, some general comments on their structure are appropriate. CMP is completely heuristic in nature, that is, no algorithmic procedures are used. CMP prefers the reversed-phase mode over others whenever possible. If significant acid–base character is present ($pK < 10$), the reversed-phase mode is divided into "submodes" of ion-suppressed, ion-pair and ionic, where the mode selected is dependent on the pK values of the strongest and weakest acid or base present in the analyte mixture. The ion-exchange mode is selected for small molecules (less than three carbons), or when it is requested by the user that bioactivity of proteins be sustained. The ion-exchange mode has "submodes" cation exchange and anion exchange, selected as appropriate. Currently, normal-phase and hydrophobic interaction modes are never selected by the basic design rules, but may be selected by the user in the redesign process described below. Although only seven rules effect selection of a separation mode, the mode in turn may influence approximately 90 other rules for selecting column and mobile phase parameters. Column selection is done partly in parallel, and partly in response to selection of a mode. Columns are specified by phase and substrate characteristics. A C_{18} bonded phase is selected for most reversed-phase separations; exceptions are a cyano phase for small molecules and C_4 for membrane proteins (without bioactivity). Strong cation- and anion-exchange phases are used for the ion-exchange mode. C_4 is used for most hydrophobic interaction separations, and the normal-

phase mode causes "no phase" to be explicitly selected. Other phases may be selected by rules triggered for very specific cases such as chiral separations.

Silica is typically selected as the substrate; exceptions are use of an organic gel for maintenance of bioactivity in proteins, or the use of low-trace-metal silica where analytes complex with metals and might cause peak tailing. Substrate particle size is restricted according to the amount of the smallest analyte, with smaller particles required for smaller amounts of analytes.

Solvents and additives making up the mobile phase are selected primarily by mode but also are influenced by the column selected. Mobile phase is specified as liquids and additives with specific concentrations. Water–acetonitrile is selected for most reversed-phase separations, with water–propanol used for proteins. The normal-phase mode recommends hexane–chloroform for mixtures with non-polar analytes, chloroform–methanol for mixtures containing all polar compounds and hexane–chloroform–methanol for mixtures with a wide polarity range. Additives are typically specified after most other separation parameters have been determined, and the rules are therefore fairly specific and complex in nature. Additives include buffers, salts and competing acids or bases.

CMP has the additional capability of allowing the selective redesign of the separation by the user. When a rule in CMP must select between two fundamental alternatives, both of which may lead to successful CMP recommendations, then the rule asserts one of the choices as preferred over the other, and the preferred choice is used to continue the inferencing process. These preferences are listed as part of the output of CMP, and the user is permitted to select the alternative (or less preferred) choice. CMP will then redesign the separation based on the alternative choice, perhaps requesting new information from the user in the process. An example of this occurs in selection of the separation mode; there is a rule in the CMP knowledge base which states a preference for reversed-phase over normal-phase separations. If the user is dissatisfied with the CMP conclusions for a reverse phase, or merely wishes to see normal-phase recommendations, then the redesign option can be selected. Of course, there are other rules in the rule base which may disallow either of these modes, or assert other modes for a particular separation.

Method optimization

Optimization procedures for LC can be categorized as either algorithmic or empirical[4]. Some examples of algorithmic methods are retention approximations, factorial design experiments, simplex approximations and chromatographic response functions. Empirical techniques include use of solvent triangles and resolution maps. This categorization scheme has some overlap, as indicated through detailed consideration of the above examples, but it is still useful within the context of expert systems. VIS is capable of encoding the knowledge of the empirical methods, while at the same time allowing implementation of algorithmic methods through procedural attachment. This is important in dealing with recent advances[21] where different algorithmic optimization methods have been related to each other through use of heuristic descriptions.

While the long-term goal of the method optimization module (MOP) is intended to assist the chromatographer in implementing a variety of optimization strategies for HPLC, it is currently limited to a small set of rules which lead the user

through a session using optimization strategies similar to those developed by Dolan *et al.*[7] and Drouen *et al.*[6], based on the linear relationship of solvent strengths. Both of these are algorithmic models for HPLC retention during gradient elution, with which one can predict isocratic retention based on one or more experimental runs.

MOP starts by requesting information about the separation method (*i.e.*, the method suggested by CMP). From these, MOP suggests a set of parameters for making an initial gradient run. When this is complete, results (retention times of specific peaks) are entered into MOP, and applied to the one-gradient model of Drouen *et al.*[6]. This results in a recommendation for the proportion of strong to weak solvent for several solvent pairs to be used with isocratic separations. The user is required to produce these isocratic chromatograms and interpret their quality; if the separation is deemed inadequate, then MOP can resort to the two gradient model of Dolan *et al.*[7], in which the increased data should give superior isocratic recommendations. Should this fail, tertiary mobile phase mixtures can be tried using extensions of the linear solvent strength model. These require more experiments and application of algorithmic methodology; MOP works with the user to accomplish this as described above. In a more complete MOP and ECAT, a failure to obtain adequate optimization at each step would result in heuristic re-evaluation of the current optimization path, resulting in either continuance of the current path, implementation of alternate optimization strategies or various levels of redesign of the separation method.

Sample preparation

Real analyses by HPLC often require significant amounts of sample preparation before chromatography can proceed[1]. This may involve removal or masking of interferences, or modification of components of interest so that they can be separated from interferences. Sample modification may also be desirable to increase the detectability of some sample components. Although sample preparation is the first step carried out in a routine analysis, it is usually considered in the later stages of method development. In practice, development and understanding of the chromatographic method and its limitations precede and lead to the appropriate sample preparation method.

The Sample Preparation module (SPR) makes recommendations for sample pretreatment by using information from the user, or alternatively from the CMP or MOP modules. The knowledge base in SPR contains sets of rules and facts for determining whether a guard column is needed, and whether the analysis can be aided by solid-phase extraction techniques. Techniques such as on-column concentration are also considered in SPR (rather than CMP) because of the their interaction with other sample preparation techniques. The user interface obtains information from the user via queries; if the user is unable to provide an answer, the knowledge base leads to a progression of queries for more fundamental information. When required, the SPR will suggest specific experiments to be performed which will provide key pieces of information. The user interface also delivers a list of recommendations at the conclusion of the session.

SPR determines the requirement for a guard column based on the characteristics of the sample matrix; if the matrix contains particulate matter or irreversibly adsorbing materials, then a guard column is recommended. The guard column packing is selected based on the column selected by CMP, with the requirement that the

1. Please enter the chemical class of the analyte: *PHTHALATES*
2. CMP cannot find information about a chemical class of type PHTHALATES.
   Enter GOODLIST for a list of CHEMICAL-CLASS's that CMP can understand,
   or press enter to describe PHTHALATES:
   <enter>

   WORKING ...

3. Enter the smallest analyte amount (in ng) of the least concentrated component
   of interest: *10*
4. Does the analyte have a MW greater than 5000? *no*
5. Does the analyte have less than 4 carbons? *no*

   WORKING ...

6. Enter the name of a specific analyte (c/r to terminate): *unknown*

   WORKING ...

7. Are PHTHALATES acidic, basic, neutral, or both? *neutral*

   WORKING ...

Fig. 3. Question and answer session for CMP analysis of phthalates.

guard column should not selectively retain the sample components. On-column concentration is recommended for reversed-phase and ion-exchange separations when less than 10 ng of a component of interest are present.

Rules for Solid-Phase Extraction (SPE) are in the early stages of preparation. Currently, SPE is recommended when a guard column fails adequately to clean up the

this system is *21FEB89*

I. User entries

THE ANALYTE CHEMICAL CLASS IS PHTHALATES
THE SMALLEST ANALYTE AMOUNT IS 10 NG
THERE ARE NO MOLECULAR WEIGHTS ABOVE 5000 DALTONS
THERE ARE NO ANALYTES WITH LESS THAN 3 CARBONS
THE USER CANNOT DETERMINE THE SPECIFIC ANALYTE
PHTHALATES ARE NEUTRAL

***** CMP RECOMMENDATIONS ******

II. LC method - Analytical column

CHOOSE REVERSE PHASE AS THE METHOD OF SEPARATION
THE DIAMETER OF THE PARTICLES SHOULD BE BETWEEN 3 AND 5 MICRONS
SELECT SILICA AS SUBSTRATE FOR THE ANALYTICAL COLUMN
CHOOSE C18 AS BONDED PHASE FOR THE ANALYTICAL COLUMN

III. Mobile phase

Reservoir A

CHOOSE WATER AS A LIQUID IN RESERVOIR A

Reservoir B

CHOOSE ACETONITRILE AS A LIQUID IN RESERVOIR B

**** ASSUMPTIONS for the previous recommendations *****

IV. Assumptions

THE PROGRAM PREFERS TO CHOOSE REVERSE PHASE AS THE METHOD OF
SEPARATION RATHER THAN TO CHOOSE NORMAL PHASE AS THE METHOD
OF SEPARATION

THE PROGRAM PREFERS TO CHOOSE ACETONITRILE AS A LIQUID IN
RESERVOIR B RATHER THAN TO CHOOSE METHANOL AS A LIQUID IN
RESERVOIR B

Available choices for your selection:

1. RUN	2. REDESIGN	3. DESCRIBE-CMP
4. REPORT	5. REVIEW	6. INPUT
7. COLUMN	8. MOBILE-PHASE	9. GUARD-COLUMN
10. ASSUMPTIONS		

Fig. 4. CMP recommendations for analysis of phthalates.

sample mixture. The SPE cartridges type is selected based on the column recommended by CMP; the rules are intended to select a cartridge with characteristics similar to but slightly less retentive than those of the analytical column, for example of C_8 cartridge for C_{18} columns. Solvents for SPE activation and washing are taken from CMP mobile-phase recommendations for strong and weak solvents, respectively, whereas the SPE elution solvent is assumed to be the same as the mobile phase solvent suggested by MOP (so that on-line elution is possible). The current rule base represents the first stages of developing SPE methods, but requires the addition of rules which specify the later stages involving experimental refinement of the parameters.

ECAT EXAMPLE: SEPARATION OF A MIXTURE OF PHTHALATES

The use of ECAT is illustrated through an analysis of a waste-water discharge containing phthalates. Although this is a simple example, and does not extensively tax the capabilities of the modules, it serves to illustrate user interactions with and the kind of information produced by the programs.

The sample was first probed through a CMP session. Figs. 3 and 4 illustrate the question and answer session and the recommendations from CMP[a]. The first query made by CMP (Fig. 3) requests the name of a chemical class for the analyte. The user response of *phthalates* is searched in the existing knowledge base for occurrences of the same word in association with a rule concerned with chemical class. Query 2 in Fig. 3 indicates that phthalates is not a known chemical class, and allows the user either to continue the session or to see a list of known chemical classes. This is required in order that the user be given the opportunity to specify the class exactly as it is entered in the knowledge base[b]. The CMP (or more precisely the VIS) response WORKING indicates that the forward chaining process is occurring; this may cause a delay of 1–20 s, depending on the number and complexity of the inferences drawn. Because there is no information available for phthalates, CMP requests information about phthalate properties in queries 3–7. Query 3 requests the smallest analyte amount, for use in determining column geometry and detector restrictions. Queries 4 and 5 ask for general molecular weight information, and categorize phthalates as being other than very small or very large molecules. Query 6 requests information about specific analytes, for which there is none available, and query 7 asks about the acid–base characteristics of the compound class. None of the rules fired in the inference chain following query 7 produce an additional query, so inferencing ends and the CMP recommendations follow. It is important to understand that all of the CMP queries after the initial question are triggered by rules in the knowledge base, and that

[a] All of the Figures showing computer screens of ECAT sessions are representations of the material appearing on the screen. The only differences are the number of characters allowed on a line, the use of italics to differentiate user responses from computer responses, and the addition of index numbers to some Figures for easy reference in the text of the paper.

[b] VIS recognizes user input by matching the typed form of the input with the internally held form. Thus, whereas a human would easily recognize the misspelling of the word "phthalates", or account for the synonym "1,2-benzenedicarboxylic acid esters", the computer relies on a specific response. The goodlist feature of ECAT is an attempt to warn the user of a possible inconsistency of symbol usage, and allow the selection of proper spelling or synonym.

the number and types of follow-up queries may vary as the responses to preceding queries change.

Section 1 in Fig. 4 shows the first portion of the CMP output, which is a list of user responses received by CMP. This is included primarily for allowing convenient documentation of the session. Section 2 lists specifications for the analytical column. Each of these specifications is a translation of an internal proposition resulting from the firing of a specific rule. The result is three separate specifications for different properties of the column packing, and a specification for the mode of separation. Section three lists the recommendations for the strong and weak solvents used for this analysis. These are the simplest recommendations possible for reversed-phase separations; if the analyte had been described as non-neutral, then additional queries would have asked the user for more specific information describing those properties. This would have lead to the inclusion of recommendations for modifiers to the mobile phase, such as buffers, competing acids or bases, etc. Note that no information concerning solvent proportioning for the mobile phase is included; this is deferred to the MOP module. Section 4 lists key assumptions used by the rules in formulation the recommendations of sections 1–3. These can be changed by selection of choice 2 (redesign) from the list of options presented by CMP at the end of the session. As an example, disallowing the reversed-phase mode as the separation mode for phthalates leads to a normal-phase separation using hexane–chloroform on a silica substrate.

Before the analyst can make use of the CMP recommendations, a decision must be made concerning the composition with time of the mobile phase. The MOP session shown in Figs. 5–8 illustrates the use of retention approximations for determining this information. The first portion of the MOP session (Fig. 5) requests information about the type of detector used, the mode of the separation and the identity of specific analytes. From this, MOP recommends (1) the use of spectra from the diode-array spectrometer for tracking individual analytes in experimental runs, (2) the preferred use of a one gradient estimation method (*i.e.*, the optimization method of Drouen *et al.*[6]) and (3) the gradient parameters to be used for that method, together with a request to return to this point for further consultation.

Fig. 6 shows the results of the gradient chromatography and the peak identifications resulting from purity parameter[22] evaluations of the spectra *versus* a library including phthalate compounds. The chromatogram provides the data needed by MOP in the next portion of the session (Fig. 7). This results in a more specific set of isocratic mixture recommendations based on the Drouen *et al.*[6] model. Fig. 8a and b show the results of both the acetonitrile and methanol isocratic separations; both are seen to be of good quality, and MOP can be terminated at this point.

```
1.  Enter the type of detector being used: diode-array
2.  Enter the mode of separation: reverse-phase
3.  Enter the name of a specific analyte: unknown

WORKING ...

Inferences:

    SPECTRA SHOULD BE USED TO TRACK INDIVIDUAL ANALYTES

    THE PROGRAM PREFERS A ONE GRADIENT ESTIMATION TO A
    TWO GRADIENT ESTIMATION

    RUN A 0-100 % METHANOL GRADIENT OVER 20 MINUTES. RETURN
    TO THIS SESSION WHEN CHROMATOGRAPHIC RESULTS ARE
    AVAILABLE.
```

Fig. 5. Initial MOP session for optimization of phthalate separation.

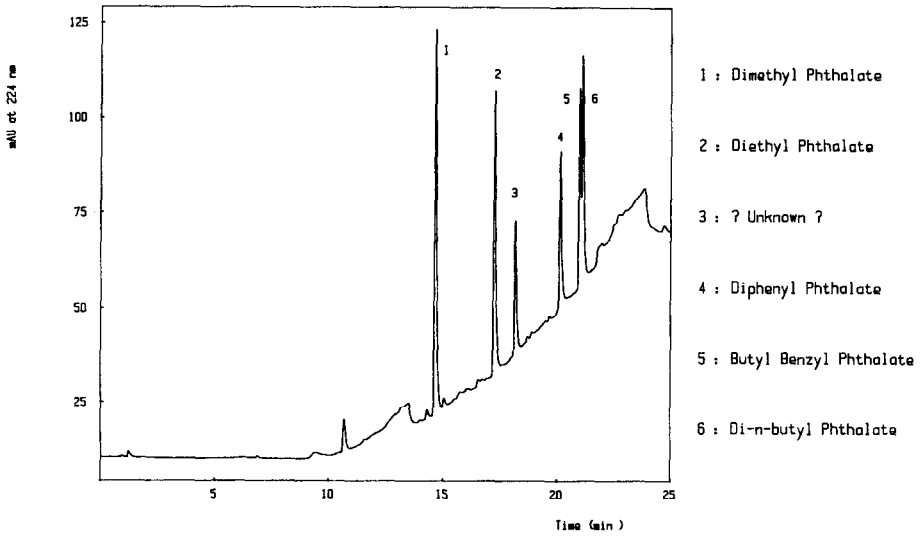

Fig. 6. Separation of phthalates according to the conditions suggested by MOP. Methanol–water gradient from (0:100) to (100:0) in 25 min.

The above examples made use of standards in aqueous solution, and thus required no sample preparation. Real samples are typically present in a sample matrix, and require pretreatment. Fig. 9 shows the initial SPR session for a real sample containing phthalates. SPR requests information concerning the sample matrix and the analytical method, and produces a set of recommendations for on-column concentration and guard columns. Questions 1–5 illustrate the triggering of subsequent questions when an entry is made which is unknown to SPR. SPR asked questions 3, 4 and 5 to determine the fundamental characteristics of the sample matrix; the questions shown are sufficient (with the present rule base) to describe the sample matrix. Questions 9–11 also illustrate an unrecognized entry, but this time the list of known modes was used for selection of a response. Note that all of the questions about the analytical method (numbers 6–11) could be removed if the information was passed internally from the CMP session. The last conclusion of the initial session recommends that the analytical method be run using a set of sample preparation techniques, and that the analyst return to this point in the session if further consultation is needed.

1. What is the retention time (in minutes) of the first peak? *14.82*
2. What is the retention time (in minutes) of the last peak? *21.57*
3. What is the void time (in minutes) of the column? *1.84*

WORKING...

Inferences:

75% MEOH / 25% WATER IS AN APPROPRIATE ISOCRATIC MIXTURE
61% ACN / 39% WATER IS AN APPROPRIATE ISOCRATIC MIXTURE
49%THF / 51% WATER IS AN APPROPRIATE ISOCRATIC MIXTURE

RUN AN ANALYSIS USING THE RECOMMENDATIONS ABOVE. RETURN TO THIS MOP SESSION IF YOU HAVE DETERMINED THAT THE SEPARATION REQUIRES FURTHER OPTIMIZATION.

Fig. 7. Follow-up MOP session for optimization of phthalate separation.

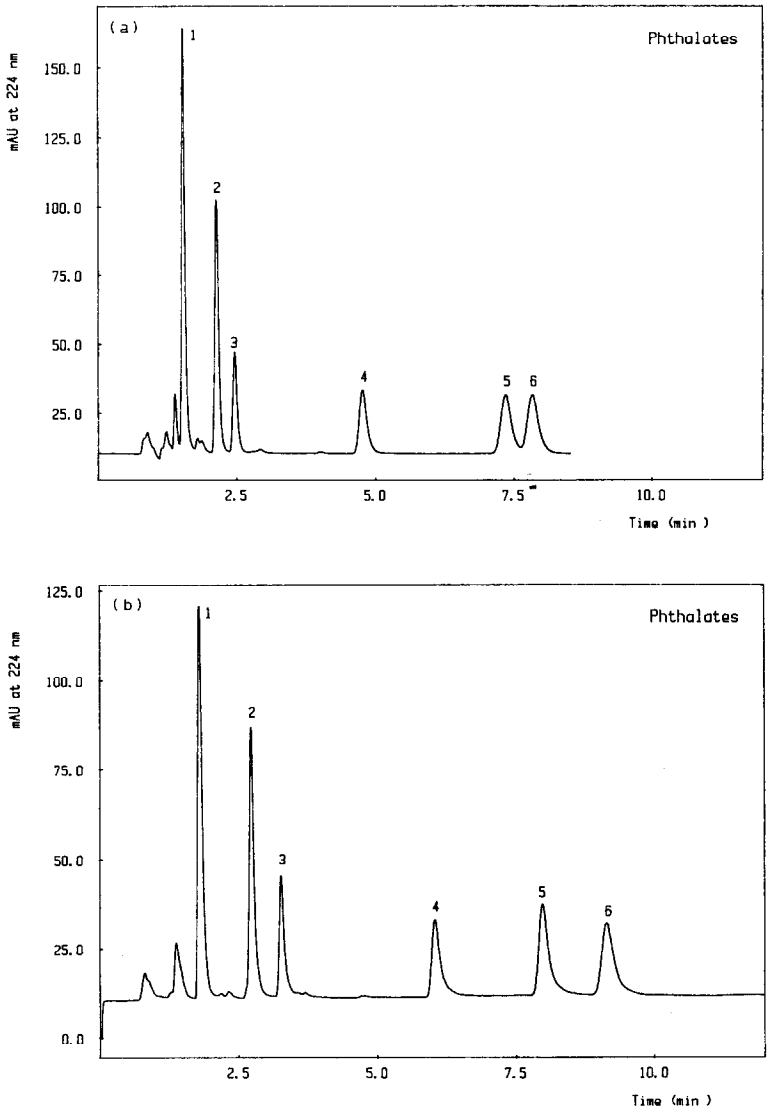

Fig. 8. Isocratic separations of phthalates suggested by MOP. (a) Methanol–water (75:25); (b) acetonitrile–water (61:39). Peak identification as in Fig. 6.

If the results are deemed by the user to be adequate, then the session may end here. Otherwise, a follow-up session (Fig. 10) may be conducted to find alternative sample preparation methods. Currently, the alternatives are limited to solid-phase extraction, and for this technique only a preliminary method. Question 1 allows the user to exert his or her expertise in analyzing the quality of the results of the preliminary experiments; if the response to the question had been "unknown", then a series of (currently simple) questions would be asked to help inexperienced users match

```
1.  Please enter a sample matrix: Industrial waste water
2.  SPR cannot find information about a sample-matrix of type INDUSTRIAL-
    WASTE-WATER. Enter GOODLIST for a list of SAMPLE-MATRIX's that
    SPR can understand, or press enter to describe INDUSTRIAL-WASTE-
    WATER: <enter>
3.  Does the sample matrix contain particulates? yes
4.  Does the sample matrix contain irreversible adsorbers? yes
5.  Is the sample matrix water-based? yes

    WORKING...

6.  Does the analytical column have a bonded phase? yes

    WORKING...

7.  Enter the type of bonded phase for the analytical column: C18 bonded phase

    WORKING...

8.  Enter the approximate amount (in ng) of the least concentrated component of
    interest: 10

    WORKING...

9.  Is any analyte of interest moderately or strongly polar? no

    WORKING...

10. Enter the mode of the separation: reversed phase
11. SPR cannot find information about a SEPARATION-MODE of type
    REVERSED-PHASE. Enter GOODLIST for a list of SEPARATION-MODE's that
    SPR can understand, or press ENTER to describe SEPARATION-MODE: goodlist
12. Select one of the following:

1. CHIRAL-REVERSE-PHASE   2. CHIRAL-NORMAL-PHASE   3. HYDROPHIBIC-INTERACTION
4. ION-EXCHANGE           5. NORMAL-PHASE          6. REVERSE-PHASE
   [d]escribe

    Selection: 6

    WORKING...

INFERENCES:
    USE OF A GUARD COLUMN IS RECOMMENDED
    C18 IS AN APPROPRIATE BONDED PHASE FOR THE GUARD COLUMN
    THE PACKING SUPPORT FOR THE GUARD COLUMN SHOULD BE LOW
      CAPACITY
    THE TECHNIQUE OF ON-COLUMN CONCENTRATION IS RECOMMENDED
    RUN AN ANALYSIS USING ANY RECOMMENDATIONS LISTED ABOVE.
    RETURN TO THIS SPR SESSION IF YOU HAVE DETERMINED THAT THE
      SAMPLE REQUIRES FURTHER PRETREATMENT, OR IF YOU NEED
      ASSISTANCE IN MAKING THIS DETERMINATION.
```

Fig. 9. Initial SPR session for pretreatment of phthalate sample.

analysis requirements with the current state of quality. The other questions request information about the analytical method. SPR then returns recommendations for basic parameters for SPE, specifically for the cartridge and solvent combinations to try. Future expansion of the SPE rule base will apply the optimization techniques of MOP to an iterative development of a complete SPE method.

```
1.  Does the sample require further pretreatment? yes
2.  What is the strong solvent of the analytical mobile phase? acetonitrile
3.  What is the ratio of strong to weak solvent in the mobile phase? 55/45
4.  Are there additives to the strong solvent of the analytical mobile phase? no

Inferences:

    SOLID PHASE EXTRACTION IS SUGGESTED AS A METHOD FOR SAMPLE
      PRETREATMENT
    C8 IS PREFERRED OVER C2 AS THE TYPE OF SPE CARTRIDGE
    ACETONITRILE IS SUGGESTED AS THE ACTIVATION SOLVENT FOR
      SPE
    WATER IS SUGGESTED FOR WASHING THE ACTIVATION SOLVENT
    WATER IS SUGGESTED AS THE WASH SOLVENT
    55% WATER / 45% ACETONITRILE IS SUGGESTED AS THE ELUTION
      SOLVENT
    THE PREVIOUS RECOMMENDATION OF GUARD COLUMN IS DISALLOWED
    THE PREVIOUS RECOMMENDATION OF ON-COLUMN CONCENTRATION IS
      DISALLOWED
```

Fig. 10. Follow-up SPR session for pretreatment of phthalate sample.

CONCLUSIONS

Many of the advances which have occurred in LC have come at the expense of increased complexity. This has been manifested in many ways; some examples are more intricate instrumentation, detectors which produce multi-dimensional data sets (diode-array) detectors, mass spectrometers and novel chemistry before, during and after the chromatographic process. Computers are used as a tool to ease the impact of this complexity, so that an overall increase in effectiveness of LC analysis takes place. The ECAT programs demonstrate a non-traditional approach to offering computer assistance for the chromatographer. They encode heuristic knowledge in a general way which, unlike most conventional computer programs, allow the programs to be applicable to situations unanticipated at the time the knowledge was coded.

Expert systems were developed originally to permit programming of non-algorithmic knowledge. However, real chromatographers make judgments through processes which interrelate data from experiments, fixed mathematical methods and general experience. Expert systems applied to chromatographic method development must take this into account, and be able to apply appropriate computer techniques to individual portions of the development. This philosophy was used in developing the ECAT programs, and has resulted in a general computing environment which is capable of meeting these needs. It can be anticipated that the expert system programming techniques demonstrated here will continue to evolve and become a part of the computer-integrated laboratory.

REFERENCES

1 B. L. Tippins, *Am. Lab.*, 19 (1987) 8.
2 J. L. Glajch, *LC · GC, Mag. Liq. Gas Chromatogr.*, 6 (1988) 120.
3 M. Verzele and C. DeWaele, in F. Bruner (Editor), *The Science of Chromatography (Journal of Chromatography Library*, Vol. 32), Elsevier, Amsterdam, 1985, p. 435.
4 J. C. Berridge, *Chemometr. Intell. Lab. Syst.*, 3 (1988) 175.
5 P. J. Schoenmakers, *Optimization of Chromatographic Selectivity. A Guide to Method Development (Journal of Chromatography Library*, Vol. 35), Elsevier, Amsterdam, 1986.
6 A. C. Drouen, P. J. Schoenmakers, H. A. H. Billiet and L. de Galan, *Chromatographia*, 16 (1982) 48.
7 J. W. Dolan, J. R. Gant and L. R. Snyder, *J. Chromatogr.*, 169 (1979) 31.
8 A. Barr and E. A. Feigenbaum (Editors), *Handbook of Artificial Intelligence*, Vol. 1, Heuristech Press, Stanford, CA, 1983, pp. 143–216.
9 P. Harmon, R. Maus and W. Morrisey, *Expert Systems Tools and Applications*, Wiley, New York, 1988.
10 R.E. Dessy, *Anal. Chem.*, 56 (1984) 1200A.
11 A. R. de Monchey, A. R. Forester, J. R. Arretteig, L. Le and S. N. Deming, *Anal. Chem.*, 60 (1988) 1355A.
12 R. Bach, J. Karnicky and S. Abbott, in T. H. Pierce and B. A. Hohne (Editors), *Artificial Intelligence Applications in Chemistry (ACS Symposium Series*, No. 306), American Chemical Society, Washington, DC, 1986, p. 278.
13 M. A. Tischler and E. A. Fox, *Comput. Chem.*, 11 (1987) 235.
14 M. De Smet, A. Peeters, L. Buydens and D. L. Massart, *J. Chromatogr.*, 457 (1988) 25.
15 G. Musch, M. De Smet and D. L. Massart, *J. Chromatogr.*, 348 (1985) 97.
16 G. Musch and D. L. Massart, *J. Chromatogr.*, 370 (1986) 1.
17 T. P. Bridge, M. H. Williams, G. G. R. Williams and A. F. Fell, *Chromatographia*, 24 (1987) 691.
18 A. Peeters, L. Buydens, D. L. Massart and P. J. Schoenmakers, *Chromatographia*, 26 (1988) 101.
19 P. J. Schoenmakers and M. Mulholland, *Chromatographia*, 25 (1988) 737.
20 A. L. Ananda, S. M. Foo and H. Gunasingham, *J. Chem. Inf. Comput. Sci.*, 28 (1988) 82.
21 A. Barthal, G. Vigh, G. K. C. Low, H. A. H. Billiet and L. de Galan, *1989 Pittsburgh Conference and exposition, Atlanta, GA, March 6–10, 1989*, paper 588.
22 T. Alfredson, T. Sheehan, T. Lenert, S. Aamodt and L. Correia, *J. Chromatogr.*, 385 (1987) 213.