Journal of Chromatography, 485 (1989) 325–340
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROM. 21 869

RETENTION PREDICTION OF ANALYTES IN REVERSED-PHASE HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY BASED ON MOLECULAR STRUCTURE

V[a]. CRIPES (CHROMATOGRAPHIC RETENTION INDEX PREDICTION EXPERT SYSTEM)

ROGER M. SMITH* and CHRISTINA M. BURR
Department of Chemistry, Loughborough University of Technology, Loughborough, Leicestershire LE11 3TU (U.K.)

SUMMARY

An expert system for retention prediction in reversed-phase high-performance liquid chromatography has been developed using a commercial expert system shell VP-Expert. The program uses the molecular structure of an analyte to calculate retention indices based on the alkyl aryl ketone scale in different eluent combinations from empirically derived quadratic expressions for the structural units. The predictions of test compounds usually show a close match with experimental values except for some analytes containing strong intramolecular interactions. The program can also calculate the resolution of pairs of analytes.

INTRODUCTION

A method has been developed for reversed-phase high-performance liquid chromatography (HPLC) which can calculate a predicted retention for an analyte based on its molecular structure and the eluent composition[1]. The system is based on the alkyl aryl ketone retention index scale to increase the reproducibility of the prediction and improve the transferability of the results between instruments and column materials.

The retention index (I) is calculated by the summation of the retention index of a parent compound (I_P, benzene) and contributions for aliphatic carbons ($I_{S,R}$), for substituents on aromatic ring ($I_{S,Ar-X}$) and aliphatic carbons ($I_{S,R-X}$) and for interactions between the substituents ($I_{I,Y-Z}$), such as steric, hydrogen bonding and electronic factors:

$$I = I_P + I_{S,R} + \Sigma I_{S,Ar-X} + \Sigma I_{S,R-X} + \Sigma I_{I,Y-Z} \quad (1)$$

Values for each of these contributions have been determined empirically over a range of methanol–buffer (pH 7) and acetonitrile–buffer (pH 7.0) compositions and are

[a] For Part IV, see ref. 3.

0021-9673/89/$03.50

expressed as the coefficients of a quadratic equation relating the value to the proportion of modifier:

$$I = ax^2 + bx + c \tag{2}$$

where x = percentage of modifier.

The coefficients have been reported so far for benzene as the parent compound[1], for substituents on aromatic[1] and aliphatic[2] carbons, for the interactions between the aliphatic substituents and the aromatic ring[2] and for branching and unsaturation in the aliphatic side-chain[3]. Studies have also been carried out to investigate the interactions between either methyl or phenolic groups and a wide range of other substituents on an aromatic ring[4]. Although individual empirical values for each of these last interactions have been included in the database, it is recognized that a more general method for their estimation will be required in the future to cover interactions between a wider range of pairs of functional groups and the application of the approaches used for the calculation of octanol–water partition coefficients (log P) by Fujita[5] will be examined. The coefficients of each of the indices are held as a database in a series of spreadsheets and the use of an expert system to act as an interface to the user has been briefly reported[6].

The application of knowledge-based or expert systems in analytical chemistry has been the subject of several recent papers[7–13]. All these systems are characterized by two parts, an inference engine or underlying logic program and a knowledge base of facts and rules that is specific to the operation or systems being examined. The expert systems fall into two main groups. The first group represents systems that examine areas of data and tries to deduce or infer the underlying relationships and expresses these as rules which can be used to characterise or analyse further samples or situations. These programs generally have a strong pattern recognition or other chemometric basis (*e.g.*, see ref. 14). No attempt appears to have been made to use this approach in retention prediction for HPLC, although it should be possible if a large database of retention properties is available. Multiple regression analysis could be used to determine the effect of each group in a similar manner to QSAR studies of octanol–water partition coefficients[15]. However, unless care is taken to normalize the data, reported capacity factors from different sources could not be used because of the wide differences in retention that can occur on different column materials even if nominally of the same type. HPLC can also use a wide range of eluent conditions, buffers, pH and ion-pair reagents.

The second type of expert system program uses a set of rules and relationships to determine the outcome or provide a conclusion from information entered during a consultation. These rules or knowledge base may be derived by a "knowledge engineer" from an analysis of the advice or comments of a "domain expert" experienced in the field under study or from existing and accepted well-established relationships. The expert system does not itself generate rules or analyse the data to derive relationships. Often the information within the program includes a database of factual values. These systems are more accurately described as knowledge-based information systems (KBIS). In this role, the expert system acts to interpret the rules by recognising values or choices entered by the user, matches these with rules and then presents a conclusion. These systems operate as interfaces enabling the user to interrogate

the database, which although possible manually would often be a time-consuming process. A major advantage of the expert system is that all possible outcomes will be considered and all the relevant rules will be applied, whereas a manual approach is more error prone.

The continuing interest in the use of computers in the laboratory has led to a number of expert systems being examined in different areas of chromatography. Often these have been specifically written using an artificial intelligence language, such as LISP or PROLOG, but more frequently commercially available expert system shells (or inference engines plus interfaces) are preferred as they enable the developer of the system to concentrate on establishing the chromatographic rules and database.

The reported programs for chromatography have included peak monitoring and deconvolution routines to examine chromatographic outputs, such as the systems described by Fell *et al.* [7]. Of particular interest because of their relationship to this study have been programs that aid retention prediction or method selection. Tischler and Fox[16] described an expert system whose aim was to aid the "inexperienced analytical chemist in choosing a separation method for HPLC". The program, ESP (Expert Separation Program), used rules drawn from a standard textbook and suggested the type of separation (reversed-phase, ion-pair, etc.) rather than the conditions required. Bach *et al.*[17] explained the development of ECAT (Expert Chromatographic Assistance Team). This major project is built around a series of modules each for a specific class of compounds (steroids, phenols, alkaloids, etc.) and recommends methods (ion-pair, normal phase, etc.). It is planned to extend it to troubleshooting and optimisation.

Lu and co-workers[18,19] have discussed the development of a chromatograph with "artificial intelligence". The expert system selects mobile phase conditions, column systems and instrumental factors such as detectors and contains information on gas chromatographic and HPLC separations. The program is based on a library of 500 "living chromatograms". The program has a strong theoretical basis and calculates detailed eluent effects. The full program also includes eluent optimisation and fault diagnosis.

A more specific application of an expert system program to determine the conditions for the analysis of selected steroids has been discussed by Gunasingham *et al.*[20]. The user was required to enter details about the sample, such as the polarity of the steroids, and about the origin as sample preparation may affect the selection of method. The same group has recently also described the basis of another approach to planning HPLC optimisation using an expert system[21].

Musch and co-workers[22,23] described a system based on KES, which can advise on the choice of spectroscopic or electrochemical detection in HPLC. This work has been extended to the expert system LABEL[24]. On entering the molecular size and polarity of the analyte, the expert program will suggest possible mobile phases conditions on a cyano column for either normal- or reversed-phase separations.

An indication of the current interest in expert systems is the EEC-sponsored project ESCA (Expert Systems for Chemical Analysis) being studied under the ESPRIT program by groups in the UK, The Netherlands and Belgium. An overview of the aims of the project was reported by Schoenmakers and Mulholland[25] and longer descriptions of the project have recently appeared[26,27]. Four possible areas for the application of expert systems to method development for pharmaceutical analysis

were identified, one of which was the prediction of "first guess conditions" for the initial separation based on the structure of the analyte and origin of the sample. Each section of the project was being implemented using a different expert system shell to compare the applicability of different approaches.

In this paper, the implementation of the expert system program CRIPES (Chromatographic Retention Index Prediction Expert System) using a commercially available shell to interrogate the database of retention coefficients and to calculate predicted retention indices and capacity factors of analytes is described in detail. The resulting predictions for a range of test compounds are compared with experimental values.

EXPERIMENTAL

The experimental methods were as described in Part I[1]. The expert system CRIPES was developed using VP-Expert (Paperback Software) and the data were held as spreadsheets written using VP Planner (Paperback Software). The program was either written using the editor in VP-Expert or as a non-document file on Wordstar (Micropro). The programs were run on an IBM-compatible OPUS II computer with 1024K RAM and dual disc drive fitted with a Hercules graphics card.

RESULTS AND DISCUSSION

The prediction system CRIPES was implemented using the expert system shell VP-Expert.

Expert system

VP-Expert is an expert system development tool written in Microsoft C to run on an IBM-compatible PC with a minimum of 300K of memory. It is a rule-based system that operates in the backward chaining or goal-driven mode. There are several features that make it particularly suitable for use with the present application. The most important is its capability to handle mathematical routines and the availability of many resident arithmetic functions. In contrast, many other expert systems shells cannot carry out even the simplest calculations unless external high-level subroutines are appended to the program. VP-Expert can also communicate with compatible external spreadsheets and databases. This means that the coefficients for the regression equations for the parent, substituent and interaction indices can be held outside the main program and therefore are easily updated as additional substituents are examined. It will also be easy to expand the program for other eluents, such as tetrahydrofuran–buffer, or eluents based on different pH buffers. This facility also enables data to be transferred between sections of the program with ease.

The rules are of a standard format:

RULE *N*
If $Y = Z$
AND $Z > W$
OR $Z < Z1$
THEN X = true

where W, Y, Z, $Z1$ and X are numerical or text variables or defined values.

The consultation of the expert system is run from an ACTION block which contains the goals which the program must satisfy (*e.g.*, Fig. 1, Actions Block for CRIPES). The goals are given using the terminology FIND variable. Sub-goals can also be given in the conclusion of the rules or buried within the Rules such that they need to be satisfied before the rule can be evaluated (examples of the form of the Rules in CRIPES are given in Fig. 2). Variables can be of several types, single, plural or dimensioned.

```
ACTIONS

PRINTOFF
Display "                    CRIPES

      Chromatographic Retention Index Prediction Expert System

       This knowledge base will calculate the retention index
       of a single compound from its molecular structure
       at either a single or multiple eluent compositions"

wks rp1,ROW= Benzene,B:Benzmecn
wks rp2,ROW = Benzene,B:Benzmeoh
FIND COLUMN_READ
FIND name
FIND num_aro
FIND check_num
FIND num_ali
FIND subali
FIND SIIaro
FIND SIIali
FIND print
FIND single_eluent
FIND RI;

ASK num_ali:"How many aliphatic substituents present in {name} ?";
ASK num_aro:"How many aromatic substituents are present in {name} ?";
ASK subali1:'Which of these aliphatic substituents are present?";
ASK subali2:"Which of these aliphatic substituents are present?";
ASK subaro1:"Which, if any, of these aromatic substituents are present";
ASK subaro2:"Which, if any, of these aromatic substituents are present";

CHOICES subali2:OH,CONH2,Br,Cl,CN;
CHOICES subali1:CHO,CO2R,OR,COR,CH_CH,ALKYL_CHAIN,ANOTHER;
CHOICES subaro1:COR,CHO,OR,CO2R,CH3,CONH2,OH,CH_CH,ANOTHER;
CHOICES subaro2: NH2,NO2,CN,Cl,Br,Ph;
```

Fig. 1. Extract of knowledge base of CRIPES showing Actions block and a selection of ASK and CHOICES statements.

During a consultation, VP-Expert attempts to satisfy a goal by looking at each rule in turn to find one containing that goal in its conclusion. If the conditions of a rule are satisfied and the variable is a single variable, then the goal is satisfied and the consultation moves on to the next goal. If the conditions are not satisfied then the program looks for the next rule containing the goal in its conclusion. The order of the rules in the knowledge base can therefore have a significant influence on the path and results of a consultation because once a goal is satisfied the search stops. The consultation then moves on to the next goal even if later rules are also satisfied. If the variable is a plural variable the program does not stop after the first value of the goal

```
RULE 1
IF column <> unknown
THEN wks col1,ROW = (column),B:MeCNcol
wks col2,ROW = (column),B:MeOHcol
column_read = done;


RULE 2
IF num_aro <= 6
THEN FIND subs
check_num = ok;


RULE 3
IF   num_aro > 0
THEN FIND r_a
FIND S1 find s1a
     y=0
     WHILEKNOWN sub2[y]
          y=(y+1)
          find sub3
          FIND S3
          FIND naro
          num = (num + n1)
          FIND check
          FIND rest_alkyl
          FIND pos1
          FIND pos2
          cg1 = (cg1 + (n1 * coeff1[1]) + ct1 + CU1)
          cg2 = (cg2 + (n1 * coeff1[2]) + ct2 + CU2)
          cg3 = (cg3 + (n1 * coeff1[3]) + ct3 + CU3)
          cg4 = (cg4 + (n1 * coeff2[1]) + ct4 + CU4)
          cg5 = (cg5 + (n1 * coeff2[2]) + ct5 + CU5)
          cg6 = (cg6 + (n1 * coeff2[3]) + ct6 + CU6)
 END
          SIIaro = done
ELSE SIIaro=0;


RULE 16
IF       c_sub=1
and      sub1[x] = prim-OH
or       sub1[x] = sec-OH
THEN IT = I-OH
     FIND w
     c_t = done
```

Fig. 2. Examples of rules used in the knowledge base of CRIPES.

is found but continues to search for as many values as possible enabling lists of values to be entered. If after all the rules have been examined and a goal has still not been satisfied then the program checks whether there is an instruction to ask the user to input a value from the keyboard (ASK variable:" ", Fig. 1). If such an instruction is found, the program asks the user to input a value via the keyboard. An ASK statement may produce a menu (CHOICES, Fig. 1) of possible responses from which the user selects the required value. If no value can be found for the goal, it remains unknown and the consultation moves on to the next goal.

The program contains a trace facility that enables the path of a consultation to be monitored and subsequently displayed either as a text file or a graphic decision tree. It is also possible to discover which rule provided a value, why the consultation

wants to know a particular value and the value of any variable at the end of the consultation.

Implementation of CRIPES

Using CRIPES, the retention index of a compound is calculated from Eqn. 1. Each term in the equation can be described using a quadratic equation (eqn. 2), which are summed to give the final equation (eqn. 3) for the analyte retention index:

$$I = \Sigma ax^2 + \Sigma bx + \Sigma c \quad (3)$$

Two equations are obtained for each compound, describing the changes with the proportion of methanol and with acetonitrile in the eluent.

The approach used to obtain the coefficients of eqn. 3 is shown in the flow chart in Fig. 3. All the regression coefficients derived from the earlier experimental studies

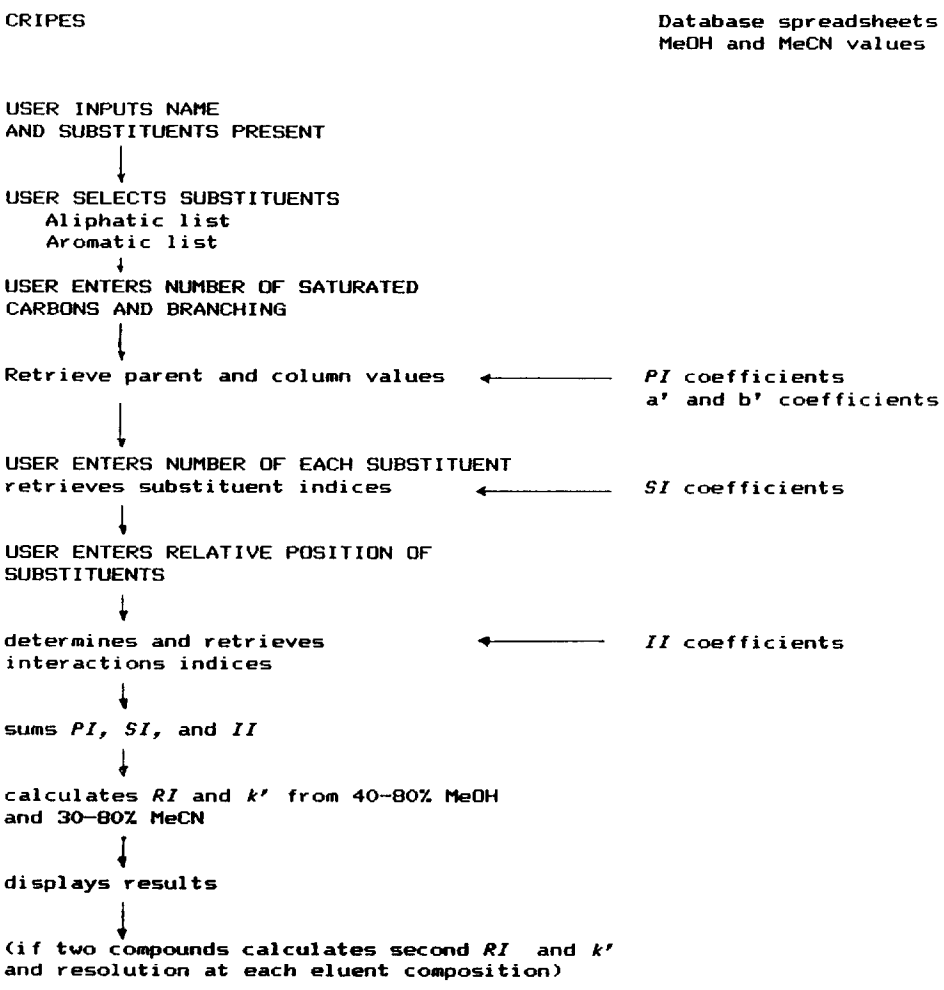

Fig. 3. Flow of information through CRIPES. User input in capitals and CRIPES actions in lower-case.

for I_P (*PI* in Fig. 3), I_S (*SI*) and I_I (*II*) are held in the external spreadsheets. First the coefficients for the parent compound[1] are taken from the database. The aromatic and aliphatic substituents are then treated separately. The user first enters the different types of aromatic substituents selected from menus presented by CRIPES (*i.e.*, Fig. 4), then the aliphatic substituents. Then the program asks the user to input the number of each aromatic substituent present in the compound and in turn the position(s) relative to a hydroxyl, amino or alkyl group if these are present. This enables CRIPES to extract the coefficients for the substituent indices[1] and to identify any interactions and extract the appropriate coefficients[4]. As a trial, the interactions for the amino group were assumed to be the same as for the phenolic hydroxyl group.

```
Which if any of these substituents is present?

COR          CHO          OR
CO2R         CH3          CONH2
OH           CHCH         ANOTHER

Use cursor key to select substituent
Press RETURN to enter and END when all choices have been made
```

Fig. 4. Menu of aromatic substituents.

These coefficients are summed to provide an overall equation describing the aromatic contribution. The program then repeats the process for the aliphatic substituents and determines any interaction with the phenyl ring[2]. Finally, the program prompts the user for information on the branching of the alkyl chain and for the length of alkyl chains in mixed alkyl–aryl groups such as PhCOR and $PhCO_2R$ to give the aliphatic chain contribution[3].

The coefficients for the aromatic, aliphatic and parent contributions are summed to give the overall equations, which are used to calculate the retention index values over the ranges 40–80% methanol and 30–80% acetonitrile at 10% intervals. The program is limited to these eluent ranges because the retentions of the model compounds used to establish the database became excessive in weaker eluents and at high proportions of modifier the retention times were often too short for accurate measurement[1]. Changes in the selectivity at high modifier compositions also suggested that extrapolation would be unreliable.

The final stage of the program is to calculate approximate capacity factors (k') to give an indication of the time required for a separation. This calculation is based on the relationship $\log k' = a'I + b'$, where a' and b' are known from the experimental regression equations for the retention of the alkyl aryl ketone standards[1]. The values of a' and b' are also dependent on the eluent composition and have been described by quadratic regression equations. As retention indices are generally independent of the brand of stationary phase[4,28], it should be possible to predict the capacity factors on any ODS-silica column if the capacity factors of the alkyl aryl ketones on that column are known. So far, CRIPES accesses a spreadsheet which contains details for Spherisorb ODS-2 from this study and for Hypersil-ODS calculated from earlier work in this laboratory[29,30].

Calculation of resolution

Using the regression equations for retention indices, in different eluents, it should be possible to calculate the conditions for the optimum separation between two compounds. However, although it is easy to determine the conditions for the maximum difference in retention indices, this does not correspond with the conditions for maximum resolution. This is because resolution is directly related to retention times whereas retention indices vary with $\log k'$. This problem has also been demonstrated by West[31] in studies based on the alkan-2-one retention index scale. It should be possible to convert the retention indices to capacity factors as indicated earlier and hence determine the conditions for maximum resolution. However, because both the indices and their conversion to capacity factors use quadratic expressions, the combined equation is too complex to solve within the expert system.

It was therefore necessary to take a simpler, more direct approach. For a pair of analytes, CRIPES calculates the combined coefficients for the retention indices of each compound and stores them in an external database file. These values are then used to calculate the predicted capacity factors and resolution at 10% intervals over the eluent ranges. The selection of experimental conditions can then be made manually, taking into account the length of analysis needed to achieve the desired resolution.

Testing CRIPES

CRIPES has been tested in two ways. First, the retention indices of a number of model compounds used to determine the substituent and interaction coefficients were predicted to check that the program was capable of extracting the appropriate data from the spreadsheets. These results matched the experimental values in each instance. Second, the retentions of a number of test compounds not previously examined were measured at selected eluent compositions and compared with those calculated by CRIPES both as retention indices (Tables I and II) and capacity factors (Table III and IV). In many instances there was close agreement between the values, suggesting that generally the prediction method was satisfactory. The compounds included in the trials include several that were selected to test specific aspects of the retention prediction. First, there were a number of compounds containing the grouping $-COCH_2Hal$ (including the phenacyl bromides) to test for aliphatic intramolecular interactions. Second, a limited number of substituted arylamines were examined. The retention indices of these compounds were calculated assuming that interaction terms were the same as those for the phenolic hydroxyl group.

In all the compounds, an assumption of additivity of the interaction indices was made. No judgement is currently exercised by CRIPES as to whether one, or more, of the interactions would dominate to the exclusion of other interactions, although it is intended to include the facility in future work. The interaction terms also do not include any interactions between aliphatic substituents (other than with a phenyl ring) or interactions between substituents on an alkyl side-chain and the aromatic substituents.

It is difficult to define for which compounds the experimental and calculated retention indices can be regarded as a close fit. The expected accuracy of an individual retention index value was determined as ± 10 units[32], but this uncertainty could apply to each substituent and interaction index value. The closeness of the fit between the experimental and calculated retention indices would therefore be expected to be

TABLE I

EXPERIMENTAL RETENTION INDICES (I_e) AND RETENTION INDICES CALCULATED (I_c) BY CRIPES FOR METHANOL-CONTAINING ELUENTS

Compound	*Methanol (%)*									
	40		*50*		*60*		*70*		*80*	
	I_e	I_c	I_e	I_c	I_e	I_c	I_e	I_c	I_e	I_c
Benzyl 2-bromoacetate			956	857	948	841	934	814	919	775
Phenacyl bromide			860	802	859	790	847	768	807	735
α-Bromo-*p*-phenylacetophenone					1169	1081			1120	1056
α-*p*-Dibromoacetophenone			1009	945	1005	932	993	915	971	892
4-Nitrophenacyl bromide			858	748	855	718	834	678	782	629
α-Chloro-3,4-dihydroxyacetophenone			547	513	547	480	(414)[a]	458	(161)[a]	448
o-Bromoaniline					840	914	840	890	818	853
m-Bromoaniline					814	828	804	815	752	794
o-Nitroaniline					764	668	760	625	711	575
m-Nitroaniline					695	696	687	675	649	647
p-Nitroaniline					648	505	642	444	556	376
N-Ethylaniline			875	858	886	860	894	854	879	841
Benzyl acetate	891	857	889	850	891	845	881	845	875	840
Benzyl chloromethyl ether			935	806	901	805	909	803	969	801
1-Bromo-2-nitrobenzene			931	993	930	995	921	998	889	1003
2-Bromo-4-methylphenol					895	922	874	916	830	908
tert.-Butylhydroquinone					771	799			637	786
p-tert-Butylphenol					998	1003	970	980	933	950
2-Chloromethyl-4-nitrophenol			427	695	507	662				
4,6-Dichloro-1,3-dihydroxybenzene					695	687			(307)[a]	594
3,4-Dimethoxyacetophenone					729	761			695	742
N,N-Dimethylbenzamide	717	803	709	791	694	779	683	766	665	753
2,6-Dimethyl-4-nitrophenol			746	830	708	812	559	787		
2,4-Dimethylphenol			863	872	838	864	848	849	792	829
2,5-Dimethylphenol			860	872	863	864	844	849	822	830
Dimethyl phthalate					785	874	757	859	716	847
Ethyl benzoate			995	1002	995	1006	998	1010	1004	1014
Ethyl 3-phenylpropionate			1046	1059	1045	1055	1040	1050	1037	1045
Ethyl phenylacetate			947	954	947	950	930	945	907	940
Ethyl phenylcyanoacetate	860	745	733	706	777	665	708	623	608	579
2-Hydroxybenzyl alcohol					605	439			554	346
4-Hydroxybenzyl alcohol					519	439			511	346
2-Hydroxy-5-nitrobenzyl bromide			429	720	510	690				
N-Methylbenzamide			636	691	623	679	616	666	591	653
m-Nitrobenzyl alcohol					688	628	688	608	607	582
p-Nitrobenzyl alcohol					678	612	665	588	601	559
4-Phenyl-1-butanol					912	931	888	915	841	893
5-Phenyl-1-pentanol					1003	1031	979	1015	937	993
n-Propyl *p*-hydroxybenzoate	940	893	914	873	882	848	844	818	809	783
Thymol	1041	1043	1035	1038	1030	1027	1001	1009	965	984

[a] The values in parentheses are considered to be unreliable because the corresponding capacity factors are less than 0.2

TABLE II

EXPERIMENTAL RETENTION INDICES (I_e) AND RETENTION INDICES CALCULATED (I_c) BY CRIPES FOR ACETONITRILE-CONTAINING ELUENTS

Compound	*Acetonitrile (%)*									
	40		*50*		*60*		*70*		*80*	
	I_e	I_c	I_e	I_c	I_e	I_c	I_e	I_c	I_e	I_c
Phenacyl bromide			891	763	878	751	857	745	832	746
Benzyl 2-bromoacetate	990	855					922	786	886	775
α-Bromo-*p*-phenylacetophenone					1135	994			1075	977
α-*p*-Dibromoacetophenone			1028	888	1015	880	994	883	972	896
4-Nitrophenacyl bromide			898	693	874	664	830	639	784	618
α-Chloro-3,4-dihydroxyacetophenone			593	449			538	393		
o-Bromoaniline			866	939	860	930	855	919	846	903
m-Bromoaniline			841	839	828	828	820	816	794	801
o-Nitroaniline			778	812	762	820	748	815	720	797
m-Nitroaniline			742	742	723	729	709	707	668	675
p-Nitroaniline			701	653	669	618	663	553	639	459
N-Ethylaniline			929	897	928	896	940	895	925	893
Benzyl acetate	883	873	877	863	869	852	859	841	844	829
Benzyl chloromethyl ether	645	788					624	758	629	750
1-Bromo-2-nitrobenzene			947	998	934	992	919	980	899	963
2-Bromo-2-methylphenol			863	897	846	889	823	887	812	892
tert.-Butylhydroquinone					978	768			974	783
p-tert.-Butylphenol			943	974	924	958	905	943	882	929
2-Chloromethyl-4-nitrophenol			432	642	567	592	(332)[a]	540	(281)[a]	484
4-Chloro-2-nitroaniline					881	925	868	906	846	876
4,6-Dichloro-1,3-dihydroxybenzene					627	697	601	673	574	658
3,4-Dimethoxyacetophenone					696	729			697	702
N,N-Dimethylbenzamide	653	735	631	714	633	714	643	737	683	781
2,6-Dimethyl-4-nitrophenol			770	791	743	759	699	724	528	686
2,4-Dimethylphenol			831	830	777	811	803	799	744	768
2,5-Dimethylphenol			829	830	817	811	795	799	783	768
Dimethyl phthalate			797	850	784	841	764	834	750	828
Ethyl benzoate			977	995	987	996	991	997	994	996
Ethyl 3-phenylpropionate	1036	1052	1030	1042	1021	1031	1011	1020	999	1008
Ethyl phenylacetate			947	963	940	952	921	941	905	929
Ethyl phenylcyanoacetate	914	763					819	655	754	618
2-Hydroxybenzyl alcohol					796	350			801	297
4-Hydroxybenzyl alcohol					520	337			463	324
2-Hydroxy-5-nitrobenzyl bromide			353	663	303	615	(264)[a]	569	(233)[a]	525
N-Methylbenzamide	589	635	560	614	552	614	549	637	558	682
p-Nitrobenzyl alcohol			638	549	628	530	624	516	599	507
m-Nitrobenzyl alcohol			668	561	643	542	541	528	611	519
4-Phenyl-1-butanol			821	856	801	830	807	829	797	836
5-Phenyl-1-pentanol			903	956	883	930	886	929	879	936
n-Propyl *p*-hydroxybenzoate			862	850			779	787	775	773
Thymol	1001	1008	984	991	968	972	955	951	941	929

[a] Values in parentheses are considered to be unreliable because the corresponding capacity factors are less than 0.2.

TABLE III

EXPERIMENTAL CAPACITY FACTORS (k'_e) AND CAPACITY FACTORS ESTIMATED (k'_c) BY CRIPES FOR METHANOL-CONTAINING ELUENTS

Compound	*Methanol (%)*									
	40		*50*		*60*		*70*		*80*	
	k'_e	k'_c	k'_e	k'_c	k'_e	k'_c	k'_e	k'_c	k'_e	k'_c
Phenacyl bromide			3.94	2.99	2.48	1.52	1.13	0.82	0.69	0.48
Benzyl 2-bromoacetate			8.71	4.48	3.64	2.07	1.81	1.02	0.89	0.55
α-Bromo-*p*-phenylacetophenone					15.92	9.00			1.96	1.30
α-*p*-Dibromoacetophenone			10.91	8.55	5.95	3.59	2.19	1.64	1.20	0.83
4-Nitrophenacyl bromide			3.88	2.01	2.42	0.98	1.07	0.53	0.64	0.33
α-Chloro-3,4-dihydroxyacetophenone			0.47	0.36	0.38	0.23	(0.16)[a]	0.19	(0.08)[a]	0.18
o-Bromoaniline					2.21	3.21	1.10	1.45	0.72	0.72
m-Bromoaniline					1.89	1.91	0.94	1.02	0.57	0.59
o-Nitroaniline					1.40	0.72	0.77	0.41	0.50	0.28
m-Nitroaniline					0.93	0.86	0.55	0.52	0.41	0.35
p-Nitroaniline					0.70	0.27	0.45	0.17	0.30	0.14
N-Ethylaniline			4.36	4.52	2.91	2.32	1.40	1.23	0.96	0.69
Benzyl acetate	13.88	9.52	5.51	4.38	2.76	2.18	1.41	1.18	0.76	0.69
Benzyl chloromethyl ether			7.47	3.09	2.98	1.66	1.61	0.96	1.07	0.60
1-Bromo-2-nitrobenzene			6.38	12.18	3.79	5.24	1.58	2.44	0.91	1.22
2-Bromo-4-methylphenol					3.08	3.37	1.28	1.65	0.75	0.88
tert.-Butylhydroquinone					1.46	1.61			0.39	0.57
p-tert.-Butylphenol					5.72	5.49	1.96	2.24	1.05	1.02
2-Chloromethyl-4-nitrophenol			0.21	1.37	0.30	0.70				
4,6-Dichloro-1,3-dihydroxybenzene					0.93	0.81			(0.13)[a]	0.29
3,4-Dimethoxyacetophenone					1.13	1.27			0.48	0.49
N,N-Dimethylbenzamide	2.82	5.97	1.46	2.76	0.85	1.42	0.55	0.81	0.36	0.51
2,6-Dimethyl-4-nitrophenol			1.82	3.68	1.00	1.73	0.30	0.89		
2,4-Dimethylphenol			4.03	5.01	2.18	2.37	1.14	1.20	0.66	0.67
2,5-Dimethylphenol			3.94	5.01	0.73	2.37	1.12	1.20	0.73	0.67
Dimethyl phthalate					1.58	2.52	0.76	1.25	0.51	0.71
Ethyl benzoate			11.76	13.01	4.80	5.60	2.44	2.58	1.27	1.27
Ethyl 3-phenylpropionate			16.84	19.75	6.44	7.55	2.99	3.13	1.38	1.41
Ethyl phenylacetate			7.11	9.14	4.21	4.00	1.65	1.90	0.97	0.98
Ethyl phenylcyanoacetate	9.92	3.62	1.72	1.48	1.32	0.71	0.62	0.41	0.29	0.28
2-Hydroxybenzyl alcohol					0.54	0.18			0.30	0.12
4-Hydroxybenzyl alcohol					0.32	0.18			0.26	0.12
2-Hydroxy-5-nitrobenzyl bromide			0.21	1.64	0.30	0.83				
N-Methylbenzamide	1.46	2.52	0.86	1.33	0.55	0.78	0.40	0.50	0.28	0.36
m-Nitrobenzyl alcohol					0.89	0.57	0.56	0.38	0.36	0.28
p-Nitrobenzyl alcohol					0.83	0.52	0.50	0.35	0.35	0.26
4-Phenyl-1-butanol					3.40	3.55	1.36	1.64	0.78	0.83
5-Phenyl-1-pentanol					5.87	6.50	2.03	2.65	0.36	1.18
n-Propyl *p*-hydroxybenzoate	20.02	13.06	6.56	5.05	2.46	2.16	1.18	1.03	0.62	0.57
Thymol	62.29	47.53	12.99	16.97	6.91	6.37	2.26	2.56	1.17	1.14

[a] As in Table I

TABLE IV

EXPERIMENTAL CAPACITY FACTORS (k'_e) AND CAPACITY FACTORS ESTIMATED (k'_c) BY CRIPES FOR ACETONITRILE-CONTAINING ELUENTS

Compound	*Acetonitrile (%)*									
	40		*50*		*60*		*70*		*80*	
	k'_e	k'_c	k'_e	k'_c	k'_e	k'_c	k'_e	k'_c	k'_e	k'_c
Phenacyl bromide			3.24	1.51	1.69	0.94	0.90	0.64	0.67	0.47
Benzyl 2-bromoacetate	9.91	4.43					1.23	0.73	0.64	0.51
α-Bromo-*p*-phenylacetophenone					4.92	2.53			1.33	0.92
α-*p*-Dibromoacetophenone			4.95	2.87	2.99	1.59	1.40	1.02	0.99	0.73
4-Nitrophenacyl bromide			2.61	1.06	1.66	0.66	0.82	0.45	0.58	0.33
α-Chloro-3,4-dihydroxyacetophenone			0.61	0.31			0.32	0.20		
o-Bromoaniline			2.41	3.71	1.57	1.95	0.97	1.14	0.70	0.74
m-Bromoaniline			2.14	2.24	1.37	1.29	0.87	0.81	0.60	0.55
o-Nitroaniline			1.57	1.95	1.04	1.24	0.69	0.81	0.49	0.55
m-Nitroaniline			1.33	1.36	0.89	0.86	0.61	0.56	0.42	0.38
p-Nitroaniline			0.52	0.87	0.71	0.55	1.09	0.34	0.39	0.21
N-Ethylaniline			3.03	3.00	2.08	1.70	1.18	1.06	0.87	0.72
Benzyl acetate	5.95	5.00	2.87	2.53	1.64	1.42	0.99	0.88	0.57	0.60
Benzyl chloromethyl ether	1.07	2.89					0.43	0.67	0.30	0.48
1-Bromo-2-nitrobenzene			3.30	5.03	2.13	2.51	1.10	1.40	0.81	0.88
2-Bromo-2-methylphenol			2.38	3.00	1.48	1.65	0.88	1.03	0.63	0.72
tert.-Butylhydroquinone					2.56	1.01			1.00	0.52
p-tert.-Butylphenol			3.51	4.45	2.04	2.19	1.15	1.24	0.77	0.80
2-Chloromethyl-4-nitrophenol			0.28	0.82	0.46	0.49	(0.16)[a]	0.32	(0.14)[a]	0.22
4-Chloro-2-nitroaniline					2.04	1.89	1.02	1.10	0.70	0.68
4,6-Dichloro-1,3-dihydroxybenzene					0.59	0.76	0.39	0.51	0.32	0.36
3,4-Dimethoxyacetophenone					0.79	0.86			0.46	0.41
N,N-Dimethylbenzamide	1.12	2.06	0.81	1.18	0.60	0.81	0.46	0.62	0.35	0.52
2,6-Dimethyl-4-nitrophenol			1.42	0.82	0.96	0.49	0.54	0.32	0.28	0.22
2,4-Dimethylphenol			1.90	2.13	1.11	1.20	0.75	0.74	0.52	0.50
2,5-Dimethylphenol			2.02	2.13	1.31	1.20	0.80	0.74	0.58	0.50
Dimethyl phthalate			1.73	2.36	1.14	1.36	0.73	0.86	0.53	0.60
Ethyl benzoate			5.07	4.95	2.70	2.55	1.57	1.48	0.83	0.97
Ethyl 3-phenylpropionate	13.39	15.76	6.18	6.29	3.12	2.94	1.69	1.60	0.89	1.00
Ethyl phenylacetate			3.30	4.20	2.18	2.13	1.11	1.23	0.82	0.80
Ethyl phenylcyanoacetate	6.09	2.47					0.86	0.48	0.43	0.32
2-Hydroxybenzyl alcohol					1.20	0.17			0.61	0.13
4-Hydroxybenzyl alcohol					0.38	0.17			0.24	0.13
2-Hydroxy-5-nitrobenzyl bromide			0.31	0.91	0.27	0.54	(0.16)[a]	0.36	(0.12)[a]	0.25
N-Methylbenzamide	0.74	1.08	0.56	0.71	0.43	0.54	0.33	0.45	0.24	0.39
p-Nitrobenzyl alcohol			0.80	0.51	0.60	0.38	0.46	0.30	0.35	0.24
m-Nitrobenzyl alcohol			0.92	0.54	0.63	0.40	0.50	0.31	0.36	0.24
4-Phenyl-1-butanol			1.95	2.23	1.22	1.30	0.83	0.85	0.61	0.61
5-Phenyl-1-pentanol			2.38	3.72	1.72	1.95	1.08	1.18	0.76	0.82
n-Propyl *p*-hydrobenzoate	4.33	4.30					0.74	0.74	0.44	0.51
Thymol	12.73	11.91	4.47	4.84	2.46	2.31	1.35	2.31	0.91	0.80

[a] As in Table II.

dependent on the number of terms used in its derivation. Because of difficulties in measuring retentions accurately (both of the model compounds used to derive the indices and of the experimental results for the test compounds), a larger uncertainty was expected with compounds with short retention times, especially at 80% organic modifier. Although the errors in capacity factors often seem large in absolute terms, they are usually much better than those found with simpler models[24].

Phenacyl halides. The calculated retention indices (I_c) of the phenacyl bromides and chlorides and benzyl 2-bromoacetate were significantly smaller than the experimental values (I_e). These observations suggest that a consistent interaction was occurring between the halogen groups and the adjacent carbonyl group and that probably an interaction index term needs to be determined.

Substituted arylamines. Five substituted anilines were studied. In both methanol- and acetonitrile-containing eluents the retention indices of *m*-bromoaniline and *m*-nitroaniline were close to the calculated values. However, for the *ortho* and *para* isomers the correlations were poorer. The deviations from the calculated values for the *ortho* substituents suggest that the interactions with the amino group may be very different in magnitude to those of the phenolic hydroxyl group. In addition, the value of the interaction index derived for *p*-hydroxy–nitro substituents may have been distorted by partial ionisation of the acidic *p*-nitrophenol model compound. However, the calculated and experimental retention indices of the secondary amine N-ethylaniline were fairly close.

Arylamides. The secondary and tertiary amides N-methylbenzamide and N,N-dimethylbenzamide showed relatively large deviations from the retention indices calculated on the basis of a primary carboxamide group (about -80 units for N,N-dimethylbenzamide and -40 to -130 for N-methylbenzamide). Separate substituent index terms for the alkylamides substituents (rather than a single amide term) will clearly be required.

Other compounds. Two alcohols, 4-phenyl-1-butanol and 5-phenyl-1-propanol, were included to test the assumption that interactions between aliphatic substituents and a phenyl group are not significant for carbon chain lengths >2. Some deviation occurred but the differences were not large. The disubstituted aliphatic compounds (benzyl chloromethyl ether and ethyl phenylcyanoacetate) both showed large deviations, suggesting that electronic and steric effects may be also important for aliphatic compounds.

The aromatic monoesters gave a good match but phthalate esters showed significant deviations, suggesting that there was steric interaction between the substituent groupings. Many simpler compounds, *e.g.*, thymol, 2,5-dimethylphenol and 3,4-dimethoxyacetophenone, had calculated retention indices that were within ± 10 units (for each substituent) of the experimental values.

However, for many, particularly disubstituted, compounds there were large differences between the experimental and calculated retention indices, showing that many interactions are not at present accounted for by the interaction indices in the database. Similar problems are encountered in the prediction of octanol–water partition coefficients, but even the imperfect models are often found to give useful results. The occurrence of deviations from calculated values can also act as markers for the presence of intramolecular interactions and encourage their study.

Although these compounds were examined as test compounds, in many in-

stances together with additional suitably selected model compounds, these polyfunctional compounds will be used to extend and refine the database of interaction terms in the future.

CONCLUSION

The expert system CRIPES can be readily used to provide an interface for the user to predict retention indices from molecular structure by combining a set of rules with retention coefficients held in an external database. The database does not yet include sufficient values for interactions 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. A comparison of the capacity factors suggested that in most instances reasonable indication of the retention times under different conditions would be given.

ACKNOWLEDGEMENTS

Thanks are due to Phase Separations for the gift of Spherisorb ODS-2 and to the Science and Engineering Research Council for a research grant and studentship to C.M.B.

REFERENCES

1 R. M. Smith and C. M. Burr, *J. Chromatogr.*, 475 (1989) 57.
2 R. M. Smith and C. M. Burr, *J. Chromatogr.*, 481 (1989) 71.
3 R. M. Smith and C. M. Burr, *J. Chromatogr.*, 481 (1989) 85.
4 R. M. Smith and C. M. Burr, *J. Chromatogr.*, in preparation.
5 T. Fujita, *J. Pharm. Sci.*, 72 (1983) 285.
6 P. B. Ayscough, S. J. Chinnick, R. Dybowski and P. Edwards, *Chem. Ind. (London)*, 15 (1987) 515.
7 A. F. Fell, T. P. Bridge and M. H. Williams, *J. Pharm. Biomed. Anal.*, 6 (1988) 555.
8 J. L. Glajch, *LC · GC, Mag. Liq. Gas Chromatogr.*, 6 (1988) 30.
9 G. J. Kleywegt, *Lab. Microcomput.*, 6 (1987) 74.
10 T. H. Pierce and B. A. Hohne (Editors), *Artificial Intelligence Applications in Chemistry*, ACS Symposium Series, No. 306, American Chemical Society, Washington, DC, 1986.
11 A. P. Wade, S. R. Crouch and D. Betteridge, *Trends Anal. Chem.*, 7 (1988) 358.
12 R. E. Dessy, *Anal. Chem.*, 56 (1984) 1200A.
13 R. E. Dessy, *Anal. Chem.*, 56 (1984) 1312A.
14 M.-P. Derde, L. Buydens, C. Guns, D. L. Massart and P. K. Hopke, *Anal. Chem.*, 59 (1987) 1868.
15 C. Hansch and A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York, 1979.
16 M. A. Tischler and E. A. Fox, *Comput. Chem.*, 11 (1987) 235.
17 R. Bach, J. Karnicky and S. Abbott, in T. H. Pierce and B. A. Hohne (Editors), *Artificial Intelligence Applications in Chemistry*, ACS Symposium Series, No. 306, American Chemical Society, Washington, DC, 1986, Ch. 22, p. 278.
18 P. Lu and X. Lu, *J. Chromatogr.*, 292 (1984) 169.
19 P. Lu and H. Huang, *J. Chromatogr.*, 452 (1988) 175.
20 H. Gunasingham, B. Srinivasan and A. L. Ananda, *Anal. Chim. Acta*, 182 (1986) 193.
21 A. L. Ananda, S. M. Foo and H. Gunasingham, *J. Chem. Inf. Comput. Sci.*, 28 (1988) 82.
22 G. Musch, M. De Smet and D. L. Massart, *J. Chromatogr.*, 348 (1985) 97.
23 G. Musch and D. L. Massart, *J. Chromatogr.*, 370 (1986) 1.
24 M. De Smet, A. Peeters, L. Buydens and D. L. Massart, *J. Chromatogr.*, 457 (1988) 25.
25 P. J. Schoenmakers and M. Mulholland, *Chromatographia*, 25 (1988) 737.

26 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.
27 A. Peeters, L. Buydens, D. L. Massart and P. J. Schoenmakers, *Chromatographia*, 26 (1988) 101.
28 R. M. Smith, *LC · GC, Mag. Liq. Gas Chromatogr.*, 3 (1984) 186.
29 R. M. Smith, *J. Chromatogr.*, 236 (1982) 313.
30 R. M. Smith, G. A. Murilla and C. M. Burr, *J. Chromatogr.*, 388 (1987) 37.
31 S. D. West, *J. Chromatogr. Sci.*, 27 (1989) 2.
32 R. M. Smith and C. M. Burr, *J. Chromatogr.*, 475 (1989) 75.