Journal of Chromatography, 465 (1989) 59–67
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROMSYMP. 1534

ARTIFICIAL INTELLIGENCE TECHNIQUES FOR PEAK HOMOGENEITY VALIDATION IN LIQUID CHROMATOGRAPHY

T. P. BRIDGE and M. H. WILLIAMS
Department of Computer Science, Heriot-Watt University, 79 Grassmarket, Edinburgh EH1 2HJ (U.K.)
and
A. F. FELL*
School of Pharmaceutical Chemistry, University of Bradford, Bradford, BD7 1DP (U.K.)

SUMMARY

A problem area in the interpretation of chromatographic data is the determination of the purity of a detected peak. The problem arises in deciding whether a peak is due solely to one component, or is caused by the simultaneous elution of two or more components. Whilst a number of multivariate statistical techniques are available to assess the number of components contributing to a given signal, they may require a significant amount of computer power to provide an answer, and are prone to noise-induced errors. As part of a programme of research aimed at developing an intelligent system for mobile phase optimisation, we have developed a knowledge-based systems approach to the problem of peak homogeneity assessment. The expert system applies a number of simple numerical tests to the three-dimensional dataset produced by a photodiode-array detector and assess the results in light of the chromatographic signal.

INTRODUCTION

The authors are engaged in the development of an expert system to assist in the development of an optimum separation for a mixture in reversed-phase liquid chromatography[1,2]. The expert system determines the appropriate solvent strength by means of a gradient elution experiment, and subsequently establishes the mobile phase composition to achieve optimum selectivity. The selectivity optimisation is achieved by adjusting the relative concentrations of three organic modifiers according to either a modified simplex[3] or iterative regression optimisation strategy[4].

The choice of optimisation strategy is guided by the amount of reliable information which can be extracted from the isoeluotropic chromatograms. The iterative regression method relies on the ability of the system to determine capacity factors for each solute under a number of different elution conditions, to provide a basis from which a mathematical model of the effect of changes in mobile-phase composition on selectivity, can be produced. The optimum composition for the mobile

0021-9673/89/$03.50

phase may then be identified by examination of the modelled response surface. The basis for identification of each solute in each chromatogram, and hence for allowing the variation in capacity factor with mobile-phase composition to be tracked, is the spectral information derived from a linear photodiode-array detector. The system compares the spectrum associated with one component with the spectra derived from other separations of the same mixture. If the matching process is successful, the variation in retention time, and hence capacity factor, can be determined for each detected solute.

However, this matching process may fail for reasons which include the variation in spectral profile caused by the change in solvent, the inability to discriminate between two or more components with similar spectral profiles, or the masking of components as a result of simultaneous elution. In such cases, the optimisation process is directed by a modified simplex algorithm, which requires no spectral information until sufficient additional information has been acquired to allow the iterative regression procedure to resume control of the optimisation process.

The reliance of the overall optimisation on the simplex method will be reduced if spectra can be extracted from peaks resulting from the simultaneous elution of two or more components. This may be achieved by the use of appropriate chemometric tools, such as factorial analysis[5]. A disadvantage of employing such techniques is the amount of computer power required to produce meaningful spectral profiles. The overall computer power requirement can be reduced significantly if the use of factorial analysis is restricted to heterogeneous peaks, where the spectral information is only available after deconvolution, and not on homogeneous peaks, from which no additional information can be derived by the application of such techniques.

To this end, the knowledge-based system is provided with the means to call upon a number of rapid homogeneity screening tests, the results of which are used in combination to assess the need for applying full matrix-based deconvolution methods.

METHODS FOR ASSESSING PEAK HOMOGENEITY

The peak homogeneity tests provided for use by the knowledge-based system were selected on the basis of their rapidity and reliability[6]. Whilst no single test is capable of detecting all cases of simultaneous elution, the use of a number of different tests in combination will increase the probability of correctly identifying heterogeneous peaks. The tests available within the system are described briefly below.

Determination of difference in spectral profile

The normalised profiles of spectra taken from the leading and trailing edges of the peak are compared. Spectra are extracted from the data matrix at times before and after the peak maximum, where the amplitude of the total absorbance signal is one third of that recorded at the peak maximum. These spectra are normalised at the wavelength of maximum absorption of the leading edge spectrum. The square root of the sum of the squares (RSS) of the differences between the two profiles is calculated, and this value is passed to the logic-processing element of the knowledge-based system.

Derivative total absorbance chromatograms

The profile of a differentiated curve is often a more sensitive indicator of the

existence of an otherwise hidden minor component than is the original zero-order curve[7]. To make use of this effect, an approximation to the time-domain derivative of the total absorbance chromatogram is calculated, and the number of zero crossings in the derivative profile is determined. For a pure peak, only one zero crossing will be produced. To reduce the effect of noise in the derivative signal in determining the true number of such crossings, a crossing is only accepted as valid if the value of the total absorbance chromatogram at that time is above a predefined threshold. First-derivative chromatograms of pure and composite peaks are illustrated in Fig. 1.

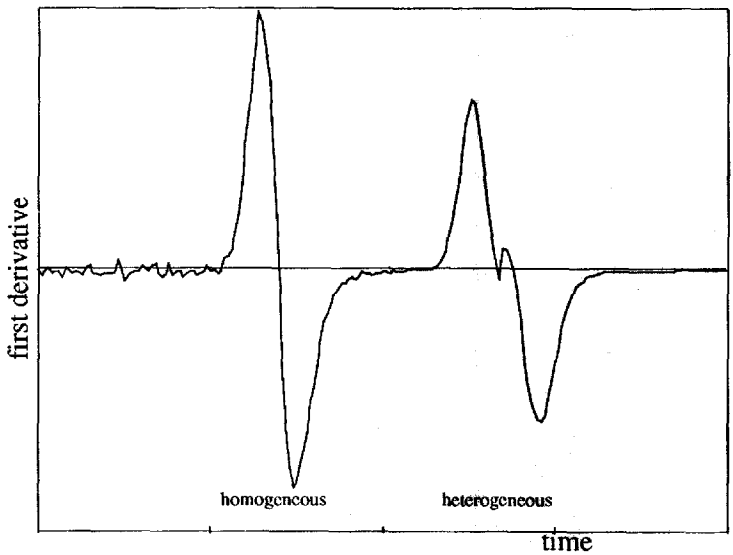

Fig. 1. First-derivative chromatogram.

As a consequence of the band-sharpening produced by differentiation, the second-derivative profile is more sensitive to the presence of minor components than is the first derivative. To produce the second-derivative profile, the first-derivative chromatogram is further differentiated numerically. As for the first-derivative profile, the number of zero crossings is used as an indicator of homogeneity. Two such crossings will be produced by a pure peak. Second-derivative chromatograms of pure and composite peaks are illustrated in Fig. 2.

Absorbance ratio chromatogram

The ratio of absorbances measured at two wavelengths should be invariant with time for a homogeneous peak[8]. This test locates the time at which maximum absorbance occurs in the total absorbance chromatogram and identifies the wavelength of maximum absorbance as well as the wavelength associated with an absorbance of half the maximum, at this time. The ratio of the signals at these two wavelengths is then calculated as a function of time across the peak. As the ratio is not evaluated for signals below a pre-determined threshold, to reduce the effect of noise on the output ratiogram, and as the ratio of the absorbances is independent of

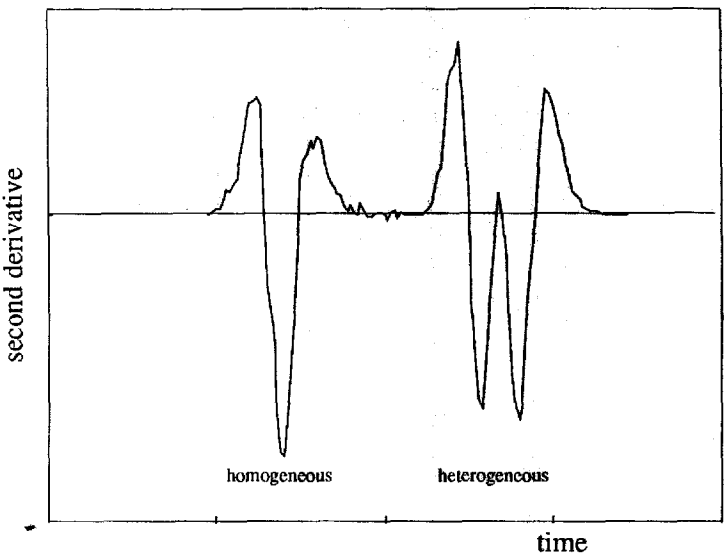

Fig. 2. Second-derivative chromatogram.

concentration, and therefore time, a pure peak produces a ratiogram with a square-wave profile. To assess the variation from this ideal profile, the RSS deviation from the value of the ratio at the midpoint of the peak is determined and reported. For a pure peak this should be zero, or for a real system it should not exceed the variation in signal due to detector noise. Ratiograms derived from homogeneous and heterogeneous peaks are illustrated in Fig. 3.

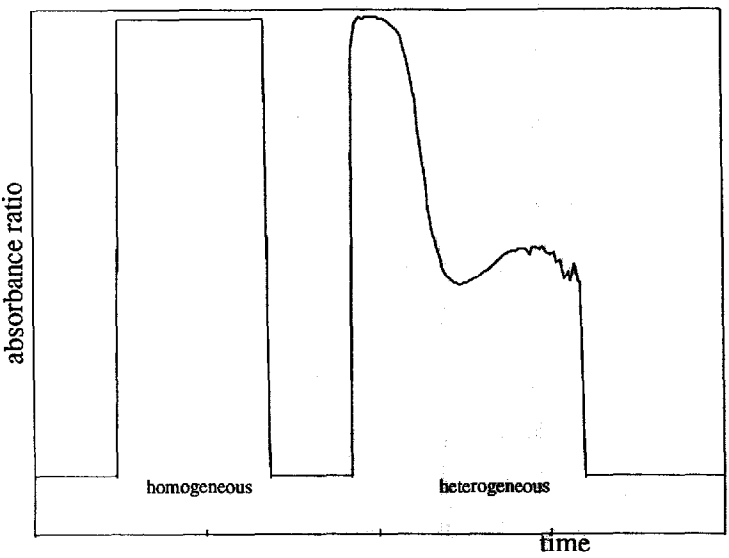

Fig. 3. Ratiogram of pure and impure peaks.

The idea of absorbance ratio evaluation may be extended into the derivative domain[9] if the chromatograms extracted, as above, at the wavelengths of maximum absorbance and half-maximum absorbance respectively, are differentiated before the ratio is evaluated. This may offer improved sensitivity towards certain cases of peak overlap, but at the cost of reduced indifference to noise in the spectrum.

Variation of apparent retention time

The observed retention time of a pure component will be independent of the wavelength used to monitor the chromatogram. However, in the case of a peak resulting from the simultaneous elution of two or more components, a variation in apparent retention time may be observed[10]. The degree of variation depends on the difference in the spectral profiles of the components, their separation and relative amplitudes. To quantify this variation for a single peak, the times of maximum absorbance are determined for each detector channel, and the difference between the largest and smallest values reported. Plots of retention time as a function of wavelength for two peaks are given in Fig. 4.

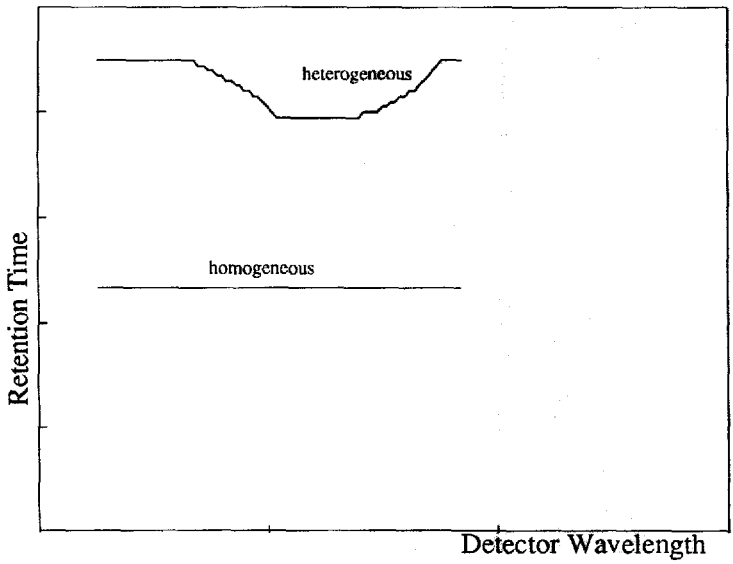

Fig. 4. Variation of retention time for pure and impure peaks.

THE EXPERT SYSTEM

The software forming the knowledge-based system is hosted on an IBM PS/2 micro-computer, operated with PC-DOS 3.30. The software is arranged in a highly modular fashion to facilitate development. Such a structure also allows the use of languages appropriate to the local task. Thus, whilst the logic processing is performed by modules written in microProlog, the homogeneity tests are coded in Pascal. The selection and execution of the various modules is controlled by a supervisory

Prolog Code for Peak Evaluation

```
The query is: ?((homogeneity assessed))

((homgencity assessed)
 (peak_is_pure yes))

((homgencity assessed)
 (peak_is_impure yes))

((peak_is_impure yes)
 (NOT derivative_check ok)
 (NOT spectrum_check ok))

((peak_is_impure yes)
 (NOT derivative_check ok)
 (NOT ratio_check ok))

((peak_is_impure yes)
 (NOT derivative-ratio_check ok)
 (NOT derivative_check ok))

 ...

((derivative_check ok)
 (first_derivative_test ok)
 (second_derivative_test ok))

((derivative-ratio_check ok)
 (Evaluate derivative_error x)
 (x LESS limit_deriv_ratio))

((first_derivative_test ok)
 (Evaluate d1_zero_crossings x)
 (x LESS 2)
 (0 LESS x))

 ...

((Evaluate x y)
 (EXECUTE x)
 (OPEN "result_file")
 (READ "result_file" y)
 (CLOSE "result_file"))
```

Fig. 5. Extract from the PROLOG code used to interpret the homogeneity tests.

microProlog program. A section of the logic routine used to interpret the results of the homogeneity tests is reproduced in Fig. 5. The logic employed by the expert system to evaluate peak homogeneity is illustrated by the flowchart of Fig. 6.

EVALUATION

The operation of the expert system was evaluated by assessing the homogeneity of 33 chromatographic peaks of known composition. The composition of each of these

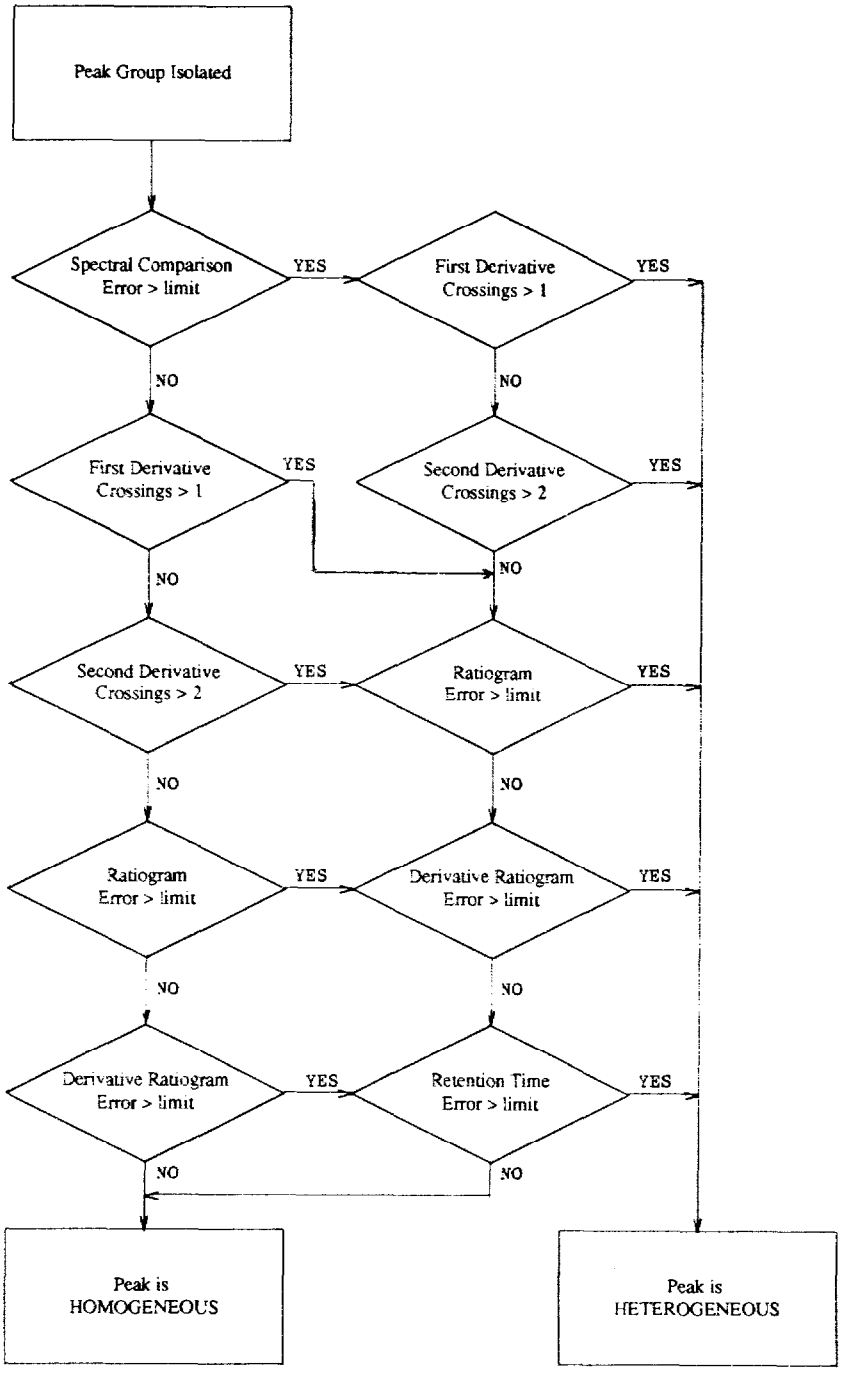

Fig. 6. Flowchart of the logic used to assess peak homogeneity.

TABLE I

COMPOSITION OF THE TEST CHROMATOGRAMS

No. of components	*Component*	*Relative amplitude*	*Separation (s)*	*No. of components*	*Component*	*Relative amplitude*	*Separation (s)*
1	Fluoranthrene	1		1	1,3-Xylenol	1	
1	Biphenyl	1		1	3,4-Xylenol	1	
1	Fluorene	1		1	2,3-Xylenol	1	
1	Pyrene	1		1	2,5-Xylenol	1	
2	Fluorene	1		2	2,3-Xylenol	1	
	Pyrene	1	20		3,4-Xylenol	1	20
2	Fluoranthrene	1		2	1,3-Xylenol	1	
	Biphenyl	1	20		2,5-Xylenol	1	20
3	Pyrene	1		3	2,3-Xylenol	1	
	Biphenyl	1	20		1,5-Xylenol	1	20
	Fluorene	1	20		1,3-Xylenol	1	20
4	Fluoranthrene	1		4	2,3-Xylenol	1	
	Biphenyl	1	20		3,4-Xylenol	1	20
	Fluorene	1	20		2,5-Xylenol	1	20
	Pyrene	1	20		1,3-Xylenol	1	20
2	Biphenyl	1		2	1,3-Xylenol	1	
	Pyrene	1	40		2,5-Xylenol	1	40
2	Biphenyl	1		2	1,3-Xylenol	1	
	Pyrene	1	20		2,5-Xylenol	1	30
2	Biphenyl	1		2	1,3-Xylenol	1	
	Pyrene	1	10		2,5-Xylenol	1	10
2	Biphenyl	1		2	1,3-Xylenol	1	
	Pyrene	1	5		2,5-Xylenol	1	5
2	Fluorene	1					
	Biphenyl	1	20				
2	Fluorene	1					
	Biphenyl	1	10				
2	Fluorene	1					
	Biphenyl	1	5				
2	Fluorene	2					
	Biphenyl	1	20				
2	Fluorene	2					
	Biphenyl	1	10				
2	Fluorene	2					
	Biphenyl	1	5				
2	Fluorene	4					
	Biphenyl	1	20				
2	Fluorene						
	Biphenyl	1	10				
2	Fluorene	4					
	Biphenyl	1	5				

peaks is indicated in Table I. As a comparative exercise the same chromatographic data were interpreted by a number of chromatographers in our laboratories. The results of the evaluation are expressed in Table II.

TABLE II

COMPARISON OF PEAK HOMOGENEITY ASSESSMENT BY THE EXPERT SYSTEM AND HUMAN EXPERTS

Analyst 1	*Analyst 2*	*Analyst 3*	*Analyst 4*	*Expert system*
Homogeneous peaks correctly assessed (from a total of 8)				
2	4	2	5	8
Heterogeneous peaks correctly assessed (from a total of 25)				
25	24	21	24	22

DISCUSSION

In the limited evaluation exercise reported above, the knowledge-based system out-performs the human analysts in the recognition of homogeneous peaks, and is of comparable performance in the detection of heterogeneous peaks caused by the simultaneous elution of two or more components. In such an exercise the human analysts might be expected to out perform the computer system, since they are using a visual interpretation of the complete chromatographic profile, rather than the limited numerical sample available to the expert system.

The use of a limited degree of artificial intelligence in the interpretation of rapid numerical tests offers a method for improving the reliability of peak homogeneity assessment in cases where the resources are not available for more comprehensive analysis by chemometric methods, such as iterative target testing factor analysis.

REFERENCES

1 T. P. Bridge, M. H. Williams, G. G. R. Seaton and A. F. Fell, *Chromatographia*, 24 (1987) 691.
2 T. P. Bridge, M. H. Williams and A. F. Fell, *Anal. Proc.*, 25 (1988) 43.
3 J. C. Berridge, *Analyst (London)*, 109 (1984) 291.
4 A. C. J. H. Drouen, *Ph.D. Thesis*, Technical University of Delft, Delft, 1985.
5 G. G. R. Seaton and A. F. Fell, *Chromatographia*, 24 (1987) 208.
6 T. P. Bridge, M. H. Williams and A. F. Fell, *Anal. Chim. Acta*m (1988) in press.
7 T. P. Bridge, A. F. Fell and R. H. Wardman, *J. Soc. Dyers Colour.*, 103 (1987) 17.
8 A. F. Fell, H. P. Scott, R. Gill and A. C. Moffa, *J. Chromatogr.*, 282 (1983) 123.
9 A. F. Fell, T. P. Bridge and M. H. Williams, *J. Pharm. Biomed. Anal.*, 6 (1988) 555.
10 A. G. Wright, A. F. Fell and J. C. Berridge, *Chromatographia*, 24 (1987) 533.