Multiple absorbance ratio correlation — a new approach for assessing peak purity in liquid chromatography

J. G. D. MARR[a,*], G. G. R. SEATON, B. J. CLARK and A. F. FELL

Pharmaceutical Chemistry, School of Pharmacy, University of Bradford, Bradford BD1 7DP (U.K.)

SUMMARY

The theory of multiple absorbance ratio correlation (MARC), a novel method for examining chromatographic peaks, is presented. The MARC hypothesis provides a method for examining chromatographic peaks for homogeneity, identity and purity, operating on multi-wavelength data sets which do not contain full spectra. Paraxanthine (PX) is an endogenous caffeine metabolite which is spectrally similar ($r > 0.99$) to theophylline (TH) and which often coelutes with it. Multi-wavelength data, collected using an HP1040A diode-array detector, for a series of binary mixtures of TH and PX ($0.7 < R_s < 2.9$; 0–50% PX), was analysed using MARC. The algorithm calculates a correlation factor between an internal (for homogeneity) or an external (for identity) reference 5-point vector and a similar 5-point vector acquired at sequential time intervals throughout the peak being investigated. For samples where the binary peak was homogeneous, limits of detection for PX of 1% (w/w) were obtained. This compared with limits of 5–10% (w/w) for the standard absorbance ratio (AR) method. MARC was found to be less sensitive to wavelength choice than AR. The sensitivity of MARC was independent of the chromatographic resolution.

INTRODUCTION

One of the main problems associated with chromatographic method development is the determination of the purity or homogeneity of the analyte peaks. With the advent of computer-aided UV–VIS detection methods, notably in the form of diode-array detection (DAD), chemometric methods have been developed which enable the analyst to verify the homogeneity and/or identity of peaks by manipulating the spectrochromatographic data set[1–4]. However, these methods are very computer intensive and confidence in the results is governed by the operators appreciation of the limitations of the algorithms employed[5,6].

The potential for using UV-absorbance data, at a defined number of wavelengths, for peak deconvolution and solute identification was first recognised in a theoretical analysis by Ostojic[7]. Due to a lack of suitable detector technology, initial

---

[a] Present address: Control Development, Upjohn Ltd., Fleming Way, Crawley RH10 2NJ, U.K.

application of the theory required repeated analysis, two detectors in series or stop-flow conditions[8-10]. However, the advent of multi-wavelength detectors permitted the necessary data to be collected simultaneously in one analysis[11,12]. Subsequently, various workers have investigated the limitations of the theory of absorbance ratioing as it is applied in the chromatographic environment[13,14]. In order to overcome the constraints imposed by the choice and number of wavelengths used without resorting to the use of computer-intensive chemometric methods, White[15,16] has proposed the use of three or more absorbance ratio values to characterise a compound. A slightly different approach has been proposed as a general screening method by Poile and Conlon[17] who utilised data extracted at nine pre-defined wavelengths from the apical spectrum as a measure of 'peak purity'. The algorithm employed by Alfredson and Sheehan[18] also utilises data collected from pre-defined areas of the apical and in-flection point spectra as a means of evaluating the 'purity' and identity of the chromatographic peak. If the identity of at least one of the eluting components is known then spectral suppression, using two discrete wavelengths[19-21], may be used. In the more general application of this principle, where N ($N \geqslant 2$) components overlap, it is appropriate to operate on an over-determined data set, *i.e.* one in which $N + 1$ wavelengths are used. This has been demonstrated using the multiple spectral suppression algorithm, where data sets incorporating up to eight discrete wavelengths can be used[22,23].

Provided there is an adequate data sampling rate, the correlation between adjacent time-point spectra collected in DAD should always be 1[24]. Any deviation from unity is indicative of a change in the retention characteristics of spectrally-dissimilar co-eluting components. Consequently, if a known spectrum is used as the correlation template, then variations in the resulting correlation coefficient, with time across the peak, will provide an indication of peak homogeneity. This approach forms the basis of the multiple absorbance ratio correlation (MARC) algorithm. If the apical spectrum is used as the reference spectrum the process is said to be internally referenced and is referred to as IMARC. However, if a spectrum of known purity and identity *e.g.* from a spectral library, is used the results from this externally referenced application (EMARC) will not only provide the analyst with an estimation of peak homogeneity but also of possible identity. As data from more than two independent domains are being examined and the results compared to those for a reference of known quality[24,25], the EMARC results will give an estimation of both the homogeneity and purity of the peak.

In order to test this hypothesis it was necessary to modify the practical application of the theory, outlined above. The memory capacity of the computer controller for DAD was not sufficient to permit whole spectra to be used in each time-point calculation. Consequently, the data used to represent the reference and test spectra had to be reduced. As the probability of five or more components co-eluting in practice is low[26,27], monitoring simultaneously at six independent wavelengths will produce an over-determined data set sufficient for most applications. This modification allows the reference and test spectra to be represented by reduced data set (RDS) vectors of up to six points each. The resultant multi-chromatographic data set is therefore a user-defined subset of the total spectrochromatographic data set, differing only in the number of wavelength components included.

Beer's law states that for a pure chemical species, there is a linear relationship

between concentration and absorbance at every wavelength. One of the fundamental principles of rapid scanning detectors, including DAD systems, is that the change in concentration with time is negligible compared to the time taken to scan between the various wavelengths. Consequently, the ratio of absorbances at two or more wavelengths, at any time point in a chromatogram, becomes the ratio of the extinction coefficients. As the extinction coefficient at each wavelength is an intrinsic property of a compound, the ratios will also be characteristic, especially if more than one pair of wavelengths is used[8,13,15]. It follows, therefore, that the normalised values in the six-point RDS vector should also be distinctive. Therefore, rather than compare individual RDS vector components between the reference and unknown compound, a more rapid initial screen can be performed using the algebraic sum of the RDS vector components (SUM).

The major objectives of the work described here were to translate the MARC hypothesis into a working algorithm on a currently available DAD system and to evaluate the subsequent performance using a binary mixture of spectrally similar compounds at different chromatographic resolutions.

EXPERIMENTAL

The liquid chromatograph consisted of an LDC Constametric 3000 dual reciprocating piston pump (LDC-Milton Roy, Riviera Beach, FL, U.S.A.) and an injection valve fitted with a 50-μl loop (Model 7125; Rheodyne, Berkeley, CA, U.S.A.) through which samples were introduced using a 100-μl syringe (SGE, Ringwood, Australia). The 100 × 8 mm I.D. Nova-Pak C_{18} Radial-Pak cartridge was held in an RCM 100 module (Waters Millipore, Millford, MA, U.S.A.).

The Hewlett-Packard (HP) 1040A DAD system was set up to monitor six wavelengths simultaneously: 204, 220, 240, 250, 268 and 284 nm (all ± 2 nm), against a common non-absorbing wavelength of 550 ± 20 nm. The HP 1040A was controlled from an HP 85A computer through an HP-IB interface. An HP 9121 dual disk drive, an HP 2225A "think-jet" printer and an HP 7470 plotter were also connected to the computer via other HP-IB connections. The original HP 85A specifications were enhanced through the addition of read only memory modules for input/output and printer–plotter communications and available memory (expanded by 16K) (Hewlett-Packard, Waldbronn, F.R.G.).

Theophylline (1,3-dimethylxanthine, TH) and paraxanthine (1,7-dimethylxanthine, PX) were used as received (Sigma, St. Louis, MO, U.S.A.). Single-component and mixed solutions were made up in and diluted with methanol–water (20:80, v/v). The mixtures contained a constant concentration of TH (52.51 μg ml^{-1}) and variable concentrations of PX (0.525–52.5 μg ml^{-1}). Methanol and acetonitrile were both HPLC grade and used as received (Rathburn Chemicals, Walkerburn, U.K.).

Based on the method proposed by Butrimovitz and Raisys[28] three different mobile phases were developed. In each case the flow-rate was 1.5 ml min^{-1} and the mobile phase comprised varying combinations of methanol–acetonitrile–sodium acetate (20 mM, adjusted to pH 5 with acetic acid), depending on the degree of separation required between the two analytes. The three specified values of resolution (R_s) were 0.7, 1.26 and 2.9, for which the respective mobile phase compositions were 0:10:90, 10:8:82 and 28:0:72 (all v/v/v).

RESULTS AND DISCUSSION

Between 1967 and 1987 over 860 papers on the subject of TH analysis were recorded by the *Chemical Abstracts* reference system, of which 33% were for HPLC-based analytical methods. TH is a drug with a therapeutic window of 5–20 μg ml^{-1}. The consequences to the patient of clinical concentrations outside of these limits have been widely discussed elsewhere[29–33]. Recently, however, concern has been raised over the interference of endogenous components in the quantification of TH by the established HPLC methods. PX is a structural isomer of TH and is a metabolite of both caffeine and TH and endogenous levels of PX have been established as one of the reasons for over-estimating the concentrations of TH in blood samples[34–38].

The wavelengths used to test the algorithm were chosen to cover the spectral range 200–350 nm. Although some reference was made to the spectra of TH and PX (Fig. 1), the detection wavelengths were not necessarily chosen to be specific for those two compounds, apart from the criteria that there should be significant absorption at all wavelengths. Spectrally TH and PX are very similar (r = 0.9947). Examination of spectrochromatographic data sets indicates that, unless baseline resolution can be assured, co-elution of TH and PX could readily go undetected or be assigned as a column malfunction (Fig. 2a–c).

As with absorbance ratio plots[14,39] it was necessary to modify the graphical scale on which the results of the MARC algorithm were viewed, in order to visually emphasize any changes in the correlation coefficient across the peak. Miller and Miller[40] suggest that a two-tailed *t*-test can be used, with $N - 2$ degrees of freedom (N = number of wavelengths monitored), to calculate a value for "*t*" below which changes in r (the correlation coefficient) are not significant. Using eqn. 1 and $N = 6$

$$t = r (N - 2)^{\frac{1}{2}} (1 - r^2)^{-\frac{1}{2}} \tag{1}$$

t = 2.776 when p = 0.05 and t = 4.604 when p = 0.01. These give limits of $r \geq 0.82$ and $r \geq 0.92$, respectively. These confines therefore provide the scale against which changes in the values of r can be assessed.
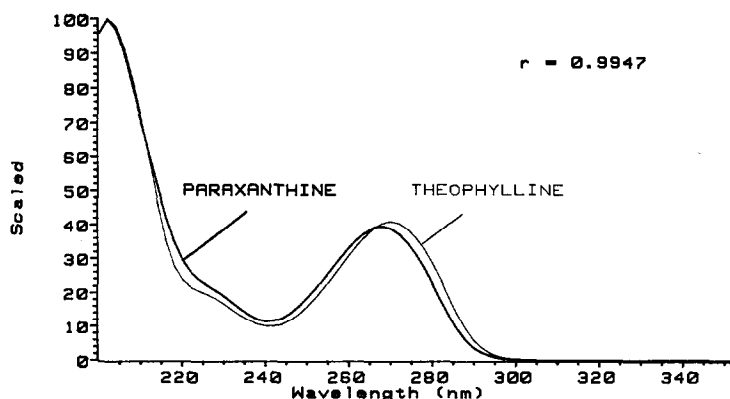

Fig. 1. Comparison of normalised theophylline and paraxanthine spectra collected on-line. See text for chromatographic conditions for R_s = 2.9.

Again, analogous to absorbance ratio[40], there are two significant values for the correlation coefficient: the maximum value obtained (r_{max}) and the mean value calculated across the peak (r_{mean}). Each of the two utilities provides a different amount of information about the content of the peak under investigation. When the IMARC algorithm is used the reference RDS vector will be that extracted at the apex of the peak under investigation. With the EMARC algorithm the identity of the reference RDS vector will be signified by EMARC(X), where X is the identity of the reference
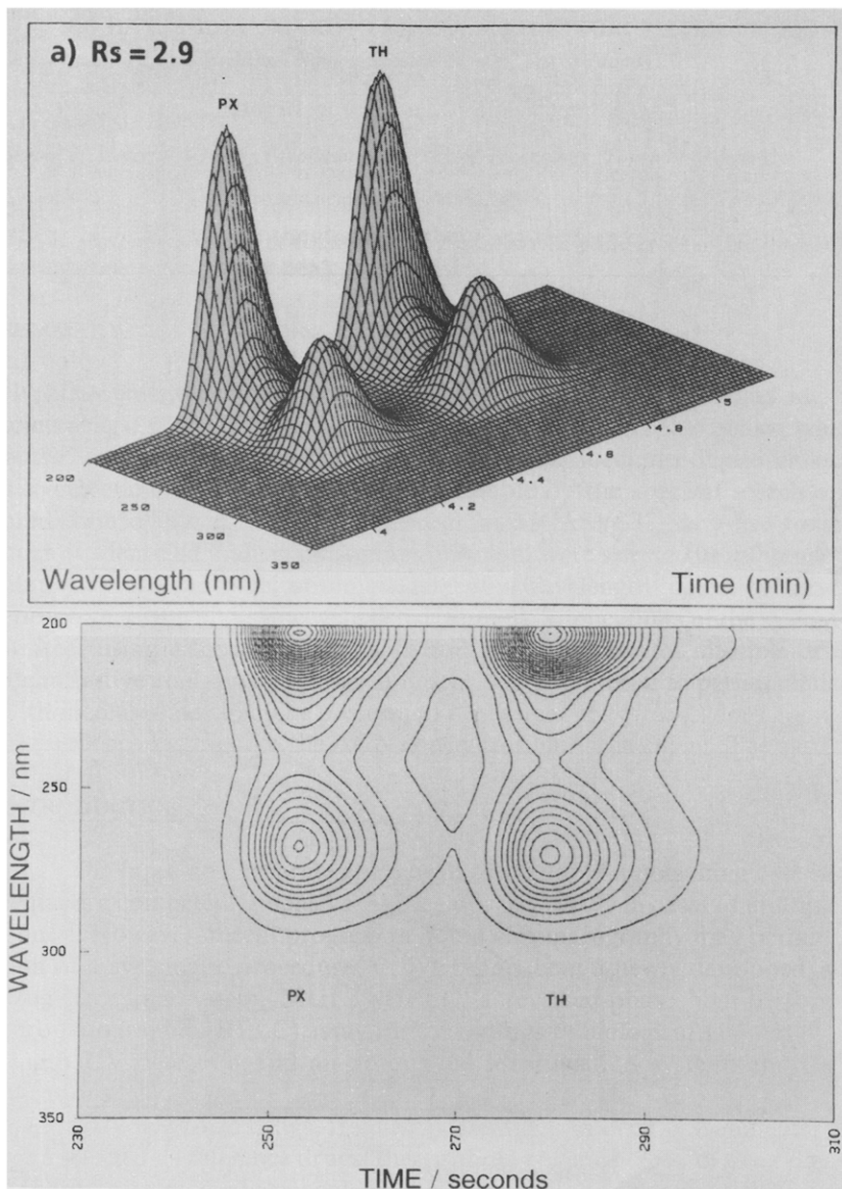
Fig. 2.

compound. The IMARC and EMARC algorithms are identical in their operation apart from the source of the reference RDS vector used in the calculation.

The selectivity of the EMARC algorithm was evaluated using the binary solutions of PX and TH chromatographed under the $R_s = 2.9$ conditions described above. Application of EMARC(PX) and EMARC(TH) to the TH peak gave the results for r_{max} shown in Fig. 3. The mean r_{max} result which was obtained following application of EMARC(PX) was $0.9950 \pm 0.16\%$ relative standard deviation
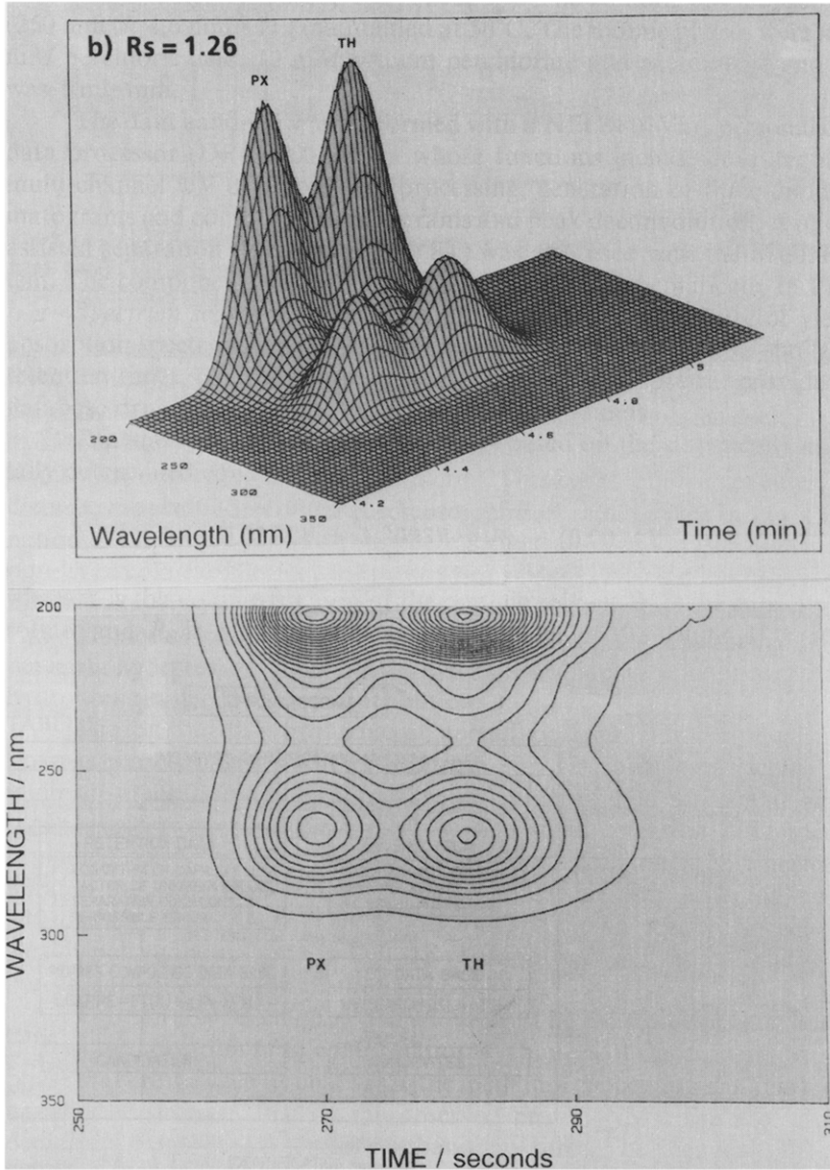
Fig. 2.

(R.S.D.), encompasses the correlation value calculated between the full spectra for TH and PX of 0.9947. This indicates that the reduction from > 70 individual wavelength points in the total spectrum to the 6 points used in the practical application of the algorithm has not proved detrimental to the performance of the algorithm.

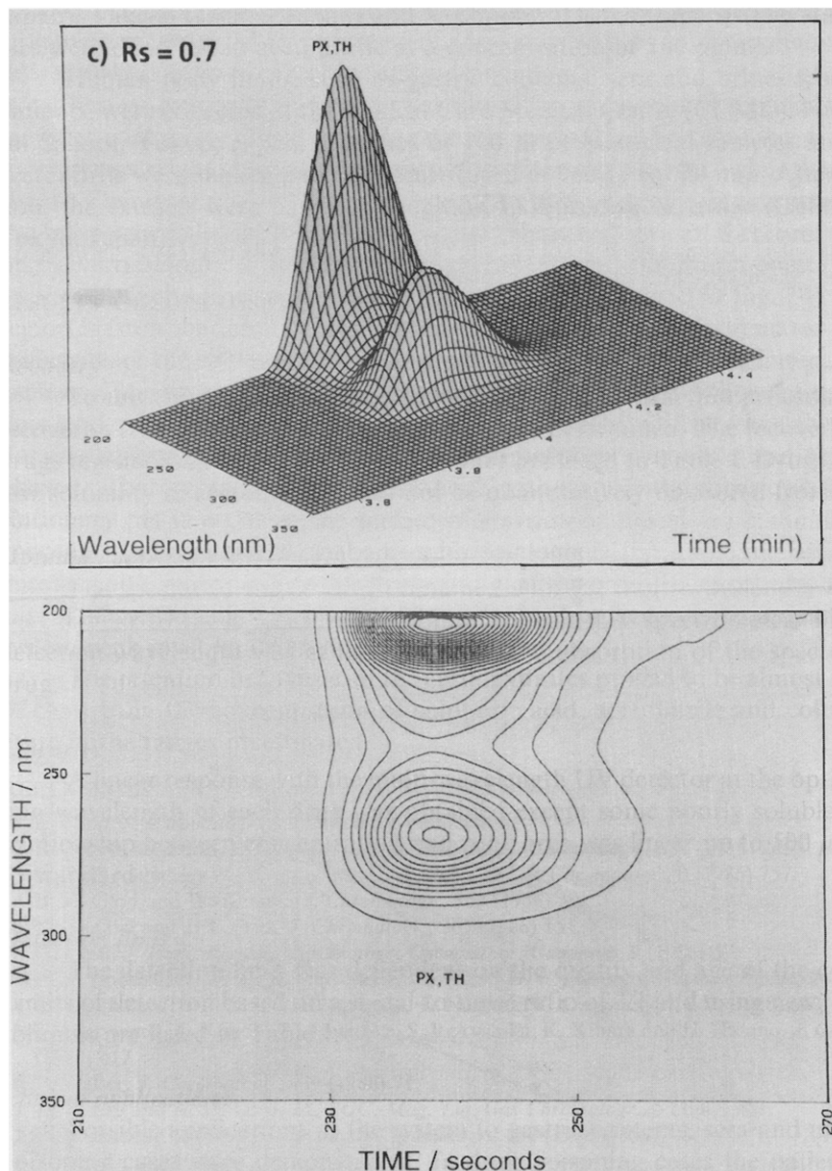When the two components partially overlap ($R_s = 1.26$) the respective

Fig. 2. Pseudo-three-dimensional and the corresponding contour plot representations of the full spectro-chromatographic data sets for a mixture of theophylline (TH; 52.51 μg ml^{-1}) and paraxanthine (PX; 52.5 μg ml^{-1}) chromatographed under three different resolution conditions : (a) $R_s = 2.9$, (b) $R_s = 1.26$ and (c) $R_s = 0.7$. See text for chromatographic conditions.
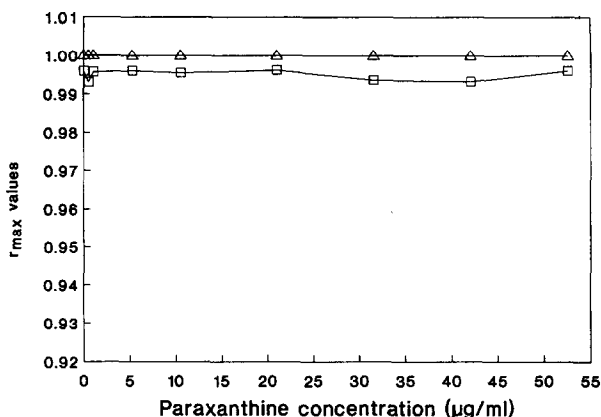
Fig. 3. The r_{max} results following application of EMARC(TH) ($\triangle$) and EMARC(PX) ($\square$) to the theophylline peak. $R_s = 2.9$; see text for chromatographic conditions.

EMARC r_{mean} values reflect the relative proportions of TH and PX in the mixed peak (Fig. 4). By definition, if there is an area in the binary peak where only one component elutes, the EMARC r_{max} value for that component will be 1. As this is the case for both components, monitoring the difference between r_{max} and r_{mean} increases the apparent discrimination of the method (Fig. 5).

With minimal resolution between the two components ($R_s = 0.7$), the conclusions drawn from the basic r_{max} values are not as distinct as in the previous applications (Fig. 6). However, discernment may be improved by monitoring a linear function of the difference between the IMARC and EMARC r_{mean} values (Fig. 7). In this latter case the crossover point of the (IMARC–EMARC) graphs represents the equiabsorptive mixture, *i.e.* an equal contribution from each of the components to the total multichromatographic absorbance of the composite peak.
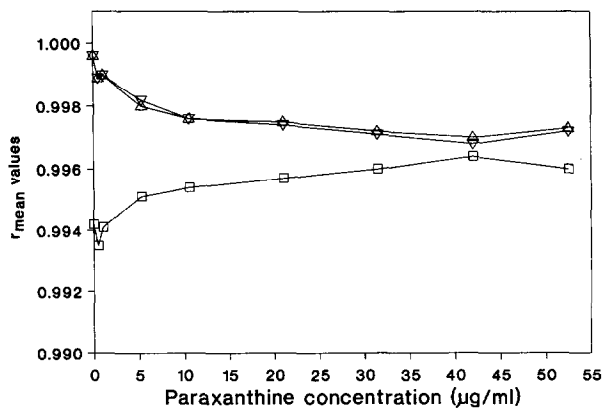

Fig. 4. The r_{mean} results following application of the MARC algorithm to binary mixtures of paraxanthine and theophylline: $\triangle$ = IMARC; ∇ = EMARC(TH); $\square$ = EMARC(PX). $R_s = 1.26$; see text for chromatographic conditions.
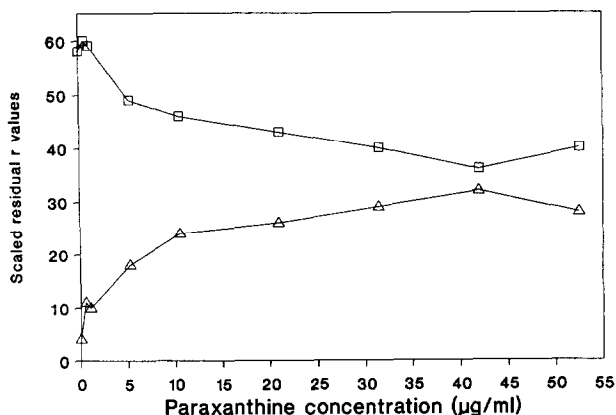
Fig. 5. Improvement in the discrimination of the EMARC(TH) ($\triangle$) and the EMARC(PX) ($\square$) obtained by monitoring the value of the difference $r_{max} - r_{mean}$. R_s = 1.26; see text for chromatographic conditions.

As well as monitoring the r_{max} and r_{mean} values it is possible to monitor the IMARC SUM values. Table I indicates that these measurements are concentration independent and analyte specific for PX and TH. Fig. 8 indicates that the IMARC SUM value is linearly proportional to the PX content of the peak, at the time point when the value is calculated. The difference in the gradients of the two graphs corresponding to the chromatographic conditions R_s = 0.7 and R_s = 1.26 is an indication of the extent to which the PX peak extends under the apex of the resultant binary peak.

Calculation of the maximum value of the absorbance ratio between the signals monitored at 220 and 268 nm (A_{220}/A_{268}) for the TH peaks (R_s = 2.9 and 1.26) and the fused peaks (R_s = 0.7) gave the results presented in Fig. 9. As the results for the TH peak under the chromatographic conditions for R_s = 2.9 and 1.26 were not significantly different only one set of results has been plotted. Whereas the peak
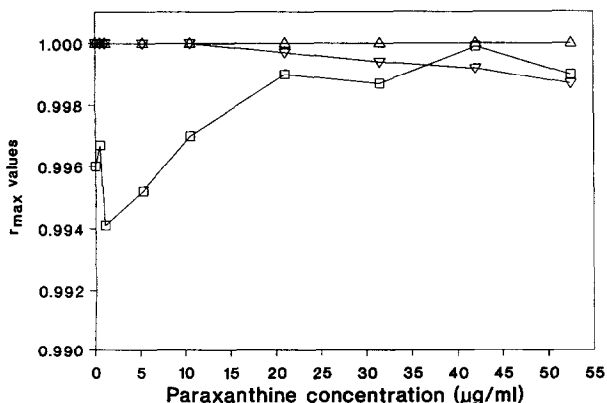

Fig. 6. The r_{max} results following application of the MARC algorithm to binary mixtures of paraxanthine and theophylline: $\triangle$ = IMARC; ∇ = EMARC(TH); $\square$ = EMARC(PX). R_s = 0.7; see text for chromatographic conditions.
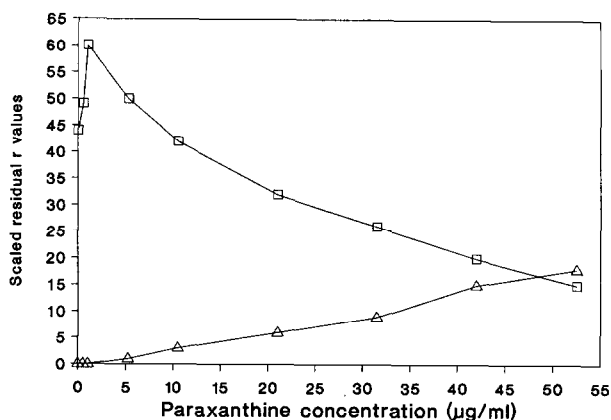
Fig. 7. Improvement in the discrimination of the MARC algorithm by monitoring the difference in the r_{mean} values obtained using IMARC and EMARC: $\triangle$ = IMARC $-$ EMARC(TH); $\square$ = IMARC $-$ EMARC(PX). R_s = 0.7; see text for chromatographic conditions.

height measurements, at 220 nm, increased linearly in proportion to the PX concentration (in μg ml^{-1}) in the mixtures, the corresponding absorbance ratio graph is curvilinear. (The reverse situation occurs if the PX concentration is expressed as "percentage total xanthine content in the mixture".) Therefore, unlike the IMARC SUM method where the comparative graph of IMARC SUM *vs.* PX concentration (μg ml^{-1}) was linear (R_s = 0.7, Fig. 8), the absorbance ratio method could only be used as a relative method for determining the concentration of PX in the mixture. In the former case there was an observable change in the IMARC SUM value of the mixture containing 0.525 μg ml^{-1} compared to that for TH alone, whereas the PX concentration must be greater than 5 μg ml^{-1} before any significant change in the absorbance ratio is recorded.

TABLE I

SELECTIVITY AND STABILITY OF SUM MEASUREMENT FOR PX (5.25–52.5 μg/ml) AND TH (52.51 μg/ml)

For chromatographic conditions see text.

Sample		N	IMARC component wavelength (nm)						SUM
			204	220	240	250	268	284	
TH									
	Mean	18	1.0000	0.2528	0.1070	0.1679	0.4200	0.1915	2.1391
	R.S.D.		0	0.12	0.56	0.92	0.43	0.47	0.18
PX									
	Mean	12	1.0000	0.3141	0.1214	0.1955	0.4108	0.1363	2.1780[a]
	R.S.D.		0	0.76	2.31	2.10	2.19	2.42	0.98

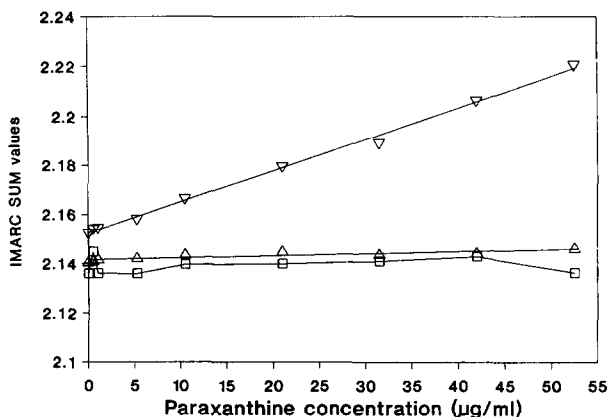[a] Mean values are significantly different; $p=0.05$.

Fig. 8. Values of the IMARC SUM for the theophylline peak resolved from or coeluting with paraxanthine in a series of binary mixtures containing constant theophylline concentrations (52.51 μg ml^{-1}) and variable paraxanthine concentrations (0–52.5 μg ml^{-1}). R_s: $\square$ = 2.9; $\triangle$ = 1.26; ∇ = 0.7. See text for chromatographic conditions.

CONCLUSIONS

In conclusion, the application of EMARC(X) is specific and most sensitive when X is the minor component in the mixture. Selectivity and sensitivity may be enhanced by a judicious combination of either the IMARC and EMARC results or the r_{max} and r_{mean} results, dependent on the degree of resolution between the two components in the binary peak. The algorithm makes no assumptions about the peak shape of any of the co-eluting components. The values of r_{mean} obtained following application of either IMARC or EMARC, give an indication of the homogeneity of the peak, while the EMARC r_{max} value may be used secondarily as an indication of identity.
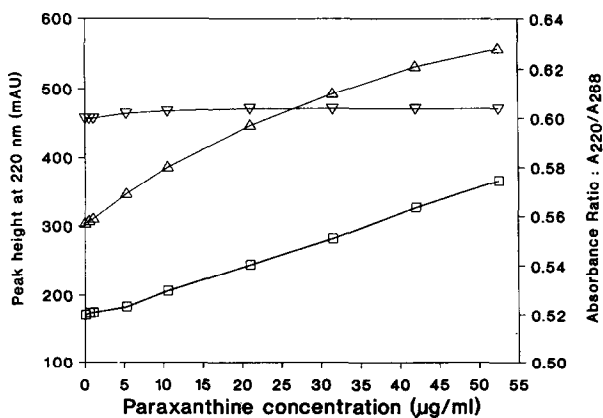

Fig. 9. Peak height at 220 nm (R_s = 0.7; $\square$) and absorbance ratio (A_{220}/A_{268}) values (R_s = 1.26, ∇; R_s = 0.7, $\triangle$) for the theophylline peak (52.51 μg ml^{-1}) in mixtures with paraxanthine (0–52.5 μg ml^{-1}). See text for chromatographic conditions.

The IMARC SUM value was found to be analyte specific and concentration independent, being dependent on the number and the uniqueness of the wavelengths used in its calculation. For a binary peak the value of the IMARC SUM was proportional to the concentration of the minor component in the mixture, with sensitivity being inversely proportional to chromatographic resolution. The limit of detection for PX in the mixture with TH was found to be 10% (w/w) using the absorbance ratio method. This compared with a 1% limit for the IMARC SUM method.

There is some indication in the results that the performance of the MARC algorithm is dependent on the number and specificity of the wavelengths used rather than on the chromatographic resolution. As the similarity of the spectra of the constituent components increases, it is expected that it may be necessary to increase the quantity and/or the selectivity of the wavelengths used in the algorithm. With the advent of more powerful computer controllers for DAD systems it should also be possible to expand the practical application of the MARC algorithm to match the original theory, *i.e.* using full spectra as the reference templates and operating on full spectrochromatographic data data sets. This would remove the need for analysts to choose "the most appropriate" wavelengths for inclusion in the algorithm. The algorithm would then become a truly operator-independent method.

REFERENCES

1 B. R. Kowalski, *Anal. Chem.*, 52 (1980) 112R–122R.
2 I. E. Frank and B. R. Kowalski, *Anal. Chem.*, 54 (1982) 232R–243R.
3 L. S. Ramos, K. R. Beebe, W. P. Carey, E. Sanchez M, B. C. Erickson, B. E. Wilson, L. E. Wangen and B. R. Kowalski, *Anal. Chem.*, 58 (1986) 294R–315R.
4 J. K. Strasters, H. A. H. Billiet, L. de Galaan, B. G. M. Vandeginste and G. Kateman, *J. Chromatogr.*, 385 (1987) 181–200.
5 W. H. Lawton and E. A. Sylvestre, *Technometrics*, 13 (1987) 617–633.
6 G. G. R. Seaton, J. G. D. Marr, B. J. Clark and A. F. Fell, *Anal. Proc.*, 23 (1986) 424–426.
7 N. Ostojic, *Anal. Chem.*, 46 (1974) 1653–1659.
8 R. Yost, J. Stoveken and W. Maclean, *J. Chromatogr.*, 134 (1977) 73–78.
9 A. M. Krstulovic, R. A. Hartwick, P. R. Brown and K. Lohse, *J. Chromatogr.*, 158 (1978) 365–376.
10 J. K. Baker, R. E. Skelton and C. Ma, *J. Chromatogr.*, 168 (1979) 417–427.
11 M. S. Denton, T. P. DeAngelis, A. M. Yacynych, W. R. Heineman and T. W. Gilbert, *Anal. Chem.*, 48 (1976) 20–24.
12 R. E. Dessey, W. D. Reynolds, W. G. Nunn, C. A. Titus and G. F. Moler, *Clin. Chem.*, 22 (1976) 1472–1482.
13 P. A. Webb, D. Ball and T. Thornton, *J. Chromatogr. Sci.*, 21 (1983) 447–453.
14 A. C. J. H. Drouen, H. A. H. Billiet and L. de Galan, *Anal. Chem.*, 56 (1984) 971–978.
15 P. C. White, *J. Chromatogr.*, 200 (1980) 271–276.
16 P. C. White and T. Catterick, *J. Chromatogr.*, 402 (1987) 135–147.
17 A. F. Poile and R. D. Conlon, *J. Chromatogr.*, 204 (1981) 149–152.

18 T. Alfredson and T. Sheehan, *J. Chromatogr. Sci.*, 24 (1986) 473–482.
19 G. T. Carter, R. E. Schiesswohl, H. Burke and R. Yang, *J. Pharm. Sci.*, 71 (1982) 317–321.
20 A. F. Fell, B. J. Clark and H. P. Scott, *J. Pharm. Biomed. Anal.*, 1 (1983) 557–572.
21 B. J. Clark, A. F. Fell, H. P. Scott and D. Westerlund, *J. Chromatogr.*, 286 (1984) 261–273.
22 P. Horvath, personal communication.
23 J. G. D. Marr, P. Horvath, B. J. Clark and A. F. Fell, *Anal. Proc.*, 23 (1986) 254–256.
24 J. Halket, *J. Chromatogr.*, 179 (1979) 229–241.
25 J. Van Rompay, *J. Pharm. Biomed. Anal.*, 4 (1986) 725–732.
26 E. Sanchez, L. S. Ramos and B. R. Kowalski, *J. Chromatogr.*, 385 (1987) 151–164.
27 E. Sanchez, L. S. Ramos and B. R. Kowalski, *J. Chromatogr.*, 385 (1987) 165–180.
28 G. P. Butrimovitz and V. A. Raisys, *Clin. Chem.*, 25 (1971) 1461–1464.
29 W. C. Bowman and M. J. Rand, *Textbook of Pharmacology*, Blackwell, London, 2nd ed., 1980, pp. 27.27–27.28.
30 T. W. Rall, in A. G. Gilman, L. S. Goodman and A. Gilman (Editors), *The Pharmacological Basis of Therapeutics*, Macmillan, New York, 6th ed., 1980, pp. 592–607.
31 P. W. Trembath, in P. Turner and D. Shand (Editors), *Recent Advances in Clinical Pharmacology*, No. 2, Churchill Livingstone, Edinburgh, 1980, pp. 55–71.
32 C. S. Frings, R. C. Keefer and J. M. Saloom, *Clin. Toxicol.*, 8 (1975) 553–561.
33 D. R. Jarvie, A. M. Thompson and E. H. Dyson, *Clin. Chim. Acta*, 168 (1987) 313–322.
34 J. R. Miksic and B. Hodes, *Clin. Chem.*, 25 (1979) 1866–1867.
35 H. A. Farrish and W. A. Wargin, *Clin. Chem.*, 26 (1980) 524–525.
36 N. Daoud, T. Arvidsson and K. G. Wahlund, *J. Pharm. Biomed. Anal.*, 4 (1986) 253–260.
37 R. H. Drost, H. V. Schultz, R. A. A. Maes and J. M. van Rossum, *Clin. Toxicol.*, 25 (1987) 231–241.
38 C. L. Fligner and K. E. Opheim, *J. Anal. Toxicol.*, 12 (1988) 339–343.
39 H. Cheng and R. R. Gadde, *J. Chromatogr. Sci.*, 23 (1985) 227–230.
40 J. C. Miller and J. N. Miller, *Statistics for Analytical Chemistry*, Ellis Horwood, Chichester, 2nd ed., 1988, pp. 104–109.