

CHROM. 8082

## AUTOMATION OF GAS CHROMATOGRAPHIC DATA COLLECTION AND ANALYSIS BY MEANS OF AN ON-LINE COMPUTER

BRUNO BARTOLI, SETTIMIO MOBILIO, NICOLA SPINELLI and FRANCESCO VANOLI

*Istituto di Fisica Sperimentale, Università degli Studi di Napoli, Via A. Tari 3, Naples (Italy)*

(Received November 14th, 1974)

---

### SUMMARY

A computer program and an interface of simple construction and operation have been developed in order to operate a gas chromatograph on-line with a medium-sized computer, thereby permitting automatic control of data collection and analysis.

The analysis program allows for variable background subtraction and resolution of overlapping peaks. The most important features of the hardware and software are described in detail.

---

### INTRODUCTION

In two previous papers<sup>1,2</sup>, the results found in radiolysis experiments on saturated hydrocarbons were discussed. The mixture of products formed by radiolysis is so rich in isomers and has such a wide range of boiling points that reliable quantitative analysis of the whole range requires the solution of some complex technical problems, as shown by the considerable spread of data obtained by different workers\*. We have already described (from refs. 3 and 4) the technical devices used in order to overcome these difficulties as far as possible. They are essentially: (1) irradiation of the sample inside capillary ampoules and its introduction into the gas chromatograph by means of a bulb-crusher suitable for capillary columns, in order to avoid losses, undesirable fractionation or inaccurate calibrations that are likely to take place when the analysis is carried out after separation of the sample into two or more fractions; (2) the use of high-resolution glass capillary columns obtained by a suitable treatment of the internal surface before impregnation; and (3) the use of a high-sensitivity flame ionization detector. Under these conditions, we were able to analyze each individual product, including those with yields as low as 0.001 molecule per 100 eV of radiation dose, and to calculate the percentage abundance of each product. The knowledge of this quantity is essential in any attempt to give a comprehensive interpretation of the phenomena that take place during radiolysis in terms of the molecular properties of the alkane under consideration. As is always the case when measurements are carried out with a very high sensitivity of detection, the value obtained in each measurement

---

\* A comparison is given in ref. 1 for *n*-pentane and *n*-hexane and in ref. 2 for *n*-heptane.

is affected by small fluctuations of a number of parameters that determine the final response of the detector; in other words, these are the typical experimental conditions under which fluctuations due to "casual" errors become evident. On the other hand, it is well known that it is this random nature of the errors which allows one to reduce their consequences by repeating the measurement a sufficient number of times, as the mean square error that affects the mean value of a group of  $N$  measurements (the standard deviation) decreases as  $N^{-1/2}$ . In our case, values of the standard deviation of 2–3% are reached with about 6–8 measurements for the largest peaks and about 10–12 measurements for the smallest peaks. Therefore, the study of the complete radiolysis spectra required the manual measurement of the areas of about 250 peaks for hexane and about 630 peaks for decane.

This explains why nobody has yet published results on the dose dependence of the yields of the various products, which would be of great interest; however, this dependence has so far been found only for hydrogen and for the total "unsaturation" products that can also be measured by other techniques. We concluded that the only reasonable way to undertake measurements of this type was the automation of the gas chromatographic (GC) apparatus.

We found a few papers<sup>5–7</sup> describing the essential features of some commercial apparatus for automatic GC analysis. Such apparatus is very expensive because of the high cost of the essential computer. For this reason, they are widely used for routine analysis only in large industrial laboratories, and only a few research groups can afford them; the programs they are supplied with are not generally accessible, are written in assembler language and are not flexible enough. On the other hand, a small research group such as ours may often share a medium or large computer with other groups, which reduces the problem to acquiring a suitable computer program and an interface that could be made with conventional components in a reasonable time.

Other papers<sup>8–10</sup> deal with programs for off-line optimization of the area measurements up to a few tenths of a per cent by means of complicated best-fit methods.

Sophisticated methods for accurate automatic control of the oven temperature and of the carrier gas flow-rate in order to achieve a high reproducibility of the retention times have also been described<sup>11,12</sup>.

The fluctuations in our results depend on a very large number of practically uncontrollable factors, such as slightly different amounts of residual air in each sample and fluctuations of the irradiation temperature and of the speed of breaking the capillary; therefore, an accuracy in the area measurement of better than 2–3% represents an unnecessary loss of computing time in our case. Also, a high reproducibility of the retention times is of no interest to us, as the very characteristic shape of the radiolysis spectrum makes each product easily recognizable.

We had to devise our own interface and program mainly on the basis of two requirements that we feel are common to a number of other groups: (1) all operations must be as simple and rapid as necessary for working on-line while obtaining an accuracy of a few per cent, even for peaks that overlap by up to 50% (see Table I); and (2) the program must be written in Fortran language and the interface must be made from conventional commercial components. In order to meet the first requirement, special care was paid to the base-line subtraction.

EXPERIMENTAL

The program is written in Fortran IV language and can be used with any computer that has at least 8 K words of core memory and a registration (disk or tape) unit. Our computer is a Digital Equipment Corporation PDP 15.

A block diagram of the hardware is given in Fig. 1. The output of the digital voltmeter, D.V., is sent to the core memory, CM, of the computer through the interface, D. Note that the control of the analogue to digital conversion is given by an external clock, E, whose rate is controlled by the computer; in our case, we program in advance the various sampling rates to be used through the whole chromatogram, but it would also be possible to allow the computer automatically to control the sampling rate so that the number of points under each peak remains approximately constant<sup>13</sup> around 30-50. In both instances the interrupt request for the computer occurs only after the analogue to digital conversion has been completed. A registration unit of the computer, F, collects simultaneously the output of the converter and any other information necessary for subsequent analysis of the data, such as the gain of the amplifier, sampling rate and starting time.

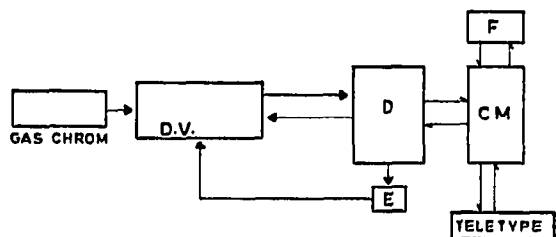


Fig. 1. Block diagram of the hardware. D.V. = Digital voltmeter; D = interface; E = external clock; F = registration unit; CM = core memory.

We will now discuss some basic aspects of the logic of our analysis program. The collection of data must start 20-30 sec before the elution of the first peak in order to collect a few hundred initial points that definitely pertain to the baseline. These points are used in each run to find the noise level,  $N$ , which is defined as twice the standard deviation of the collected points.

All of the incoming data points are then submitted to filtering and smoothing operations before being examined in the search for peaks. These operations are performed on groups (of variable length) of points into which the points enter and leave one by one according to the flow diagram shown in Fig. 2. The filtering operation is

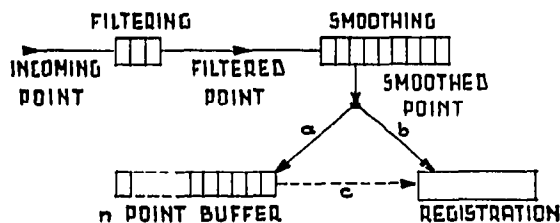


Fig. 2. Scheme of the points storage. (a) Before peak start; (b) after peak start; (c) transfer of the buffer after an "0-type" point.

carried out in order to remove spikes; for this purpose, the differences between three neighbouring points are examined and, when the first difference is found to be greater than a properly chosen function of the noise,  $f(N)$ , and the other difference is opposite in sign, the second point is set equal to the first (in other words, it is brought down to the noise level). The function  $f(N)$  should in principle be defined in each individual case, depending on the signal-to-noise ratio of the particular run. The temperature, type and state of the column and of the detector as well as other operating conditions determine the value of this ratio. The expression

$$f(N) = 5 N$$

proved to be satisfactory in many instances in our experiments.

After removal of the spikes, the next operation is smoothing, which is achieved by fitting a third-order curve to a number (usually nine in our case) of consecutive points following the method outlined by Savitzky and Golay<sup>14</sup> and developed by Steiner *et al.*<sup>15</sup>. The smoothed points enter a buffer (the capacity of which is discussed later) where the difference between the maximum and minimum values is continuously examined; when this difference becomes larger than  $f(N)$ , the points of the buffer are transferred to the registration unit and the search for a peak is started. For this purpose, the first and second derivatives are examined; when  $m$  ( $m = 3$  in our case) consecutive second derivatives are greater than zero and the last point is the maximum in the buffer, then the  $(n - m)$ th point is considered to be an "0-type" characteristic point, namely the starting point of a peak or of a group of peaks (see Fig. 3b). The average of these  $m$  values is recorded as  $M$ , while the average of the last  $m$  first derivative is recorded as  $H$ .

After this, the differences in the buffer are no longer calculated until the baseline is reached again. Instead, the second derivative is continuously followed; when  $p$  ( $p = 5$  in our case) consecutive second derivatives are less than zero, the  $(n - p)$ th point is considered to be a "1-type" point; next, when  $q$  ( $q = 5$  in our case) consecutive second derivatives are greater than zero, a "2-type" point is found.

Half of the time interval between a "1-type" and a "2-type" point is the half-width of the peak and the median point of this interval gives the elution time of the corresponding compound.

If a "2-type" point does not appear after a "0-type" point before a time interval fixed in advance (on the basis of the largest predictable peak width), the program starts again the search for peaks, taking the previously collected points as the baseline.

After a "2-type" point, the program goes on following the values of second derivatives: if  $r$  ( $r = 5$  in our case) such derivatives are less than zero, this is taken to indicate the presence of a new peak that partially overlaps the preceding one and the  $(n - r)$ th point is defined as a "4-type" point.

In order to specify the "separation point" of the two peaks, the program looks for a minimum in the zone between the "2-type" and "4-type" points; if no minimum is found, the point which corresponds to the maximum of the second derivative is chosen as separation point.

A "3-type" point (end of a peak or of a group of peaks) is found when the values of the second derivatives become smaller than  $M$  while the first derivatives

are close to  $H$ . When the last "3-type" point has been reached, the program starts the integration operations.

Each peak area is calculated by a number of consecutive operations; the method is described with the help of the example illustrated in Fig. 3b. First the program finds the baseline by fitting a second-order curve to a number of points preceding the "0-type" point and following the "3-type" point.

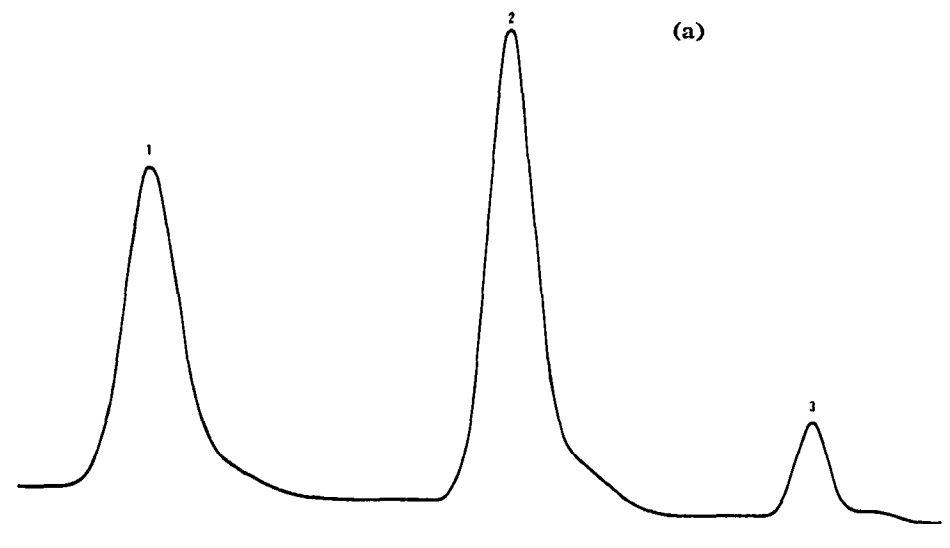
By subtracting the baseline, a new peak or group of peaks is found in which the ordinate  $y$  corresponding to any abscissa  $x$  is given by the difference  $y_{\text{chrom}} - y_{\text{baseline}}$ . The total,  $A_{\text{tot}}$ , is now calculated by the equation

$$A_{\text{tot}} = \int_a^b y \, dx$$

where  $a$  and  $b$  are the abscissae of the "0-type" and "3-type" points.

The partition of  $A_{\text{tot}}$  among the various peaks is now carried out. By means of successive approximations, the program evaluates the tangents to the curve in the vicinity of the "0-type" point and of the first separation point for the first peak of the group, of two successive separation points for any intermediate peak, and of the last separation point and the "3-type" point for the last peak. Next, the areas  $A_1$ ,  $A_2$ , etc., of the single peaks delineated by the curve and the tangents are calculated.

When the overlapping of the peaks is such that no minimum exists between them, *i.e.*, in the case of a peak with a shoulder, the segment of straight-line tangent to the curve at the beginning and at the end of the shoulder peak is found by the successive approximation method. From the upper tangency point, the tangent to the curve in the vicinity of the "0-type" point (or of the "3-type" point) is now drawn and the two areas thus obtained are calculated as mentioned above.



(Continued on p. 56)

Fig. 3.

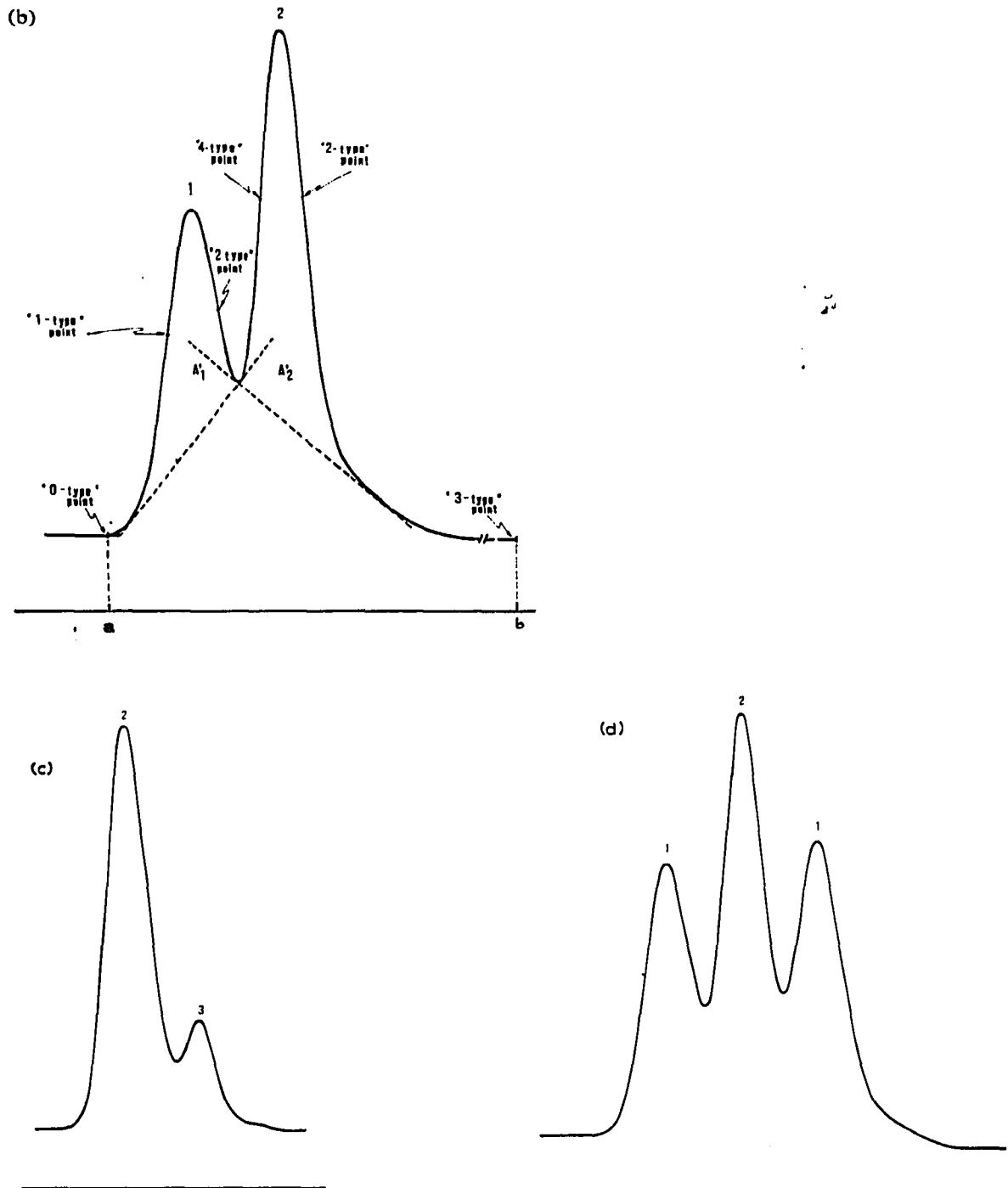


Fig. 3. Typical peaks. (a) Separate peaks. (b) 47% Fused peaks. (c) 75% Fused peaks. (d) 57% Fused peaks.

The difference

$$A_{\text{tot}} - (A_1 + A_2 + \dots) = R$$

is now "re-allocated" to the various peaks by assigning to the  $i$ th peak a supplementary area given by the equation

$$A_i' = A_i \cdot \frac{R}{\sum A_i}$$

so that any peak area is given by the sum

$$A_i'' = A_i + A_i'$$

In this way, a better approximation is obtained than with the more widely used method of perpendiculars<sup>16</sup>, while still maintaining simplicity and speed of calculation.

At the end of the chromatogram, the computer gives the areas  $A_i$ ,  $A_i'$  and  $A_i''$  of each peak and the corresponding retention times.

In order to control the overall reliability of the method, we constructed a number of peaks similar in shape to typical GC peaks by summing two gaussian curves of different heights and widths properly shifted so as to obtain reasonable tails.

We made the following test:

- (1) added each peak to a curve of known shape simulating a variable baseline;
- (2) constructed groups of two peaks by overlapping the single peaks to different extents and always adding them to a decreasing baseline; and
- (3) combined peaks in various ways so as to obtain groups of three peaks, always adding the group to a variable baseline.

The peaks and groups of peaks obtained were submitted to the computer for analysis. In Table I we give as an example the areas obtained in the cases shown in Fig. 3a-d.

The agreement found in the case of large overlap may appear rather rough to workers who are unfamiliar with measurements of this kind; however, it suffices

TABLE I  
AREAS OF PEAKS IN FIG. 3

Fig.	Peak	Overlapping	Predicted area	Evaluated area
3a	1	Separated	1.00	1.00
	2		1.23	1.27
	3		0.17	0.16
3b	1	}47%	1.00	1.04
	2		1.23	1.28
3c	2	}75%	1.23	1.22
	3		0.17	0.20
3d	1	}57%	1.00	1.03
	2		1.23	1.20
	1	}57%	1.00	1.07

to try carrying out the same operations manually in order to understand that the result obtained is satisfactory.

In working out the program, we took care to give it great flexibility with regard to different requirements. For this purpose, we introduced the possibility of adjusting the values of all of the most important parameters in order to optimize them in each instance. Such parameters are: (1) the function  $f(N)$  of the noise level,  $N$ ; (2) the extension of the buffer used for peak detection (the number,  $n$ , of points in Fig. 2); (3) the number of second derivatives ( $m, p, q, r$ ) to be taken into account in order to find the position of the various characteristic points; (4) the number of points to which smoothing is applied; and (5) the number of points from which the value of the first derivative is calculated.

The first parameter is adjusted to the noise level of the background, which can vary markedly depending on the type of detector, its working temperature, etc. The value of the second parameter fixes the smallest slope that the signal must have in order to be considered as the rising portion of a peak. The choice of the best values for  $m, p, q$  and  $r$  can be made for each independently; the values of these four important parameters control the sensitivity of the program in the detection and separation of the various peaks and shoulders. We usually obtain good results with the use of nine and five points for the operations in (4) and (5), respectively.

## CONCLUSION

In this description of our system of automatic analysis we have tried to emphasize two basic features that make it different from most of the commercial systems.

The first feature is the accuracy in following the baseline through continuous interpolation, whether it is increasing or decreasing, or even in the presence of a step. This allows, for instance, measurements of a small peak on the tail of a much larger one to be made. Similarly, the area "re-allocation" operation allows the resolution of overlapping peaks of any kind, provided that the two inflection points are at least 15 points apart. With a sampling rate of 10–30 points/sec, this means a separation time of the two inflection points of 0.5–1.5 sec. These times are, in general, satisfactory. If narrower peaks are to be analyzed, a faster sampling rate is necessary. In this case, it would be advisable to perform the operations of data collection and analysis at different times; this is easily done by accumulating the data on a registration unit before submitting them to the analysis. In this way, sampling rates of up to 500–1000 points/sec can easily be achieved.

The second feature is the great flexibility of the system, due to the possibility of adjusting the most important parameters to meet the requirements of any kind of GC detector or of any apparatus that has an analogue output.

As already stated, we developed the interface and program for a PDP 15 computer; slight technical modifications (essentially of the hardware schemes) would be necessary for use with other computers. We shall be glad to give more detailed information and suggestions on request.

## ACKNOWLEDGEMENTS

This work was carried out with financial support from the Italian Consiglio



Nazionale delle Ricerche. The authors also thank Dr. V. Santoro and the staff of the Computing Centre of the Laboratori Nazionali di Frascati del C.N.E.N., Frascati, Italy.

## REFERENCES

- 1 E. Pancini, V. Santoro and G. Spadaccini, *Int. J. Radiat. Phys. Chem.*, 2 (1970) 147.
- 2 M. Armenante, G. F. Calicchio, V. Santoro and N. Spinelli, *Int. J. Radiat. Phys. Chem.*, 6 (1974) 177.
- 3 M. Armenante, V. Santoro and G. Spadaccini, *J. Chromatogr.*, 52 (1970) 227.
- 4 M. Armenante, V. Santoro and G. Spadaccini, *J. Chromatogr. Sci.*, 9 (1971) 510.
- 5 D. A. Craven, E. S. Everett and M. Rubel, *J. Chromatogr. Sci.*, 9 (1971) 541.
- 6 J. P. Hettinger, J. R. Hubbard, J. M. Gill and L. A. Miller, *J. Chromatogr. Sci.*, 9 (1971) 710.
- 7 R. A. Landowne, R. W. Morosani, R. A. Herrmann, R. M. King, Jr., and H. G. Schmus, *Anal. Chem.*, 44 (1972) 1961.
- 8 S. M. Roberts, D. H. Wilkinson and R. L. Walker, *Anal. Chem.*, 42 (1970) 886.
- 9 H. A. Hancock Jr., L. A. Dahm and J. F. Muldoon, *J. Chromatogr. Sci.*, 8 (1970) 57.
- 10 A. H. Anderson, T. C. Gibb and A. B. Littlewood, *J. Chromatogr. Sci.*, 8 (1970) 640.
- 11 R. S. Swingle and L. B. Roger, *Anal. Chem.*, 43 (1971) 810.
- 12 R. G. Thurman, K. A. Muller and M. F. Burke, *J. Chromatogr. Sci.*, 9 (1971) 77.
- 13 K. Kishimoto and S. Musha, *J. Chromatogr. Sci.*, 9 (1971) 608.
- 14 A. Savitzky and M. J. E. Golay, *Anal. Chem.*, 36 (1964) 1627.
- 15 J. Steiner, Y. Termonia and J. Deltour, *Anal. Chem.*, 44 (1972) 1906.
- 16 K. Kishimoto, H. Miyauchi and S. Musha, *J. Chromatogr. Sci.*, 10 (1972) 220.