A Review of Expert Systems for Chromatography

C.H.Bryant*, A.Adam (Computation Department)

D.R.Taylor (Chemistry Department)

University of Manchester Institute of Science and Technology,

PO Box 88, Manchester, M60 1QD, United Kingdom.

R.C.Rowe

Zeneca Pharmaceuticals, Alderley Park, Macclesfield, Cheshire,

SK10 2NA, United Kingdom.

Abstract

Expert systems for chromatography are reviewed. A taxonomy is proposed that allows present (and future) expert systems in this area to be classified and facilitates an understanding of their inter-relationship. All the systems are described focusing on the reasons for their development, what their purpose was and how they were to be used. The engineering methods, knowledge representations, tools and architectures used for the systems are compared and contrasted in a discussion covering all the stages of the development life cycle of expert systems. The review reveals that too often developers of expert systems for chromatography do not justify their decisions on engineering matters and that the literature suggests that many ideas advocated by knowledge engineers are not being used.

1   Introduction

Section 2 compares this work with previous reviews; section 3 introduces some of the concepts and terminology of chromatography and describes how a chromatography method is developed; section 4 gives a definition of 'expert system' and specifies the scope of this review; section 5 explains why expert systems for chromatography are needed; and section 6 proposes a taxonomy for these systems, describes its purpose and how it was developed.

Sections 7, 8, and 9 are arranged in a taxonomic fashion corresponding the taxonomy and explain why each system was developed, what its purpose was and how it was to be used.

Sections 11, 12, 13, and 14 discuss the engineering methods, knowledge representations, tools and architectures used in the development of expert systems for chromatography.

2   Previous Reviews

Many papers have been published that review artificial intelligence for chemistry but their scope is so wide as to prohibit the inclusion of a review of expert systems for chromatography as detailed as that given here. For an example see [1]. In addition to artificial intelligence for chemistry, the more specific topic of expert systems for analytical chemistry has been reviewed in a number of papers. [2] provides an introduction to knowledge representation and explanation facilities for expert systems for analytical chemistry but only gives four references for expert systems for chromatography. [3] also discusses topics such as knowledge representation but it does not mention any expert systems for chromatography. [4] compares the use of artificial intelligence languages and expert system shells for building expert systems for limited domains but only reviews one expert system for chromatography.

Three papers [5] [6] [7] have been published that specifically review expert systems *for chromatography*. [5] reviews expert systems for liquid chromatography very briefly. It lacks the authoritative tone that usually accompanies an academic paper: the style of writing is imprecise, some of the analogies used are weak and only two references are given. [6] does not give any references and only reviews four systems of which only one is an expert system, the other three being simulation programs. [7] mentions more than sixty [1] computer systems but only about a quarter of these are expert systems for chromatography. This paper reviews all the expert systems for chromatography that [7] reviews and a further eleven expert systems for chromatography that [7] does not mention. [7] does not make the relationship between different expert systems for chromatography clear; this paper proposes a taxonomy which facilitates an understanding of their inter-relationship. [7] has a useful explanation of how expert systems in general are structured and introduces some aspects of knowledge engineering. However its analysis of how expert systems *for chromatography* have been engineered is superficial and, in contrast to this paper, no conclusions are drawn on this engineering. This paper reviews expert systems for chromatography comprehensively, covering more than twenty-five systems and gives over eighty references.

---

[1] This number does not include subsystems of expert systems for chromatography.

3   An Brief Introduction to Chromatography

The aim of this section is to explain *briefly* some concepts and terminology of chromatography and to describe how a chromatography method is developed, a knowledge of which is assumed later in this paper. Complete definitions, including any relevant mathematical equations, can be found in the references given.

Chromatography is a separation process in which the sample mixture is distributed between two phases in the chromatographic bed (column or plane). One phase is stationary whilst the other passes through the chromatographic bed. [8] Thus the former is referred to as the stationary phase and the latter as the mobile phase (or eluent). Substances to be separated by a chromatographic system must have different relative affinities for these two phases. Thus, a substance with a relatively higher affinity for the stationary phase moves with a lower velocity through the chromatographic system than does a substance with lower affinity. This difference in migration velocity ultimately leads to physical separation of the components in a sample. [9]

A component of the sample mixture that leaves the stationary phase is said to be eluted in a process known as elution. The eluting power is the power of the eluent, that is the mobile phase, to elute the components remaining on the stationary phase.

If the substances to be separated do not have widely differing affinities for the stationary phase then the column can be eluted with the same solvent (mobile phase) all the time. However if the affinities vary widely then the composition of the eluting solvent can be gradually changed. This technique is known as gradient elution. [10]

Stationary phases are either a solid, porous, surface-active material in small-particle form or a solid support covered with a thin film of liquid [8]. The particle size is related to the efficiency. Column efficiency is inversely related to the rate at which solute molecules spread out as they travel through the stationary phase [10]. The potential of a chromatographic system to separate two compounds is referred to as its selectivity. Most expert systems for chromatography concern chromatographic systems in which the eluted compounds are transported to a detector and recorded as Gaussian (bell-shaped) curves. The signals are known as peaks and are shown on the chromatogram. [8] The chromatogram is a record of the concentration or mass profile of the sample components as a function of the movement of the mobile phase [11]. The degree of separation of successive solute bands or peaks is referred to as the resolution [10]. The peaks give qualitative and quantitative information on the mixture in question [8]:–

Qualitative A peak can be identified by injecting the relevant substance and then comparing retention times. The retention time is equal to the period between sample in-

troduction and the detector sensing the maximum of the retained peak. The retention time of a component is always constant under identical chromatographic conditions. The column dimensions, type of stationary phase, mobile phase composition and flow velocity, sample size and temperature provide the chromatographic conditions. Since these conditions vary the capacity factor is better for characterising a compound, although its measurement does require that an unretained component is timed through the column. The capacity factor is the ratio of the time spent by the solute in the stationary phase to the time spent in the mobile phase (or eluent).

Quantitative The area of a peak is proportional to the amount of a compound injected. Calibration graphs can be drawn to determine unknown concentrations of identified samples.

3.1 Chromatographic Modes

Gas chromatography is the branch of chromatography in which the mobile phase is a gas. In HPLC (High Performance Liquid Chromatography) the mobile phase is a liquid.

There are a number of methods, or modes, for liquid chromatography. Adsorption, Reversed-phase and Ion-pair are three which are relevant to this paper. In adsorption chromatography a relatively polar material is used as the stationary phase and a relatively non-polar solvent as the mobile phase. The different rates at which the various types of molecules in the mixture are adsorbed on the stationary phase provide the separation effect. In reversed-phase chromatography the polarity is reversed: the stationary phase is very non-polar and the mobile phase is relatively polar. Ion-pair chromatography can be used for the separation of ionic compounds. Ionic sample molecules are 'masked' by a suitable counter ion; hence the term pair. [8]

3.2 Method Development

Developing a method for chromatography involves five steps:–

1. **Method Selection** This step is also known as the first guess. The appropriate chromatographic method, materials and instrumentation are selected.

2. **Optimisation** The three essential characteristics of any chromatographic separation (retention, selectivity and efficiency) are optimised.

    (a) **Retention Optimisation** A chromatogram is obtained in which the components of interest appear as sharp, symmetrical peaks with a retention time in an optimum range. The retention of components needs to be sufficiently high to achieve

4

separation, but sufficiently low to maintain a reasonable analysis time and good sensitivity.

(b) **Selectivity Optimisation** The best possible selectivity is sought within the constraints of optimum retention times for each peak.

(c) **Method (or Chromatographic or System or Efficiency) Optimisation** After the selectivity has been optimised the efficiency of the chromatographic system is optimised by selecting the most appropriate column, operating conditions and instrumentation.

3. **Method Validation** Finally the method is validated to prove that it meets the analytical requirements.

4 Definition of an Expert System

An expert system is a computer program that represents and reasons with knowledge of some specialist subject with a view to solving problems or giving advice. An expert system should exhibit all of the following features to some degree. [12]

- It *simulates human reasoning* about a problem domain, rather than simulating the domain itself.

- It performs reasoning over *representations of human knowledge*, in addition to doing numerical calculations or data retrieval.

- It solves problems by *heuristic or approximate methods* which unlike algorithmic solutions, are not guaranteed to succeed.

- It deals with subject matter of realistic complexity that normally requires a considerable amount of human expertise.

- It must exhibit high performance in terms of speed and reliability in order to be a useful tool.

- It must be capable of explaining and justifying solutions or recommendations to convince the User that its reasoning is in fact correct.

[12] gives a broad introduction to expert systems. The literature abounds with books on the various aspects of engineering relevant to this paper. Useful texts include [13] for knowledge acquisition, [14] for knowledge representation and [15] for expert system tools and languages.

As stated above an expert system does not simulate the domain itself but human reasoning about the domain. Hence this paper does not review literature that describes work in

which the primary aim was to model, quantitatively or qualitatively, aspects of chromatography. Thus [16] is not discussed because it is concerned with *model* based reasoning and Drylab [17] is not included because it *simulates* HPLC development.

5   Why Expert Systems are needed for Chromatography

The introduction of workable expert systems for chromatography would be of great benefit because of the wide range of analytes amenable to chromatography and the complexity of this field with regard to the choice of materials and instruments [18].

An example of the wide range of amenable analytes is seen in the pharmaceutical industry: there is an ever increasing volume of diverse novel compounds which have to be screened in order to develop compounds with diagnostic or therapeutic properties. In principle, each compound needs its own method of analysis; thus the method development process must be repeated in its entirety for each new compound. Most of these new compounds are analysed using some form of chromatography, mainly HPLC. [18]

The choice of materials for HPLC is complex. Many factors affect the choice of mode of separation, packing material, mobile phase, and instrumental operating conditions. The development of an optimum method and the interpretation of the results usually requires a lot of expertise and experience to solve the problems that arise for each particular case.

It is impossible for any individual analyst to become an expert in all the various areas of chromatography. Hence individual analysts working in areas in which they are not experts have to seek assistance from the experts in those areas. Human, as opposed to computer, experts have several drawbacks however. They are often in short supply or unavailable in a particular laboratory. They are usually busy and have little time to spare to help non-experts. Occasionally their expertise is lost because they leave an organisation or die. Sometimes their expertise is temporarily unavailable because they are ill or on leave. When they are available they may not always be in a helpful mood and may not offer consistent advice. Expert systems offer an opportunity to encode chromatographic expertise in computer systems which can be made widely available and which provide advice which is consistent, User-friendly and easily documented.

6   The Taxonomy

This section discusses the proposed taxonomy of expert systems for chromatography.

The purpose of the taxonomy is to:−

- facilitate the readers understanding of how the systems are related. The diagrams that show the taxonomy allow it to be quickly assimilated and they can be used for reference.

- allow readers to focus their attention on a class of expert system for chromatography that is of particular interest to them, if they should so wish.

- allow future expert systems for chromatography to be classified and hence illustrate how they are related to previous systems.

The taxonomy was developed as follows. A number of attributes that were candidates for classifying the expert systems were selected. They included the type of expert system, who developed it and where, brief details of how it was implemented and what the domain was. For each system a list of the values of these attributes was compiled. Attempts were made to group the systems into classes where class membership was determined by the values of one of these attributes. It proved impossible to develop a taxonomy that unambiguously classified all the systems if the type of expert system was used to determine class membership: the classes generated were not mutually exclusive for all the systems. Attempts to classify the systems by considering who developed them or where did not give rise to a taxonomy that would, in any way, facilitate the readers understanding of the inter-relationship between the systems. Classifying the systems in terms of how they were implemented was not possible because their implementations could not be accurately compared and contrasted: the range of aspects of the implementations described in the literature varies for different systems. However it was discovered that all the systems could be unambiguously classified if class membership was based on the domain of a system. An analysis of the resulting classes showed that in some cases two or more of these classes were subclasses of a more general class. An examination of these more general classes showed that some of them were subclasses themselves of a yet more general class. Thus a taxonomy was developed, as shown in Figures 1, 2 and 3.

Sections 7, 8, and 9 describe the functionality of the various expert systems for chromatography, presenting them in a taxonomic fashion corresponding to the taxonomy outlined above.

7 Systems Not Specific to a Chromatographic Method or Class of Compounds

This section discusses all the expert systems for chromatography that were not specifically designed for a particular method or a particular class of chemical compounds.

7

7.1 Systems Designed for Several Stages of the Process

This section corresponds to the branch of the taxonomy which is entitled 'Multiple Stage' on Figure 1 and describes three projects (ESCA, ECAT and ESC/ESLC) which all attempted to implement systems that would give advise to an analyst on several stages of the process of developing a successful HPLC separation.

7.1.1 ESCA

The expert systems in chemical analysis (ESCA) project, Esprit project 1570, started in May 1987, officially finished in May 1990, and involved five partners:–

- Philips Scientific, Cambridge, U.K.

- Catholic University Nijmegen, The Netherlands

- Organon International B.V., Oss, The Netherlands

- Philips Research Eindhoven, The Netherlands

- Philips Research Hamburg, F.R.G.

- Brussels Free University, Belgium.

ESCA "was set up to evaluate the merits of expert system technology for use in industrial chemical analysis" [18]. Its aim was to provide expert systems that would illustrate the benefits and shortcomings of expert system technology. ESCA resulted in the publication of around 25 papers. Some of these give general descriptions of the project. (See [19] [18] [20] [21]).

The knowledge domain chosen for ESCA was HPLC method development in pharmaceutical analysis. "In the pharmaceutical industry an ever increasing volume of diverse novel compounds has to be searched in order to develop compounds with diagnostic or therapeutic properties. In principle, each compound needs its own method of analysis. ...Most of these ...are analysed using some form of chromatography, mainly HPLC. " [18] It was believed that an expert system for HPLC method development "would speed up the process; making it more consistent and better documented" [18] while offering expertise that was not generally available in every laboratory.

Initially the following four domains which covered the entire field of method development in HPLC were selected. Each domain represented a discrete step in the process. Thus the intention was to build four experts systems that could be integrated later.

1. **Selection of Initial Conditions**

2. **Selection of Selectivity Optimisation Criteria** It was decided to build an expert system for the *selection of* optimisation criteria because "it is difficult to select the most appropriate criterion in different situations" [18] and the choice greatly affects the outcome of the optimisation. The resulting system helped the User to use a conventional package of computer optimisation programs.

3. **Optimisation of Chromatographic Parameters** Chromatographic (or Method) optimisation follows selectivity optimisation. The latter results in a method that yields adequate separation in an acceptable amount of time, for the given instrumental conditions. The former optimises these conditions. The aim of the optimisation of the chromatographic parameters is to reduce analysis time and increase the sensitivity of the method. An expert system was developed for this task because the "relations between these parameters are complex. Finding the optimal settings requires the evaluation of a number of equations that are difficult to see through, even after a long period of study." [18]

4. **Validation of the Development Method** The expert system developed for this domain was limited to precision testing.

A successful integration of the work of all the participants in the ESCA project required that the project was managed effectively: there was a large number of participants from five organisations that were geographically isolated from one another. Although there were no less than four proposals on how various parts of the work could be integrated, not one of these described how all of the parts could be combined. Furthermore the literature does not provide any evidence that workers on the ESCA project understood that the success of the integration of a project as large as ESCA could not be left to chance and that consequently there was a need for a plan that described how all the parts would be integrated. This may be because the actual goal of the ESCA project was the development of stand-alone systems [21].

The four parts of the domain covered by the four proposals on how the various parts of the work could be integrated are outlined below.

1. The selection of initial conditions, retention optimisation and selectivity optimisation.

2. Selectivity and chromatographic (or method) optimisation.

3. Method validation.

4. Repeatability testing and troubleshooting.

Figure 4 shows the ESCA subsystems implemented for each of the ESCA subdomains and shows the four proposals for integrating the subsystems.

PROPOSAL ONE FOR INTERGRATION OF ESCA WORK; SELECTION OF INITIAL CONDITIONS, RETENTION OPTIMISATION AND SELECTIVITY OPTIMISATION.

Some of the workers on the ESCA project performed a feasibility study [22] [23] for the construction of an integrated system for the selection of initial conditions, retention optimisation, and selectivity optimisation. They proposed to construct the system from the stand-alone expert systems (described in the remaining part of this section) and some additional knowledge which was needed to direct the user through the integrated system.

In the proposed integrated system first-guess conditions were to be selected by one of the expert systems LABEL, DASH or LIT, the choice depending on the application field. After carrying out the first experiment, the retention time range of the solutes was to be evaluated. If there were solutes with capacity factors outside the desired range, then one of the corresponding retention optimisation expert systems LABEL$^/$, DASH$^/$ or LIT$^/$ was to be consulted. The result would be a chromatogram in which all the solutes would have eluted within a reasonable time but two or more peaks may still have overlapped. The selectivity optimisation expert system SLOPES would then have been consulted.

DASH (Drug Analysis System in HPLC) [79] [24] was originally developed for chromatographic method selection and retention optimisation for the purity control of basic compounds, namely CNS (central nervous system)-active and cardiovascular drugs. The system determined the initial conditions to obtain a capacity factor between 3 and 10. "In more than 75% of all cases, correct predictions were obtained for the original family of substances" [22].

DASH$^/$ [24], an extension of DASH, was specifically designed for retention optimisation. DASH$^/$ was a good-second-guess system whereas DASH was a first-guess system. DASH had proved satisfactory when limited to basic drugs. However the integrated ESCA system had to allow for a wider range of compounds. DASH$^/$ was proposed because DASH alone would not have performed satisfactorily for this wider range.

LABEL [25] [26] [27] was an expert system for selecting the initial chromatographic conditions for the label claim analysis of pharmaceutical formulations on a cyanopropyl column used in different chromatographic modes. LABEL was developed by one of the partners involved with the project before ESCA started. "It was included in the project because it covers the situation that one sample must be analysed for different compounds. This is in contrast to DASH" [21].

LABEL contained knowledge for the selection of a suitable detection mode. It only considered UV or electrochemical detection, the latter in the oxidation mode. [27] [25] [28]

LABEL assumed the use of a single stationary phase type, namely a nitrile or cyanopropyl column, which could be used in both normal-phase and reversed-phase chromatography. LABEL selected the mobile phase system, either normal-phase, reversed-phase with water or reversed-phase with buffer, using rules. It then decided if the addition of ion-suppressing agents to the eluting agent was necessary and finally gave the starting composition of the mobile phase. [26]

In 36 of 44 tests performed on LABEL "...success was achieved in a manner that probably could not have been done better by a human chromatography expert" [26].

LABEL$^/$, like DASH$^/$, was specifically designed for retention optimisation: the main task of the expert system was to situate the capacity factor in a suitable range. The optimisation was performed by increasing or decreasing the percentage of organic modifier in the mobile phase, starting from the first-guess composition. [22]

LIT was a small expert system that helped to select all the important parameters of a literature method and checked whether the method could be treated by SLOPES. [21]

SLOPES, the SeLectivity OPtimisation Expert System, was to comprise three different modules: VARIABLES, DESIRE, and CRISE. VARIABLES selected the relevant optimisation parameters (the *variables* that need to be considered) and their boundaries (minimum and maximum limits). DESIRE determined the type of the experimental design (the pattern according to which the necessary experiments will be performed such as Simplex or Doehlert) and the number and location of the experiments. Finally, the most suitable optimisation criterion to describe the quality of separation in a chromatogram was selected by CRISE.

CRISE (CRIteria SElection) [29] was an expert system for the selection of objective criteria for systematic optimisation of selectivity. Such criteria must characterise the quality of separation in a chromatogram.

The system was developed for several reasons, but principally because numerous criteria have been suggested, each of which yield different results and the choice depends on a large number of factors. "It is genuinely difficult to select the most suitable criterion in a particular situation" [29].

CRISE consisted of four modules. In the first the most suitable elemental criterion was selected to quantify the extent of separation between two adjacent peaks in the chromatogram. Resolution (R_s), separation factors (S), separation factors corrected for plate counts (S_N) but also expressions such as peak-valley ratios could be used. As R_s, S_N and S do not take into account a possible loss in resolution due to non-ideal situations module 2 investigated whether any corrections were required. Module 3 assisted the User in the selection of weighting-factors to make a difference between peaks according to their relevance

or relative importance. In the last module, the elemental criteria were incorporated in the global optimisation criterion that was best suited to quantify the separation quality of the entire chromatogram taking into account the purpose of the chromatogram. [30]

The main conclusions of the validation of CRISE were that " i) the expert system provided clear and unambiguous answers for each consultation and ii) the expert system and the human expert provided the same answers for all ten cases (19 consultations) considered during the validation" [31].

A hypermedia version of CRISE, **CRISEBOOK** [30], is discussed in Section 13.2.

PROPOSAL TWO FOR INTERGRATION OF ESCA WORK; SELECTIVITY AND CHROMATOGRAPHIC (OR METHOD) OPTIMISATION.

Some of the workers on the ESCA project proposed an integrated system for just two of the domains: selectivity optimisation and chromatographic (or method) optimisation. They proposed to construct the system from three programs: CRISE (see above), Diamond and SOS. [31]

Diamond was a package of conventional computer programs which required some expertise to use it. The most difficult decision that a Diamond User needed to make was the selection of the most appropriate optimisation criterion. In the integrated system this decision was to be made with the help of the expert system CRISE. Another expert system, SOS, was to be used to transform the chromatogram with optimum selectivity produced as a result of Diamond into the optimum overall method by establishing the best column conditions, instrumentation, injected amount, etc.

SOS (System Optimisation System) optimised separations in terms of i) sufficient separation ii) sufficient sensitivity and iii) shortest possible time [32].

To run the system, an initial chromatogram was needed, together with the relevant information of how it was recorded. The system used two databases, one describing the available columns, and the other the available detectors. (These were created once for a given laboratory, but could be modified at any time.) The system used this information to:−

- recommend the optimum column, detector cell, time constant, flow-rate and sample size.

- predict the required analysis time, the 'critical resolution' (that is the lowest value observed for the resolution between a relevant pair of peaks) and the pressure drop over the column.

- provide an explanation of its reasoning in the form of a bar chart and some additional advice to the user.

The optimum result was defined as:–

1. The resolution for all relevant pairs of peaks had to exceed a minimum value specified by the User.

2. The signal-to-noise ratio for the smallest relevant peak had to exceed a minimum value specified by the User.

3. The required analysis time had to be as short as possible.

A prototype was implemented in Knowledge Craft [2] but the final system was implemented in Pascal. The developers described the conventional explain facilities provided by expert-system development tools such as Knowledge Craft as " generally not very helpful" [32]. Thus when SOS was reimplemented in Pascal [33] a new set of help and explain facilities was implemented. As a result bar charts could be used to establish which factors were limiting the speed of analysis, which factors caused a particular combination to be invalid and whether or not a major reduction in the analysis time was still feasible.

PROPOSAL THREE FOR INTERGRATION OF ESCA WORK; METHOD VALIDATION.

The ESCA expert systems discussed so far have not dealt with method validation. A prototype for an integrated expert system for intralaboratory precision testing was implemented [34] [35]. It integrated two expert systems described below, one for repeatability testing and the other for ruggedness testing. The architecture of the integrated system is described in Section 13.4.

The purpose of a precision test is to establish the random deviation from the mean in a certain analysis. Precision testing normally consists of repeatability and (interlaboratory) reproducibility tests. In the former, the same sample is analysed under the same conditions by the same analyst a number of times. In the latter, the same sample is tested in different laboratories to examine the precision of the method under slightly changing conditions. Since reproducibility tests involve more than one laboratory they are relatively costly; to reduce these costs a ruggedness test can be performed after the repeatability test but before the reproducibility test. Like a repeatability test, a ruggedness test can be performed in just one laboratory but the effects of using different laboratories are simulated; this detects some of the problems that would usually be discovered during the relatively costly reproducibility test.

The system was developed because method validation was becoming increasingly important as stricter rules were applied by regulatory authorities. Precision testing was a vital

---

[2]Knowledge Craft is described in Section 13

step in this validation. The system concentrated on precision tests that could be done in the laboratory where an LC method was developed. For most analysts in a routine laboratory, the performance of a precision test was not straightforward. The system was intended to given the analyst validating the method as much certainty as possible that the method would not fail in a collaborative interlaboratory test. [34]

The expert system for repeatability testing was called REPS (REPeatability testing System) [36] [37]. The expert system was used to select suitable test procedures for particular applications and to interpret their results. All the algorithms for the calculation of variance were set up in spreadsheets which could interact with the expert system. This is described in Section 13.3.

The system was reimplemented in Pascal in multiple windows environment [84]. This version worked as follows. It started by consulting a system optimisation module to provide the fastest analysis time within the required resolution. The method features to be tested were then defined as the sample preparation and the injection procedure. The User had to input a description of the HPLC method together with information on its expected usage. The system recommended an experimental design based on this input. The User then carried out the experiments and collected data which was input to the system. The system then diagnosed problems with the repeatability of the HPLC method. Each problem diagnosed had a list of actions that could be taken to try to solve the problem. The order in which the actions were presented to the User was such that those which were most likely to solve the problem appeared first.

The expert system for ruggedness testing was called RES (Ruggedness Expert System) [38] [39] [40]. RES comprised six modules that represented the major steps in the set-up and interpretation of the ruggedness test. The modules were controlled by a supervisor.

The Factor Choice Module This selected the factors to be included in the ruggedness test and the factor levels. "Normally, some 50 factors may influence method performance in liquid chromatography. For every method, however, only a small group of factors is expected to be critical in that small changes can cause large effects." [41]. The relevant factors will be different for every chromatographic method. All the relevant factors must be tested but the number of experiments required increases dramatically with the number of factors. Thus the number of factors chosen by the system had to be minimal but sufficient.

The module was based on a rule based system [41]. The expert system approach yielded acceptable results to the problem: it had proved satisfactory for ten of the eleven cases of pharmaceutical formulations tested. This was in contrast to an al-

gorithmic approach; "...the problem of factor choice is difficult to handle in normal programming languages because no algorithms are available, and they are difficult to design because of the many parameters involved and their complex interactions." [41] The expert system could cope with the problem because the relations between the parameters were organised in frames, rules, handlers and demons [41].

The Design Selection Module The module selected the experimental design from a set stored in the module. However the User could overrule the decision of the module and even add new designs to the system together with the rules for their selection.

After the User had consulted the factor and design selection modules the experimental design could be stored in a file for reference. The User then performed the experimental work and measured the necessary parameters.

The Statistical Results Module This performed statistical calculations using algorithms.

The Chemical Results Module This contained heuristic knowledge. The output from this module consisted of a set of warnings that had to be included in the final method description. The module decided whether one of the two repair modules (described below) should be activated.

The Reselect Factor Levels Module This was a rule based module which modified the factor levels to a narrower range. The module produced a list of factors and levels that could serve as a *new* input to the design selection module to repeat the ruggedness testing procedure.

The Method Improvement Module This was derived from SOS, which was described above.

During the validation of RES the factors chosen by an unbiased expert were compared with those suggested by RES for 11 test cases. In all but one, RES performed satisfactorily.

PROPOSAL FOUR FOR INTERGRATION OF ESCA WORK: REPEATABILITY TESTING AND TROUBLESHOOTING.

The fourth proposed integration combined REPS, SOS, and three other modules which were built to add flexibility to the integrated system. [21] This allowed the User to consult the system in three different situations. It could be used to access the repeatability of a new method, to check the repeatability of a previously validated method and as a troubleshooting tool. The latter "...turned out to be a valuable feature" [21].

CONCLUSIONS ARRISING FROM THE ESCA PROJECT

Despite the publication of around 25 papers on the ESCA project, the implementation of over a dozen ESCA subsystems and no less than four proposals on how various parts of

the ESCA work could be integrated there is no evidence in the literature to suggest that a single integrated system was implemented. Future projects to develop expert systems for chromatography which tackle domains as large as that covered by the ESCA project must be managed more effectively.

7.1.2 ECAT

Wokers at Varian Associates in the USA developed ECAT (Expert Chromatographic Assistance Team) [42] [43] [44]. ECAT was a collection of expert system programs intended to assist the inexperienced chromatographer in HPLC method development. The recommendations of the system were intended to lead the User logically through the entire method development process towards a workable HPLC separation.

ECAT comprised the following four modules connected to an inference engine:–

Column and Mobile Phase This received inputs about the analyte characteristics and made recommendations for the column packing, column geometry, mobile phase liquids and mobile phase modifiers. The module used heuristics which were represented by a set of rules. The module forward chained from the initial factual input about the analyte class(es), querying the User where necessary, in order to develop a more general identity for the class(es), or to conclude the existence of properties associated with the class(es). When the chain of inferencing was exhausted the module searched for key facts to determine which of its recommendations it was to output. The module then gave the User the option of changing some of the decisions that the module made when the rules were initially fired. If the User decided to change any decisions then the module redesigned the separation based on the alternative choice.

Sample Preparation This helped the User to determine whether or not the sample required pre-treatment in order to enhance some quality of the separation. The module's knowledge base contained sets of rules and facts for determining whether a guard column was needed, and whether the analysis could be aided by solid-phase extraction techniques. A list of recommendations was delivered at the end of the session.

Method Optimisation This led the User through a series of experiments in which the separation was optimised in both total time and acceptable resolution. This module was limited to a small set of rules which led the User to use a limited range algorithmic optimisation strategies.

Database of Chemical Properties This contained factual information about specific chemicals and classes of chemicals.

16

7.1.3   ESC/ESLC

A chromatography expert system was developed by researchers at the Dalian Institute of Chemical Physics in China [45] [46] [47] [48]. This system was named ESC (Expert System for Chromatography) in [45] but was referred to as ESLC (Expert System for Liquid Chromatography) in [47].

The strategy for the development of the system was to develop a knowledge and chromatogram base, an inference engine and a user interface. The chromatogram base could be searched by compound, analyte class or author. (The base included literature references for the chromatograms.) Thus the User could either enter the structure of the sample or the sample name or the name of the class to which the analyte belonged. [45] The chromatogram played an important role in that it not only provided a reliable knowledge base for the system but it was also used to verify the results recommended by the system [47].

The system had five modules:−

Separation Mode, Mobile Phase and Stationary Phase Selection

Selection of the Sample Pretreatment and Detection Method

Optimisation of the Operating Conditions

Peak Identification Four approaches for identifying peaks were considered, one was online and the other three were off-line.

Diagnosis of the Hardware System This was developed because "almost all chromatographs do not possess self-diagnosis at the heart of the system, the column" [45]. It proposed a general strategy for hardware diagnosis involving comparison of the hardware to be used with standards. Any differences found resulted in rules being fired and a diagnosis being proposed.

7.2   Systems Designed for One Stage of the Process Only

This section describes those systems shown on Figure 2 and is organised in a taxonomic fashion corresponding to Figure 2. All the systems were designed for one stage of the chromatographic process only.

7.2.1   Method Selection Systems

Researchers at Virginia Tech in the USA developed ESP (Expert Separation Program) [49]. This was an expert system designed to aid the inexperienced analytical chemist in choosing a HPLC method. It did *not* suggest what conditions were required.

ESP began by conducting an interview to gather information on the separation problem. The program then attempted to reduce the number of possible candidate solutions by decomposing the separation problem into independent sub-problems. ESP was menu driven and incorporated how and why explanation facilities. The system's architecture is described in Section 13.4.

The validation of ESP showed that in nine out of 12 test cases ESP suggested a reasonable method. [49]

7.2.2 Retention Prediction Systems

An expert system called CRIPES (Chromatographic Retention Index Prediction Expert System) was developed by researchers at Loughborough University [50] [51].

The system used the molecular structure of an analyte to calculate retention indices from empirically derived quadratic expressions for the structural units. The program could also calculate the resolution of pairs of analytes.

During the validation stage the retention of a number of test compounds not previously examined were measured and compared with those calculated by CRIPES. "In many instances there was close agreement between the values, suggesting that generally the prediction method was satisfactory ... The database does not yet include sufficient values for interaction between a wide range of substituents to permit a consistently high accuracy in the calculation of retention indices, although in most instances the predictions are within experimental error" [50].

7.2.3 Optimisation Systems

Researchers from Bradford and Heriot-Watt Universities developed a system for eluent optimisation in reversed phase HPLC [52] [53] [54] [55]. The input to the system was spectral information from a multichannel diode array detector which provided retention information. This type of detector "provides an extra dimension of information from a single chromatogram ... a three dimensional spectrochromatogram is produced ... in which the axes are time, *wavelength* and absorbance" [52].

The approach taken by the system was to perform a gradient elution experiment to determine the appropriate initial solvent strength, followed by response-surface modelling on the resulting spectral information using an iterative regression method to determine the mobile phase composition for optimum resolution.

The iterative regression approach relied upon the correct and unambiguous identification of each solute in each of the chromatograms. If this was not possible, perhaps due to co-

elution of two or more peaks, the system then employed a modified Simplex procedure which makes no assumptions based on spectral data.

Peak homogeneity (the extent to which each peak corresponds to just one component of the sample) was assessed by a number of independent modules, the output from which was interpreted by the expert system and used to validate the response surface model constructed by the optimisation procedure.

The system required the column to be preselected, and an operator to prepare the mobile phase, operate the pumps and perform the injections; the developers stated their intention to automate these tasks in the future [52].

Researchers at Zhejiang University in China developed a system for optimisation of the mobile phase composition in reversed phase HPLC [56].

Workers in Hungary produced an expert system for the prediction of initial HPLC conditions for selectivity optimisation in pharmaceutical analysis [57].

The input data corresponded to the structural formulas of the compounds of interest. Using this data and a database, the system calculated partition coefficients which were correlated with HPLC retention.

7.2.4 Troubleshooting Systems

Three troubleshooting systems, HPLC-Doctor [5], CATHIE and PECOD, diagnosed problems with instrumentation whereas another, Upjohn's system, diagnosed problems with methods.

Workers at the Upjohn company in the USA produced a knowledge-based expert system for troubleshooting 21 of the HPLC assay methods frequently used in their laboratories [58]. The aim of the project was to construct an expert system that would advise an entry-level laboratory technician to diagnose problems in a HPLC system.

The problem solving approach of the system was as follows:−

1. It gathered the symptoms and other relevant information from the User.

2. It derived the selectivity, capacity factor and resolution from the data.

3. It analysed the assay problem, obtaining more data if necessary.

4. It diagnosed the most probable causes and then ranked them. The results were shown to the User and justifications given.

5. In depth troubleshooting followed for the highly ranked causes in order to identify the cause of the problem.

Milne, from Intelligent Applications Limited in Livingston, produced an expert system called CATHIE [59] for the automatic interpretation of gas chromatographic data and for the provision of an expert analysis of the state of the instrument in order to detect possible failures or deteriorations.

Workers at Tsingham University in China developed a prototype diagnostic expert system called PECOD. "The domain-specific knowledge on the performance, design and operation of packed extraction columns was reviewed, structured and encoded" [60]. The knowledge was represented by combining a production system with conventional programs.

8 Systems Specific to a Particular Class of Compounds

This section describes those systems that are shown on Figure 3 and is organised in a taxonomic fashion corresponding to it.

8.1 Metabolites

Two expert systems for metabolites were developed: SPES and HPLC-METABOLEXPERT. SPES [61] was an expert system for capillary gas chromatography analysis of human ester metabolites.

HPLC-METABOLEXPERT [62] was developed by researchers at the Hungarian Academy of Sciences and the CompuDrug company. The system became commercially available from CompuDrug. The system simultaneously predicted the metabolites of an organic compound and their retention data.

The system was developed because HPLC had acquired a major role in the identification of metabolites in medicinal chemistry which had been growing in importance. The identification procedure was complicated by the problem that compounds with unknown or incompletely known structures and retention times needed to be identified in a chromatogram that contained mainly the peaks of non-metabolites.

The User had to enter the parent compound via a graphical interface. The system then output a tree-like picture which represented the metabolic transformations. This was produced using a knowledge base of generalised metabolic transformations taken from a standard text.

To obtain the retention predictions the User had to input the chromatographic conditions and retention times of the parent compounds. On requesting the HPLC retention prediction for metabolites, the necessary mobile phase composition and predicted retention times were displayed together with a note about the necessary pH and detection wavelength changes.

The prediction was based on the structural differences between the parent compound and the metabolites: the system used a database that contained the retention changes caused by a substituent of the molecule that appeared or disappeared in the physiological metabolic route.

The power of the expert system to predict retention was investigated by measuring synthetic mixtures of parent compounds and metabolites. The differences between the measured and predicted retention times was always less than eight minutes and the average difference was 1.8 minutes.

8.2 Steroids

Two similar expert systems for planning separations of steroids by HPLC were designed by workers in Singapore and Australia [63] [64].

The first system that was developed tried to select an appropriate separation system for a given number of compounds. The User had to input a list of compounds which s/he suspected to be present in the sample, and a list of compounds which had to be especially well separated. Candidate separations were created by:–

1. establishing the characteristics of the sample, that is the compounds of interest and their polarity.

2. finding a stationary phase of similar polarity.

3. finding an eluent mixture of opposite polarity.

4. finding a compatible detector.

However this strategy would have produced a large number of candidate solutions so the expert system imposed constraints. These constraints were determined from the list of compounds which had to be especially well separated and other additional queries made to User during the consultation. Candidate solutions that did not meet the constraints were ignored.

The system first searched for a case in the knowledge based in which some or all of the compounds of interest in the sample had been separated. If this failed it searched for a case where similar compounds or class of compounds had been separated. Finally if this failed then the system planned the separation from first principles.

The implementation described was limited to quite simple separations. However this limitation was due to insufficient information in the knowledge base rather than to deficiencies in the system structure or the approach to problem solving. [63]

This system is referred to in this paper as the the decision tree system for steroids, as explained in Section 12.1.2.

The other system for planning separations of steroids was designed to tackle the problem in the same way but using a different formalism to represent the knowledge. Section 12.1.2 explains why this was attempted and discusses the results. This system is referred to in this paper as the APN (Augmented Planning Network) system for steroids because it used ATNs. ATNs are discussed in Section 12.1.2.

8.3 Biological Fluids

Workers at Glaxo have investigated the feasibility of building an expert system for method development in HPLC for bioanalysis, that is the analysis of drugs in biological fluids [66]. They developed a structured approach to method development in this area and they envisaged that the data resulting from it would be used to build the knowledge base for an expert system.

8.4 Proteins

Protein purification involves several stages from the extraction from the source to the final purification with high-resolution chromatography techniques. Three expert systems for planning the purification of a protein were developed: PROTEIN, PPA and P8.

A prototype expert system called PROTEIN was developed at the University of Reading to assist in the preliminary selection of operations for the recovery and purification steps in the manufacture of proteins. The aim was "... to design large scale protein purification processes with a very high recovery, a virtually pure product and minimum cost ... Expert knowledge was obtained partially from the literature but mainly from industrial experts ..." [67].

"The work showed that expert systems can be a helpful tool to assist in solving the knowledge intensive and heuristic based problem ..." [67]. However the system could only amplify the effectiveness of an expert, it could not replace the expert.

Workers in Sweden developed a knowledge-based system called PPA (Protein Purification Advisor) for planning protein purifications [68].

PPA was developed because most researchers had been adopting a trial-and-error approach to protein purification "... without thought for improvement or optimisation of the method, leading to unpredictable and irreproducible results, and rapid exasperation" [68].

PPA was completely User driven in that the Users told the system what they knew about the sample, the source material, and the protein they wanted to isolate, and then

asked the system for a purification plan. The system then presented a plan. The idea of PPA was not to get a 'perfect' plan but rather to avoid making the most obvious and time consuming errors.

PPA was intended to support Users with a varying degree of experience: it was designed to facilitate learning by graduate students and for senior researchers seeking a second opinion. Not only could the User request PPA to formulate a plan but s/he could also describe a complete or partial plan and ask PPA for comments. Further the User could ask for alternative plans and ask why or why not a certain operation had or had not been chosen.

PPA was validated by testing it with 20 different proteins selected by scientists. The resulting plans were discussed with the respective scientists. "In every case the plans were found to be acceptable with some minor variations" [68]. When compared to published procedures for the proteins the plans were found to be very similar.

Three of the people who worked on PPA also worked on P8. P8 was another system for planning protein purifications but P8 used a different planning technique to PPA and was designed for a different type of User.

P8 was a planner "...that planned and provided recommendations for the chromatographic stage of protein purification. The system was developed to meet the requirements for information and decision support among *workers in biochemical research laboratories.*" [69] P8 provided advice on which technique to use, the running conditions for each operation, and measurements to be taken between each purification step. The system was restricted to liquid chromatographic techniques and membrane-bound proteins.

P8 was primarily intended as a research prototype to explore knowledge representation and algorithms for protein purification planning. Thus only "...a limited effort was spent on the User interface" [69].

The techniques for protein purification affect the physical properties of a sample differently but a protein sample requires carefully controlled conditions. Thus a sample has to be adjusted between the steps of a purification but these adjustments need to be kept to a minimum. One of the aims was that P8 would design a plan that involved the minimal number of adjustments.

P8's knowledge base was structured around a set of partial plans. These were represented by a 'technique and decision-point tree', which covered different outcomes of operations in the plan. The knowledge in each decision-point was represented as a rule set. The nodes of the tree corresponded to purification techniques. P8 acted as follows:–

1. P8 began by asking questions about the detergent, the amount of sample and protein stability. A rule base then determined what strategy to use. A strategy was essentially

an ordering of partial plans.

2. For each partial plan purification techniques were selected, using the corresponding decision tree, until the User acknowledged that the target protein was sufficiently pure. For each purification technique that was selected :−

   (a) A check was made that the sample met the requirements for the technique. Constraint violations were passed on to a module which recommended the adjustments that were necessary before the technique could be employed.

   (b) The running conditions were then selected using a rule base.

   (c) The technique was then performed.

3. Once the target protein was sufficiently pure P8 halted.

Section 13.5 contrasts the two planning techniques of PPA and P8.

8.5 Pesticides

Researchers at the University of Missouri produced an knowledge-based system called PIA (Pesticide Identification Assistant) [70] designed to assist an analytical chemist with the identification of pesticide residues in food products using gas-liquid chromatography. [70]

The intention was that PIA would be able to identify approximately 300 different potential residues found in 234 different food items based on retention data, analysis information and heuristic rules based on food types. The analysis information was used to constrain possible identifications of a chromatographic peak. There was a possibility that a chromatographic response under investigation was due to a residue from outside the set of compounds that the system recognised. Therefore a definite identity could not be established and a variety of heuristic rules allowed the selection of the most probable identity from among the possibilities. Help functions were provided throughout the system. Explanations and suggestions were only available after the probable identification was made. [70]

PIA operated by using Prolog's inference engine directly to perform a number of searches [70] :−

1. Prior to attempting residue identification, it performed a search based on the food item in question. This had the effect of gathering heuristics relevant to the sample type into working memory.

2. It tried to identify the residue(s) by:−

   (a) searching for a single residue that could account for all the data.

(b) searching for multiple residues, if no single residue could account for all the data, and finding a possible residue for each peak on the chromatogram.

Rather than always performing a complete search of possible residues, the system first performed a search among only those residues previously identified in the sample food type, continuing on to a complete search only if the restricted search failed to generate an identification.

3. Finally it searched for information to use in explanations of results, and for suggestions for confirming analyses. However the system was incapable of expanding upon these suggestions or explaining why they were appropriate.

"Currently, eleven of the 234 food items are represented in the system, as well as 18 of the 300 possible compounds. Identification of items drawn from this initial sample is 100%. Obviously, such performance will degrade as the complete database of food items and compounds is incorporated into the system." [70]

8.6   Fossil Fuels

Four of the expert systems for chromatography were limited to fossil fuels. Two of these tried to automate the interpretation of data from gas chromatography-mass spectrometry. This is a highly skilled, labour intensive task which is used to identify compounds in fossil fuels.

Researchers at the Georgia Institute of Technology and the University of Alabama, both in the USA, suggested design criteria for a gas chromatography-mass spectrometry (GC/MS) based expert system for arson analysis. Since petroleum distillates are "the overwhelming choice of the arsonist" [71] a system specific to this group of compounds was designed. "Advantage was taken of the high selectivity of mass spectrometry towards aliphatic and aromatic hydrocarbons" [71].

"In the USA about one out of five reported fires is of suspicious origin. ... The chemical analysis of residual accelerant is an integral part of most investigations into the origin and cause of a fire. ... The power of GC-MS in accelerant analysis has been demonstrated convincingly on real world samples." [71] The system was developed because of the limitations of human chromatogram interpretation. Data interpretation had almost always been carried out by pattern recognition of chromatograms. This had involved the analyst looking for groups of peaks that were characteristic for fuels commonly encountered. In many cases, chromatograms of suspect samples had been simply arranged side by side to chromatograms of standards. There are many human and technical limitations of chromatogram interpre-

25

tation. For example it was obvious that factors such as retention time shifts or a mismatch of the chromatographic scale could have caused problems. [71]

Thus a system was designed which tried to automate the process. The system began by acquiring GC/MS data. It tried to categorise chromatographic profiles by establishing the presence of key components. Comparison was made to data from standards stored in the memory of the computer. Retention time information as well as spectral criteria was used. The system then printed a table giving extensive information about the presence or absence of the key compounds and denoting the accelerant standard which was the best match to the unknown sample. The system could then be instructed to print out ion chromatograms for the sample and simultaneously display the best match. [71]

The researchers were aware that their system had limitations: "Chemical analysis of fire debris remains a challenging task. There is no substitute for human intuition and a machine will never replace a chemist. The program does fail on occasion for samples that produce overwhelming interferences or have other peculiarities. ... The program ... is of significant help in routine analysis." [71]

Workers in the USA at the University of Houston and at the company Exxon developed an expert system for the interpretation of the GC/MS spectra of fossil fuel distillates.

Interpretation "... of GC/MS spectra is still very labour intensive, in particular for non-routine samples containing hundreds of components, such as fossil fuel distillates .... Manual identification ... requires a highly skilled professional and is extremely time consuming." [72] Even when a commercially available data system is used, identification "... has to be based on a large number of experienced based, inter-connected decisions, that include a detailed examination of the mass spectra, considering the relative order of elution and a very critical evaluation of the library response" [72].

The Exxon system used the commercially available identification capabilities but it extended them by defining limits of acceptability for the library identification yielded by those systems. These limits were based on *automatic* examination of the spectra, on retention time considerations and on the quality of the matching itself. The system was shown "... to be highly reliable in correctly identifying components in complex hydrocarbon streams, and doing so at a fraction of the time that would be required by a human expert" [72].

Two expert systems for diagnosing the insulation condition of power transformers were developed: TOGA [73] and a system from Xian Jiaotong University in China [74]. The latter worked from the measured chromatograph for the transformer oil. The insulation condition of power transformer was diagnosed as either normal, in need of checking and repair, or the system recommended that the diagnostic period needed shortening.

9 Systems Specific to a Particular Chromatographic Method

This section corresponds to the branch of the taxonomy which is entitled 'Specific to Chromatographic Method' on Figure 1.

9.1 Systems for Ion Pair Chromatography

Three expert systems were developed for ion pair chromatography. One suggested a suitable set of method conditions whereas the other two were mainly concerned with optimisation.

The expert system for suggesting a set of suitable method conditions was developed at the University of New South Wales for ion chromatographic methods using dynamically coated ion-interaction (also known as ion-pair) chromatography [75].

Ion-pair chromatography involves the use of apolar stationary phases with eluents containing a hydrophobic ion of opposite charge sign to that of the analyte ions, that is a counter ion. Two alternative methods exist: 'permanent coating' or 'dynamic coating'. The former requires that the column is first equated with the counter ion, which is then absent from the eluent during the analysis step. The latter uses an eluent containing the counter ion for both the column conditioning and analysis steps. The knowledge acquisition process for the expert system revealed that permanent coating represented less than 20% of ion-pair applications and it was therefore decided to concentrate on the dynamic coating procedure.

The actual rules for defining the method conditions were divided into four stages:–

Selection of the method type This part ascertained the class of ion to be analysed and selected a suitable preliminary method.

Selection of the column The column was selected by examining the nature of the sample and its matrix.

Selection of the eluent The eluent was determined chiefly by the class of ions for analysis.

Selection of the detector The detector was chosen on the basis of the properties of the solute ions and the availability of detector types.

A report was shown to the User which summarised the chosen method conditions. On viewing this it was possible for the User to modify any answers and the system would produce new conclusions. Throughout the consultation help was available for each query that was made of the User.

The system allowed the User to change the conditions for activating some of the rules and the conclusion made by these rules. If any such modifications were made then the rule

base was reconfigured to account for these changes. Thus the User could customise the system to include specific preferences.

Researchers in Brussels developed one of the two systems concerned mainly with optimisation [76]. The main purpose of the research was to explore how to introduce optimisation procedures into the expert system rather than to build a system for the domain. Thus only some basic drugs were considered, only one ion-pair reagent was investigated and the detector mode was limited to UV.

The system was based on LABEL. Like LABEL it gave advice on the detection and mobile phase but it also included an experimental design module. The systems comprised four modules:−

Introductory Module This checked whether ion pair chromatography was necessary and possible, and selected the detector settings.

First Guess Module This advised on the mobile-phase conditions. It began by deciding whether a first guess was indeed possible. If the situation was too complex then it decided that an experimental design should be applied immediately.

Optimisation Module This used an experimental design when an initial guess approach was likely to give unsatisfactory results, that is when more than one substance was present. A simple design was chosen because the aim of the expert system was not to establish the best optimisation method but to investigate the feasibility of intergrating optimisation methods in to an expert system.

The User then tested the method and entered the results in to the system. The system evaluated the results and either advised that that the method be used, advised that a good separation could not be obtained using the system selected or adapted the method by calling the fourth module.

Adaptation Module The input to this module was the chromatographic results of the method proposed by the first guess or optimisation modules. The module either adapted the method or advised the User that a good separation could not be obtained using the system selected.

The system was validated by following the advice of the system for sixty synthetic mixtures created using a random number generator. "The rate of success was satisfactory" [76].

A rule based system was developed at Texas A&M University in the USA for the determination of solute types in unknown sample mixtures as a first step of selectivity optimisation parameter selection in reversed-phase ion-pair chromatography [77a].

This work illustrated that the nature of the solutes in totally unknown aqueous sample mixtures can be determined by the combination of an experimental procedure and a rule-based retention-shift evaluation strategy that is implemented in a computer program. The approach suggested did not require peak tracking or the use of any other extra-chromatographic information. Other expert systems for selectivity optimisation need extensive *a priori* chemical information about the sample components in order to make predictions about the expected retention behaviour.

The system comprised three parts:–

Data Input The User entered the retention times of all the peaks observed, together with the respective pH and pairing-ion information. The retention times and chromatograms could be entered in any order.

Type Evaluation The program tried to label each peak as either SA, SB, WA, WB or N where N=neutral, S=strong, W=Weak, A=acid, and B=base. The retention data of all peaks from all the chromatograms were compared. (At least two chromatograms were required.) Initially all five possible solute types were assigned to all peaks in all chromatograms. Then, recursively using the retention-shift rule set, the program eliminated the impossible solute-type designations for all the peaks.

List Results The solute-type assignments were then displayed to the User. The User could request the list of rules (and their hierarchical sequence) which were used in the solute-type assignment process, for each individual peak.

When tested with a variety of complex samples, the program correctly identified the type of the first and last eluting peaks and excluded the impossible solute-type designations for the rest of the peaks.

9.2   System for Thin–Layer Chromatography

A system was developed for the identification and separation of samples by thin-layer chromatography(TLC) [77b]. The system was based on an extensive database of pharmaceutical compounds which could be modified by the User. The system was useful for TLC screening and for selecting a TLC separation system for a given sample. [7]

10   Synopsis of Behaviour and Users

Table 1 compares the predominant behaviour of expert systems for chromatography [3] and the types of User for which they were intended. It indicates that only three out of the 18 systems were targeted at Users who were novices in the particular domain of each of the systems. Thus most expert systems for chromatography were designed for scientists with some experience of chromatography. The table also shows that the systems exhibit four predominant behaviours: just over a third of them (7 of 18) plan the whole or part of a chromatographic separation, whist the remainder either predict the values of experimental data, diagnose problems or identify sample components.

11   How the Knowledge was Acquired for the Systems

The process of acquiring the knowledge needed for an expert system is called knowledge acquisition. The knowledge acquisition process is usually divided into three stages: deciding what knowledge is needed, variously referred to as the definition stage or initial analysis; getting knowledge, predominantly from human experts, and interpreting it, usually called elicitation; and 'writing' the knowledge in the internal language of the system, encoding it, usually called representation. Knowledge acquisition is a notoriously slow process and has become known as the 'bottle-neck' in the process of developing expert systems. [78]. It has accounted for a significant proportion of the total work involved in developing expert systems for chromatography. For example the knowledge acquisition phase for the PPA project required approximately six person-months of expert time and in terms of calendar time the project required one year. Future developers of expert systems for chromatography should allow for the knowledge acquisition 'bottle-neck' when planning their projects.

11.1   Sources of Expertise and Knowledge

Some of the systems relied on the literature as the main source of expertise. This was the case for the expert system for planning separations of steroids that was implemented as a decision tree; the rules were induced from specific examples of validated results of successful separations reported in the literature. Examples of how the expert system should behave were taken and then generalised to a higher-level rule that could guide the inference procedure [63]. ESP used rules drawn from a standard text book [49].

---

[3]The systems that were designed for several stages of the chromatographic process (ESCA, ECAT and ESC/ESLC) are not included because they do not have a predominant behaviour: they comprise several subsystems which behave differently.

However the knowledge and expertise for most of the systems was acquired with the help of one or more experts. For example the sources of expertise for the PPA system were specialists on individual separation techniques and researchers on particular classes of proteins [68]. The structure of the knowledge for ECAT was " determined through informal interviewing of experts, and by accumulation of ideas and experience from chromatographers through the literature" [75].

One of the problems of working with experts is that they are usually busy people with little spare time to devote to the development of an expert system. The developers of some of the systems found ways of minimising this problem. For example the researchers, based in New South Wales, that developed the expert system for ion chromatographic methods using dynamically coated ion-pair chromatography [75] developed their system from an extensive database of previously published ion chromatography methods. "The database was searched systematically to find the most commonly used method conditions for different applications. These conditions were then examined by the expert and rules generated for the expert system ... This method considerably reduced the amount of time required from the expert but still resulted in the development of a competent expert system." [75]

When help is sought from experts for an expert system project it is important that the most suitable experts are chosen. For a company project this may involve deciding whether to use experts from within the company or from outside. Such a decision was made by the developers of the Upjohn troubleshooting system. The sources of the expertise for the Upjohn troubleshooting system were publications and chromatographic experts in the Pharmaceutical Quality Control Division of the Upjohn company [58]; since the system was to be an 'in-house' system, 'in-house' expertise was adequate.

Using more than one expert in a project can lead to problems. Experts may disagree or use different approaches. For some of the systems the number of experts involved was deliberately limited to one. The developers of CRISE stressed the importance of consistency in the knowledge. They achieved this for CRISE by obtaining almost all the knowledge from a single expert. Knowledge from another source was included in the *periphery* of the system. [29] In the ESCA project four experts separately contributed knowledge for each of the four subdomains. "Although more than one expert could have contributed to a domain, it was decided to avoid any discussion between experts."[18]

Thus previous experience suggests that developers of expert systems for chromatography must carefully consider the number and choice of experts to be consulted, and decide whether 'in-house' or external expertise is to be sought.

11.2 Elicitation Techniques Employed

For many of the systems the literature does not describe which elicitation techniques were used. However interviewing was used for the majority of those for which the techniques used are described. For a few systems other techniques were employed but they were not described in any detail. For example the knowledge elicitation for the ESCA DASH system was mainly done by interview but other techniques were also used. However they were only described in the vague statement "Sometimes information was exchanged in written form" [79].

The literature on knowledge engineering describes in detail a wide range of techniques for knowledge elicitation that can supplement or replace interviewing [80] [78] and gives advice on how and when a given technique should be employed. However the literature on expert systems for chromatography does not mention any of these techniques. Despite the emphasis placed on the techniques by the literature on knowledge elicitation there is no evidence that developers of expert systems for chromatography have used them, considered but rejected using them or were aware of them. Future developers of expert systems for chromatography should consider using the various techniques and describe and justify their choice of technique(s) in any published work.

11.3 Machine Induction and Neural Networks

There is an alternative to the developer of an expert system trying to elicit the expertise from an expert; a computer can be programed to develop the expertise. This can involve a computer being trained to develop a set of rules or a neural network.

Rule Induction Rule induction involves a computer system being provided with a number of cases and the system using inductive techniques to create its own rules from the cases.

Neural Networks A neural network comprises a layer of input elements connected to one or more hidden layers of elements, the last of which is connected to a layer of output elements. Data from the input layer are transmitted through the hidden layer(s) to the output layer in a manner determined by the strengths of the connections between elements. Training modifies the connection strengths, and it is the set of their values in the trained network which represents the relationship between the input and the output.

There are no reports in the literature to suggest that induction was considered as a method for acquiring the expertise for any of the expert systems for chromatography that

had been implemented.

In the conclusion of the paper on the expert system developed in New South Wales for ion chromatographic methods using dynamically coated ion-pair chromatography [75] the authors stated their intention to investigate the potential of neural networks and the Quinlan rule generating system [81] for knowledge acquisition. They proposed to compile rule bases for the ion exclusion and ion exchange methods of chromatography They stated their belief that since these methods are used more frequently the task of manually performing a statistical analysis on a database of literature methods, as was done for the ion pair mechanism, would be extremely difficult due to the large amount of data.

12   The Knowledge Representations Adopted

12.1   Formalisms for Knowledge Representation

The literature on knowledge representation describes a number of formalisms that can be used such as productions rules, frames, semantic nets and the predicate calculus [14] [82]. [3] includes lucid examples showing how all four formalisms can be used to represent chemical knowledge.

Nearly all the expert systems for chromatography used rules. Despite the emphasis placed on the alternatives to purely rule based systems by the literature on knowledge engineering [14] [12] there is no evidence that developers of the rule based systems for chromatography considered using them or were aware of them. Future developers of expert systems for chromatography should consider using the various formalisms and describe and justify their choice of formalism(s) in any published work.

The following sections discuss those few systems which did use other formalisms.

12.1.1   Frames

Frames [83] "... as data structures for storing expectations about typical objects and events have become quite widespread in artificial intelligence applications" p169 of [12]. The "... additional structure which can be represented or imposed by frames has considerable value. One demonstration of this is the reconstruction of knowledge bases originally expressed in unstructured production rules into frame systems and the consequent improvement in system understanding and ease of maintenance. ... The use of frames to represent knowledge about structured domains or structured knowledge continues to increase in popularity ... " p99 of [14] Frames "... are mostly used in conjunction with other representations, such as production rules" p169 of [12].

33

It is therefore surprising that the use of the increasingly common combination of production rules and frames was only reported for a few of the expert systems for chromatography. The ESCA system for precision testing was one of these. The usual features of frames such as instantiation, inheritance and demons were used [37] [36] [41].

12.1.2 Augmented Transition Networks

The two systems for planning separations of steroids were designed to tackle the same problem in the same way but using different formalisms to represent the knowledge. The first represented a decision tree by a series of rules. The second used an augmented transition network (ATN) [4] and was developed to show that the ATN formalism was "...a more efficient structure for representing the knowledge base of the ...system" [64]. However no evidence was presented that proved this. Furthermore the comparison of the ATN and rule formalisms ignored the possibility of using some of the main control strategies for rule based systems, such as allowing rules to be fired more than once, and using meta and context sensitive rules. These omissions render the work as irrelevant to the contemporary debate as to which formalism is most efficient for knowledge representation. The work failed to prove that ATNs should be used for future expert systems for chromatography.

12.2 Uncertainty

Uncertainty exists whenever expert systems have to make non-categorical decisions. Uncertainty can be represented using probability theory, certainty factors, fuzzy logic and the Dempster-Shafer theory of evidence [12]. The literature mentions the representation of uncertainty in only a few of the systems. The Upjohn troubleshooting system developed by [58] was implemented in M.1 which is rule-based expert system software which deals with uncertainty by the use of MYCIN style certainty factors. The capability of M.1 in dealing with uncertainty was used to weigh evidence and thereby establish priority for efficient troubleshooting. PROTEIN used certainty factors as well. PIA handled uncertainty "...in an ad-hoc fashion, basically by noting any causes for uncertainty in working memory, assigning a subjective weight, and accumulating the weights associated with each proposed identification" [70]. The Pascal reimplementation of the ESCA REPS also incorporated weighting factors [84].

---

[4]For a description of ATNs see [64].

12.3　Representations not Specific to Expert Systems

Some standard software development representations were used. Data flow diagrams and state transition diagrams were used for the ESCA REPS [36].

Flow charts were used in the development of ESC/ESLC [45] [46]. In academic computer science circles flow charts are associated with poor implementation practice [85]. They "...are very limited in their capability for modularisation and structuring ..." [86]. It is therefore surprising that flow charts were used in the late 1980s to design a computer program, especially one which was an expert system. Flow charts should not be used to design future expert systems.

13　The Software Tools and Architectures Used

This section lists the software tools and architectures used to implement expert systems for chromatography.

13.1　Software Languages and Expert System Shells and Tools

Table 2 shows which expert systems for chromatography used which languages [5] and the aspects of the each system for which a language was used. The table shows that those systems in which the coding was not in one language alone, the AI languages were mostly used for reasoning, and never for numerical processing, where as conventional languages were mostly used for numerical processing and never for reasoning. This reflects the strengths of the AI languages: they make it possible to represent knowledge quickly and conveniently p95 of [87] making it easier to code reasoning. The way in which the conventional languages were used also reflects one of their strengths: they usually provide better facilities than the AI languages for numerical processing.

The aim of Table 3 is to show which expert system shells and tools were used in the development of the systems. [6]

---

[5]Prolog and LISP are referred to as artificial intelligence (AI) languages. Third generation languages, such as BASIC, FORTRAN, C and Pascal, will be referred to as conventional languages.

[6]It does not try to review the shells and tools used or to classify them; they are presented in alphabetical order. No description of the shells and tools is given because of the differences between the various versions used for different chromatography systems.

13.2 Hypermedia

The developers of CRISEBOOK [30] discussed three problems with 'traditional' expert systems development software:–

1. Updating and maintaining systems based on this type of software is difficult. "With the conventional present shells it is difficult for the User to make changes owing to the way rules are connected ... Adding or changing rules could have repercussions all over the system."
   [30]

2. To use some of this software requires advanced skills of knowledge engineering.

3. Systems based on this type of software often have a poor User interface.

The developers of CRISEBOOK investigated whether Hypermedia could be used as an alternative to conventional expert systems development software. "Hypermedia is the combination of multimedia and hypertext. In a hypertext document the User does not have to read all information in a sequential way, but is guided by his needs and interests by pieces of information that are linked to each other. Multimedia is a collection of tools producing graphics, sound, and animation, to present data in a more flexible way." [30].

The work on CRISEBOOK did not show whether Hypermedia would make update and maintenance of expert systems easier. The work was inconclusive as to whether advanced skills would be needed to build expert systems with Hypermedia, although the developers of CRISEBOOK reported that the 'scripts' that are developed for Hypermedia are "... entirely readable even for those unfamiliar with programming." [30] CRISEBOOK suggested that Hypermedia would enable better interfaces to be produced because data can be presented in a more flexible manner and because much of the programming associated with interfaces is already present in Hypermedia.

Problems with expert system development software can not be considered independently. Changing the development software to ease one problem may affect the severity of others or even create new ones. Although CRISEBOOK showed that Hypermedia could help to ease two of the afore mentioned problems, it did not prove that Hypermedia would be a better alternative *overall* to 'traditional' expert systems development software.

13.3 Spreadsheets

Two systems were developed by combining expert systems development software and spreadsheets packages.

In REPS spreadsheets were programmed to produce the necessary statistics. The spread-sheet package used was Lotus 1-2-3 (1986) (Lotus Development Corporation, Cambridge, MA, USA); the expert system tool used, Goldworks, was able to communicate with this. Rules, defined using Goldworks, set up spreadsheets into which the analyst input data, and the rules then interpreted the data processed in the spreadsheet. [36]

In CRIPES VP-Exert communicated with spreadsheets written in VP-Planner (Paper-back Software) [50].

13.4 Architectures

It was vital for the proposed integration of the ESCA repeatability and ruggedness systems that all the modules of the integrated system used the same concepts as the basis for reasoning so that flexible communication was possible. Simple transfer of files between modules would have been insufficient for two reasons [34]:−

1. Most of the facts produced by one module had to be available to all other modules.

2. The modules were not to be consulted in a standard sequence. Trying to implement all possible consultation sequences would have become unrealistically complex.

It was decided to merge all the existing concepts in one common data structure that formed the basis for all the systems. "The blackboard architecture was chosen [7] because it allows integration of modules which use different interfacing or problem solving techniques" [34].

"Blackboards are well known artificial intelligence techniques for the integration of expert systems" [34]. They allow several modules to communicate with each other via a common data structure, the blackboard. In a blackboard architecture "...the knowledge sources trigger themselves when the state of the blackboard is such that they can contribute to the solution of the problem. ...In an ideal situation several knowledge sources can be activated at the same time. The blackboard architecture offers the possibility for parallel processing." [34]

The ESCA integrated system did not utilise the possibility for parallel processing offered by a blackboard. In fact all but one of the expert systems were designed for serial processing. The exception was ESP which was implemented as two concurrent processes. The interface, implemented in C, was a parent process and the 'intelligence' of the system, the Prolog

---

[7] The architecture used for the ESCA integrated system varied slightly from a traditional blackboard in that a supervisor module controlled when and which knowledge sources were to be activated.

code, was the child process. Calls to the UNIX operating system kernel allowed concurrent processes to execute, communicating through UNIX pipes. [49]

13.5   Expert Planners

Some of the systems can be thought of as planners. For example PPA was a planner that planned a purification process from the original source to the final product. Thus it was a deductive planner as opposed to a reactive planner. Reactive planners, such as P8, immediately respond (without much deduction) to input without forming a complete plan on how the goal should be reached. A deductive planner was developed because reactive planners are "faster but in many senses 'dumber' than deductive planners" [68]. A deductive planner had the advantage that it was possible to plan in advance and, where possible, optimise the plan, and the ability to deal with unanticipated purification tasks by reasoning from general rules. However the designers of the system were aware of the major drawback of deductive planners: they can not provide dynamic responses to different outcomes of *applied* operations. P8 was designed as a reactive planner because reactive planners do not have this drawback. "Armed with a certain amount of knowledge . . . one can devise a separation plan in advance. The efficiency of the plan will depend on the actual outcome of the *individual* steps. When the results are known, one is better equipped to plan subsequent steps" [69].

14   Validation and Evaluation

"The increasing application of expert systems in routine tasks, and the consequent gradual automation of knowledge intensive work, has led to a need for some kind of assurance of the quality of . . . " p174 of [88] the resulting systems. Hence expert systems need to be validated and evaluated.

14.1   How the Systems were Validated

Validation involves testing by the developers to estimate the faithfulness with which the expert knowledge has been captured. This involves asking questions such as the following. Does the system offer advice that is appropriate as judged by experts? Is the reasoning correct, consistent and complete? Are the explanations adequate? Do the test problems cover the domain?

Validation is an important stage in the development of an expert system. Unfortunately, due to the nature of expert systems, the standard principles and practices of software

engineering for validation are not directly applicable. The workers on the ESCA project stressed this: "Expert systems can only be expected to be useful in practice if they are reliable. Conventional software products can be tested thoroughly through a number of standard procedures. However, for testing expert systems no standard procedures exist" [21].

The validation of only a minority of the expert systems was reported in the literature. For those systems for which a validation was reported one of three approaches was adopted:–

1. Advisory systems for planning were validated by comparing the advice given by the systems with the advice given by experts. For example RES was validated by using 11 test cases, each of which was a liquid chromatography method for the separation of a pharmaceutical sample. For each method an unbiased expert's factor choice was listed before the expert system was consulted. The factors chosen by the expert and system were then compared.

   A slight variation of this approach was used for PPA. Various domain experts were asked to pose problems for the expert system and then decide whether the resulting advice given by the system was acceptable. Twenty different proteins were selected by as many researchers. PPA was then run for each one to produce a plan for its purification. Each of the plans were then discussed with the respective scientist. [68]

2. Advisory systems for novice Users were validated by comparing the performance of the User with and without the help of the system. For example the designers of ESP suggested that "Future [8] validation could take place by asking two similar student groups to solve a set of representative separation problems with and without the aid of ESP" [49].

3. Predictive expert systems were validated by comparing the the results predicted by the systems with corresponding experimental results. To validate CRIPES the retention of a number of test compounds not previously examined were measured and compared with those calculated by CRIPES. The predictive power of HPLC-METABOLEXPERT was investigated for seven parent "... molecules and their eight main metabolites by measuring their reversed-phase retention data using various reversed-phase columns and conditions" [62]. The measured and predicted retention data were then compared.

All three approaches involved using some metric of similarity to decide whether two sets of results were similar enough to allow the system to be judged a success. For example for

---

[8] The performance of ESP had been measured by a HPLC technician answering ESP's interview questions for a series of compounds from the literature. [49]

RES the criterion for deciding whether or not a factor choice made by the expert system was acceptable was that the choices of the expert and expert system did not differ by more than two factors. The number of factors selected was allowed to differ only by one: the number of factors was important for the final acceptance of the system because the number had to be small but sufficient.

All the approaches to validation required that a set of tests was selected. The tests were selected in different ways:–

Random sets The tests for the validation of LABEL were randomly chosen. "Initially[9], 50 pharmaceutical formulations were selected at *random* from the Belgium Drug repertory 1987" [26].

Sets chosen by various experts The set of tests for the validation of PPA was chosen by a number of domain experts.

Representative sets The compounds selected from the literature for use in the validation of ESP (see above) were "...*representative* of cases covering the range of possible conclusions that ESP can arrive at" [49].

Sets including specific tests For example the compounds used in the trials of CRIPES "...included some that were selected to test *specific aspects* of retention prediction" [50].

The 'random sets' and 'sets chosen by various experts' can considered as black box testing and the others can be viewed as clear (or white) box testing.

A notorious problem encountered when developing software is that when it is repaired, following the discovery of an error, the repair itself may introduce new, undetected, errors. This is particularly relevant for expert systems incorporating production rules where adding or changing rules can have repercussions all over the system. The developers of the ESCA expert systems were conscious of this when planning the validation of their systems. To ensure that changes to a knowledge base did not produce unexpected side effects, it was necessary to define a so-called regression test. The expert selected a number of test cases, representing a broad range of possible cases, that were solved by the expert system after each major revision. If the solutions to the test case did not remain the same it was assumed that changes had been made that adversely affected previously evaluated knowledge. [18] "The limited nature of the regression test is of course no guarantee that unexpected results will not appear but the choice of a good regression test set will minimise the chance" [18].

---

[9]Six of these were ignored because they contained compounds that could not be determined by UV or electrochemical detection. [26]

14.2　How the Systems were Evaluated

Evaluation is an investigation of the extent to which the system is useful and acceptable to the users. It involves asking questions such as the following. Is the input and output convenient? Is the nature of the interaction appropriate? Is the system fast enough? Does it fit in with the normal work pattern? Does it threaten the User? Is it being used? Will it be used? Have the goals been achieved? Are there hidden costs?

Given that these questions are so fundamental it seems remarkable that for all but one of the expert systems for chromatography there were no reports in the literature to suggest that any attempt had been made to evaluate the systems. Unless the results of evaluations are published it is impossible for people who were not involved in the development of a system to know how effective a given system was. Future developers of expert systems for chromatography should evaluate their systems and publish the results.

Only the evaluation of the ESCA systems was reported in the literature. It involved testing the expert systems in practical situations in order to evaluate their performance in daily practice. Generally, these tests were performed by external evaluators [21]. An example of such an evaluation, for two of the ESCA subsystems DASH and DASH$^/$, is described in [7]. The literature on the ESCA evaluations included some interesting comments on the process. For example the evaluators of REPS felt the measured performance was ". . . acceptable for this method, but were pleased to find that the system suggesting actions that could further improve the method. This is a typical advantage of expert systems over conventional software packages. . . . The evaluation of this system proved extremely valuable as it resulted in several additions which enhanced the software considerably" [84].

The ESCA evaluations were carried out by different persons, ranging from experts in method development to students with little or no experience. During the evaluation phase of all the expert systems it became clear that the attitude towards expert systems is strongly dependent on the expertise level of the evaluator. The accessibility of the specialists' expertise was clearly appreciated by inexperienced Users. Experienced Users could appreciate the quality of advice given by the systems. However when the strategy implemented in a system did not agree with the expertise of experienced Users they became dissatisfied with the system because their own experience, probably better adapted to their specific situation, was not considered by the system. [21]

15   Conclusions

Expert systems for chromatography have been reviewed. Most are designed for scientists with some experience of chromatography. The predominant behaviour of each is either to plan the whole or part of a chromatographic separation, predict the values of experimental data, diagnose problems or identify sample components.

The domains of the systems vary in their coverage of:−

- stages of the chromatographic process.

- compounds to be separated by the chromatograph.

- chromatographic methods.

Thus the systems cover a large number of different subdomains of chromatography. A taxonomy has been proposed that allows expert systems for chromatography to be classified and facilitates an understanding of their inter-relationship.

The literature suggests that the most successful expert systems for chromatography are those which tackle specific aspects of chromatography rather than those which tackle large parts of this domain. Expert system projects which tackle domains as large as that covered by the ESCA project must be managed effectively, particularly if the objective is to produce an integrated system.

Previous expert systems for chromatography show that spreadsheets can be embedded within expert systems for chromatography but fail to prove that Hypermedia is a better alternative *overall* to 'traditional' expert systems development software or that ATNs offer a formalism which is more efficient than rules for knowledge representation.

15.1   Conclusions on the Engineering

There are a number of aspects of the software and knowledge engineering that were approached in a similar way for most of the expert systems for chromatography. From this emerges the following stereotypical portrayal of the engineering for a previous expert system for chromatography. The knowledge acquisition accounts for a significant proportion of the total work, and involves interviewing one or more experts. Rules are used to represent the knowledge and they are encoded using an artificial intelligence language or expert system shell or tool. The resulting system is validated but is not evaluated.

Despite the literature on knowledge engineering stressing the importance of evaluating systems and advocating a wide range of techniques for knowledge elicitation and a wide range of formalisms for knowledge representation, developers of expert systems for this 'real world'

domain have not responded: nearly all of the developers use interviewing for elicitation and rules for representing knowledge and they do not evaluate their systems. Further research is needed to determine why because developers of expert systems for chromatography often do not justify their decisions on engineering matters in their published work.

15.2 Summary of Main Conclusions

- A taxonomy has been proposed that allows present (and future) expert systems for chromatography to be classified and facilitates an understanding of their inter-relationship.

- The literature suggests that many of the ideas advocated by knowledge engineers are not being used by developers of expert systems for chromatography.

- Too often developers of expert systems for chromatography do not justify their decisions on engineering matters.

16   Recommendations

Future developers of expert systems for chromatography should:–

- Either plan their projects so as to allow sufficient time for 'the bottle-neck' in the process of developing expert systems, the knowledge acquisition phase, or consider using other techniques such as neural networks or rule-induction to automatically acquire the expertise.

- Carefully consider the number and choice of experts to be consulted, and decide whether 'in-house' or external expertise is to be sought.

- Consider using the various techniques for knowledge elicitation and formalisms for knowledge representation; and justify their choice in any published work. At the very least they should consider the use of the increasingly common combination of production rules and frames and avoid out-dated formalisms such as flow charts.

- Evaluate their systems, in addition to validating them, and publish the results. Unless the results of evaluations are published it is impossible for people who were not involved in the development of a system to know how effective a given system was.

References

[1]   V.Jakus, Collect. Czech. Chem. Commun., 57 (1992) 2413.

[2]   T.P. Bridge, M.H. Wiliams, and A.F. Fell, Chemistry in Britain, (November 1987) 1085.

[3]   J.Klaessens, and G. Kateman, Fresenius Z Anal Chem., 326 (1987) 203

[4]   A.P. Wade, S.R. Crouch, and D. Betteridge, TRAC, Trends in Analytical Chemistry, 7 (1988) No.10, 358.

[5]   S.A. Borman, Analytical Chemistry, 58 (1986) No.12, 1192.

[6]   J.L. Glajch, LC-GC 6 (1988) No.1, 30.

[7]   T.Hamoir, and D.L. Massart, Advances in Chromatography, 33 (1993) 97.

[8]   V.R. Meyer, Practical High-performance Liquid Chromatography, John Wiley & Sons, 1988.

[9]   J.A. Jonsson, Chromatographic Theory and Basic Techniques, Dekker, New York 1987.

[10]   A. Braithwaite, and F.J. Smith, Chromatographic Methods, Chapman and Hall, London, 1985.

[11]   C.F. Poole, and S.A.Schuette, Contemporary Practice of Chromatography, Elsevier, Amsterdam, 1984.

[12]   P. Jackson, Introduction to Expert Systems, 2nd Ed., Addison-Wesley, 1990.

[13]   A. Hart, Knowledge Acquisition for Expert Systems, Kogan Page, London, 1989.

[14]   G.A. Ringland, and D.A. Duce, Approaches to Knowledge Representation. An Introduction, Research Studies Press Ltd, Taunton, Somerset, England, 1988.

[15]   P. Harmon, R. Maus, and W. Morrissey, Expert Systems. Tools & Applications, John Wiley & Sons, 1988.

[16]   I. Dickinson, Laboratory Robotics and Automation, 4 (1993) 85.

[17]   L.R. Snyder, J.W. Dolan, and D.C. Lommen, Journal of Chromatography, 485 (1989) 65.

[18]   J.A. van Leeuwen, L.M.C. Buydens, B.G.M. Vandeginste, and G. Kateman, TRAC, Trends in Analytical Chemistry, 9, No.2, 49.

[19]   D. Goulder, T. Blaffert, A. Blokland, L. Buydens, A. Chhabra, A. Cleland, N. Dunand, H. Hindriks, and G. Kateman, Chromatographia, 26 (1988) 273.

[20] M. Mulholland, N. Walker, J.A. van Leeuwen, L. Buydens, F.Maris, H. Hindriks, and P.J. Schoenmakers, Mikrochim Acta [Wien], II (1991) 493.

[21] L. Buydens, P. Schoenmakers, F. Maris, and H. Hindriks, Analytica Chimica Acta, 272 (1993) 41.

[22] T. Hamoir, M. De Smet, H. Pyrins, P. Conti, N.V. Driessche, D.L. Massart, F. Maris, H. Hindriks, and P.J. Schoenmakers, Journal of Chromatography, 589 (1992) 31.

[23] P. Conti, T. Hamoir, M. De Smet, H. Piryns, N.V. Driessche, F. Maris, H. Hindriks, P.J. Schoenmakers, and D.L. Massart, Chemometrics and Intelligent Laboratory Systems, 11 (1991) 27

[24] F. Maris, R. Hindriks, J. Vink, A. Peeters, N.V. Driessche, L. Massart, Journal of Chromatography, 506 (1990) 211

[25] G. Musch, M.M. De Smet, and D.L. Massart, Journal of Chromatography, 348 (1985) 97.

[26] M. De Smet, A. Peeters, L.Buydens, and D.L. Massart, Journal of Chromatography, 457 (1988) 25.

[27] M. De Smet, G. Musch, A. Peeters, L. Buydens, and D.L. Massart, Journal of Chromatography, 485 (1989) 237.

[28] G.Musch, and D.L. Massart, Journal of Chromatography, 370 (1985) 1.

[29] A. Peeters, L. Buydens, D.L. Massart, P.J. Schoenmakers, Chromatographia, 26 (1988) 101.

[30] B. Bourguignon, P. Vankeerberghen, and D.L. Massart, Journal of Chromatography, 592 (1992) 51.

[31] P.J. Schoenmakers, A. Peeters, and R.J. Lynch, Journal of Chromatography, 506 (1990) 169.

[32] P.J. Schoenmakers, N. Dunand, A. Cleland, G. Musch, and T. Blaffert, Chromatographia, 26 (1988) 37.

[33] P.J. Schoenmakers, and N. Dunand, Journal of Chromatography, 485 (1989) 219.

[34] J.A. van Leeuwen, L.M.C. Buydens, B.G.M. Vandeginste, G. Kateman, and M. Mulholland, Analytica Chimica Acta, 235 (1990) 27.

[35] L.M.C. Buydens, J.A. van Leeuwen, M. Mulholland, B.G.M. Vandeginste, and G. Kateman, TRAC, Trends in Analytical Chemistry, 9 (1990) No.2, 58.

[36] M. Mulholland, N.Dunand, and A.Cleland, Journal of Chromatography, 485 (1989) 283.

[37]   M. Mulholland, J.A. van Leeuwen, and B. Vandeginste, Analytica Chimica Acta, 223 (1989) 183.

[38]   J.A. van Leeuwen, L.M.C. Buydens, B.G.M. Vandeginste, G. Kateman, G.P.J. Schoenmakers, and M. Mulholland, Chemometrics and Intelligent Laboratory Systems, 10 (1991) 337.

[39]   J.A. van Leeuwen, L.M.C. Buydens, B.G.M. Vandeginste, G. Kateman, P.J. Schoenmakers, M. Mulholland, Chemometrics and Intelligent Laboratory Systems, 11 (1991) 37.

[40]   J.A. van Leeuwen, L.M.C. Buydens, B.G.M. Vandeginste, G. Kateman, P.J. Schoenmakers, and M. Mulholland, Chemometrics and Intelligent Laboratory Systems, 11 (1991) 161.

[41]   J.A. van Leeuwen, B.G.M. Vandeginste, G. Kateman. Analytica Chimica Acta, 228 (1990) 145.

[42]   R. Bach, J. Karnicky, and S. Abbot, in T.H. Pierce, and B.A. Hohne, Artificial Intelligence Applications in Chemistry, American Chemical Society, Washington, 1986, p278.

[43]   S.S. Williams, J.F. Karnicky, J. Excoffier, and S.R. Abbot, Journal of Chromatography, 485 (1989) 267.

[44]   S.S. Williams, TRAC, Trends in Analytical Chemistry 9 (1990) No.2, 63.

[45]   P. Lu, and H. Huang, Journal of Chromatography, 452 (1988) 175.

[46]   Z. Yukui, Z. Hanfa, and L. Peichang, Journal of Chromatography, 515 (1990) 13.

[47]   N. Chen, X. Liang, Y. Zhang, and P. Lu, Chinese Science Bulletin, 37 (1992) No.21, 1809.

[48]   X. Liang, H. Huang, Y. Zhang, and P. Lu, Chromatogram, (1990) 215.

[49]   M.A. Tischler, and E.A. Fox, Computers and Chemistry, 11 (1987) No.4, 235.

[50]   R.M. Smith, and C.M. Burr, Journal of Chromatography, 485 (1989) 325.

[51]   C.M. Burr, and R.M. Smith, Analytical Proceedings (London), 26 (1989) No.1, 24.

[52]   T.P. Bridge, M.H. Wiliams, and A.F. Fell, Journal of Liquid Chromatography, 12 (1989) (1&2), 23.

[53]   T.P. Bridge, M.H. Wiliams, and A.F. Fell, Journal of Chromatography, 465 (1989) 59.

[54]   T.P. Bridge, M.H. Wiliams, G.G.R. Seaton, and A.F. Fell, Chromatographia, 24 (1987) 691.

[55] T.P. Bridge, M.H. Wiliams, and A.F. Fell, Journal of Pharmaceutical & Biomedical Analysis, 6 (1988) No.s6-8, 555.

[56] X. Wu, S. Hu, J. Hou, and X. Xu Journal of Chemical Engineering of Chinese Universities, 5 (1991) No.2, 136.

[57] G. Szepesi, and K. Valko, Journal of Chromatography, 550 (1991) 87.

[58] K.M. Kiyoshi Tsuji Jenkins, Journal of Chromatography, 485 (1989) 297.

[59] R. Milne, Journal of Chromatography, 485 (1989) 341.

[60] X. Wen, and W. Fei, Acta Petrolei Sinica (Petroleum Processing Section) 5 (1989) No.4, 85.

[61] P.B. Ayscough, S.J. Chinnick, R. Dybowski, and P. Edwards, Chem. Ind. (Lond.), 15 (1987) 515.

[62] K. Valko, G. Szabo, J. Rohricht, K. Jemnitz, and F. Darvas, Journal of Chromatography, 485 (1989) 349.

[63] H. Gunasingham, B. Srinivasan, and A.L. Ananda, Analytica Chimica Acta, 182 (1986) 193.

[64] A.L. Ananda, S.M. Foo, and H. Gunasingham, Journal of Chemical Information and Computational Science, 28 (1988) 82.

[65] C.T. Mant, T.W. Lorne Burke, N.E. Zhou, J.M.R. Parker, and R.S. Hodges, Journal of Chromatography, 485 (1990) 365.

[66] J.C. Pearce, A. Churchill, and A.C. Terry, Chemometrics and Intelligent Laboratory Systems: Laboratory Information Management, 17 (1992) 213.

[67] J.A. Asenjo, B. Byrne, and L. Herrera, Journal of Biotechnology, 11 (1989) 275.

[68] H. Eriksson, K. Sandahl, G. Forslund, and B. Osterlund, Chemometrics and Intelligent Laboratory Systems, 13 (1991) 173.

[69] H. Eriksson, K. Sandahl, J. Brewer, and B. Osterlund, Chemometrics and Intelligent Laboratory Systems, 13 (1991) 185.

[70] P. Schneider, and S.K. Graham, in H. Berghel, J. Talburt, D. Roach (Eds.) , Proceedings of the 1990 Symposium on Applied Computing. 101.

[71] G. Holzer, W. Bertsch and Q.W. Zhang, Analytica Chimica Acta, 259 (1992) No.2, 225.

[72] T. Aczel, S.G. Colgrove, and L. Le, ASTM Spec. Tech Publ., 1019 (Novel Tech. Fossil Fuel Mass Spectrometry), (1989) 159.

[73]   C.E. Riese, and J.D. Stuart, in T.H. Pierce, and B.A. Hohne, Artificial Intelligence Applications in Chemistry, American Chemical Society, Washington, 1986, p18.

[74]   Z. Yan, Y.Z. Wu, and Y. Zhou, in Proceedings of the 3rd International Conference on Properties and Applications of Dielectric Materials, 1 (1991) 63.

[75]   M. Mulholland, R. Haddad, and D.B. Hibbert, Journal of Chromatography, 602 (1992) 9.

[76]   H. Yuzhu, G. Peeters, G. Musch, and D.L. Massart, Analytica Chimica Acta, 223 (1989) 1.

[77a] A.Bartha, and G. Vigh, Journal of Chromatography, 485 (1989) 383.

[77b] H. Moll, Ph.D. thesis, University of Bern, 1991.

[78]   D. Diaper, Knowledge Elicitation. Principles, Techniques and Applications. Ellis Horwood, 1989.

[79]   R. Hindriks, F. Maris, J. Vink, A. Peeters, M. De Smet, D.L. Massart, and L. Buydens, Journal of Chromatography, 485 (1989) 255.

[80]   M. Meyer, J. Booker, Eliciting and Analysing Expert Judgement: A Practical Guide. Academic Press, London, 1991.

[81]   J.R. Quinlan in D. Michie, (Ed.) Expert Systems in the Micro Electronic Age, Edinburgh University Press, Edinburgh, 1979.

[82]   B. McGee, in [87] 69.

[83]   M. Minsky, in P.H. Winston, The Psychology of Computer Vision, McGraw-Hill, New York, 1975.

[84]   M. Mulholland, N. Walker, F. Maris, H. Hindriks, L. Buydens, I. Blaffert, and P.J. Schoenmakers, Journal of Chromatography, 550 (1991) 257.

[85]   A. Macro, Software Engineering. Concepts and Management. Prentice Hall, Hertfordshire, UK, 1990.

[86]   A. van Mayrhauser, Software Engineering. Methods and Management. Academic Press, London, 1990.

[87]   M.F. McTear, and T.J. Anderson, (Ed.s) Understanding Knowledge Engineering, Ellis Horwood, 1990.

[88]   E. Hollnagel, The Reliability of Expert Systems, Ellis Horwood, Chichester, 1987.

Figure 1: A Taxonomy of Expert Systems for Chromatography.

Figure 2: A Taxonomy of Single Stage Expert Systems for Chromatography.

Figure 3: A Taxonomy of Expert Systems for Chromatography that are specific to a class of compounds.

Figure 4: A mapping between the subdomains of ESCA, the subsystems implemented for each of those subdomains and the proposals for integrating these subsystems. The four polygons correspond to the four proposed integrations and each polygon surrounds those systems that were to be integrated by that proposal.

Figure 1:

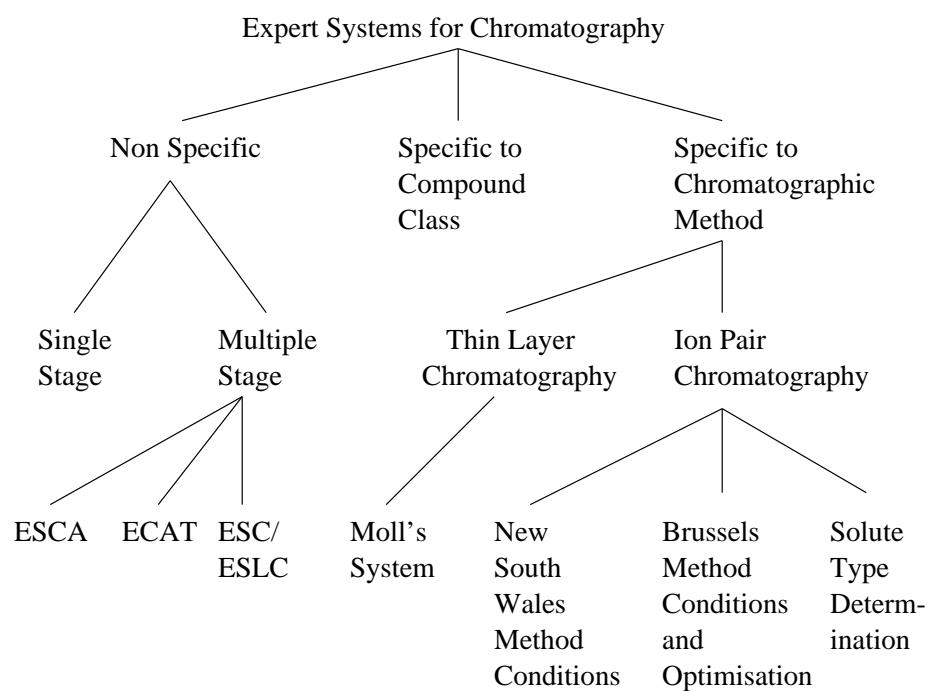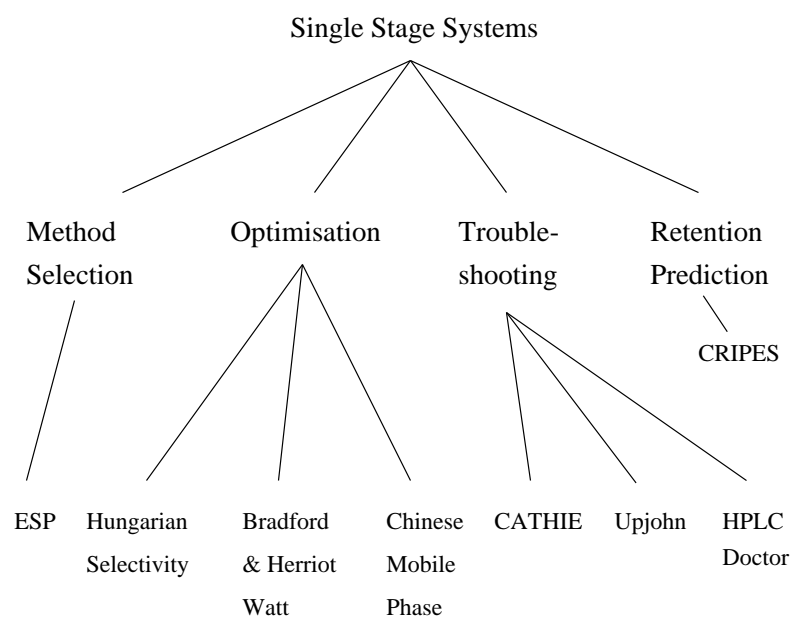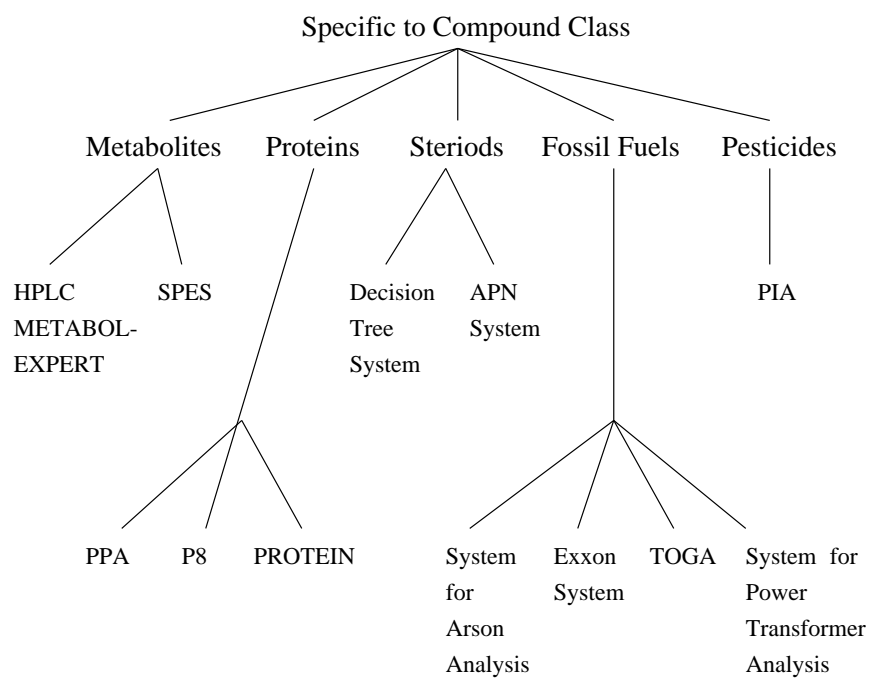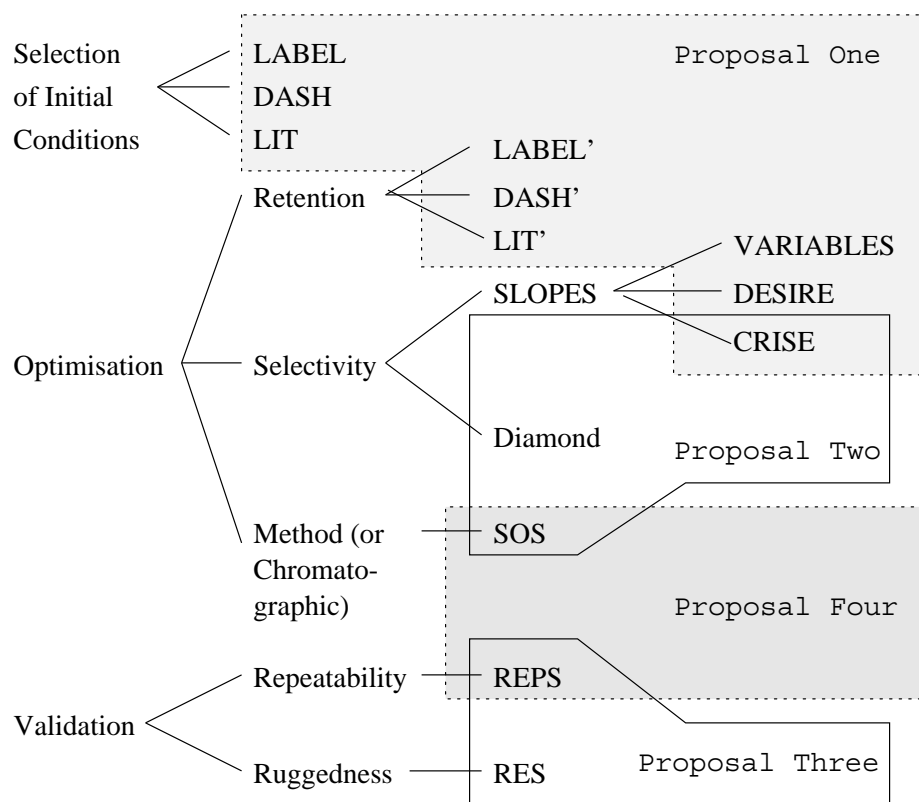Single Stage Systems

Method Selection — Optimisation — Trouble-shooting — Retention Prediction

ESP   Hungarian Selectivity   Bradford & Herriot Watt   Chinese Mobile Phase   CATHIE   Upjohn   HPLC Doctor

CRIPES

Figure 2:

Figure 3:

Figure 4:

Name of Expert System	Predominant Behaviour				Intended Users		
	Plans	Predicts	Diag-noses	Ident-ifies	Experts	Novices	Unspe-cified
Arson Analysis				Y	Y		
Brussels Method Conditions and Optimisation	Y						Y
CATHIE			Y				Y
CRIPES		Y					Y
ESP	Y					Y	
Exxon				Y			Y
Hungarian Selectivity Optimisation		Y			Y		
HPLC METABOL EXPERT		Y					Y
New South Wales Method Conditions	Y						Y
PECOD			Y				Y
PIA				Y	Y		
PPA	Y				Y	Y	
PROTEIN	Y				Y		
P8	Y				Y		
Power Transformer			Y				Y
Solute Type				Y			Y
Upjohn			Y			Y	
Steroid Systems	Y						Y
18	7	3	4	4	6	3	10

Table 1: Behaviour and Users of Expert Systems for Chromatography.

Language	Systems that used Language	Used for whole system?	Reasoning	Numerical processing	Interface	Data capture
				Aspects for which the language was used		

ARTIFICAL INTELLIGENCE LANGUAGES

Language	Systems that used Language	Used for whole system?	Reasoning	Numerical processing	Interface	Data capture
Prolog	ESP	N	Y			
MicroProlog	Bradford Herriot Watt System	N	Y			
	Steroid Decision Tree System	Y				
TurboProlog	Exxon System	Y				
	PIA	Y				
	Solute type determination system	Y				
Common LISP	ECAT	Y				
	PPA	N			Y	
SCHEME LISP	ESC/ESLC	N	Y			

CONVENTIONAL LANGUAGES

Language	Systems that used Language	Used for whole system?	Reasoning	Numerical processing	Interface	Data capture
BASIC & FORTRAN	Brussels ion pair system	N		Y		
	LABEL	N		Y		
C	Chinese mobile phase system	Y				
	ESP	N			Y	
	RES	N		Y		
Lattice C	APN Steroid system	Y				
Pascal	SOS reimplementation	Y				
	REPS reimplementation	Y				
Turbo Pascal	Exxon System	Y				
Unspecified conventional language	Bradford Herriot Watt system	N		Y		Y
	ESC/ESLC	N		Y		

Table 2: Languages used by Expert Systems for Chromatography.

Name of Tool	Supplier	Systems that used Tool	Notes
Epitool	Epitec AB	PPA	Except interface
Expert Systems Environment (ESE)	IBM	PROTEIN	
Goldworks	Gold Hill Computers, Cambridge, MA, USA	REPS	First implementation only
		RES	Factor choice module only
Knowledge Craft	Carnegie Group, Pittsburgh, PA, USA	SOS	Prototype only
Knowledge Engineering System (KES)	Software Archi− tecture and Engineering, Arlington, VA, USA	CRISE	
		DASH	
		LABEL	
		Brussels ion pair system	
M.1	Cimflex-Teknow− ledge, Palo-Alto CA. USA	Upjohn system	
Nexpert Object	Neuron Data, Palo Alto, CA, USA	SOS	An attempt was made to use it for the reimplementation
Personal Consul− tant Plus:PC Plus	Texas Instruments	PROTEIN	
Xi-Plus	Expertech, UK	S.Wales method conditions system	
VP-Expert	Paperback Software	CRIPES	

Table 3: Expert System Tools and Shells used by Expert Systems for Chromatography.