

CHROM. 22 030

EXPERT SYSTEM FOR THE SELECTION OF INITIAL HIGH-PERFORMANCE LIQUID CHROMATOGRAPHIC CONDITIONS FOR THE ANALYSIS OF PHARMACEUTICALS

RIK HINDRIKS*, FRANS MARIS and JAN VINK

Akzo Pharma Division, Organon International BV, Analytical R&D, P.O. Box 20, 5340 BH Oss (The Netherlands)

ANNE PEETERS, MARINA DE SMET and DESIRE L. MASSART

Farmaceutisch Instituut VUB, Laarbeeklaan 103, B-1090 Brussels (Belgium)

and

LUTGARDE BUYDENS

Catholic University of Nijmegen, Toernooiveld, 6525 ED Nijmegen (The Netherlands)

SUMMARY

The high-performance liquid chromatographic (HPLC) behaviour of basic pharmaceutical substances is strongly influenced by the type of column packing, the pH of the mobile phase and the concentration and type of buffer ions. This results in many choices to be made by the chromatographer. In order to assist the chromatographer, an expert system has been developed for the selection of initial HPLC conditions. For this purpose, HPLC data for about 600 basic compounds were used. These compounds belong to the class of CNS-active or cardiovascular drugs.

On the basis of this knowledge, which was completed with literature data, rules were defined and a knowledge base was built. The knowledge was implemented in KES (Knowledge Engineering System), a mid-sized expert system shell which runs on an IBM-PC. The system asks for information with respect to the substance(s) to be analysed. This information is given in the form of a table of structural elements. The output of the system specifies the chromatographic characteristics which should be used, *i.e.*, type of stationary phase, mobile phase composition, buffer pH, flow-rate and method of detection.

INTRODUCTION

The development of a new drug is a long and costly process. At the various stages, from the first synthesis of a new compound to drug monitoring in clinical studies and quality control of the finished product, analytical methods play an important role. Among the analytical methods used, high-performance liquid chromatography (HPLC) is most widely applied. HPLC is used to establish the purity of a drug candidate and to determine the drug in pharmaceutical formulations, in samples from stability testing and in pharmacokinetic and clinical studies.

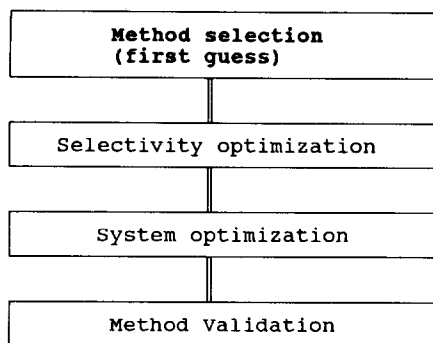


Fig. 1. Outline of an integral expert system for method development in HPLC. The present paper concerns method selection (first guess), which is the first (highlighted) box.

In the first step of drug development, a large number of chemical compounds are synthesized. Before these compounds are screened in pharmacological tests, they are subjected to HPLC analysis to check their purity. As most of these compounds are submitted for analysis only once, optimization of either the selectivity or the analysis time is not required. However, the "first guess" HPLC conditions should preferably result in capacity factors (k') between 3 and 10 in order to obtain optimal resolution in an acceptable time.

The choice of the initial HPLC conditions is the first step in method development (Fig. 1). The selection of these initial conditions requires specific knowledge and expertise. For example, several studies¹ were directed to finding the relationship between chemical structure and chromatographic retention. However, the selection of the correct percentage of modifier in the mobile phase in order to obtain a certain retention is still an heuristic process.

In recent years, expert systems have been proposed to advise the chromatographer. Varian was the first to announce work on expert systems in chromatography, in 1984. Under the acronym ECAT (Expert Chromatographic Assistance Team), the system has since been introduced in detail and explained^{2,3}.

Within ESPRIT (European Strategic Programme for Research and Development in Information Technology), a programme supported by the EEC, a large group of scientists is working on a project on the "Application of Expert Systems in the Chemical Analysis" (ESCA)⁴. The aim of this project is to demonstrate the applicability of expert systems in HPLC, particularly applied to pharmaceutical analysis. The project covers the whole field of method development⁵⁻⁷. The scheme shown in Fig. 1 became the basis of our work within ESCA. Based on this scheme, four different expert systems were developed.

Other groups are also working on the applicability of expert systems in (liquid) chromatography⁸⁻¹². This illustrates the broad interest of the industry and universities in these types of computer systems. One expert system, called LABEL^{13,14}, has been developed for the selection of the initial HPLC method in pharmaceutical analysis. The system is suitable for a broad range of compounds, but only a cyanopropyl column is used. The knowledge of LABEL is, like the present expert system, also implemented in KES (Knowledge Engineering System; Software Architecture and Engineering, Arlington, VA, U.S.A.).

In this paper we present details of an expert system for the first step of method development in HPLC.

SELECTION OF CHEMICAL AREA

The expert system, called by us a "first guess" system, was developed in the first instance for the purity control of compounds in development at Organon International. These compounds, which are submitted to a purity check, are generally fairly pure, having a total concentration of substance-related compounds of less than 5%. The polarity of these related compounds is usually close to that of the main substance. The strongly more polar and/or less polar compounds will generally be removed at an earlier stage, *e.g.*, by (fractional) crystallization and, if still present, will be traced by thin-layer chromatography. Because, within one sample, only compounds with a limited range of polarities have to be analysed by HPLC, an expert system has been developed that predicts conditions for isocratic elution only. Further, the resolution obtainable, which is determined by the length of the column and the retention of the compound, is more important than the analysis time. Generally, we use a long column and a capacity factor (k') between 3 and 10.

Traditionally, chromatographers have encountered problems in the reversed-phase chromatography of basic nitrogen-containing substances. Among the undesirable effects are severe tailing, band broadening and low plate numbers. In recent years we have carried out much work on the reversed-phase HPLC analysis of basic compounds which belong to the CNS-active drugs (drugs acting on the central nervous system) or cardiovascular drugs. Some typical examples are shown in Fig. 2.

Generally, these drug molecules are small and contain one or more nitrogen atoms. Many of these nitrogen-containing compounds are basic and can be protonated. Important exceptions are the barbiturates, which are slightly acidic, and

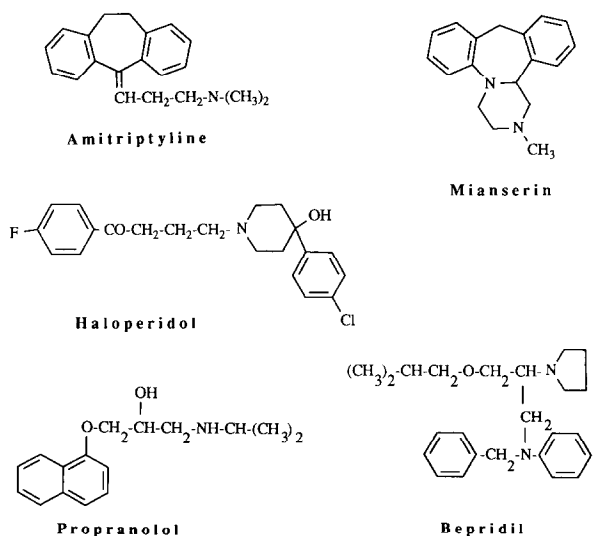
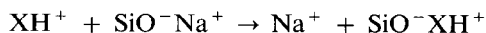


Fig. 2. Some typical representatives of basic pharmaceutical compounds.

amino acids, which are amphoteric. It is assumed that the protonated drugs (XH^+) interact with the acidic free silanol sites ($\text{SiO}^- \text{Na}^+$) on the reversed-phase packing by an ion-exchange process^{15,16}:



As a result, the peaks of the basic solutes have a relatively broad and tailing shape.

Basic drugs are extremely important for the pharmaceutical industry. Therefore, great attention has been paid to the often poor results of the reversed-phase HPLC of these compounds and many suggestions have been made for improving this¹⁷⁻¹⁹.

Our extensive experience with the HPLC of basic drugs, together with literature data, form the knowledge base for this expert system. Elements from the knowledge base will be explained. Rules that have been derived from the knowledge are implemented in an expert system shell and this process is also outlined.

SELECTION OF EXPERT SYSTEM SHELLS

A large number of expert system shells (tools) are available today, and the number is still steadily growing. A useful possibility for classification of these tools is to divide them on the basis of the size of the tool, *i.e.*, small, mid-size and large. An overview of the tools tested in ESCA is given in Table I. Small tools generally run on PCs whereas mid-size tools need expanded PCs or workstations. The large tools run only on large workstations.

One of the aims of ESCA has been to select suitable tools for the project. This was done by judging the relevant aspects, such as ease of implementation, inference facilities, support facilities, price and possibility of running on a personal computer.

Apart from Delphi 2 and MYLOG, the tools have all been developed in the U.S.A. Delphi 2 was apparently too small for the job. On the other hand, the large tools can be regarded as oversized for our purpose. Often they are less easy to explore by the knowledge engineers because of their large potential, and the high price can also be prohibitive. For us, the main disadvantage was the fact that these large tools cannot be run on a personal computer. The mid-size tools were clearly the best for our project. Within ESCA we chose Goldworks, KES and NExpert Object for further evaluation⁴. They all run on an IBM/PC and compatible systems. The knowledge is represented by

TABLE I

SOME EXPERT SYSTEM TOOLS INVESTIGATED WITHIN ESCA AND THEIR SIZE AND ORIGIN

<i>Name</i>	<i>Size</i>	<i>Origin</i>
Delphi	Small	The Netherlands
Goldworks	Medium	U.S.A.
MYLOG	Medium	France
KES	Medium	U.S.A.
NExpert Object	Medium	U.S.A.
KEE	Large	U.S.A.
Knowledge Craft	Large	U.S.A.

a set of rules. The selected systems offer backward and forward chaining as inference mechanisms. Goldworks and KES can also use frames and a network of frames can be built by defining relations between them, *e.g.*, by inheritance. Goldworks also features pattern matching. External links to databases, spreadsheets and other external processes are provided. With respect to the user-knowledge engineer interface, Goldworks has the best facilities.

Goldworks was tested for the method validation expert system⁷ and NExpert Object was used in the development of the system optimization expert system⁶. For the expert system which we describe here, we worked exclusively with KES (release 2.4), a tool which was already well known in Massart's group^{13,14} and which has been used for many applications. Moreover, KES matches closely our formulated selection criteria and was shown to be a very flexible tool. In the final phase of the ESPRIT project we shall evaluate the possibilities and limitations of the three shells.

KNOWLEDGE BASE

Before an expert system can be consulted, the knowledge base has to be filled with rules by the knowledge engineer. In our case the most important input parameters are related to information on the chemical structure, chromatographic and chemical information, detector parameters and other information.

Chemical structure information

Information about the polarity of the sample component(s) is essential for the chromatographer. A relative measure that was introduced some years ago is the retention index (RI)²⁰, which can be used to characterize the polarity of the molecule and to link the structure elements to some type of polarity descriptor.

Numerous methods have been described for the determination of RI values²⁰. Throughout we used the method of Baker and Ma²¹, which is based on a homologous series of 2-ketoalkanes. Under strict conditions it was shown that the RI values are reasonably constant. RI values for 300 compounds in combination with chromatographic data on the purity analyses of more than 300 compounds were the basis of the knowledge.

We chose to estimate the polarity of a given molecule on the basis of the presence of polar and apolar groups. Therefore, the molecular structure of the sample is an essential feature. The expert system has to calculate the polarity of a new compound from its structure and expresses the result as a percentage of organic modifier (methanol) in the mobile phase. Before this can be done, the structure is subdivided into fragments or structural elements. These elements are so defined that they can describe a structure in a simple and unambiguous manner. Examples of such elements are phenyl, methyl, hydroxyl and tertiary nitrogen. All initially selected structural elements are shown in Table II. As in QSAR (quantitative structure-activity relationship), each element can be linked to a fragment value²² or, as here, to a percentage of methanol. For apolar groups this is a positive contribution and for polar elements a negative contribution.

The percentages listed in Table II are essentially derived from experimental data. As an example, one could chromatograph a drug molecule and an analogue having an additional methyl or hydroxyl group. By comparing the methanol percentages at the

TABLE II

SOME STRUCTURAL ELEMENTS AND THEIR EFFECTS ON THE PREDICTED PERCENTAGE OF METHANOL AT pH 7.4 AND 4.0

Structural element	Methanol (%)	
	pH 7.4	pH 4.0
(1) Phenyl, monosubstituted (C ₆ H ₅)	+11	+11
(2) Phenyl, disubstituted (C ₆ H ₄)	+10	+10
(3) Phenyl, trisubstituted (C ₆ H ₃)	+9	+9
(4) Cl on aromatic group	+7	+7
(5) Cl on aliphatic group	+1	+1
(6) OH on aromatic group	-2	-2
(7) OH on aliphatic group	-10	-10
(8) O atom in ether. The oxygen positioned between:		
(a) two aromatic groups	-5	-5
(b) an aromatic and an aliphatic group	-5	-5
(c) two aliphatic groups	-10	-10
(9) O atom in ketone. The carbon connected to the oxygen is positioned between:		
(a) two aromatic groups	-5	-5
(b) an aromatic and an aliphatic group	-6	-6
(c) two aliphatic groups	-10	-10
(10) S atom. The sulphur positioned between:		
(a) two aromatic groups	+3	+3
(b) an aromatic and an aliphatic group	+1	+1
(c) two aliphatic groups	-3	-3
(11) Pyridine	+3	-5
(12) CH ₃	+5	+5
(13) CH ₂	+3	+3
(14) CH	+2	+2
(15) C	+1	+1
(16) N atom in ring plus double bond	-5	-5
(17) N atom in two rings	0	0
(18) Other N atoms: first one	-5	-30
every next one	-5	-5

same k' , the contribution of that group can be derived for a given column. Using our large database, this calculation can be done fairly accurately. As can be seen from Table II, most fragment contributions are independent of the pH of the mobile phase. It can easily be understood that this is not the case for fragments that can be protonated.

The percentage of organic modifier (methanol) can be calculated by adding up the fragment contributions together with a zero level. We determined experimentally the zero level for a NovaPak C₁₈ column at 43% of methanol. In other words, methanol (%) = Σ (fragment contributions) + 43%. The same holds for a μ Bondapak C₁₈ column, except that the zero level is 2% lower.

For every new compound, the present prototype expert system needs a list of structural elements to calculate the polarity. Although in principle this is a simple task, errors are being made especially by workers who are not trained in "reading"

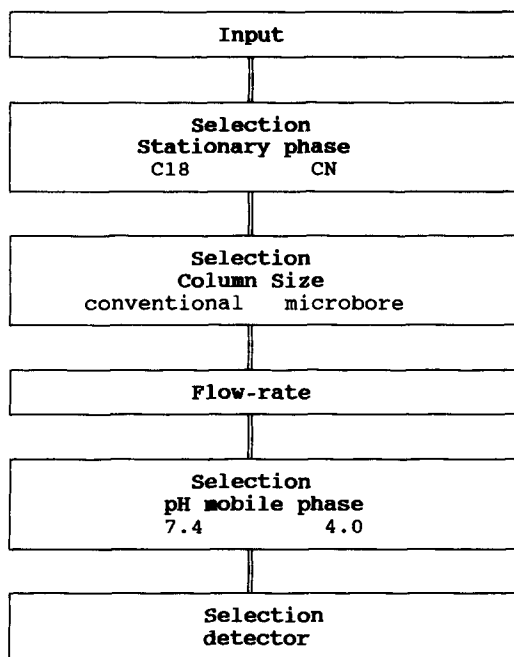


Fig. 3. Method selection in the present expert system.

structures. Therefore, the number of structural elements is limited. On the other hand, in order to increase the applicability of the system, it is clear that the number of structural elements has to be increased. It was found that the number of structural elements shown in Table I is a good compromise between usability and applicability. For future developments, we are trying to implement a program that makes use of so-called connectivity tables. These tables are generated by advanced structure representation software programs, such as DARC (Télesystèmes, Paris, France).

Chromatographic information

Part of the chromatographic knowledge is included in the method selection section of the expert system, as shown in Fig. 3. For most applications we routinely use tetramethylammonium phosphate buffer to block the free silanol sites on the reversed-phase material¹⁷. Methanol is preferred as the organic modifier.

In our experience, a reversed-phase system provides the required selectivity for purity analysis of drugs. We use a cyanopropyl column for compounds containing quaternary nitrogen atoms and a C₁₈ column for other basic compounds. The buffer pH of the mobile phase is merely dictated by the character of the separation. In addition to the chemical structure of the compound, other information can also be useful, such as related compounds that might be present and remaining products from the last step of the synthesis. We have observed that the selectivity for small lipophilic differences, *e.g.*, for *cis-trans* isomers, is generally optimal at a pH between 7 and 8. Therefore, we prefer a high pH for separations that might be critical in this respect.

Detector parameters

For a purity check of new compounds, it is important to detect all kind of by-products that can be expected, *e.g.*, starting materials, related products from the synthesis and reagents. Normally, these products have different UV absorptivities. As they are often present in small amounts (<1%), universal detection by an insensitive refractive index detector is impractical. A more practical solution is to use a UV detector (variable wavelength or diode-array) operating at a short wavelength (205–210 nm). Because most pharmaceutical compounds have at least some UV absorptivity, UV detection is the best compromise between a reasonable universal

CHROMATOGRAPHIC SETUP

- mobile phase system {RP; NP; IPC}
- column
- detection {UV detector; RI detector; unknown}
- flow rate {0.00 .. 5.00}
- temperature {0 .. 70}

FIRST TRIAL CHROMATOGRAPHIC SETUP

is a CHROMATOGRAPHIC SETUP

- mobile phase system {RP; NP} [default: RP]
- column
- detection {UV detector; RI detector}
- flow rate {0.2 or 0.3; 1.5 or 2.0}
- temperature {30}

RP

- organic modifier
- buffer

NP

- organic modifier
- basic solvent

ICP

- organic modifier
- PIC

BUFFER

- nature {TMA-phosphate}
- pH {4; 7.4}
- % {00 .. 100}

ORGANIC MODIFIER

- nature {methanol; acetonitrile; mixture} [default: methanol]
- % {00 (default) .. 100}
- % final {00 .. 100}

MIXTURE

- nature {methanol + acetonitrile}
- % methanol {00 .. 100}
- % acetonitrile {00 .. 100}

BASIC SOLVENT

- nature {hexane}
- % {00 .. 100}

PIC

- nature {hexane sulphonate, heptane sulphonate, octane sulphonate}

Fig. 4. Frame structure showing how the mobile phase is related to the chromatographic set-up.

detection method and good sensitivity. Provided that methanol of good quality is used, impurities at the 0.1% level can easily be detected.

Other parameters

Other characteristics of the sample molecule, which have been determined previously, can be very helpful for the chromatographer. One of the important characteristics of a drug molecule is its dissociation constant (pK). Initially, the expert system was in principle developed for basic Organon compounds with a pK_a value between 4 and 10 (measured for the protonated drug). We are now investigating whether the system can also be applied to a broader range of pharmaceutical compounds.

IMPLEMENTATION

The implementation of the expert system was preceded by two important steps, the knowledge acquisition and the formalization of the knowledge.

The knowledge acquisition was mainly done by interview, but other techniques were also used. Sometimes information was exchanged in written form. The results from this knowledge acquisition are that the important features in the domain are defined and the relationship between those features are determined.

The next step in the process was to formalize the knowledge. This was done independently of the selected tool. A kind of series of frames was built. Some of these frames have one or more subframes which reflect part of the relationship between the frames. Different sets of rules were built to translate the complete relationship between the frames. Part of this structure is shown in Fig. 4 and an example of a rule group is shown in Fig. 5. Another way to formalize the knowledge is the use of state transition diagrams and data flow diagrams. The representation of the knowledge kept in this expert system is shown in Fig. 6 in a data flow diagram. After this formalization step, the real implementation of the prototype started.

The selected tool is KES.PS (Software Architecture and Engineering, release 2.4), which is written in C. The use of frames is not explicitly available in this tool, but by combining the classes (abstract objects, models on which attributes are defined) and

```
F1
IF      Column.type = CONVENTIONAL
      and Column.brand = novapak
      THEN flow rate = 1.5ml/min

F2
IF      Column.type = CONVENTIONAL
      and Column.brand = microbondapak
      THEN flow rate = 2ml/min

F3
IF      Column.type = MICROBORE
      and Column.length = 100mm
      and Column.diameter = 2.1mm
      THEN flow rate = 0.2ml/min

F4
IF      Column.type = MICROBORE
      and Column.length = 200mm
      and Column.diameter = 2.1mm
      THEN flow rate = 0.3ml/min
```

Fig. 5. Rules showing how the flow-rate is determined.

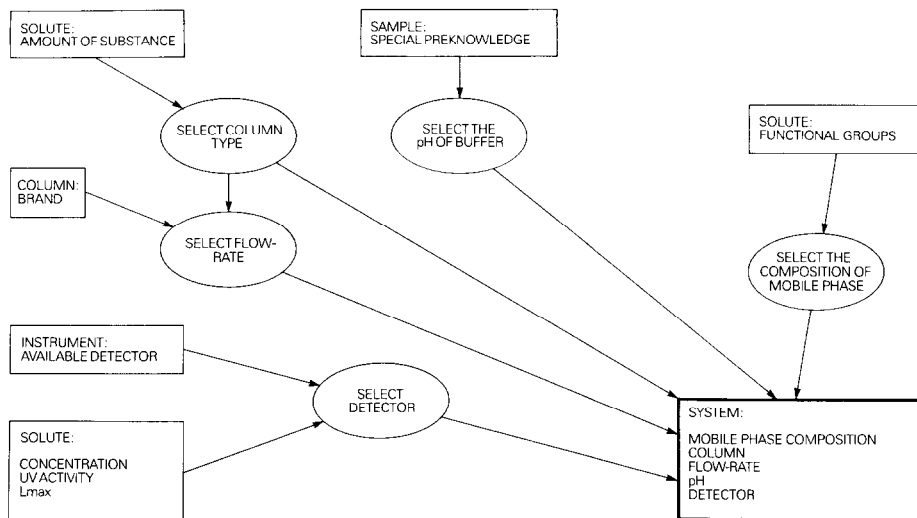


Fig. 6. Data flow diagram provided with input and output.

the attributes (objects used to define characteristics and features of a problem), a frame-like structure can be built.

After the definition of the attributes and the classes, the rules must be defined in the rule section. Here the rules are listed that are used by a domain expert to infer a value for an attribute based on the values of other attributes. In KES it is impossible to define groups of rules. However, by defining to which class a rule is applicable the rules are more or less split up in rule groups. The next part in the knowledge base is the action section, where the order and the character of an end-user session are controlled. The knowledge base is guided by the action part to seek appropriate information and to inform the end-user about relevant pieces of information generated by the system.

CONCLUSIONS AND FUTURE DEVELOPMENTS

Reversed-phase HPLC of basic pharmaceuticals requires extensive experimental and theoretical expertise. Therefore, it was seen as a challenge to develop an expert system that can assist the chromatographer in the selection of initial HPLC conditions.

The most difficult task proved to be to describe correctly and yet in a simple way the polarity of a sample molecule on the basis of its chemical structure. Our approach, *viz.*, to divide the molecule into structural elements, proved to be successful, but also showed some limitations.

On the basis of expert knowledge rules were formulated, which were implemented in KES. A prototype expert system showed its usefulness for the selection of initial HPLC conditions for the analysis of basic drugs.

We are now working on an extension of this project in two directions: (i) to establish a second (or next) guess when the initially selected conditions proved to be not in the correct range, *viz.*, a capacity factor between 3 and 10; and (ii) to use a computer program for structure representation such as DARC, and write an

additional computer program for it so that the subdivision of a chemical structure into defined fragments can be carried out by the computer.

Work has also been planned to validate the implementation of the chemical knowledge and to evaluate the accuracy of the advice of the present expert system.

ACKNOWLEDGEMENT

This research is supported by the European Commission as part of ESPRIT Project 1570 (ESCA).

REFERENCES

- 1 R. R. Kaliszan, *Quantitative Structure–Chromatographic Retention Relationships*, Ellis Horwood, Chichester, New York, 1987.
- 2 R. Bach, J. Karnicky and S. R. Abbott, *ACS Symp. Ser.*, No. 306 (1986) 278.
- 3 S. S. Williams, J.-L. Excoffier and S. Abbot, presented at *Pittsburgh Conference, Atlanta, GA, 1989*, abstract No. 1162.
- 4 D. Goulder, T. Blaffert, A. Blokland, L. Buydens, A. Chhabra, A. Cleland, N. Dunand, H. Hindriks, G. Kateman, H. van Leeuwen, D. Massart, M. Mulholland, G. Musch, P. Naish, A. Peeters, G. Postma, P. Schoenmakers, M. de Smet, B. Vandeginste and J. Vink, *Chromatographia*, 26 (1988) 237.
- 5 A. Peeters, L. Buydens, D. L. Massart and P. J. Schoenmakers, *Chromatographia*, 26 (1988) 101.
- 6 P. J. Schoenmakers, N. Dunand, A. Cleland, G. Musch and Th. Blaffert, *Chromatographia*, 26 (1988) 37.
- 7 J. A. van Leeuwen, B. G. M. Vandeginste, G. Kateman, M. Mulholland and A. Cleland, *J. Chromatogr.*, in press.
- 8 M. A. Tischler and E. A. Fox, *Comput. Chem.*, 11 (1987) 235.
- 9 A. F. Fell, T. P. Bridge and M. H. Williams, *J. Pharm. Biomed. Anal.*, 6 (1988) 555.
- 10 J. W. Dolan and L. R. Snyder, presented at *Pittsburgh Conference, Atlanta, GA, 1989*, abstract No. 1160.
- 11 J. I. Kirkland, J. L. Glajch, S. W. Rementer, T. G. Jones and L. R. Snyder, presented at *Pittsburgh Conference, Atlanta, GA, 1989*, abstract No. 1161.
- 12 R. S. Hodges, J. M. R. Parker and C. T. Mant, presented at *Pittsburgh Conference, Atlanta, GA, 1989*, abstract No. 1163.
- 13 M. De Smet, A. Peeters, L. Buydens and D. L. Massart, *J. Chromatogr.*, 457 (1988) 25.
- 14 M. De Smet, G. Musch, A. Peeters, L. Buydens and D. L. Massart, *J. Chromatogr.*, 485 (1989) 237.
- 15 E. Papp and Gy. Vigh, *J. Chromatogr.*, 282 (1983) 59.
- 16 B. Law, *J. Chromatogr.*, 407 (1987) 1.
- 17 F. A. Buytenhuys and F. P. B. van der Maeden, *J. Chromatogr.*, 149 (1978) 489.
- 18 R. W. Roos and C. A. Lau-Cam, *J. Chromatogr.*, 370 (1986) 403.
- 19 M. A. Stadalius, J. S. Berus and L. R. Snyder, *LC · GC, Mag. Liq. Gas Chromatogr.*, 6 (1988) 494.
- 20 R. M. Smith, *Adv. Chromatogr.*, 26 (1987) 277.
- 21 J. K. Baker and C.-Y. Ma, *J. Chromatogr.*, 169 (1979) 107.
- 22 R. F. Rekker, *The Hydrophobic Fragmental Constant*, Elsevier, Amsterdam, 1977.