Journal of Chromatography, 485 (1989) 237–253
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 21 939

EXPERT SYSTEM FOR THE SELECTION OF HIGH-PERFORMANCE LIQUID CHROMATOGRAPHIC METHODS FOR THE ANALYSIS OF DRUGS

MARINA DE SMET, GREET MUSCH, ANNE PEETERS, LUTGARDE BUYDENS and DESIRE L. MASSART*

Pharmaceutical Institute, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Brussels (Belgium)

SUMMARY

An expert system for the selection of high-performance liquid chromatographic methods is described for the label claim analysis of drugs in pharmaceutical formulations. The system contains knowledge for the selection of a suitable detection mode (UV detection or electrochemical detection in the oxidative mode), an appropriate chromatographic mode (reversed-phase with water, reversed-phase with buffers or normal phase) and the starting mobile phase compositions in each chromatographic mode. The chromatography is performed on a single type of column, namely a cyanopropyl column. The implementation of the knowledge in the commercially available expert system shell (KES) is also described. As a knowledge representation method, production rules are used.

INTRODUCTION

In the last few years, several workers have proposed the introduction of expert system (ES) technology for the selection of analytical methods. An overview of existing or developing expert systems in analytical chemistry was given by Bridge *et al.*[1]. In this domain, some more general method selection systems have been developed. Settle and Pleva[2] developed an expert system for selecting methods of chemical analysis in general. We have described an expert system that derives a procedure for the determination of pharmaceutically active substances in tablets by UV spectrophotometry[3,4].

The area in which there has been the most activity with expert systems is in chromatography. Dolan and Snyder[5] developed software, called DryLab, which can be described as modular expert systems. By carrying out computer simulations, it helps the user to select a suitable solvent strength, solvent selectivity, pH and buffer concentrations for the mobile phase. All these software programs were then combined into an global expert system, called LCHELP[6]. ECAT is an expert system which is composed of four modules[7]: sample preparation, column and mobile phase selection, column diagnostics and method optimization. It gives general information about method development in high-performance liquid chromatography (HPLC).

0021-9673/89/$03.50

Hodges *et al.*[8] described an expert system, called PRO-DIGEST-LC, for predicting the separation of peptide mixtures. It gives very detailed advice about the chromatographic method to be applied for a certain peptide separation. For instance, it selects the appropriate chromatographic mode for peptide analysis, *i.e.*, size exclusion, cation-exchange or reversed-phase chromatography, the column type, the mobile phase, the flow-rate, the gradient rate and the sample size.

This paper describes an expert system, called LABEL, using production rules as a knowledge representation method. As with most other systems, it is designed to contain the knowledge of the authors and, more specifically, the knowledge and experience developed in our laboratory for the selection of HPLC methods for drug analysis. Depending on the components to be determined, the ES gives advice about the detector, the type of mobile phase and the starting mobile phase composition to use. The stationary phase used is always a cyanopropyl column. At present, the system is restricted to label claim analysis, *i.e.*, the verification of the contents as given on the manufacturer's label for the product. It is our intention, however, to expand it into other areas, particularly into biopharmaceutical analysis.

As we apply HPLC to many very different drugs, our first approach for method selection has been to develop a strategy. This permits the reduction of the large number of possible choices to a smaller number. We have, for instance, shown that nearly all drug analyses can be carried out with a single stationary phase, the cyanopropyl (CN) bonded phase[9]. An overview of the complete strategy followed in our laboratory in recent years is given in ref. 10.

In this paper, it is also explained how this knowledge is incorporated in an expert system shell, called KES.

EXPERIMENTAL

Software and hardware

KES [Knowledge Engineering System, Release 2.3. © 1986, Software Architecture & Engineering] is a software tool for building expert systems. It is written in C and runs on, among others, Apollo, Vax and IBM/AT computers. Apollo was used at the development stage and the ES thus obtained was then translated to the IBM/AT. KES is a structured rule-based tool. The tool uses essentially backward chaining. In the actual release of KES (release 2.5), forward chaining is introduced although it was not present in release 2.3 in which the described expert system is built. Production rules are used to represent the knowledge.

External programs are written in Fortran (Fortran 77) when the ES runs on Apollo and in Basic [version A3.10, © IBM, 1981, 1985] for the version running on IBM. The database program used is Dbase 3.

KES can execute an external program by two mechanisms: the first is when a command is given to execute the program in the expert system and the second is when it needs information to continue its inferencing and when a predefined external program provides this required information. After consultation of the external program, the expert system consultation continues.

The KES knowledge base itself can also be integrated into other software written in C. A knowledge base used in this environment is said to be embedded within the program. A KES expert system which is embedded becomes part of a single

executable C program. KES supplies a set of run time functions and data types that are used to control the expert system and to send, receive and manipulate data from a parsed KES knowledge base, although this feature of embedding a KES expert system is not used in the work described in this paper.

RESULTS AND DISCUSSION

Description of the chromatographic knowledge incorporated in the ES

The first decision made by the ES concerns the detector. At present it considers only UV and electrochemical detection (ED) in the oxidation mode. The user is asked whether he has already decided on detection. If not, it is the ES's task to do so. Because of the user friendliness and wider range of application of UV detection, this is considered to be the preferred method. When only a single compound is present, the system first checks whether, given the amount injected on-column, the molar absorption coefficient and the path length of the detector (a 254 nm fixed-wavelength detector) can be used. It is calculated with the following equation:

$$\text{conc.} \cdot \text{inj. vol.} \geqslant 4 \cdot \text{MW} / pl254 \cdot E254 \quad (1)$$

where conc. = the concentration of the injected solution, inj. vol. = the injection volume, MW = molecular weight, $pl254$ = the path length of the detector at 254 nm and $E254$ = molar absorption coefficient at 254 nm of the solute being analysed. All these parameters have to be given by the user as input data.

When the condition stated in eqn. 1 is not fulfilled, the ES investigates if detection at λ_{max} of the component with a programmable UV detector yields a solution and, finally, if this fails the system investigates detection at 220 nm, which is considered to be the cut-off wavelength, at least in the reversed-phase (RP) mode [235 nm in the normal-phase (NP) mode]. In each instance eqn. 1 is used by taking the appropriate molar absorption coefficient at the wavelength used into account. It is also asked if the path length of the variable-wavelength detector is the same as that of the detector used at 254 nm.

When this, too, is not possible, ED in the oxidation mode is investigated. The ES proposes a list of oxidizable functions[11] and, if the user answers that one of these is present, it considers the limit of detection. It is, of course, impossible to predict accurately the limits of detection for a substance without experimental data. However, in earlier work it was found that the median gain in detection limit on going from UV detection to ED is a factor of 20 (ref. 12). The ES uses this to state that it is probable or not probable that ED will help.

When there is more than one compound present, the ES considers the following alternatives (see Fig. 1):

(i) no detection method available for all the substances considered;

(ii) UV detection for all the substances [at 254 nm, at λ_{max} of the respective substances with a programmable detector, at 220 nm (235 nm in NP) for all substances);

(iii) ED for all substances; or

(iv) ED and UV detection in series.

Again, UV detection is the preferred method and one first investigates whether this is

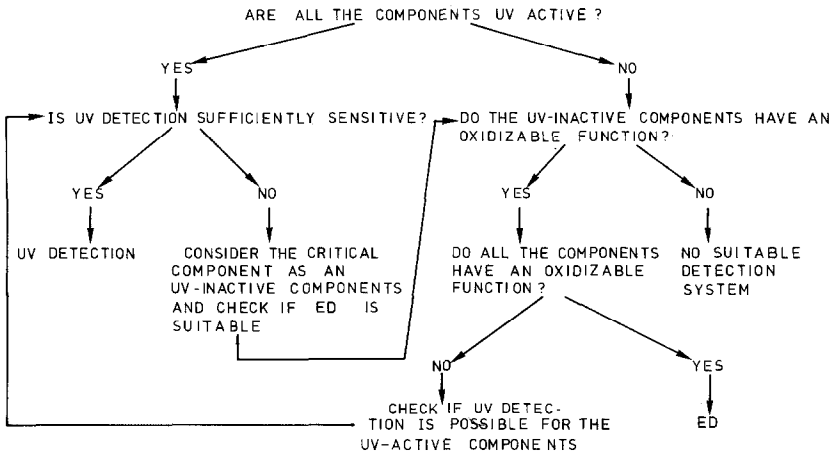


Fig. 1. Decision tree used by the chromatographic expert for the selection of the detection mode.

possible. To obtain the so-called critical component, one multiplies the molar absorption coefficients and the concentrations of all the components. The component with the lowest product is then the critical component. When one of the UV modes is found to be suitable for this critical component by using eqn. 1, then the ES decides that the UV mode will also be suitable for all the other components. When UV detection is not possible for this component, ED is examined for the whole mixture. The use of ED is examined in the same way as when only one component is present, *i.e.*, by looking at the presence of oxidizable functions. When the critical compound which was not UV-detectable is found to be ED-detectable, the ES investigates whether all substances can be detected by ED. If the critical component is ED-detectable but not all the others, one investigates the possibility of carrying out ED and UV detection in series. This means that one decides to do ED for one component and then to consider UV detection for all the remaining components using the same reasoning as above, *i.e.*, by now considering the second lowest absorbing component as the critical one. This is a recursive way of reasoning and KES does not allow this. How this was solved is discussed under *Logical structure.*

Once the detector has been chosen, one can proceed with the selection of one of the four mobile phase systems. The mobile phase systems considered are (1) RP with water–methanol, (2) RP with buffer–methanol, (3) NP with hexane–dichloromethanol and (4) ion-pair chromatography (ICP) with buffer–methanol. The selection of one of these possibilities depends on the hydrophobicity of each substance, their acid–base characteristics, the number of compounds to be separated in the mixture and the detector chosen.

If UV detection is selected the mobile phase selection first requires a knowledge of the hydrophobicity and the acid–base characteristics of the solutes. A rough estimate of the hydrophobicity is obtained from the number of carbon atoms: < 10 is considered as non-hydrophobic, 10–20 as medium hydrophobic and > 20 as hydrophobic. This is a very rough estimate, but it seems that it is sufficient in practice[13]. Four acid–base categories are considered, namely acidic, basic, amphoteric and neutral. As it is not always clear to the non-expert how to classify the substance in one of

these categories, the ES helps the user to decide. The user is asked whether the drug belongs to certain important groups for which the ES has the acid–base properties or even the chromatographic conditions available in its rule base. These groups are xanthine derivatives, corticosteroids, male hormones, fat-soluble vitamins, barbiturates, salicylates, female hormones, phenothiazines, β-blockers, tricyclic antidepressants, alkaloids and benzodiazepines. When a drug is a barbiturate, for instance, the system decides it is acidic or, when one of the components of a mixture is a fat-soluble vitamin, the system decides that NP should be used. If this does not lead the ES to an answer, the user is asked if he knows whether one component is acidic, basic, neutral or amphoteric. If this is unknown, further classification into the acid–base categories is done on the basis of either the anion or cation accompanying the active form (*e.g.*, codeine is recognized as a base when it is present as codeine phosphate, because the system knows that phosphate is an anion) or of the presence of certain functional groups. Very similar acid–base rules have been described in detail previously[3,4], together with an expert system for UV spectrophotometry.

Next, one determines how complex the mixture is. When a single solute is present, one considers only RP with water–methanol and NP as possible solutions. RP is chosen for polar substances, *i.e.*, for a substance with a carbon number $\leqslant 20$ and NP for apolar substances, *i.e.*, for a substance with a carbon number > 20. If the substance is basic, in both instances one adds propylamine (PA) to mask residual silanol groups and as an ion suppressor (0.01% PA in RP, 0.1% PA in NP), and when it is acidic one adds 1% acetic acid as an ion suppressor.

When there are two or more substances present, one first checks whether both acids and bases are present. If this is found to be so, one carries out RP chromatography with a buffer–methanol mobile phase. If one would use water instead of a buffer, one would need to add an ion suppressor. Acetic acid results in long retention of the bases with bad peak shapes and PA in non-retention of the acids. The ES now recommends a phosphate buffer of pH 3, ionic strength 0.05 (except for small molecules, where an ionic strength of 0.02 is recommended). The detailed reasons for this have been published elsewhere[14].

When there are only neutral, only basic (+ neutral) or only acidic (+ neutral) solutes present, one makes a decision between either RP with water–methanol or else NP and IPC. NP and IPC are always recommended together. With mixtures with two solutes with very polar properties, *i.e.*, solutes with a carbon number < 10, or with mixtures with more than two substances containing two or more very polar substances, both NP and IPC will be advised. Otherwise, one uses RP with water–methanol. This may appear surprising. The decision is based on the reasoning that very polar substances have low retention values in RP on a CN column. This is no problem when there is only one substance present, but it is when several substances have to be separated.

There are two situations where RP with a buffer is always used and where consequently all the rules mentioned above are abrogated, namely when electrochemical detection is used and when the sample contains amphoteric compounds. Again the ES recommends a buffer of pH 3 and an ionic strength of 0.05, except when only or mainly very polar compounds are present, in which event the ionic strength is 0.02.

The volume percentage of organic modifier for the starting mobile phase composition in the different chromatographic modes (*i.e.*, RP with water–methanol, RP

with buffer–methanol or NP) is also proposed. Of course, the isocratic conditions so obtained are considered to be only starting conditions for further optimization. Moreover, these rules have been developed for use in our own laboratory with the specific type and brand of stationary phase, namely LiChrosorb CN columns. More work is needed to make rules that are generally applicable. This work is in progress. The rules as they are now depend on the number of carbon atoms. The mean number of carbon atoms in all the compounds present in the sample is calculated. In the ES, rules are incorporated that relate this mean value to the volume percentage of organic modifier in the starting mobile phase composition. These rules are different for acidic and basic compounds. The reason is that different chromatographic behaviour is observed for acidic and basic compounds. For instance, residual silanol groups present at the surface of the column packing exhibit a stronger interaction with basic compounds so that the retention behaviour of these solutes is more influenced than that of the acidic compounds. Consequently, a larger volume percentage of organic modifier is used for basic compounds in order to obtain suitable capacity factors (k'), even when PA is added to mask these residual silanol groups. All details of the rules for the selection of the volume percentage of organic modifier in the starting mobile phase composition are given in ref. 13, where the validation of the whole expert system is performed on 50 pharmaceutical formulations.

The ES considers also whether gradient elution is indicated or not. This depends on the substances with the largest and smallest numbers of carbon atoms. When the difference in carbon number is 15 or more, gradient elution is considered preferable. However, when electrochemical detection has to be applied, gradient elution is never advised.

Logical structure

In this section, consideration is given to the explicit way in which the previously described knowledge is incorporated in the expert system shell KES. For the representation of knowledge in KES, attributes are used which are divided into two groups, namely the global attributes and the class attributes. Global attributes are used to represent general information which is valid for the whole chromatographic system, for instance the path length of the detector. Class attributes represent information about a certain characteristic of a group of objects, for instance each component has its own molar absorption coefficient. Within each group, one can distinguish input attributes, intermediate attributes and goal attributes. Input attributes contain information given by the end user. The goal attributes are the variables that provide solutions for the questions asked by the user. For instance, the goal attribute DETECTION describes which detection modes are considered by the system. Intermediate attributes are used by the inference mechanism to link the input and goal attributes. Appendix I describes the input attributes and the goal attributes with their possible values.

In KES, the logical structure of the attributes, *i.e.*, the hierarchy or the relationship between them, is implicitly built by means of rules. Here, an overview is given of the hierarchy between the most important attributes.

The ES first asks whether a certain detection system has already been selected by the end user or not. When this is not so, the system looks for the best detection system for the given mixture. The reasoning followed by the ES to select the detector,

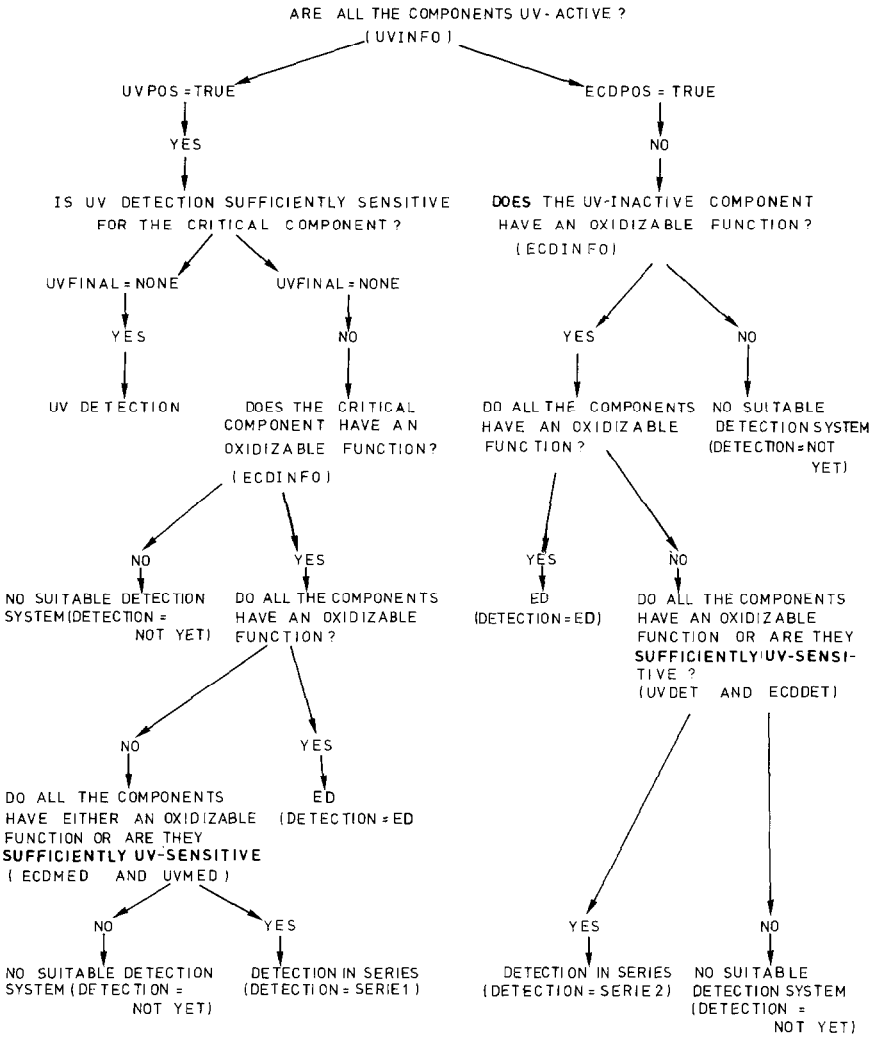


Fig. 2. Pathway followed by the ES for the selection of the detection mode.

which is the first step performed in the method selection process, is shown in Fig. 2. This seems a longer tree than necessary. However, as explained above, one must avoid the recursive reasoning of Fig. 1.

The two goal attributes determining the detection are UVFINAL and DETECTION (see Appendix). Because domain experts prefer UV detection to ED, the inference engine looks for a value of UVFINAL before looking for a value for the attribute DETECTION. UVINFO is an input class attribute. The user is asked to state for each component of the mixture whether it is UV-active or not. Two attributes UVPOS and ECDPOS depend on UVINFO. These two attributes take as values true or false. If all the components are UV-active then UVPOS = true, and if not then ECDPOS = true.

When all the components are UV-active, the ES checks if UV detection is

suitable for the whole mixture. To conclude whether UV detection is possible at 254 nm, a critical component, the component for which $E254 \cdot$ conc. is lowest, is determined. If this is not so, the ES looks whether it is possible to carry out UV detection at the λ_{max} of this component. When this also is not possible, the ES checks UV detection at 220 nm for RP and 235 nm for NP as last solution. This leads to a value for UVFINAL (see Appendix).

If it is found that UV detection is suitable for the whole mixture (*i.e.*, that UVFINAL has a value different from none), the best final detection system is found. To give additional information, the system checks if ED would also be possible. This is done by asking for each component if it has an oxidizable function and if it has, and if the concentration is higher than the detection limit (which is verified by means of some intermediate attributes), the attribute DETECTION has a value 'both', which means that in addition to UV detection ED can also be carried out.

If it has been decided that UV detection is not a good detection method for the whole mixture, the system checks the oxidizability of all the components and, when they are all oxidizable, the recommended detection method is ED (attribute DETECTION = ED). When it is found that one of the compounds is UV-inactive and has no oxidizable function, no detection method is available and DETECTION = not yet. If some components are sufficiently UV-active and others have an oxidizable function, detection in series will be advised.

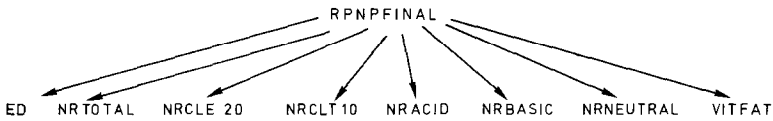
As explained above, after having decided that the critical component for UV detection must be determined by ED, one would like to consider among the remaining products that for which $E254 \cdot$ conc. is lowest. The inference engine of KES does not allow this. Therefore, after having determined the critical component, the ES immediately checks if eqn. 1 is fulfilled for all the remaining components. These calculations are needed when detection in series is required.

The attributes UVMED and ECDMED are class attributes. UVMED takes the same value as UVFINAL and ECDMED, and is yes or no depending on whether ED is possible for the considered component or not. When all the components are UV-active but the ES derived that not all the components can be detected in a sufficiently sensitive way, detection in series is needed and the value of DETECTION is series 1.

When reasoning started with ECDPOS = true and when detection in series is needed, the class attributes ECDDET (is ED possible: values yes or no) and UVDET (which can take the same values as UVFINAL) are considered, and the final value of DETECTION is series 2 when it is found that all substances can be determined either with UV detection or ED.

The selection of the mobile phase is performed in two stages. The first is to decide which of the following three HPLC-systems will be used: RP with water; RP with buffer solution; or NP. The second stage consists in determining the volume ratio of the solvents in the starting mobile phase for the HPLC system selected. The knowledge needed could be structured at first sight as shown in Fig. 3. To be able to apply short-circuit evaluation (see programming considerations), it is better to split up the structure of the tree as in Fig. 4. The goal attribute containing the decision (see Appendix) is RPNPFINAL, which is based on two intermediate attributes, RPNP and RPMAND.

The attribute RPNP can have the values NP and RP with water. Its default value is RP with water. The rules give all the conditions that lead to the value NP. If



WHERE RPNPFINAL = THE NAME OF THE ATTRIBUTE, CONTAINING THE DECISION ABOUT THE MOBILE PHASE.
ED = THE FACT IF ED IS USED OR NOT.
NRTOTAL = THE NUMBER OF SOLUTES IN THE MIXTURE.
NRCLE20 = THE NUMBER OF COMPONENTS WITH LESS THAN OR EQUAL TO 20 C ATOMS.
NRCLT10 = THE NUMBER OF COMPONENTS WITH LESS THAN 10 C ATOMS.
NRACID = THE NUMBER OF ACIDS.
NRBASIC = THE NUMBER OF BASES.
NRNEUTRAL = THE NUMBER OF NEUTRAL SUBSTANCES.
VITFAT = THE FACT IF THERE IS A FAT SOLUBLE VITAMINE IN THE MIXTURE OR NOT.

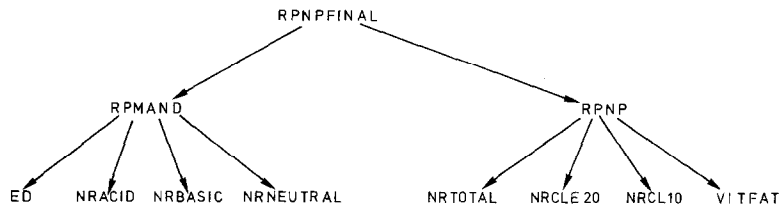
Fig. 3. Parameters determining the mobile phase selection.

none of the given conditions is true, the attribute takes the value RP. RPMAND is true when DETECTION = ED or series 1 or series 2. When RPMAND is true, then RPNPFINAL is RP with buffer. When it is false, RPNP gives its value to RPNPFINAL.

The system also gives preliminary, not detailed, advice on the use of IPC. IPC is recommended as an alternative to NP when the mixture contains only acidic (and neutral) or only basic (and neutral) products.

After the logical structure of the attributes has been built by means of the rules, the inference engine must be guided to operate and use the rules in a logical sequence. Therefore, one must write an action part. The action section is used to direct the operation of the ES. It guides the inference engine through the different rules. In this instance, the action part carries out the following operations in sequence: obtain UVFINAL, if UVFINAL = none obtain DETECTION, obtain RPNPFINAL, obtain GRADIENT, obtain TSF, obtain MURP, obtain PCTM.

Finally, it is the action part that determines the overall interaction of the ES with the end user, *i.e.*, the values obtained for the goal attributes must be shown to the end user. The action part therefore also contains the display commands required to make this possible.



WHERE RPMAND = DECISION TO USE REVERSED PHASE WITH BUFFER.
RPNP = DECISION TO USE REVERSED PHASE WITH WATER OR NORMAL PHASE.

Fig. 4. Pathway followed by the ES for the selection of the mobile phase parameters in the ES.

Programming considerations and implementation

In this section, the philosophy followed to implement the knowledge in KES is given. KES permits different means of knowledge representation to be applied. The most important aspects in this context are production rules and frames. Frames describe classes of objects and associated knowledge. In the present instance, there could be a frame about phenothiazines, how they can be detected and chromatographed, whether they are to be considered polar or not, etc., or there could be a frame about ion-pair chromatography (IPC), the kind of solvents to be used when applying IPC, the situation in which it is helpful, etc. These frames can be considered more or less as data in a database and the AI system links these data, fills in missing slots, etc. Since in the present problem many decisions are based on data such as maxima and polarities, it seems probable that frames could be used to advantage. However, we chose to apply production rules [*i.e.*, the use of IF antecedent(s) THEN consequent(s) type rules]. This knowledge representation system is more usual, more ES shells have it and more experience is generally available.

One of the questions that has to be addressed when one writes an ES is who is to be the principal user. In this instance, it was decided that the system would be written to help scientists and technicians with some experience in HPLC. This category of user wants solutions without needing much text that explains how the solution was obtained. In fact, too much text irritates the expert. However, to learn how the ES functions and why certain conclusions are made, it is necessary to have these explanations. For this reason, the system was written in such a way that the user can choose between a verbose and a terse version. First-time users or users with little experience will choose the verbose version and regular users and chromatographers with more experience would opt for the terse one.

In the same way that verbose systems irritate experts, systems that ask 'stupid' questions are not appreciated. However, it is again necessary to think of less experienced users or occasional holes or lapses in the expert user's knowledge. For this reason, one should create bypasses. For instance, the user of the system is required to say whether a certain substance is an acidic, basic, neutral or amphoteric substance. The fifth alternative is: I do not know. Answering this leads to several new questions (about functional groups) and lets the system decide on one of the first four alternatives. Experts can give immediately one of the first four answers and can therefore bypass what they would consider as too simple questions about acid–base properties.

In theory, one can enter the rules in such an ES without paying any attention to the order in which they will be used and leave the chaining and logic structuring to the inference machine. In practice, this would at least be wasteful. Therefore, the knowledge engineer has to structure the knowledge. Another reason why it is necessary to structure the knowledge is to shorten the reasoning process. The ES always checks all the rules in the order in which they appear. Why this is important can be explained with an example. Suppose a situation in which the inference engine has already decided to use ED and the following rule must be evaluated:

If the total amount of products is 1 and
there are more than 20 carbon atoms in the compound and
RP with buffer is not used
then NP is needed.

The inference engine will then first check how many solutes there are. When there is

only one, it looks at the number of carbon atoms in the product and whether it is more than 20. Only then it will investigate whether RP with buffer is used. As ED is used, this is the case, and the condition of the antecedent is not met. This could have been decided immediately. The investigation concerning the number of carbon atoms and substances was therefore superfluous. In such instances, it is necessary to use so-called short-circuit evaluation. Suppose the inference engine requires a value for RPNPFINAL, the goal attribute containing the final decision about the mobile phase system to be used. The inference engine then collects and evaluates all the rules which determine a value for it in their consequent conditions. If there are undetermined attributes in the antecedent conditions, these antecedents are first considered as subgoals and the inference engine looks for the rules to determine their value (backward chaining). To obtain the decision for RPNPFINAL, the inference engine finds two subgoals: RPMAND and RPNP (see Fig. 4). It first tries to determine their value. The first rule is:

If ED is used
then RPMAND is true.

This rule removes the default value (false) of the attribute RPMAND and changes its value to true. Once the subgoal RPMAND is true, the ES defines immediately the RPNPFINAL value with the following rule:

If RPMAND is true
then RPNPFINAL is RP with buffer.

Short-circuit evaluation then consists in letting all other rules to determine a decision for the HPLC system begin with one of the following conditions:

If ED is not used.
or
If RPMAND is false.

Wasteful inquiries are then avoided. The newest version of KES solves this in a different and better way by using so-called demons.

One way of structuring is to introduce the use of classes. To avoid having to write rules for each component separately one defines a class in the knowledge base, in this instance called SAMPLE. When needed, the inference engine asks for each member of that class (the components of the mixture) and for some or all of the characteristics defined for them. For a characteristic or attribute, the number of members of the class does not need to be defined while building the knowledge base. It is variable and differs for each end user session. KES permits the use of classes. However, its inference engine cannot count how many members there are in a class or how many members of a class have a certain characteristic. For instance, how many components of the mixture examined have less than 10 carbon atoms? Are all the compounds acidic? External programs must then be used. Such external programs needed to be written to count the number of certain characteristics such as the total number of components, the number of acids, bases, neutral components, components with oxidizable functions, UV-active components and the number of carbon atoms of all the components together. An external program is also written to calculate for each component the number of carbon atoms and the molecular weight from the molecular formula. These properties cannot be calculated within the KES system. Therefore, programs external to KES have been written (*e.g.*, in FORTRAN 77) that calculate these values, pass the result back to KES, which captures it in an attribute and continues its reasoning process.

To make the communication between KES and the externals possible, one needs to define in the knowledge base what external programs the expert system has access to, what attribute values serve as input to those externals and in what attributes the output will be captured. The communication between KES and the externals is indirect: KES writes the input that the external needs in a communication file that has to be used by the external program. After execution, the external puts its obtained result in a communication file where KES will read it and capture it in an attribute.

Communication with external programs is also needed for another reason. The end user should work with a run time version that does not allow him to change the rule base or action part. However, he will want to store certain data he has put in into the system. To allow him to do that, a database (DBase3) was attached to the ES. The end user can add certain data such as maximum values to the database, which can be kept in that database and used by the ES when needed.

CONCLUSION

The prototype ES, as it stands now, consists of about 150 rules. Expert systems must be validated. In this instance, several sets of rules were validated separately. For instance, the rules which are used to decide whether a substance can probably be detected by ED have been tested on nearly 100 substances[11,12]. The system as a whole and as it now stands has been applied to 44 commercial drug formulations, selected with a random number generator from the Belgian Drug Compendium[13]. These samples contained drugs from nearly all major pharmacological and chemical classes in very different galenical forms and therefore it is representative of the situation one encounters in practice in label claim analysis. To judge whether the advice of the expert system is correct, the following criteria were used: in first instance, the k' values recorded in the selected mobile phase composition for the drugs have to be situated between 0.5 and 10. Second, the asymmetry factor calculated at 10% of the peak height should not exceed 2. These criteria are also those applied in practice to consider a first-guess system as succesful. In 82% of the cases the first-guess system is succesful, and in 9% of the not immediately succesful trials a simple adjustment of one of the parameters permits succes to be achieved[13]. These final results are at least as good as human experts would obtain.

Expert systems are dynamic systems, however, which are continuously being upgraded. In theory the system should be thoroughly validated each time it changes but, in practice, this is of course not possible. Validation, as is the case with upgrading, is therefore an on-going process. The present expert system will be upgraded in several respects in the near future. For instance, rules will be included that permit solvent systems to be selected that allow measurements below 220 nm, when this is indicated. Two major additions that will be introduced concern the exact composition of the mobile phase. Rules that permit the volume percentage of organic modifier to be derived for different, *i.e.*, more than one, stationary phases are under development. The other major addition is now the subject of a separate small expert system. This does not give a single composition of the mobile phase but proposes compositions that form experimental designs, the complexity of which depends on the complexity of the separation to be carried out.

The distinction between small, medium size and large tools is fuzzy. It may occur that a tool has some features which are typical of a large tool, but that some other important features that are more typical of a medium-size tool are missing. It is also important to check how well certain features of the tool are elaborated. KES can be described as a small to medium-size tool. One of its very strong points is the use of structured rules. It was mentioned above that certain rules can be easily added. This makes it easy to include new knowledge in the expert system and the expert system can be easily updated.

ACKNOWLEDGEMENTS

The authors thank J. Crommen for helpful advice and comments on the IPC. This work was financially supported by EEC (ESCA project number P1570), EGWO (project number 9.0012.85) and Concerted Actions (project number 83/89-58).

APPENDIX: INPUT AND GOAL ATTRIBUTES AND THEIR POSSIBLE VALUES

A. Global input attributes

VERBOOS:
- with full explanation: when a consultation with explanation about the method of reasoning is required.
- with minimum explanation: when no explanation is required.

DETECTIONMODE:
- UV detection for all substances: when the expert has decided that UV detection is to be used.
- ED for all substances: when ED is to be used.
- UV detection with ED in series: when UV detection and ED in series are required.
- unknown: when the end user needs the advice of the expert system about the detection mode.

SPECGRYN:
- all: all the components of the mixture belong to a special group (see SPGRY).
- some: some components of the mixture belong to a certain class or when all the components belong to different classes.
- none: no components belong to a special group.

PL254Q:
- path length of the detector at 254 nm.

SAMEPL:
- yes: the path length at λ_{max} and at 220 nm are the same as at 254 nm.
- no: another path length is used at λ_{max} and 220 nm than at 254 nm (the system then also asks for this path length: PLMAXQ and PL220Q).

INJVOL:
- injection volume.

SPGRY:
- all the solutes of the mixture belong to one of the following groups: xanthine derivative, corticosteroid, male hormone, fat-soluble vitamin, barbiturate, salicylate, female hormone, phenothiazine, β-blocker, tricyclic antidepressant, alkaloid, benzodiazepine.

B. Class input attributes

BRUTO:
- molecular formula of the considered compound.

UVINFO:
- uv-active: the considered component is UV-active.
- uv-inactive: the considered component is UV-inactive.
- undetermined: the UV-activity of the considered compound is unknown.

CONCQ1:
- yes: the expected concentration of the compound is known.
- no: when it is not known.

EM254Q:
- molar absorption coefficient at 254 nm.

EMMAXQ:
- molar absorption coefficient at λ_{max}.

EM220:
- molar absorption coefficient at 220 nm.

ECDINFO:
- yes: the component has an oxidizable function.
- no: the considered component does not have an oxidizable function.

SPECGR:
- asks for each component of the mixture if it is a xanthine derivative, coricosteroid, etc.

PRODUCT DETERMINING FUNCTIONS:
- weak acidic functions.
- weak basic functions.
- strong acidic functions.
- strong basic functions.
- none: when none of the above functions are present.

COUNTERION:
- anion: the counter ion of the active substance is an anion.
- cation: the counter ion of the active substance is a cation.
- none: no counter ion is present.

C. Global goal attributes

UVFINAL:
- uv254: UV detection at 254 nm is recommended for the whole mixture.
- uvmax: UV detection at variable wavelength is recommended for the whole mixture.
- uv220: UV detection at 220 nm is recommended for the whole mixture.
- none: UV detection for the whole mixture is not possible.

DETECTION:
- none: there are no rules which determine a value for detection (default value).
- ED: ED is used for the whole mixture.
- serie1 or serie2: detection in series with different detectors is needed.
- both: UV detection is recommended for the whole mixture but when ED is also possible.
- notyet: no suitable detection system for a certain mixture is known yet by the ES. This is the case when there are components which are both UV-inactive and have no oxidizable functions.
- notlim: ED is not sensitive enough.
- becfat: UV detection is not possible and there is a fat-soluble vitamin in the mixture (a fat-soluble vitamin requires NP but ED requires RP with buffers).

RPNPFINAL:
- RPbuffer: RP with buffer is the selected system.
- RP H2O: RP with water is the selected system.
- NP: NP is the selected system, which will also be IPC when there is more than one product in the mixture.

TSF:
- none: no tailing suppressing factor is used.
- 0.1% PA: 0.1% propylamine is added.
- 0.01% PA: 0.01% propylamine is added.
- 1% acetic acid: 1% acetic acid is added.

MURP:
- not applicable: no buffer is needed.
- pH3 mu = 0.05: the pH of the buffer used is 3 and the ionic strength is 0.05.
- pH3 mu = 0.02: the pH of the buffer used is 3 and the ionic strength is 0.02.

PCTM and PCTB:
- the percentage of methanol and water or buffer, respectively, in the mobile phase.

PCTH and PCTDM:
- the percentage of hexane and dichloromethane, respectively, in the mobile phase.

GRADIENT:
- true: gradient elution is suitable.
- false: gradient elution is not suitable.

MU ION:
- pH7.5, 0.05: IPC is done at pH 7.5 and with ionic strength = 0.05.
- pH2.5 to 3, 0.05: IPC is done at pH 2.5 to 3 and with ionic strength = 0.05.

COUNTERI:
- heptane sulfonate 0.005M: IPC is done with heptanesulphonate as counter ion at a concentration of 0.005 *M*.
- tetrabutylammonium 0.005M: IPC is done with tetrabutylammonium as counter ion at a concentration of 0.005 *M*.

D. Class goal attributes

UVDET and UVMED:
- none: detection in series is needed but UV-active compounds are not UV-active enough.
- uv254: detection in series is needed and the UV-active compounds of the mixture are detectable at 254 nm.
- uvmax: detection in series is needed and the UV-active compounds of the mixture are detectable at variable wavelength.
- uv220: detection in series is needed and the UV-active compounds of the mixture are detectable at 220 nm.

ECDDET and ECDMED:
- yes: detection in series is needed and for the considered compound ED is done.
- no: detection in series is needed and for the considered compound ED is not done.

REFERENCES

1 T. P. Bridge, M. H. Williams and A. F. Fell, *Chem. Br.*, 87 (1987) 1085.
2 F. A. Settle and M. A. Pleva, presented at *Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy March 9–13, 1987, Atlantic City, NJ*, abstract 29.
3 L. Buydens, M. Detaevernier, D. Tombeur and D. L. Massart, *Chemometrics Intell. Lab. Syst.*, 1 (1986) 99.
4 M. R. Detaevernier, Y. Michotte, L. Buydens, M. P. Derde, M. De Smet, L. Kaufman, G. Musch, J. Smeyers, A. Thielemans, L. Dryon and D. L. Massart, *J. Pharm. Biomed. Anal.*, 4 (1986) 297.
5 J. W. Dolan and L. R. Snyder, presented at *Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, March 6–10, 1989, Atlanta, GA*, abstract 1160.

6 J. J. Kirkland, J. L. Glajch, S. W. Rementer, T. G. Jones and L. R. Snyder, presented at *Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, March 6–10, 1989, Atlanta, GA*, abstract 1161.
7 S. S. Williams, J. L. Excoffier and S. Abboth, presented at *Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, March 6–10, 1989, Atlanta, GA*, abstract 1162.
8 R. S. Hodges, J. M. R. Parker and C. T. Mant, presented at *Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, March 6–10, 1989, Atlanta, GA*, abstract 1163.
9 M. De Smet, G. Hoogewijs, M. Puttemans and D. L. Massart, *Anal. Chem.*, 56 (1984) 2662.
10 M. De Smet and D. L. Massart, *Trends Anal. Chem.*, 6 (1987) 266.
11 G. Musch, M. De Smet and D. L. Massart, *J. Chromatogr.*, 348 (1985) 97.
12 G. Musch and D. L. Massart, *J. Chromatogr.*, 370 (1986) 1.
13 M. De Smet, A. Peeters, L. Buydens and D. L. Massart, *J. Chromatogr.*, 457 (1988) 25.
14 M. De Smet and D. L. Massart, *J. Chromatogr.*, 410 (1987) 77.