Vol. 157 No. 2, March 1, 1984.
Completing Vol. 157

ISSN 0003-2670

ANALYTICA CHIMICA ACTA

International journal devoted to all branches of analytical chemistry

ELSEVIER

ANALYTICA CHIMICA ACTA

International journal devoted to all branches of analytical chemistry
Revue internationale consacrée à tous les domaines de la chimie analytique
Internationale Zeitschrift für alle Gebiete der analytischen Chemie

PUBLICATION SCHEDULE FOR 1984

	J	F	M	A	M	J	J	A	S	O	N	D
Analytica Chimica Acta	156	157/1	157/2	158	159	160	161	162	163	164	165	166

Scope. *Analytica Chimica Acta* publishes original papers, short communications, and reviews dealing with e aspect of modern chemical analysis, both fundamental and applied.

Submission of Papers. Manuscripts (three copies) should be submitted as designated below for rapid and effic handling:

Papers from the Americas to: Professor Harry L. Pardue, Department of Chemistry, Purdue University, West Lafaye IN 47907, U.S.A.

Papers from all other countries to: Dr. A. M. G. Macdonald, Department of Chemistry, The University, P.O. Box 3 Birmingham B15 2TT, England. Papers dealing particularly with computer techniques to: Professor J. T. Cl Universität Bern, Pharmazeutisches Institut, Baltzerstrasse 5, CH-3012 Bern, Switzerland.

Submission of an article is understood to imply that the article is original and unpublished and is not being conside for publication elsewhere. Upon acceptance of an article by the journal, authors will be asked to transfer the copyr of the article to the publisher. This transfer will ensure the widest dissemination of information.

Information for Authors. Papers in English, French and German are published. There are no page char Manuscripts should conform in layout and style to the papers published in this Volume. Authors should con Vol. 150/2 for detailed information. Reprints of this information are available from the Editors or from: Elsevier Edit Services Ltd., Mayfield House, 256 Banbury Road, Oxford OX2 7DH (Great Britain).

Reprints. Fifty reprints will be supplied free of charge. Additional reprints (minimum 100) can be ordered. An o form containing price quotations will be sent to the authors together with the proofs of their article.

Advertisements. Advertisement rates are available from the publisher.

Subscriptions. Subscriptions should be sent to: Elsevier Science Publishers B.V., Journals Department, P.O. 211, 1000 AE Amsterdam, The Netherlands. Tel: 5803 911, Telex: 18582.

Publication. *Analytica Chimica Acta* appears in 11 volumes in 1984. The subscription for 1984 (Vols. 156–166 Dfl. 2145.00 plus Dfl. 231.00 (p.p.h.) (total approx. U.S. $950.40). All earlier volumes (Vols. 1–155) except Vols and 28 are available at Dfl. 200.00 (U.S. $80.00), plus Dfl. 15.00 (U.S. $6.00) p.p.h., per volume.

Our p.p.h. (postage, packing and handling) charge includes surface delivery of all issues, except to subscribers in U.S.A., Canada and India who receive all issues by air delivery (S.A.L. — Surface Air Lifted) at no extra cost. For rest of the world, airmail and S.A.L. charges are available upon request.

Claims for issues not received should be made within three months of publication of the issues. If not they canno honoured free of charge.

For further information, or a free sample copy of this or any other Elsevier Science Publishers journal, readers in U.S.A. and Canada can contact the following address: Elsevier Science Publishing Co., Inc., Journal Informa Center, 52 Vanderbilt Avenue, New York, NY 10017, U.S.A., Tel: (212) 867-9040.

27

631 Ee

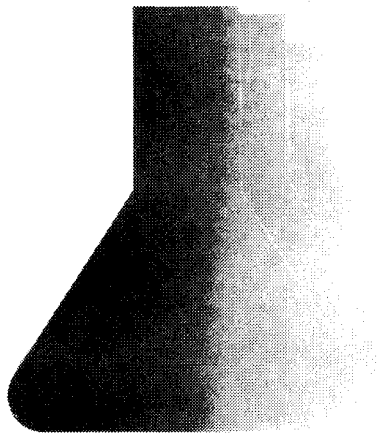
ANALYTICAL METHODS AND PROBLEMS IN BIOTECHNOLOGY

AN INTERNATIONAL SYMPOSIUM

17–19 April, 1984, Leeuwenhorst Congress Center, Noordwijkerhout, The Netherlands

Organised by:
Royal Netherlands Chemical Society (KNCV), Section for Analytical Chemistry
Netherlands Biotechnology Society (NBV) under the sponsorship of the Federation of European Chemical Societies

Scope:
- The development of analytical methods for biotechnological applications is an area of growing importance.
- Analytical methods currently available are now being adapted for practical use in biotechnological research, development and industrial production. But a large gap remains to be bridged between experts in analytical methodology and experts in biotechnology.
- It is the purpose of this Symposium to outline the problems faced in this field and to describe the rapid developments taking place.
- The Symposium is aimed at an interdisciplinary audience of analytical chemists and those involved in industrial and academic biotechnology.



The conference language will be English. An exhibition of products and instruments within the scope of the symposium will be held.

For full details and registration form please write or send coupon to

W. A. Scheffers
ANABIOTEC
Delft University of Technology
Laboratory of Microbiology
Julianalaan 67A,
2628 BC Delft, The Netherlands
Telephone (15)-782411

Analytical Methods and Problems in Biotechnology
Please submit detailed information

Name: _____

Address: _____

_____

_____

Analytica Chimica Acta, 157 (1984) 211—226
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

DECONVOLUTION OF MULTICOMPONENT ULTRAVIOLET/VISIBLE SPECTRA

PETER JOCHUM

Mathematisches Institut der Universität, Theresienstrasse 39, D-8000 München 2 (West Germany)

ERICH L. SCHROTT*

Botanisches Institut der Universität, Menzingerstrasse 67, D-8000 München 19 (West Germany)

(Received 29th August 1983)

SUMMARY

A reliable method based on non-negative least squares is suggested for the deconvolution of spectra of chemical and biochemical mixtures into their individual components. The efficiency of the method is demonstrated by use of artificial inorganic salt mixtures of defined composition as well as of real carotenoid samples of considerable complexity, the results for which are compared with those of conventional chromatography. The problem of error propagation is discussed and compared to the sensitivity of standard least-squares and n-wavelengths algorithms.

It has been known for half a century that the validity of Beer's law for mixtures allows the determination of the relative amounts of the participating chemical species by simply measuring and deconvoluting the corresponding spectrum of the mixture (for references, see [1]). However, application of the method in chemistry, and particularly in analytical biochemistry, has been hampered by the high sensitivity of the method to the quality of the instrumentation and by the large amount of data to be handled.

If $a_i(\lambda)$, $i = 1, \ldots, n$, is used to denote the absorbances of the pure compounds and $y(\lambda)$ denotes the absorbances of the mixture, then by Beer's law $y(\lambda)$ depends linearly on $a_i(\lambda)$

$$y(\lambda) = \sum_{i=1}^{n} a_i(\lambda)x_i + r(\lambda) \tag{1}$$

The unknown factors x_i are the relative concentrations, and the term $r(\lambda)$ covers systematic errors, noise and model deviations which, under ideal conditions, should be negligible. If $r(\lambda) = 0$, the unknowns x_i can be determined by picking n suitable (see below) wavelengths to establish the following system of linear equations

0003-2670/84/$03.00

$$y(\lambda_j) = \sum_{i=1}^{n} a_i(\lambda_j)x_i \quad (j = 1, \ldots, n) \tag{2}$$

Because the absorbances $a_i(\lambda_j)$, $i, j = 1, \ldots, n$, are known and the corresponding absorbances $y(\lambda_j)$ can be measured, this system determines the vector x of unknown relative concentrations uniquely provided that the determinant, det(A), of the coefficient matrix $A = (a_{ij})$, is different from zero. Although Eqn. 2 seems attractively simple, its solution poses many problems in chemical reality.

The selection of n good, or at least suitable, wavelengths can be difficult, especially because there is no simple criterion for establishing the quality of the picked wavelengths. The size of the determinant of A, as suggested by Ebel [2], is not a reasonable measure because multiplying a matrix by a constant c leads to a multiplication of the determinant by c^n. Thus, a matrix with a large determinant may be nearly singular. Moreover, the number of operations to compute the determinant for one set of n wavelengths is of the same order of n as is required for solving Eqn. 2. Thus, the computational effort for picking n optimal wavelengths is excessive compared to the original problem. A good measure for quantifying the nonsingularity of a matrix is the condition number which will be introduced in the theoretical section. It is a suitable criterion because it estimates the error propagation in solving Eqn. 2 (see below). However, minimizing the condition as a function of the n picked wavelengths is even more tedious because the computation of cond(A) requires about n-times the work needed to compute det(A).

To avoid the optimization problem, it is common usage to pick the maxima of the individual components. That this choice does not necessarily lead to a "good" matrix A is demonstrated in the Results section (see Table 3). Moreover, this method is incapable of evaluating the error r involved in the measurements. Thus there is no way of evaluating stochastic or deterministic error limits for x_i (i.e., random errors or essential bias). Although some of these cumbersome properties have been noted [1, 3, 4], this method is still very commonly applied in routine work. Most of the commercially available multicomponent systems in u.v.-visible spectrophotometry are based on the n-wavelengths method and there are still papers discussing its advantages (cf. [2] and literature cited therein).

THEORY

For a proper determination of the unknown relative concentrations x_i, as much information should be used as can be extracted from Eqn. 1. As a first step in this direction, the least-squares method uses Eqn. 1 at m wavelengths ($m \geqslant n$), minimizing the sum of squared deviations in each of the m linear equations

$$\sum_{j=1}^{m} [y(\lambda_j) - \sum_{i=1}^{n} x_i a_i(\lambda_j)]^2 \rightarrow \text{Minimum} \tag{3}$$

As a generalization of the least-squares method, the accuracy by which each equation is assumed to be satisfied can be taken into account by introducing appropriate weighting factors $w_j > 0$ leading to the following optimization problem

$$\sum_{j=1}^{m} w_j [y(\lambda_j) - \sum_{i=1}^{n} x_i a_i(\lambda_j)]^2 \rightarrow \text{Minimum} \tag{4}$$

The weights w_j are usually chosen according to the estimated (proportional to the reciprocal) size of the errors $r(\lambda_j)$ (see below).

Differentiating the left-hand side with respect to the n variables x_i and setting the partial derivatives equal to zero, yields the (equivalent) system of linear equations, in the usual matrix notation [5].

$$\mathbf{A}^T\mathbf{W}\mathbf{A}x = \mathbf{A}^T\mathbf{W}y, \; \mathbf{W} = \text{diag}(w_1, \ldots, w_m), \; y_j = y(\lambda_j) \tag{5}$$

which are well known as (generalized) Normal Equations. Such equations possess a unique solution x if (and only if) the n columns of the "standard" matrix $\mathbf{A}$ corresponding to the n pure spectra are linearly independent. In this case, the coefficient matrix $\mathbf{C} = \mathbf{A}^T\mathbf{W}\mathbf{A}$ is positive and the linear system (Eqn. 5) can be solved by any linear equation solution method (e.g., Cholesky's algorithm [5]). It is worth noting that under realistic conditions ($m \gg n$) most of the computing time is needed for establishing the coefficient matrix $\mathbf{C}$ and the right-hand side $b = \mathbf{A}^T\mathbf{W}y$ (i.e., $mn(n + 1)/2$ operations, 1 operation = 1 multiplication + 1 addition).

In many applications, it is not known previously whether or not the mixture in question does in fact contain a certain component. The absence of such a component should always be expressed by a corresponding zero entry in the resulting vector x of relative concentrations. Because of departures from the model and measurement errors this, unfortunately, does not happen for a great variety of real samples when analyzed by one of the methods described above.

As a somewhat artificial but illustrative example, the situation shown in Fig. 1 may be considered. Curves A, B and C represent the standard spectra. If it is assumed that the 3-component mixture actually contains only standard B with relative concentration 2, then under ideal experimental conditions each of the above methods computes the vector $x = (0, 2, 0)^T$. If, however, the spectrum of the mixture (M) is affected by a slightly curved baseline (curve Bl, Fig. 1b) caused, for example, by unmatched cells, the pure spectra A and C will try to compensate each other. This could typically result in a solution vector of the form $x = (0.1, 2.05, -0.14)^T$. If the negative part of x is ignored and simply set to zero, the "reconstructed" spectrum $s(\lambda)$ =

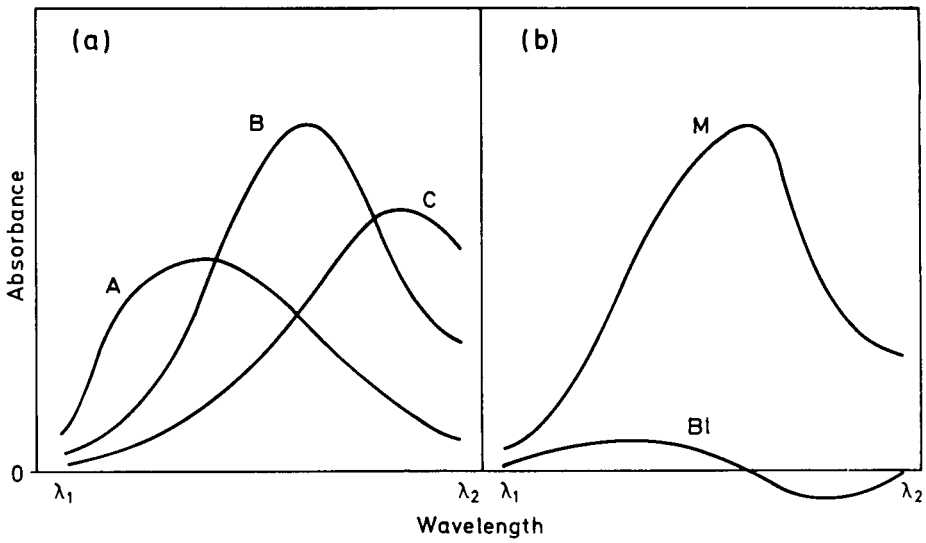


Fig. 1. Schematic explanation of compensation effects caused by a curved baseline (Bl). (a) Spectra of standards; (b) mixture spectrum (M). For further explanation, see text.

0.1 A(λ) + 2.05 B(λ) is significantly different from the measured spectrum of the mixture and cannot be considered as a reasonable approximation to M(λ).

There are three main reasons why the application of mathematical methods of multicomponent analysis developed so far is restricted to well-behaved samples: (1) the appearance of negative concentrations; (2) the lack of reliable error limits for the computed partial concentrations; and (3) the problem of data collection and storage together with the considerable amount of computational work if $m \gg n$, which is necessary if the standard spectra are similar.

To overcome these difficulties, the following three modifications were selected. First, non-negativity of the variables x_i is introduced as an additional constraint for the minimization problem described by relation 4; this has been discussed earlier [3, 6]. Secondly, a sharp sensitivity evaluation based on recent results in numerical mathematics is incorporated [7—10]. Thirdly, the hardware combination chosen consists of a microprocessor-controlled spectrophotometer equipped with an automatic $\Delta\lambda$ mode on line with a fast microcomputer. In the $\Delta\lambda$ mode, the absorbances are not measured continuously over the range $\lambda_0 - \lambda_m$ but are measured on a discrete "wavelength grid" $\lambda_i + i\Delta\lambda$ ($i = 0, \ldots, m$ with increment $\Delta\lambda$) and transmitted to the computer. The monochromator stops at each spectral band during the measurement period.

The first two modifications are now described in some detail, the third is treated in the experimental section. The method of non-negative least squares

(NNLS) was investigated by Gayle and Bennet [6], who concluded that NNLS is not significantly superior to conventional methods. This was not verified in the present experiments (cf. Table 3). It is well known [10, 11] that the introduction of physically induced constraints reduces the error amplification factor of so-called incorrectly posed problems (highly sensitive to measurement errors), sometimes by an order of magnitude. Even the examples given by Gayle and Bennet [6] seem to demonstrate the advantages of NNLS. But there is also a serious disadvantage of NNLS. The improved reliability must be paid for by considerably increased storage requirements and by a comparatively complicated mathematical problem. Instead of the $n \times n$ system of linear equations (5), the following quadratic optimization problem with n linear constraints must be solved

$$\sum_{j=1}^{m} w_j \left[y_j - \sum_{i=1}^{n} a_{ij} x_i \right]^2 \rightarrow \text{Minimum} \tag{6}$$

subject to $x_i \geqslant 0, i = 1, \ldots, n$. The conditions that must be satisfied by a solution to this problem (Kuhn—Tucker conditions [8]), do not lead to a system of linear equations in contrast to the unconstrained problem 4. Hence the minimization 6 has to be solved iteratively. However, experiment showed that the computational time needed to solve problem 6 exceeded that needed for problem 4 by a factor between 2 and 10, even when a finite iteration process was applied.

Several methods were tested, starting from the nonlinear SOR method [12] which is the simplest to program but which theoretically involves an infinite iteration. The NNLS algorithm suggested by Lawson and Hanson [8] is finite and fast, but the storage requirement is excessive for a microcomputer; it restricts the number m of wavelengths that can be used. The best compromise found between storage requirement and speed was Beale's algorithm [13] for which only a $2n \times n$ array must be kept in the central memory (independent of m) and which is nearly as fast as Lawson and Hanson's NNLS. The increased computation time versus normal least squares may be restrictive in some situations, if, for example the relative concentrations of the individual participants of a chemical reaction are to be traced in real time. In most applications, however, the computation time of 5—80 s can be neglected compared to the time needed to prepare (and measure) the sample.

Usually of greater interest in analytical work, other than accuracy, is reliable information about error limits for the results. Hence, a thorough evaluation of sensitivity is needed simultaneously with the analytical process. Some important results in the field of mathematical perturbation theory for least-squares problems have been published [7—9], but few of these have found their way into chemical practice. There are two main problems with commonly used statistical error considerations [6] in multicomponent analysis. First, most of the experimental error is not random but systematic (contaminations, cell differences, baselines etc.). Secondly, the (usually negligible)

random error does not satisfy the statistical hypotheses of normal distribution and independence.

The first statement is proven by experimental evidence that the fit error (minimum value of function 4) remains constant after a certain integration period t and does not tend to zero when the time t for the absorbance readings is extended (in the case of random errors, it should decrease to zero, proportionally to $(1/t)^{1/2}$). The second assertion is established by the well known physical fact that the error trends of absorbance readings at two neighbouring wavelengths are not independent. Thus a deterministic evaluation of sensitivity should be preferred to statistical error considerations. In the following paragraphs, standard notations for n-dimensional real space, euclidian vector and matrix norm, condition numbers, etc. [8] are used.

The simple case of a quadratic system of linear equations is considered first: $\mathbf{A}x = y$, $\mathbf{A} \in \mathbf{R}^{n,n}$, $x, y \in \mathbf{R}^n$ with a nonsingular symmetric coefficient matrix. If it is assumed that $\mathbf{A}$ is known exactly and that y is subject to experimental errors δy, then it can easily be proved (Eqns. 4 and 5) that the relative perturbation $\|\delta x\|/\|x\|$ of the solution of $\mathbf{A}(x + \delta x) = y + \delta y$ can be estimated by

$$\|\delta x\|/\|x\| \leqslant \mathrm{cond}(\mathbf{A})\|\delta y\|/\|y\| \tag{7}$$

where the condition number, cond(A), is equal to the ratio of the largest and the smallest absolute eigenvalues of $\mathbf{A}$. This estimate is sharp in the sense that there is always a right-hand side y and a perturbation δy such that relation 7 holds with equality. An illustrative numerical example is available [4]. Thus the condition number is a reasonable measure of experimental error propagation. For an asymmetric matrix $\mathbf{A}$, $\mathrm{cond}(\mathbf{A}) = \sigma_1/\sigma_n$ where σ_1 and σ_n are the smallest and largest singular values of $\mathbf{A}$ [8]. If the coefficient matrix $\mathbf{A}$ is also subject to perturbations, an estimate similar to Eqn. 7 can be established [4]. Sensitivity evaluation of an overdetermined system, such as arises in the LS and NNLS problem, is considerably more complicated [8, 9], details are beyond the scope of this paper. Here, only a typical result [5] is cited

$$\|\delta x\|/\|x\| \lessapprox \mathrm{cond}(\mathbf{A})\,\|\delta \mathbf{A}\|/\|\mathbf{A}\| + \mathrm{cond}^2(\mathbf{A})(\|r\|/\|\mathbf{A}\|\,\|x\|)(\|\delta \mathbf{A}\|/\|\mathbf{A}\|)$$

$$+ \mathrm{cond}(\mathbf{A})(\|y\|/\|\mathbf{A}\|\,\|x\|)(\|\delta y\|/\|y\|) \tag{8}$$

In this inequality, x is the least-squares solution of $\mathbf{A}x = y$, $\delta \mathbf{A}$, δy, and δx are the perturbations of the system matrix, the right-hand side, and the solution, respectively. The vector r denotes the residual $\mathbf{A}x - y$ and cond(A) now stands for the generalized condition number because $\mathbf{A}$ is rectangular [14] (in case of a rectangular matrix $\mathrm{cond}(\mathbf{A}) = \sigma_1/\sigma_n$, where σ_1 and σ_n are the square roots of the largest and smallest eigenvalues of $\mathbf{A}^T\mathbf{A}$). The dotted inequality sign indicates that higher-order error terms have been neglected. Although a first-order approximation is valid for small perturbations only, no difficulties arise for multicomponent applications because large residuals invalidate the results of multicomponent analysis anyway and do not require

a sophisticated evaluation of sensitivity. For an evaluation of the right-hand side of relation 8, the unknown terms **A**, y, and x have to be replaced by the (possibly perturbed) measured terms. It is easily seen by a Taylor expansion of the denominators that this causes a second-order perturbation of the right-hand side and is therefore justified. The size of $\delta\mathbf{A}$ and δy is either fixed by instrument specifications or can be tested by typical sample series. It is worth noting that the amount of work needed for an exact computation of cond(**A**) usually exceeds that for the determination of the (ordinary) least-squares solution itself. The computation of cond(**A**) requires the determination of the two extreme eigenvalues of $\mathbf{A}^T\mathbf{A}$ [14]. To overcome this difficulty, it is sometimes suggested that cond(**A**) be replaced by an easily attainable lower limit [8]. Experimental experience, however, shows that in most cases these simple formulae underestimate the real condition to such an extent that a considerable degree of uncertainty is introduced in the estimate of relation 8. Therefore, an exact computation of cond(**A**) by the methods suggested by Stewart [14] is highly recommended.

In the following section, the function $i(\mathbf{A}) = \text{cond}(\mathbf{A}) - 1$ is called "independence of standards"; $i(\mathbf{A})$ ranges between zero for non-overlapping pure spectra and infinity for linearly dependent pure spectra. The introduction of weighting factors [e.g., $w_j = 1/(1 + s_j)$, where s_j is the relative standard deviation of repeated measurements at a particular wavelength) did not cause significant changes of the results (s_j was close or equal to zero). Therefore, $w_j = 1$ was chosen in these experiments.

EXPERIMENTAL

Equipment

The Kontron Uvikon 820 spectrophotometer used had a photometric accuracy of ±0.2% at an absorbance of 1; the wavelength reproducibility was ±0.02 nm and the photometric linearity was $\pm 10^{-4}$%. Through RS-232 interfaces, the photometer was connected on-line with a Kontron PSI-80 microcomputer [CPU Z80A (4 MHz), central memory 64 kbyte, two integrated floppy discs 616 kbyte]. The computer controlled all photometric functions. The mathematical algorithms were written in Assembler and compilable Microsoft-BASIC; the program developed is distributed by Kontron.

Standards

Inorganic salts and picric acid. Analytical-grade salts (copper, cobalt, nickel and uranyl nitrates as their usual hydrated salts) and picric acid were dissolved in deionized water without pH adjustment, and kept at room temperature until the absorbance of the solution remained constant (at least 24 h before use). The same solutions were used for establishing the spectra of standards and for preparing the "unknown" test mixtures by pipetting aliquots of these standard solutions at 20°C (pipetting error for each solution ±0.3—0.5%).

Carotenoids. Carotenoids were extracted from the fungi *Neurospora crassa* and *Fusarium aquaeductuum* as described elsewhere [15]. After the removal

of the constituent neurosporaxanthin by phase partition with methanol/water (9 + 1) as the alternative solvent, they were separated by liquid chromatography as described by Bindl et al. [16] and modified by Reger [17] on a calcium hydroxide/Hyflo-SuperCel (7:3, w/w) column (8 × 1.7 cm) with petroleum benzine as developing solvent, containing increasing volumes of toluene (0—100%). Chromatography was done at room temperature under nitrogen pressure (ca. 150 kPa) until γ-carotene reached the lower end of the column. Then the compacted column filling was extruded, the zones of the individual carotenoids were sliced off, and the carotenoids were eluted with acetone (except for 3,4-dehydrolycopene which had to be eluted with chloroform) followed by filtration. This procedure turned out to be less time-consuming and of higher resolving power than ordinary processing. The coloured carotenoids running in front of the γ-carotene and eluted prior to extrusion of the column (β-carotene, β-zeacarotene and ζ-carotene) were separated by t.l.c. on activated (45 min at 120°C) Merck (type 60) silica gel glass plates using petroleum benzine/toluene (92+8) as developing solvent.

The absorbances of the carotenoids in n-hexane were measured at their individual absorption maxima and related to their concentration. Whenever possible, the carotenoids were handled under a nitrogen atmosphere and kept at 0°C or stored at −20°C.

RESULTS AND DISCUSSION

Analysis of inorganic salt mixtures

Before real samples were analysed, the reliability of the above-described NNLS method was tested on an artificial mixture satisfying all the hypotheses necessary for optimal performance of this method. For the preparation of "unknown" test mixtures, a modification of the aqueous solutions of inorganic salts described by Jochum et al. [4] was used; the copper, cobalt and nickel nitrates were used instead of the chlorides, and to avoid disproportionation, potassium dichromate was replaced by uranyl nitrate. The anions had to be all of the same kind, because the spectra also depend on the anion present. The potential complexity of the resulting "unknown" test mixtures was increased by introducing picric acid as a fifth substance, without any noticeable interference with the metal nitrates. The spectra of these analytes are shown in Fig. 2.

The mixtures were prepared by pipetting aliquots of the following standard solutions: 0.2 M $Cu(NO_3)_2$, 0.5 M $Co(NO_3)_2$, 0.5 M $Ni(NO_3)_2$, 0.3 M $UO_2(NO_3)_2$ and 0.1 mM $C_6H_2OH(NO_2)_3$, their absorbance at the absorption maxima being in the same range. The absorbance of the freshly prepared solutions decreased 3—4% within some hours, therefore the standards were not used until their absorbance remained stable.

In one of the mixtures, only two of the five components were present, namely copper and nickel, in order to check whether the method worked reliably under these conditions. Further, in this mixture the relative

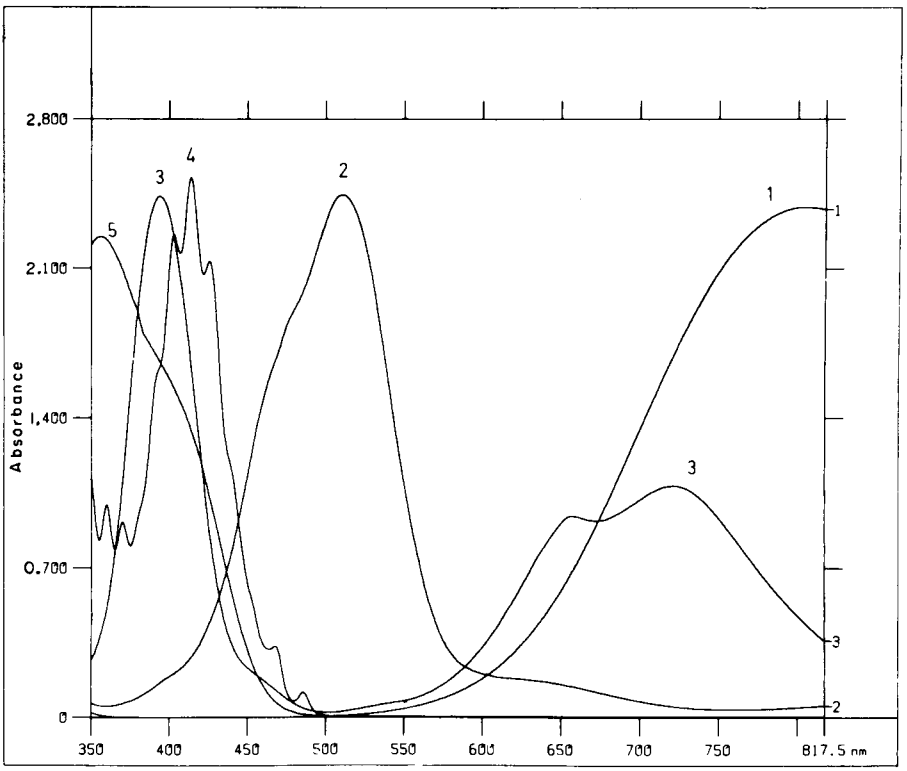


Fig. 2. Spectra of the standard solutions: (1) copper nitrate; (2) cobalt nitrate; (3) nickel nitrate; (4) uranyl nitrate; (5) picric acid.

concentrations were chosen as 10:1 (i.e., 10:1 Cu/Ni) to establish the tolerance to considerable differences in the concentrations of the analytes. Tests were also made to establish whether the wavelength range can be restricted if necessary (e.g., to avoid unknown contamination absorbing in parts of the wavelength range covered by the analytes) to the unaffected part of the spectra. In this case [10:1 Cu/Ni(r)], the absorbance of nickel in the particular reduced wavelength range (620—817.5 nm) was only about half that of copper, so that the multicomponent analysis had to deal with an absorbance ratio of about 20:1. However, when the restricted wavelength region was used, the extremely high dependence of standards (because of the lack of absorbance of copper and nickel in that region) allowed the computation with only three of the five standards.

Evidently, as shown in Table 1 and Fig. 3, the NNLS method can be used to analyse these mixtures reliably. The increment $\Delta\lambda$ was the same for all samples, therefore only a reduced number of data points was available for the analysis of the 10:1 Cu/Ni(r) sample. The results for the two mixtures containing all the components in different relative amounts show that the

TABLE 1

Results of multicomponent analysis of synthetic mixtures of standard solutions of copper, cobalt, nickel and uranyl nitrates and picric acid (Pic). The theoretical values for the relative concentrations are given in parentheses

Sample	Cu/Co/Ni/U/Pic 1+1+1+1+1	Cu/Co/Ni/U/Pic 2+10+20+1+4	Cu/Ni 10+1	Cu/Ni(r) 10+1
Wavelength range (nm)	350—817.5	350—817.5	350—817.5	620—817.5
Increment (nm)	2.5	2.5	2.5	2.5
Number of data points	188	188	188	80
Dependence of standards	3.6	3.6	3.6	36.9
Fit error (%)	0.665	1.190	0.498	0.089
Estimated maximum error (%)	3.04	5.45	2.28	4.84
True error (%)	2.60	2.10	0.44	0.94
Computation time (s)	30	32	30	22
Rel. concs.: Cu:	0.2065 (0.2)	0.0522 (0.0541)	0.9123 (0.9090)	0.9143 (0.9090)
Co:	0.1941 (0.2)	0.2717 (0.2703)	0.0000 (0)	0.0000 (0)
Ni:	0.2034 (0.2)	0.5467 (0.5405)	0.0886 (0.0909)	0.0841 (0.0909)
U:	0.2047 (0.2)	0.0373 (0.0270)	0.0000 (0)	n.d.[a] (0)
Pic:	0.1950 (0.2)	0.1038 (0.1081)	0.0003 (0)	n.d. (0)

[a]Not determined.

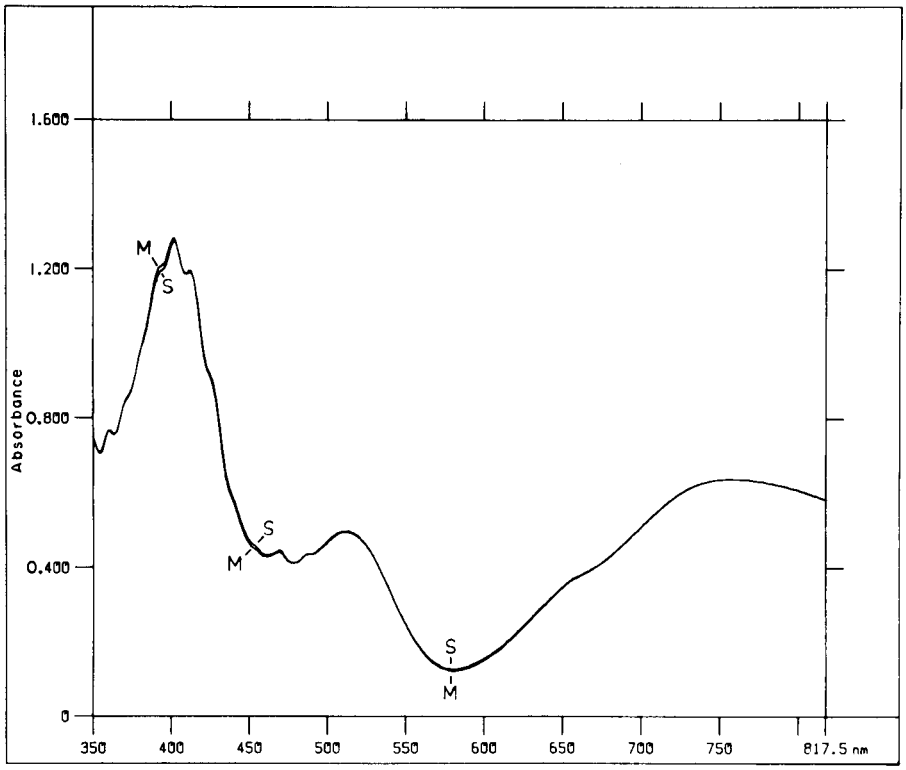


Fig. 3. Sample spectrum (M) and superposition (S) of the spectrum resulting from the computer calculation, for a 1:1:1:1:1 mixture of the standard solutions (Cu, Ni, Co, U, picric acid).

accuracy of the analysis, which is in the order of magnitude of the overall pipetting error, is consistent and not random. The low error also shows that pH adjustment of the solution prior to measurement is not mandatory with the salts used. The error was smallest for cobalt nitrate, most likely because its spectrum overlapped only slightly with the spectra of the other constituents.

The performance of the NNLS method was also demonstrated using hemoglobin/bilirubin mixtures, chloroplast pigments (chlorophylls and carotenoids) as well as u.v.-absorbing material (noradrenalin/lidocaine). The latter real mixture was contaminated by unknown material absorbing in sections of the wavelength range comprising the spectra of noradrenalin and lidocaine. It was found that a wavelength range of only 30 nm was sufficient to guarantee reliable results.

Analysis of carotenoid mixtures

The analysis of carotenoid mixtures may be regarded as a challenge for the method, in that the spectra of carotenoids are very similar (Fig. 4). Their

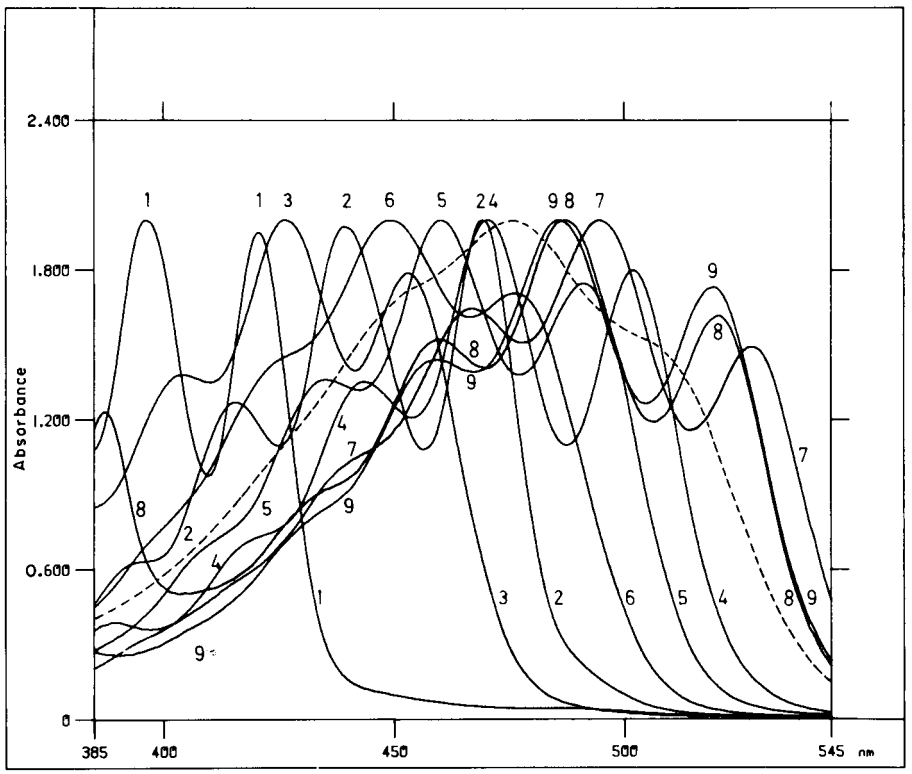


Fig. 4. Spectra of the carotenoid standards in n-hexane: (1) ξ-Carotene; (2) neurosporene; (3) β-zeacarotene; (4) lycopene; (5) γ-carotene; (6) β-carotene; (7) 3,4-dehydrolycopene; (8) *cis*-3,4-dehydrolycopene; (9) torulene; (— — —) neurosporaxanthin.

sensitivity to light and oxygen causes additional problems. The major constituent of the *Neurospora* and *Fusarium* carotenoids, neurosporaxanthin, which amounts to 50—70% of the whole [16, 18] is troublesome because it displays a rather diffuse spectrum (Fig. 4). Neurosporaxanthin, however, can easily be separated by phase partition, an essential procedure in the preparation of the pigments for column chromatography. As shown in Table 2, the remaining nine components of the *Neurospora* carotenoids were determined satisfactorily, although, as expected, the dependence of standards is relatively high. With 3,4-dehydrolycopene, *cis*-isomerisation is prominent, even though care was taken to prevent this during extraction and sample preparation. This is also true for torulene but was neglected because of the trace amounts present. Inevitably, *cis*-isomerisation is markedly enhanced when column chromatography is used, compared to spectrum deconvolution; the total amounts are presented in the Table for comparison.

Although acetone, which is normally used for elution of 3,4-dehydrolycopene, was replaced by chloroform, some of the pigment remained irreversibly

TABLE 2

Results of the multicomponent analysis of natural mixtures of coloured carotenoids extracted from *Neurospora* and *Fusarium*, respectively, compared to the results of conventional chromatography (values in parentheses)

Sample	*Neurospora*		*Fusarium*	
Wavelength range (nm)	385—545		405—530	
Increment (nm)	0.9		0.9	
Number of data points	178		139	
Dependence of standards	33.5		60.4	
Fit error (%)	0.284		0.164	
Estimated maximum error (%)	10.14		10.72	
Computation time (s)	67		48	
	Absorbance	*Percent*	*Absorbance*	*Percent*
ζ-Carotene	0.231	10.7 (8.2)	0.026	1.7 (2.4)
Neurosporene	0.296	13.7 (12.8)	0.033	2.2 (2.4)
β-Zeacarotene	0.037	1.7 (1.6)	n.d.[a]	— (traces)
Lycopene	0.346	16.0 (21.8)	0.639	42.3 (41.5)
γ-Carotene	0.563	26.1 (26.8)	0.469	31.0 (28.4)
β-Carotene	0.017	0.8 (0.8)	0.006	0.4 (2.2)
3,4-Dehydrolycopene	0.555	25.7 (18.2)[b]	n.d.	— (traces)
cis-3,4-Dehydrolycopene	0.074	3.4 (8.0)[b]	n.d.	— (traces)
Torulene	0.041	1.9 (1.8)	0.082	5.4 (3.1)[c]
cis-Torulene	n.d.	— (traces)	0.259	17.1 (19.8)[c]

[a]Not determined. [b]Total of the dehydrolycopene isomers is 29.1% (26.2%). [c]Total of the torulene isomers is 22.5% (22.9%).

bound to the column material which remained reddish. Consequently, the total amount of 3,4-dehydrolycopene found by means of column chromatography is lower than that found by NNLS; thus the chromatographic values are not a wholly reliable basis for judging the NNLS because of their own particular errors. Synthetic test mixtures produced from the standard solutions were evaluated reliably (data not shown).

The *Fusarium* carotenoids were computed more precisely than the *Neurospora* mixture because they did not contain measurable amounts of 3,4-dehydrolycopene and β-zeacarotene; these components were therefore neglected in the analysis (Table 2). The greater precision is most likely due to the smaller number of standards although the dependence of standards is higher because of the narrower wavelength range which turned out to be optimal for the analysis of the *Fusarium* carotenoids.

The same *Neurospora* mixture was used to demonstrate the performance of NNLS compared to conventional methods, such as standard least-squares (LS) and the *n*-wavelength method (NWM). Table 3 shows the results of these methods compared with the results of chromatography (see also Table 2). The advantage of NNLS over LS is obviously the reliability of the computed amounts of constituents at low concentrations. The fit errors for LS

TABLE 3

Comparison of the results of different computation methods with results obtained by conventional chromatography. The same *Neurospora* mixture as in Table 2 was analyzed, but its complexity was increased by using *cis*-torulene as an additional standard in order to facilitate discrimination of the performance of the different methods[a]

Method	CHROM	NNLS	LS	NWM	NWM(max)
Wavelength range (nm)	—	385—545	385—545	385—547	—
Increment (nm)	—	0.9	0.9	18	(λ max)
Number of data points	—	178	178	10	10
Dependence of standards	—	97.4	97.4	639.4	1158.5
Fit error (%)	—	0.584	6.089	15.3	—
Computation time (s)	—	69	55	11	(7)
ζ-Carotene (%)	8.2	10.7	10.1	8.7	—[b]
Neurosporene (%)	12.8	13.7	14.1	8.9	—[b]
β-Zeacarotene (%)	1.6	1.7	1.5	4.0	—[b]
Lycopene (%)	21.8	16.0	20.1	28.3	—[b]
γ-Carotene (%)	26.8	26.1	24.1	16.2	—[b]
β-Carotene (%)	0.8	0.8	−2.6	1.4	—[b]
3,4-Dehydrolycopene (%)	18.2	25.7	23.2	22.0	—[b]
cis-3,4-Dehydrolycopene (%)	8.0	3.4	4.0	8.7	—[b]
Torulene (%)	1.8	1.9	2.9	1.6	—[b]
cis-Torulene (%)	traces	0.0	−2.6	−12.0	—[b]

[a]CHROM chromatography; NNLS non-negative least-squares method; LS least squares method; NWM *n*-wavelength method (wavelengths selected equidistantly); NWM(max) same method but wavelengths selected according to the absorption maxima of the constituents. [b]Arthmetic alarm. Computation impossible.

and NWM were computed after the negative concentrations had been set to zero. Already the fit error of LS is worse by an order of magnitude than that of NNLS. The fit error of NNLS is only slightly above the unrestricted minimum of 0.385%. The advantage of NNLS over NWM is striking.

General assessment of the method

The discussion of the advantages of photometric multicomponent analysis should not cause confusion about the role of conventional chromatography: both methods support rather than replace each other. This is clear from the observation that the proper performance of multicomponent analysis requires correct preparation of the pure standards which normally needs chromatographic methods.

Yet, when standards can be purified, the application of deconvolution methods reduces the amount of experimental time, material, and manpower by an order of magnitude. Like many other analytical methods, its applicability depends on a few specific conditions such as characteristic standard spectra, the validity of Beer's law, and knowledge of the qualitative composition of the mixture. Hence, in some cases, there is no alternative to chromatography.

If the mathematical and chemical hypotheses discussed in the theoretical part are satisfied, photometric multicomponent analysis becomes very well suited for biochemical routine purposes. The success of the method and its reliability is strongly dependent on two factors, first the quality of the spectrophotometer (wavelength reproducibility, linearity, resolution, signal-to-noise ratio), and second the sophistication of the mathematical model and method. To optimize the first factor, it is highly recommended to use a fully digital signal processing (no logarithmic amplifiers) together with a microprocessor-controlled wavelength adjustment of the monochromator. The photometer used in these experiments has a specified wavelength reproducibility of 0.02 nm.

Among the tested mathematical methods, the non-negative least-squares algorithm programmed in double precision yielded the best results. The criticism [6] that the use of restrictions invalidates the statistical meaning of the covariance matrix $(\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}$ is irrelevant insofar as a meaningful error evaluation based on the covariance matrix relies on the assumption that model deviations are random. That this assumption is hardly satisfied in general was pointed out in the theoretical part.

A thorough (deterministic) sensitivity evaluation should not be considered as a sometimes helpful appendix but as an intrinsic part of the method itself. It is no exaggeration to say that photometric multicomponent results without error limits are nearly meaningless. Thus the standard NWM seems to be restricted to applications where either model deviations and measurement errors can be excluded or a reference method can be used to check the accuracy.

In spite of the considerable complexity of the numerical method and the sensitivity evaluation through eigenvalue computation, it was possible to run the NNLS program on a 64 kbyte microcomputer coupled on-line to the photometer. Thereby, the system was fully automated and could be handled even by unskilled personnel.

The authors express their gratitude to G. Hämmerlin and W. Rau for making this teamwork possible.

REFERENCES

1 R. Wodick, Ph.D. Thesis, University of Marburg, West Germany, 1968.
2 S. Ebel, Fresenius Z. Anal. Chem., 313 (1982) 452.
3 D. J. Legget, Anal. Chem., 49 (1977) 276.
4 C. Jochum, P. Jochum and B. R. Kowalski, Anal. Chem., 53 (1981) 85.
5 J. Stoer, Einführung in die Numerische Mathematik I, 2nd edn., Springer, Berlin, 1976, p. 171.
6 J. B. Gayle and H. D. Bennet, Anal. Chem., 50 (1978) 2085.
7 A. van der Sluis, Numer. Math., 23 (1975) 241.
8 L. Lawson and R. Hanson, Solving Least Squares Problems. Prentice-Hall, Englewood Cliffs, NJ, 1974.

9 L. Eldén, SIAM (Soc. Ind. Appl. Math.), J. Numer. Anal., 17 (1980) 338.
10 U. Eckhardt, Computing, 17 (1976) 193.
11 L. Eldén, Report LiTH-MAT-R-1977-20, University of Linköping, Sweden.
12 J. Ortega and W. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1972.
13 E. M. L. Beale, J. R. Stat. Soc. Ser. B, 17 (1955) 173.
14 G. W. Stewart, Introduction to Matrix Computations, Academic Press, New York, 1973.
15 E. L. Schrott, A. Huber-Willer and W. Rau, Photochem. Photobiol., 35 (1982) 213.
16 E. Bindl, W. Lang and W. Rau, Planta, 94 (1970) 156.
17 A. Reger, Masters Thesis, University of Munich, 1982.
18 U. Mitzka and W. Rau, Arch. Microbiol., 111 (1977) 261.

Analytica Chimica Acta, 157 (1984) 227—243
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

AN AUTOMATED LIBRARY SEARCH SYSTEM FOR ^{13}C-N.M.R. SPECTRA BASED ON THE REPRODUCIBILITY OF CHEMICAL SHIFTS

R. W. BALLY, D. VAN KRIMPEN, P. CLEIJ and H. A. VAN 'T KLOOSTER*

State University of Utrecht, Laboratory for Analytical Chemistry, Croesestraat 77A, 3522 AD Utrecht (The Netherlands)

(Received 2nd August 1983)

SUMMARY

A library search system for ^{13}C-n.m.r. spectra, based on a statistical description of the reproducibility of chemical shifts, is presented. A similarity index in the form of a significance probability (*P*-value) is developed from a previously introduced general concept. The applied data base of some 6000 spectra originates from the NIH-EPA Chemical Information System (CIS). The reproducibility model and the retrieval system are developed on a CDC Cyber 175 computer, with PASCAL as the programming language. The performance of the system is evaluated by using recall/reliability and confusion/recall plots. In a tentative comparison with the Clerc search method (included in the CIS package), the Utrecht ^{13}C-n.m.r. reproducibility-based retrieval system shows a better identification performance. The system is adaptable for use on a microcomputer.

In recent years, several systems have been proposed for computer-aided library search of ^{13}C-n.m.r. spectra, aiming at the identification of organic compounds and based on various coding and comparison algorithms [1—13]. Collections of several thousands of spectra, from different sources (i.e., recorded on different instruments, under various experimental conditions) are usually applied as data bases. Such data bases, which are often commercially available, generally suffer from poor interlaboratory reproducibility of the spectral data involved. In a previous paper [14], a new similarity index for "straightforward" library search methods was introduced, primarily for application to this type of data base. The main object of straightforward search methods is to retrieve the reference data (if available) of the unknown compound. This contrasts with "interpretative" methods, which aim principally at retrieving data of compounds similar in structure to the unknown. The proposed index has the form of a significance probability and is developed from a statistical model of the reproducibility of the quantities used for the comparison of unknown and reference data. Defined in general terms, this index is applicable to different types of analytical data, provided that the variables used to characterize unknown and reference spectra (feature quantities) are continuous in nature. In ^{13}C-n.m.r. spectra, the chemical shift and the peak intensity are such variables, in contrast to multiplicities. When a large multisource data base such as the CIS collection is used, how-

0003-2670/84/$03.00

ever, peak intensities are specified only in a minority of the spectra and show very poor reproducibility, which makes this information unsuitable for library search.

This paper reports the development and evaluation of a straightforward ^{13}C-n.m.r. library search system based on chemical shifts as the features. As a criterion of matching, a similarity index is developed based on a statistical description of the reproducibility of chemical shifts. The system thus takes into account any systematic and random errors occurring in chemical shifts. As a first step in the development of the retrieval system, a reproducibility model was derived from some 200 pairs of "alternative spectra"; each pair was for the same compound, but the spectra were recorded under different experimental conditions. This model describes the statistical behaviour of differences in chemical shifts in these alternative spectra. The resulting reproducibility function forms the basis of the definition of the similarity index (S_I). The main function of this index is to permit a separation of the reference compounds into two classes, i.e., compounds that could be the unknown and compounds that cannot. This corresponds to the acceptance or rejection of the null hypothesis that the unknown and the reference compound are identical. By specifying a threshold value of the index, all references with an index value above this threshold should be retrieved as belonging to the "possible" class. The precise value of this threshold depends on the pre-accepted risk of misclassifying (not retrieving) the correct reference compound. The result is a list of reference compounds (if any), one of which could be identical to the unknown, instead of the usual 5—10 "best" matches (which may in fact be very bad matches). The main features in the development of the ^{13}C-n.m.r. reproducibility-based retrieval (C13RR) system are schematically indicated in Fig. 1.

EXPERIMENTAL

For the design, development and evaluation of the C13RR system, a CDC Cyber 175 (60 bits) computer was used. Programs are written in PASCAL, with the exception of a few subroutines; a Wiswesser line notation/connectivity table conversion program (kindly supplied by Prof. J. Zupan, Kemijski Institute Boris Kidrič, Ljubljana, Yugoslavia) was written in FORTRAN. The data base originated from the NIH-EPA Chemical Informa-

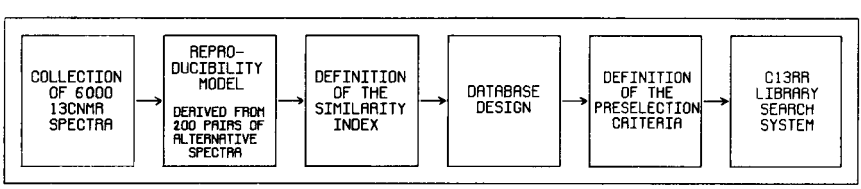


Fig. 1. Main features in the development of the ^{13}C-n.m.r. reproducibility-based retrieval (C13RR) system.

tion System (CIS; 1982 version) and contained some 6200 spectra of 5800 different compounds. The original data base, coded in Bremser exchange format [7], was transformed and reorganised, resulting in B-tree-structured and indexed sequential subfiles [15]. An important adaptation was the removal of redundant shift values; the original data base contained multiple shift values of equal magnitude, obviously referring to the same peak. In the data base ultimately applied, a chemical shift was stored with a precision of 0.01 ppm.

A REPRODUCIBILITY-BASED SIMILARITY INDEX

The similarity index proposed by Cleij et al. [14] implies the use of so-called difference quantities, representing the differences in value, for two spectra, of feature quantities. The use of chemical shifts as the feature quantities leads to a set of n difference quantities, $\Delta q_1 \ldots \Delta q_n$, defined by

$$\Delta q_i = \delta_{U,i} - \delta_{R,i} \quad (\text{for } i = 1 \ldots n) \tag{1}$$

where $\delta_{U,i}$ and $\delta_{R,i}$ are the shift values of the ith peak in the unknown and reference spectrum, and n is the number of peaks (assuming an equal number of peaks in both spectra).

Reproducibility of chemical shifts

For the development of the model of reproducibility of chemical shifts, a set of 202 pairs of alternative spectra (each pair for the same compound but consisting of different spectra), originating from the CIS data base, was used. This set was selected from 441 such pairs, by exclusion of pairs of spectra which were identical, or contained less than 3 peaks, or had different numbers of peaks. The 202 pairs of alternative spectra were analysed in the form of "difference spectra". A difference spectrum of two spectra A and B is a plot of the differences in shift values for corresponding peaks, versus the shift values of spectrum A, as shown in Fig. 2.

The differences in shift values, as shown by the difference spectra, consist of a random and a systematic part. The random part is demonstrated by the scatter of the data points in a difference spectrum. The magnitude of this scatter ranges from very small (Fig. 3), to relatively large (Fig. 4). The systematic part of the differences in shift values is demonstrated by difference spectra in which the average difference significantly deviates from the zero (no difference) line (Fig. 5). Evaluation of some 50 difference spectra showed that systematic differences can adequately be described by a straight line, parallel to the zero line. These systematic deviations of the shifts are mainly caused by variation in the (absolute) position of the TMS peak.

The reproducibility function, for the set of n difference quantities denoted by $p_0[\Delta q_1 \ldots \Delta q_n]$, should provide a description of the formal difference spectra. As shown above, the total difference in a chemical shift (Δq_i) is considered to consist of the random part (Δq_i^r) and the systematic part (Δq^s), i.e.:

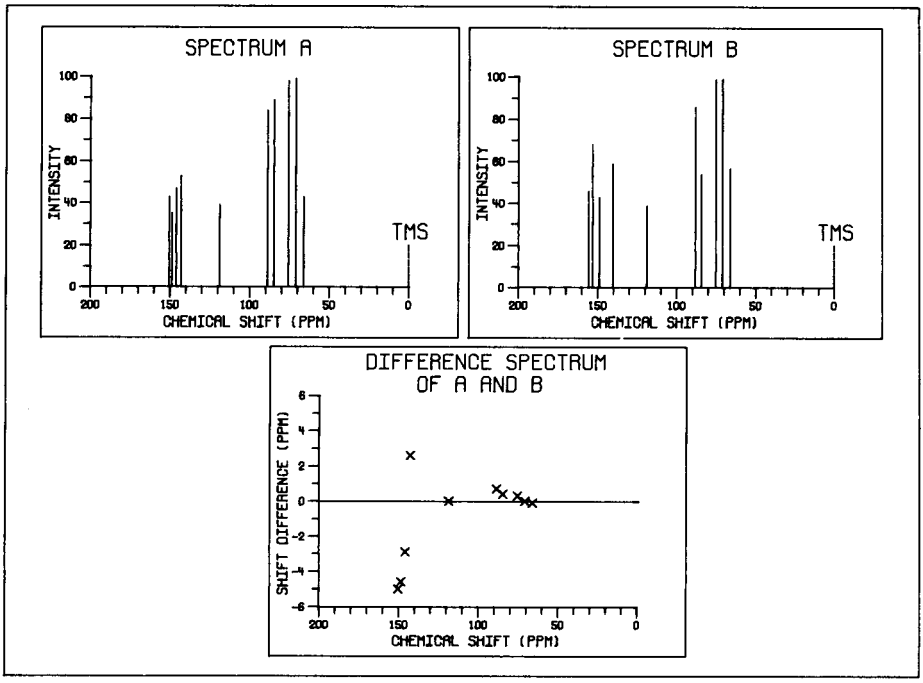


Fig. 2. Two ^{13}C-n.m.r. spectra A and B of the disodium salt of adenosine 5'-(tetrahydrogen-triphosphate), dissolved in two different basic solutions [16]. The "difference spectrum" of A and B comprises the differences in the shift values of the peaks in spectrum A and B.

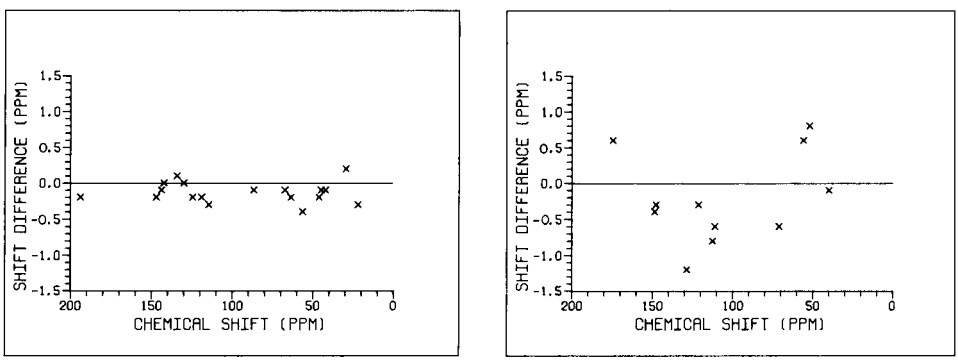


Fig. 3. Example of a difference spectrum, illustrating random differences with a relatively small variance. Differences in the shift values of two spectra of 3,14-bis(acetyloxy)-7,8-didehydro-4,5-epoxy-17-methyl-, (5-α)-morphinan-6-one. The two spectra were measured on different spectrometers [17, 18].

Fig. 4. Example of a difference spectrum, illustrating random differences with a relatively large variance. Differences in shift values of two spectra of benzenepropanoic acid, α-hydroxy-3,4-dimethoxymethyl ester. The two spectra were measured on different spectrometers and in different solvents [19].

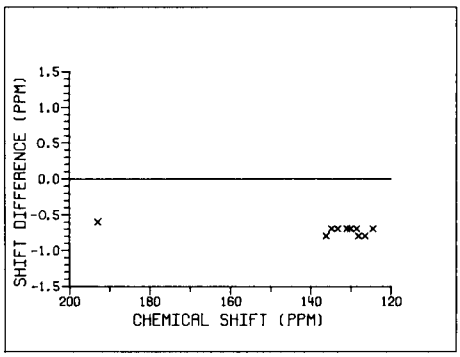


Fig. 5. Example of a difference spectrum, demonstrating a systematic difference in the shift values of two spectra of 1-naphthalenecarboxaldehyde, measured by two different operators [16, 20].

$$\Delta q_i = \Delta q_i^{\mathrm{r}} + \Delta q^{\mathrm{s}} \tag{2}$$

This is illustrated by Fig. 6.

The random differences are assumed to be normally distributed about zero and with variance σ^2. Hence

$$\mathrm{p}(\Delta q_i^{\mathrm{r}}) = N(\Delta q_i^{\mathrm{r}}{:}0, \sigma^2) \tag{3}$$

[$N(x{:}E, V)$ is used in this paper as the general notation for a normal density function with variable x, expected value E and variance V.]

In order to meet the large variations in the magnitude of the scatter observed in difference spectra, the variance σ^2 is considered as a stochastic quantity, varying with the difference spectrum. Preliminary studies showed that the probability density of σ^2 is adequately described by a log-normal function:

$$\mathrm{p}(\sigma^2) = [\sigma^2 \cdot (2\pi \cdot V \ln \sigma^2)^{1/2}]^{-1} \exp[-0.5(\ln \sigma^2 - E \ln \sigma^2)^2 / V \ln \sigma^2] \tag{4}$$

where $E \ln \sigma^2$ is the expected value of $\ln \sigma^2$ and $V \ln \sigma^2$ is the variance of $\ln \sigma^2$.

The systematic difference is a constant in a particular difference spectrum but a variable for various difference spectra. Assuming that the systematic difference follows a normal probability function with the same variance as applied to the random differences, the probability distribution of the systematic differences can be described by:

$$\mathrm{p}(\Delta q^{\mathrm{s}}) = N(\Delta q^{\mathrm{s}}{:}0, \sigma^2) \tag{5}$$

For a set of n difference quantities (differences in shift values) the reproducibility function can then be formulated as

$$\mathrm{p_o}[\Delta q_i] = \int_{\sigma^2=0}^{\infty} \mathrm{p}(\sigma^2) \cdot \int_{\Delta q^{\mathrm{s}}=-\infty}^{\infty} \mathrm{p}(\Delta q^{\mathrm{s}}) \cdot \prod_{i=1}^{n} \mathrm{p}(\Delta q_i^{\mathrm{r}})\, \mathrm{d}\Delta q^{\mathrm{s}} \mathrm{d}\sigma^2 \tag{6}$$

in which n is the number of peaks of the spectra involved, $[\Delta q_i]$ is the set of n difference quantities, $p(\sigma^2)$, $p(\Delta q^s)$, $p(\Delta q_i^r)$ are given by Eqns. 4, 5 and 3, and where $\Delta q_i^r = \Delta q_i - \Delta q^s$ (see Eqn. 2). An elaboration of Eqn. 6 is given in the Appendix.

The similarity index (S_I)

The similarity index [14] is defined as:

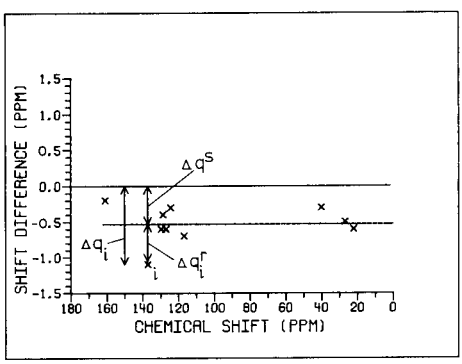
$$S_I = \int_{\Delta q_1} \dots \int_{R[\Delta Q_i]} \int_{\Delta q_n} p_0[\Delta q_i] \, d\Delta q_1 \dots d\Delta q_n \qquad (7)$$

where n is the number of chemical shifts, $p_0[\Delta q_i]$ is given by Eqn. 6 and $R[\Delta Q_i]$ is the region in the space of difference quantities, defined by the condition $p_0[\Delta q_i] \leqslant p_0[\Delta Q_i]$, with ΔQ_i being the "observed" value of Δq_i, i.e., the difference of the shift values of the ith peak in the unknown and reference spectrum.

An elaboration of Eqn. 7 into a more workable expression of the S_I is given in the Appendix.

THE RETRIEVAL SYSTEM

The ^{13}C-n.m.r. reproducibility-based retrieval (C13RR) system consists of a data base, the (modular) set of search programs and job control software of the computer being used. In the standard mode, the system retrieves



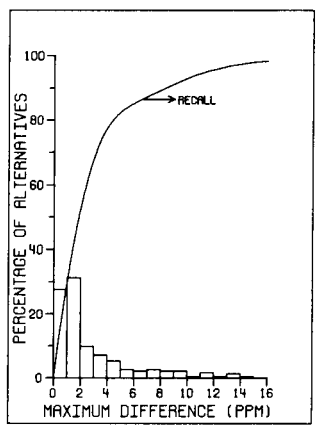


Fig. 6. The total difference (Δq_i) is the sum of the random difference (Δq_i^r) and the systematic difference (Δq^s). This difference spectrum results from two spectra of 3,4-dihydro-1(2H)-quinolinecarboxaldehyde, measured in different solvents [21].

Fig. 7. Histogram of the maximum differences in shift values as occurring in 230 pairs of alternative spectra. The curve represents the theoretical recall as a function of the maximum difference. For 15 ppm (default value), the recall is 98%.

all the reference spectra with a number of peaks equal to that of the unknown spectrum and with an S_I value higher than the threshold level, usually preset at 2%. In order to deal with "missing" peaks in ^{13}C-n.m.r. spectra, an alternative mode, allowing a difference of one peak between unknown and reference spectrum, is optional. (Missing peaks are mainly caused by differences in instrumental and recording conditions, sample impurities and careless copying of spectral data.) Table 1 gives a survey of missing peaks in 316 pairs of alternative spectra, from which it can be concluded that more than 1 peak is missing in only 9% of the spectra. The fact of a missing or extra peak in the reference spectrum is neglected in the calculation of the S_I. In the alternative mode, three kinds of hit lists are produced for an unknown spectrum with n peaks, i.e., lists containing reference spectra (if any) with $n-1$, n and $n+1$ peaks.

Preselections

If the algorithm for the calculation of the S_I were applied to all the reference spectra, this would lead to an inacceptably large response time, especially with large data bases. A way to solve this problem is the use of preselection criteria, which should meet the following requirements: (1) all reference spectra with an S_I greater than the applied threshold level should pass the preselection; (2) the preselection should have a high total selectivity; (3) the (computer) time consumption should substantially be decreased.

In the C13RR system the following preselection criteria are applied sequentially. First, in the standard mode, the number of peaks in the unknown and the reference spectra must be equal; in the alternative mode a difference of one peak is allowed. Secondly, the maximum allowable difference in corresponding shift values is 15 ppm. The second criterion was derived from an evaluation of the set of pairs of alternative spectra. The distribution of the maximum differences between corresponding shift values occurring in the pairs of alternative spectra is given in Fig. 7; it may be interpreted as the recall [22] as a function of the tolerance. A tolerance of 15 ppm (the default value in the system) corresponds to 98% recall.

TABLE 1

Frequency of differences in the total number of peaks for 316 pairs of alternative spectra

Difference in number of peaks	Pairs of alternative spectra	Percentage of total
0	230	72.8
1	57	18.0
2	15	4.8
3	8	2.5
4	4	1.3
5	1	0.3
6	1	0.3
>6	0	0

Organisation of the data base

The data base is organised according to the preselection criteria. An outline of the data base is given in Fig. 8. As shown, the original data base is divided into two parts: the first contains the spectral data, and the second holds all further information such as names, molecular formulae, CAS registration numbers, sources of origin, etc. This division is made because the retrieval algorithm only uses the shifts of the spectra. The part containing the shifts is divided into subfiles, each containing spectra with the same number of shifts. Within a subfile the spectra are ranked according to the highest (most significant) shift values of the spectra. The reason for this ranking is the use of an indexed sequential file organisation [15], which we consider to be an efficient access strategy in combination with the second preselection criterion. The other part of the data base consists of index files on CAS registration number and molecular formula, which allows the

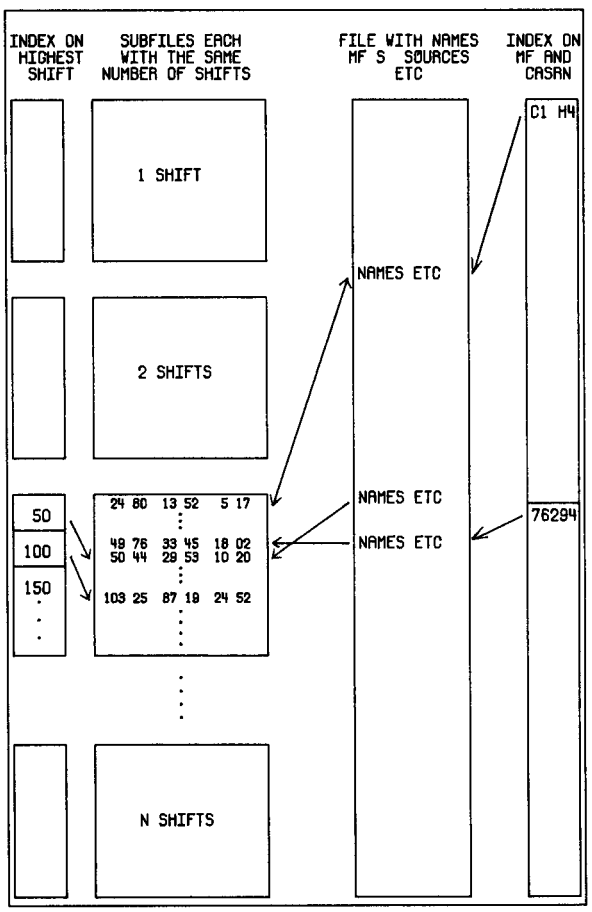


Fig. 8. Organisation of the data base used in the C13RR system.

establishment of the presence or absence of a specified compound as a reference in the data base. These files contain a "dense" index and are organised as a B-tree [15] with 128 keys per node, resulting in at most two disk accesses in order to find all the keys that belong to a given entry.

The retrieval algorithm

The main steps of the retrieval algorithm are as follows.

1. The subfile corresponding with the number of shifts in the unknown spectrum is selected.
2. On this subfile a window is defined by the highest shift in the unknown spectrum ± the preset tolerance.
3. The selection starts with the first spectrum that is "seen" through the window. The position of this spectrum in the subfile is calculated with the use of the index file. From this point the sequential access plan is executed.
4. Each difference between a shift of the unknown and the corresponding shift of the reference spectrum is compared with the applied tolerance; a spectrum is discarded when a difference greater than the tolerance is found.
5. If a reference spectrum has not been discarded in step 4, the S_I is calculated. The reference spectrum is selected for the hit list if the S_I value is greater than the applied threshold value. A binary tree is used to rank the reference spectra according to the S_I values.
6. If the highest shift in a subsequent reference spectrum falls within the specified window, it is submitted to step 4, etc., otherwise the program terminates and the results of the search are displayed.

As mentioned above, it is possible to allow a difference of one peak in the total number of peaks of the unknown and the reference spectrum. In this case, the retrieval algorithm deletes from the "larger" spectrum a shift in such a way that the maximum S_I value for that reference is calculated.

Modular structure of the C13RR system

The C13RR system consists of five modules. The program starts in the module CONTROL, which contains the system commands and controls the input and storage of the unknown spectra. Options for further processing of the input data are activated by specifying the first character or the whole option label. The user-friendly operation of the system includes the display of warnings after wrong statements have been typed. Typical outputs of each module are illustrated in Fig. 9.

EVALUATION OF THE C13RR SYSTEM

An essential criterion in the evaluation of a retrieval system is the discriminating power of the comparison index in separating the target spectrum (spectra) from the remaining reference spectra. Two aspects of this property can be distinguished, viz., the "absolute" and the "relative" discriminating power. A good absolute behaviour is demonstrated when the distribution of

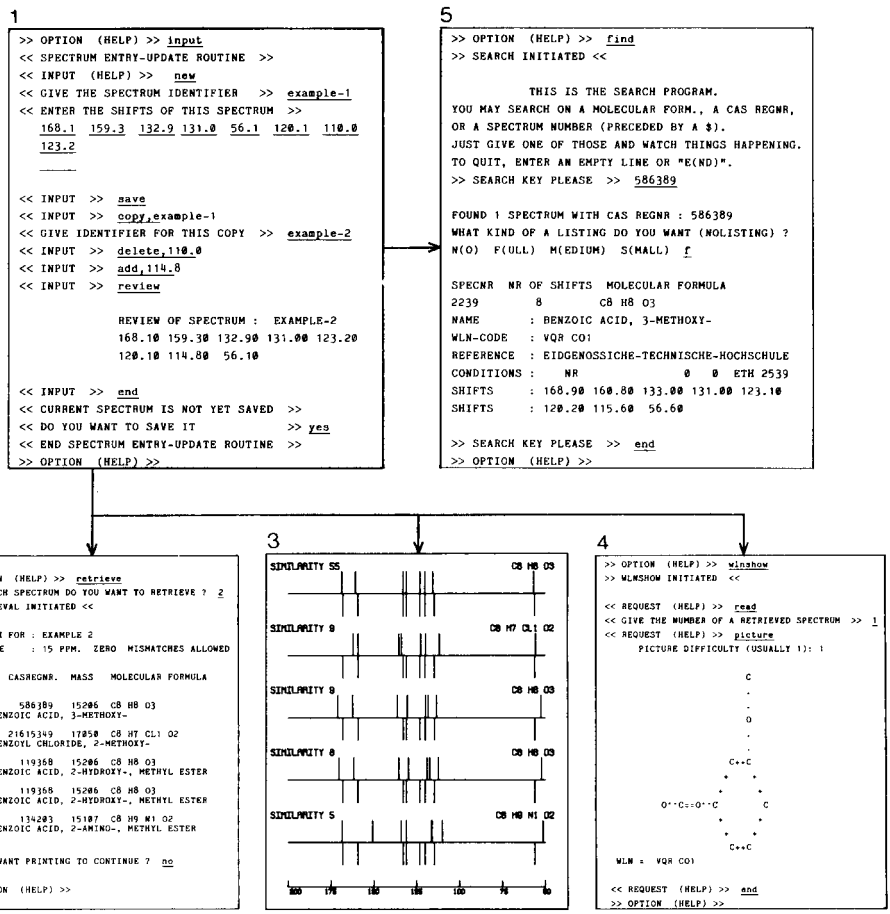


Fig. 9. Typical outputs of the C13RR system modules. The system starts in the module CONTROL, which allows the operator to enter the spectral information of the unknown. Included are several options such as: add and delete a shift, copy and review a whole spectrum or save a spectrum. In CONTROL it is also possible to select one of the other modules. RETRIEVE (2): the system starts the library search and presents the result of the search in the form of a hit list, containing reference spectra (if available) with a similarity greater than the specified threshold (default 2%); a typical response time for one search is 10 s. PLOT (3): the system produces plots of the unknown spectrum (pointed downwards) versus the retrieved reference spectra (pointed upwards), allowing a visual comparison. WLN-SHOW (4): the system produces a structure formula from a Wiswesser line notation [23] of a compound present in the hit list. FIND (5): after entering a search key (CAS reg. no. or molecular formula), the system gives a list of reference compounds containing the search key.

index values for correct matches (reference compound = unknown compound), determined for a large set of test spectra, shows only a small overlap with the distribution for incorrect matches. In the case of an optimal absolute behaviour of the comparison index, it is possible to define a single threshold value that divides all matches (for all searches) into correct and

incorrect ones (or correct positives and false positives), respectively. A retrieval system with a good relative discriminating power generally assigns, within one search, a "better" index value to the target spectrum (or spectra) than to the remaining reference spectra. An optimal system in this respect always places the target spectrum (or spectra) on top of the list of retrieval results. Such a system, however, is not necessarily an optimal or even a good retrieval system with respect to the absolute behaviour of the comparison index. In contrast, a comparison index with an optimal behaviour in the absolute sense, also has an optimal relative behaviour. It may seem, at first sight, that a good relative discriminating power of the comparison index is a sufficient guarantee for an adequate performance of the retrieval system. This is, however, not wholly true, especially not with regard to an important property of a good retrieval system, i.e., the possibility for the user of the system to judge, on the basis of the index values, whether the reference listed first is correct, or that no reference spectrum is available for the unknown compound. When a comparison index with a good discriminating power is used in the absolute sense, it is relatively easy to distinguish between the case of the first retrieved reference being the target spectrum, and the situation in which no target spectrum is available. With a comparison index that demonstrates a good relative behaviour, yet shows poor performance as an absolute index, these situations are much more difficult to distinguish.

For these reasons the absolute discriminating power of the similarity index is chosen as the main criterion in the evaluation of the C13RR system. The relative behaviour of the S_I, however, is also considered, because the relative and absolute behaviours are correlated only to a certain extent.

The absolute discriminating power of a comparison index can be evaluated with recall/reliability plots [22]. This type of evaluation fits with a retrieval plan in which all reference spectra having an index value above (or below) some prespecified threshold value, are retrieved. The recall for a test set of "unknown" spectra is defined as the number of actually retrieved target reference spectra (correct positives), divided by the total number of target spectra, available for the test set. Provided that the unknown compound has a spectrum in the reference file, the recall represents the estimated probability that this target spectrum is actually retrieved. The relationship between the recall and the threshold value of the S_I [14] is Recall = $(1 - Th)$.

The reliability is defined as the number of retrieved target spectra, divided by the total number of retrieved reference spectra (correct and false positives) for the "unknown" spectra of the test set. Both the recall and the reliability are functions of the index threshold value. Increasing the threshold value of the similarity index will decrease the recall and increase the reliability. For a good relative behaviour, the increase in reliability will exceed the decrease in recall. A recall/reliability plot can thus be obtained by varying the threshold values and plotting the reliability values as a function of the recall values. Recall/reliability values can, in principle, be calculated from the distributions

of the index values for correct and false positives, which explains their utility in evaluating the absolute behaviour of comparison indices.

A variation of the recall/reliability evaluation method is the extension of the concept "correct" for the evaluation of the reliability, by considering a retrieved reference spectrum not only as correct if the reference compound is identical to the test compound but also if this reference compound is, within well-defined limits, similar in structure to the test compound. This method of evaluation has been applied to the PBM retrieval system for mass spectra, using four definitions of a similar compound [2]. The performance of a retrieval system with regard to the relative behaviour of the comparison index, should be evaluated on the basis of the ranking of correct and incorrect matches for individual searches, executed with the spectra of the test set. As a measure of this behaviour, the recall as a function of the "confusion" [24] has been used. The confusion is defined here as the number of false positives with a S_I equal or better than the S_I of the target reference spectrum.

The C13RR system was compared with the CIS/Clerc CNMR search system [2, 7] with regard to the relative behaviour of the comparison index. (A recall/reliability plot of the Clerc system is not feasible, because of the nature of the applied matching criterion.) The (commercially accessible) CIS/Clerc system was chosen, as it has the same data base and is also meant to be used as an aid in the identification of organic compounds.

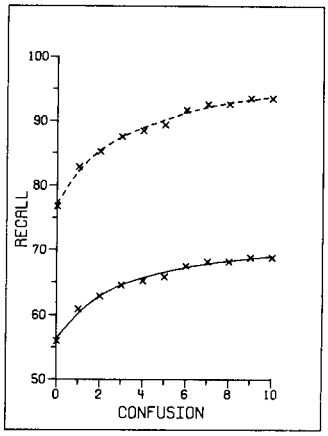
Test spectra

A test set (A) of 264 spectra was selected from the data base, each having one or more alternative spectra (no exact duplicates) as target spectra in the data base. For the whole set, 304 target spectra were available. All 264 test spectra were processed using the complete reference file of 6200 spectra minus the test spectra. To establish the influence of "missing" peaks, a second test set (B) was composed of 194 spectra having only target spectra in the reference file with the same number of peaks.

For a comparison with the CIS/Clerc system a third test set (C) of 50 spectra was randomly selected from test set A.

Results

The confusion plots of test sets A and B, given in Fig. 10, clearly show the negative effect of missing peaks. In processing the spectra of test set A, both preselection criteria are effective, resulting in a maximum attainable recall of 71% (27% of the correct positives are "lost" in the first, 2% of the remaining in the second preselection). The plot for test set B has a maximum of 94%, which is close to the limit of 98%, caused by the second preselection criterion (the first one has no effect in this case). From Fig. 10 it can be concluded that the correct positive is generally among the first retrieved references, or it is not retrieved at all. The recall/reliability plots of test sets A and B, for threshold values of the S_I ranging from 90% to 2%, are given in Fig. 11. Both curves can be divided into two regions of the recall, i.e., a first



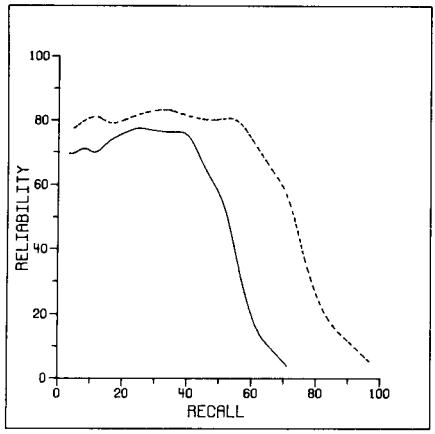
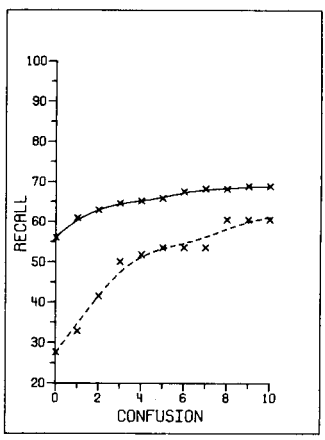


Fig. 10. Confusion plots for two test sets. Test set B (– – –) consisting of 194 spectra with target spectra having equal numbers of peaks. Test set A (——) containing test set B plus 70 spectra with target spectra having different numbers of peaks. (cf. Table 1.)

Fig. 11. Recall/reliability plots of test set A (——) and test set B (– – –). (cf. Fig. 10.)

region with about constant reliability (the plateau) and a region with strongly decreasing reliability. The plateau can be explained by the fact that there are some sets of compounds with (in principle) identical spectra. For instance, if it is assumed that (a) a reference file of 10 000 spectra consists of 5000 pairs of compounds, (b) the spectra of each pair differ only within measur-



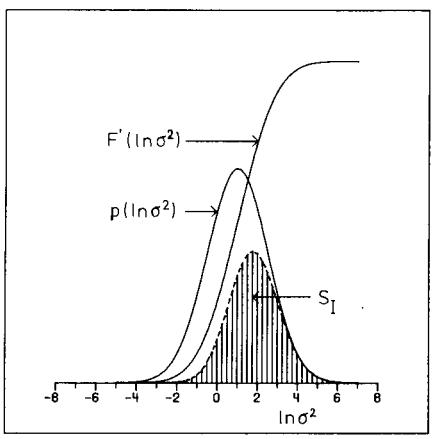


Fig. 12. Confusion plots of the C13RR system (——) and the CIS/Clerc CNMR search system (– – –).

Fig. 13. Graphic representation of the similarity index and related functions, according to Eqns. A14—16.

ing errors, and (c) spectra of different pairs are significantly different, this would result in a reliability of 50% for the full recall range. The tail of the recall/reliability curve can be explained by the fact that measuring errors may be about equal to, or even exceed, the real differences (differences corrected for measuring errors) between spectra of different compounds. The recall/reliability plots of Fig. 11 show the negative effect of missing peaks on the absolute behaviour of the similarity index: the plot for test set A indicates a lower reliability over the full range of the recall, as 27% of the correct positives are not retrievable, owing to the first preselection.

For a comparison of the C13RR system with the CIS/Clerc system, confusion plots for both systems, based on test set C, were generated (Fig. 12). These plots clearly demonstrate the better performance of the C13RR

TABLE 2

Results of a library search with the C13RR system and the CIS/Clerc system
[Test spectrum: "unknown" (one of two alternative library spectra of 1,3,3-trimethylbicyclo-2,2,1-heptan-2-one). Specified chemical shifts (ppm) were 222.3, 53.9, 47.2, 45.3, 41.6, 31.8, 24.9, 23.3, 21.6, 14.5. The trivial 100% matches are deleted]

Result of the C13RR system

In search for: unknown
Tolerance: 15.00 ppm (no missing peaks allowed)

No.	SI	Cas reg. nr.	Mol. formula/name
1.	38.2	1195795	C10 H16 01
	Bicyclo 2.2.1 heptan-2-one, 1,3,3-trimethyl-		
2.	8.5	76222	C10 H16 01
	Bicyclo 2.2.1 heptan-2-one, 1,7,7-trimethyl-		
3.	7.1	51100006	C10 H16 01
	Bicyclo 2.2.1 heptan-2-one 3,3,6-trimethyl-, *ENDO-*		
4.	5.3	56764320	C10 H16 01
	Bicyclo 3.2.1 Octan-6-one, 7,7-dimethyl-		
5.	4.2	4613461	C10 H16 01
	Bicyclo 2.2.1 heptan-2-one, 3.3.6-trimethyl, *EXO-*		

Result of the CIS/Clerc system

No.	Fit	Cas reg. nr.	Mol. formula/name
1.	97.29	41781-62-8	C16 H35 N1
	1-Hexanamine, *N*-butyl-*N*-(1,3-dimethylbutyl)-		
2.	97.12	20352-67-4	C8 H19 N1
	1-Hexanamine, *N*-ethyl-		
3.	96.85	4178-39-9	C12 H27 N1
	1-Hexanamine, *N*-(1,3-dimethylbutyl)-		
4.	96.68	1195-79-5	C10 H16 O1
	Bicyclo[2.2.1]heptan-2-one, 1,3,3-trimethyl-		
5.	96.24	123-82-0	C7 H17 N1
	2-Heptanamine		

system. As an illustration, the retrieval results obtained with both systems for an "unknown", which was one of two alternative library spectra for 1,3,3-trimethylbicyclo-2,2,1-heptan-2-one, are given in Table 2. It should be noted that the CIS/Clerc system includes the option of specifying multiplicities, in which case better results may be obtained.

Conclusions

An effective library search system should include a similarity index with a high discriminating power in the absolute as well as in the relative sense. The reproducibility based similarity index of the retrieval system for ^{13}C-n.m.r. spectra, reported in this paper, meets this requirement. The same concept of the similarity index has been used successfully in the development of a library search system for mass spectra. On the basis of the C13RR system, a microcomputer version (in UCSD Pascal) has also been developed. Results of this MC-C13RR system, as well as of the mass spectral reproducibility-based retrieval (MSRR) system, will be reported later.

The situation of "missing peaks" in unknown or reference spectrum causes a problem, which is dealt with in the present version of the library system by a simple modification of the search algorithm. Research aimed at further optimization of the retrieval systems, including a more adequate handling of missing peaks, is in progress.

The authors thank Dr M. J. A. de Bie for valuable discussions.

Appendix

Development of the reproducibility function. To rewrite Eqn. 6, the following function is introduced

$$p_0([\Delta q_i]/\sigma^2) = \int_{\Delta q^s=-\infty}^{\infty} p(\Delta q^s) \prod_{i=1}^{n} p(\Delta q_i^r)\, d\Delta q^s \qquad (A1)$$

This function is a normal probability function with non-zero correlations and can alternatively be expressed by

$$p_0([\Delta q_i]/\sigma^2) = A \exp(-0.5\, \chi^2) \qquad (A2)$$

where $A = (n+1)^{-1/2}\, (2\pi\sigma^2)^{-1/2\,n}$ and

$$\chi^2 = \sum_{i=1}^{n} (\Delta q_i - \overline{\Delta q^s})\Delta q_i/\sigma^2 \qquad (A3)$$

with

$$\overline{\Delta q^s} = \sum_{i=1}^{n} \Delta q_i/(n+1) \qquad (A4)$$

Here, χ^2 is a quantity distributed as chi-squared with n degrees of freedom. The quantity χ^2 can also be written as a function of S^2, being the estimator of σ^2

$$\chi^2 = n\, S^2/\sigma^2 \qquad (A5)$$

where

$$S^2 = \sum_{i=1}^{n+1} (\Delta q_i - \overline{\Delta q^s})^2/n \qquad \text{(A6)}$$

with $\Delta q_{n+1} = 0$. The $(n + 1)^{th}$ (artificial) difference in shift value can be considered as originating from the TMS peak in the reference and unknown spectra, and is, by definition, equal to zero. It may be noted that $\overline{\Delta q^s}$ can now also be expressed as a normal average value (cf. Eqn. A4): $\overline{\Delta q^s} = \Sigma_{i=1}^{n+1} \Delta q_i/(n + 1)$.

From Eqns. A1 and A2, the reproducibility function of Eqn. 6 can be rewritten as

$$p_0[\Delta q_i] = \int_{\sigma^2=0}^{\infty} p(\sigma^2)\, A \exp(-0.5\, \chi_n^2)\, d\sigma^2 \qquad \text{(A7)}$$

According to Eqn. 4, the reproducibility function is characterised by the empirical parameters $E \ln \sigma^2$ and $V \ln \sigma^2$. From Eqn. A5 it follows that $\ln \sigma^2 = \ln S^2 - \ln(\chi^2/n)$. The expected value of $\ln(\chi^2/n)$ can be approximated [26] by $(-1/n - 1/3n^2)$. This implies that $E \ln \sigma^2$ can be estimated by averaging the quantity $\ln S^2 + 1/n + 1/3n^2$ for all difference spectra of the test set. If $E\sigma^2$ is defined as the expected value of σ^2, then from the assumption that σ^2 is lognormally distributed, it can be derived that

$$E\sigma^2 = \exp[E \ln \sigma^2 + 1/2\, V \ln \sigma^2] \qquad \text{(A8)}$$

given that

$$V \ln \sigma^2 = 2[\ln(E\sigma^2) - E \ln \sigma^2] \qquad \text{(A9)}$$

The value of $E\sigma^2$ can be approximated [25] by averaging S^2 for all difference spectra. The value of $V \ln \sigma^2$ can be found by substitution of the calculated values for $E\sigma^2$ and $E \ln \sigma^2$ in Eqn. A9.

Development of the similarity index. Elaboration of Eqn. 7 leads to

$$S_I = \int_{\sigma^2=0}^{\infty} p(\sigma^2)\, F(\sigma^2)\, d\sigma^2 \qquad \text{(A10)}$$

with

$$F(\sigma^2) = \int_{\chi^2 < X^2} \ldots \int p_0([\Delta q_i] \mid \sigma^2)\, d\Delta q_1 \ldots d\Delta q_n \qquad \text{(A11)}$$

where X^2 is the value of χ^2 (see Eqn. A3), calculated from the observed values of the difference quantities. This means that X^2 can be written as

$$X^2 = K/\sigma^2 \qquad \text{(A12)}$$

where $K = \Sigma_{i=1}^{n} (\Delta Q_i - \overline{\Delta Q^s})\, \Delta Q_i$ with $\overline{\Delta Q^s} = \Sigma_{i=1}^{n} \Delta Q_i/(n + 1)$.
Because χ^2 is distributed as chi-squared with n degrees of freedom, it follows from Eqn. A11 that

$$F(\sigma^2) = 1 - C(X^2) \qquad \text{(A13)}$$

where C is the (cumulative) chi-squared distribution function with n degrees of freedom. Replacing, in Eqn. A10, the integration over σ^2 by an integration over $\ln \sigma^2$, and using Eqns. 4, A12 and A13 yield the following expression for S_I:

$$S_I = \int_{-\infty}^{+\infty} p(\ln \sigma^2)\, F'(\ln \sigma^2)\, d \ln \sigma^2 \qquad \text{(A14)}$$

where $p(\ln \sigma^2) = N(\ln \sigma^2 : E \ln \sigma^2, V \ln \sigma^2)$ (A15)

and $F'(\ln \sigma^2) = 1 - C(K/e^{\ln \sigma^2})$ (A16)

This expression of S_I is the basis of the algorithm for the calculation of S_I values. The integration over $\ln \sigma^2$ (see Fig. 13) is done by numerical integration, approximating the functions $F'(\ln \sigma^2)$ and $p(\ln \sigma^2)$ by a fixed number of straight lines.

REFERENCES

1 W. Voelter, G. Haas and E. Breitmaier, Chem. Ztg., 97 (1973) 507.
2 P. R. Naegeli and J. T. Clerc, Anal. Chem., 46 (1974) 739A.
3 B. A. Jetzl and D. L. Dalrymple, Anal. Chem., 47 (1975) 203.
4 W. Bremser, M. Klier and E. Meyer, Org. Magn. Reson., 7 (1975) 97.
5 R. Schwarzenbach, J. Meili, H. Koenitzer and J. T. Clerc, Org. Magn. Reson., 8 (1976) 11.
6 J. Zupan, M. Penca, D. Hadzj and J. Marsel, Anal. Chem., 49 (1977) 2141.
7 D. L. Dalrymple, C. L. Wilkins, G. W. A. Milne and S. R. Heller, Org. Magn. Reson., 11 (1978) 535.
8 J. E. Dubois and J. C. Bonnet, Anal. Chim. Acta, 112 (1979) 245.
9 V. Mlynárik, M. Vida and V. Kellö, Anal. Chim. Acta, 122 (1980) 47.
10 W. Bremser, H. Wagner and B. Franke, Org. Magn. Reson., 15 (1981) 178.
11 M. Zippel, J. Mowitz, I. Kohler and H. J. Opferkuch, Anal. Chim. Acta, 140 (1982) 123.
12 A. P. Uthman, J. P. Koontz, J. Hinderliter-Smith, W. S. Woodward and C. N. Reilley, Anal. Chem., 54 (1982) 1772.
13 J. Kwiatkowski and W. Riepe, Anal. Chim. Acta, 135 (1982) 293.
14 P. Cleij, H. A. van 't Klooster and J. C. van Houwelingen, Anal. Chim. Acta, 150 (1983) 23.
15 E. Horowitz and S. Sahni, Fundamentals of Data Structures, Pitman Press, London, 1976.
16 L. F. Johnson and W. C. Jankowski, Carbon-13 NMR Spectra, Wiley-Interscience, New York, 1972.
17 F. I. Carroll, C. G. Moreland, G. A. Brine and J. A. Kepler, J. Org. Chem., 41 (1976) 996.
18 Y. Terui, K. Tori, S. Meada and Y. K. Sawa, Tetrahedron Lett., (1975) 2853.
19 C. J. Kelley, R. C. Harruff and M. Carmack, J. Org. Chem., 41 (1976) 449.
20 D. A. Forsyth, R. J. Spear and G. A. Olah, J. Am. Chem. Soc., 98 (1976) 2512.
21 H. Fritz and T. Winkler, Helv. Chim. Acta, 59 (1976) 903.
22 F. W. McLafferty, Anal. Chem., 49 (1977) 1442.
23 E. G. Smith, The Wiswesser Line—Formula Chemical Notation, McGraw-Hill, New York, 1968.
24 S. L. Grotch, Anal. Chem., 43 (1971) 1362.
25 M. Abramowitz and I. A. Segun (Eds.), Handbook of Mathematical Functions, Dover Publications, New York, 1968.

Analytica Chimica Acta, 157 (1984) 245—260
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

FEATURE EXTRACTION FROM SPECTRAL AND OTHER DATA BY THE PRINCIPAL COMPONENTS AND DISCRIMINANT FUNCTION TECHNIQUES

J. ARUNACHALAM and S. GANGADHARAN*

Analytical Chemistry Division, Bhabha Atomic Research Centre, Trombay, Bombay 400 085 (India)

(Received 15th July 1983)

SUMMARY

Principal component analysis is applied to the interpretation of ^{13}C-n.m.r. spectra and to the resolution of mass spectral data. A procedure is given for determining the relative amounts of pure components, with and without the use of pure mass lines, in mass spectra of mixtures. The use of the Fisher discriminant method in combination with the principal components technique is demonstrated in the treatment of trace element data on hair for environmental purposes. The importance of feature generation and selection is emphasized.

Advances in multi-element determinations and developments in the application of mathematical methods for the treatment of multi-variate data have enabled valuable information to be extracted from analytical data, that would not have been readily accessible otherwise. Pattern recognition approaches have a significant role in the extraction of information from measurements. The various approaches and their application to chemical problems have been periodically reviewed [1, 2] and discussed in monographs [3, 4].

Feature extraction

The most important step in the pattern recognition approach is feature extraction/selection which generates an optimum number of variables. The term "feature" refers to a combination of measurements describing a certain pattern aspect, relatively invariant within one class, but assuming different values outside this class. Unlike classification, which may be based on a linkage approach or a threshold logic, feature extraction/selection is highly problem-dependent; thus, the choice of initial variables measured and the preprocessing of the data set assume importance. Primarily, feature-extraction techniques aim at two objectives: first, dimensionality reduction as in principal component analysis (Karhunen—Loeve transformation, factor analysis) or non-linear mapping (NLM) and, secondly, class separation as in discriminant function techniques, e.g., the Fisher discriminant and linear-

0003-2670/84/$03.00

learning machine (LLM) procedures. Practical problems require a judicious combination of both dimensionality reduction and class separation, as this should offer a better interpretation of the data set than any of these techniques in isolation.

The SIMCA method proposed by Wold [5] can be considered as an intra-class feature extraction procedure; it uses a supervised learning approach to classification.

In many of the publications on feature extraction, the emphasis has been primarily on the dominant features; for example, in the principal components method, two components (accounting for the largest fraction of the total variance) only are used to obtain the eigenvector plot, the other components being generally ignored. However, experience in dealing with different data sets has demonstrated the desirability of using all available information on the samples and to examine the less dominant features as well.

This paper reports on the application of the principal components and Fisher discriminant methods for feature extraction. The data sets processed include mass spectral data on mixtures of hydrocarbons reported by Ritter et al. [6], ^{13}C-n.m.r. spectra (chemical shift values) of 2-substituted norbornanes reported by Grutzner et al. [7], and trace element data for human-head hair samples, generated in this laboratory as part of an IAEA research program on methods of monitoring environmental exposure. The required computer algorithms were developed on the BESM-6 and PDP-11/34 computer systems at the centre.

PRINCIPAL COMPONENTS ANALYSIS

Principal components analysis (p.c.a.) is widely used. The object of factor analysis is to describe the variability in a multivariate data set by using abstract causal variables (factors) which are derived from the observed variables. The covariance matrix of the original data set is taken to represent the dispersion of the data. The new variables are obtained through the matrix rank analysis of the covariance matrix, which extracts the eigenvalues and corresponding eigenvectors. If the variables had been autoscaled, the correlation matrix is used for the assessment. The number of factors required to describe the data is given by the number of eigenvalues significantly different from zero. Some authors have proposed empirical functions to identify the number of significant eigenvalues [8, 9]. Usually, the number of factors derived is much less than the original number of observed variables and provides a parsimonious description of the data. Whenever possible, a lower-dimensional representation of the data (2 or 3 dimensions) can be obtained of the multi-dimensional data, affording visual examination. Detailed procedures for p.c.a. have been described [10—12].

Associations between the variables can be discerned from the plots of the eigenvector coefficients; if the variance/covariance matrix is used, the highly (positively) correlated variables form a set of vectors collinear from the

origin, with their distances from the origin proportional to their variances [13]; such correlated variables should form clusters if the correlation matrix is used.

The communalities of the variables in the derived factors provide information as to how many factors are required to account for their variability. Communality of a variable in a factor is the fraction of its total variance accounted for by that factor. Because p.c.a. should account for 100% of the total variance, the number of components required for each variable can be easily identified.

The desirability of preprocessing the data prior to p.c.a. is still debated. Autonormalization gives equal weighting to all variables, and results in the use of a correlation matrix for the p.c.a. Another consideration is the inclusion of outliers in the data analysis. The sample may be an outlier with respect to only one variable, but its inclusion will disturb the eigenvector analysis [14]. While it might be expedient to exclude such samples, sampling restrictions, or the paucity of samples, may require their use in data analysis; one way of obviating this problem is to replace the outlier values with the mean of the rest of the samples for that variable. This avoids giving undue weighting to the given variable, while allowing the inclusion of other variables measured on such samples. Replacing the outliers, or the missing observations, with zero, grossly distorts the variance measure. Some authors have preferred to use non-linear (e.g., logarithmic) transformations but this requires the assumption of some monotonic relationship between the variables, which may not obtain in the actual data.

Principal components analysis of ^{13}C-n.m.r. data

Several theoretical treatments of ^{13}C-n.m.r. chemical shifts have been described [15]. Experimentally, the chemical shifts are influenced by features such as hybridization of the carbon atom, direct substitution by nuclei with electron-withdrawing characteristics and steric effects. The extent to which these factors influence chemical shifts should offer a better insight into the data.

Grutzner et al. [7] reported ^{13}C-n.m.r. chemical shift values for the seven carbon atoms of norbornane and its fourteen derivatives (7 *exo* and 7 *endo* 2-substituted norbornanes). The substituents were CH_3, NH_2, OH, CN, COOH, $COOCH_3$ and CH_2OH. A p.c.a. was done on these (15×7) chemical shift values (the chemical shifts for *exo*-2-fluoronorbornane [7] were not included, as values for the corresponding *endo* compound were not reported).

The p.c.a. showed that the variability in the data set could be mostly accounted for by three components. Table 1 gives the eigenvector coefficients, communalities and the percentage of variance accounted for by the three components. The communalities show that component 1 is mostly influenced by C-2, the α-carbon relative to the substituent, component 2 is influenced by C-1 and C-3 (the β-carbons), and component 3 is influenced by C-4, C-6 and C-7 (the γ-carbons). Examination of the percentage of variance

TABLE 1

Results of p.c.a. for the ^{13}C-n.m.r. data on 2-substituted norbornanes

Carbon No.	Eigenvector coefficients			Communalities			Percentage of variance accounted for in		
	Comp 1	Comp 2	Comp 3	Comp 1	Comp 2	Comp 3	Comp 1	Comp 2	Comp 3
2	0.975	−0.205	0.063	185.152	0.659	0.045	99.62	0.36	0.02
1	0.101	0.418	0.072	1.997	2.734	0.058	34.64	47.42	1.00
3	0.156	0.844	−0.025	4.743	11.159	0.007	29.56	69.54	0.04
4	−0.007	0.016	−0.167	0.009	0.004	0.313	1.92	0.82	67.15
6	−0.109	−0.130	0.841	2.318	0.264	7.919	21.98	2.50	75.10
7	−0.051	−0.233	−0.495	0.503	0.855	2.749	11.62	19.74	63.52
5	−0.005	0.016	−0.102	0.004	0.0	0.118	0.93	0.00	24.50

for each variable accounted for in component 1 suggests that this component may be an "inductive factor". Component 1 accounts for almost all the variance of shift values in C-2; 34% and 29%, respectively, of the variance in the shift values for C-1 and C-3 (the β-carbons) and 22% and 12%, respectively, of the variances in the shift values for C-6 and C-7 (the γ-carbons). This pattern is very similar to that of an inductive effect which diminishes with increasing distance from the position of substitution.

Table 2 gives the component scores on each compound. Component 3 shows excellent discrimination between the *endo* and *exo* forms. Irrespective of the substituent in the C-2 position, the *exo* and *endo* forms could be discriminated by component 3, (which itself accounts for only 5% of the total variance), which suggests that this component is a "structural factor". From the communality values of this factor, the orientation of substituents seems to have a major influence on the chemical shift values of the γ-carbons.

Stothers [15] has commented that it has not been possible to decide whether the effect of orientation is electronic in origin or is caused by slight differences in molecular geometry, because the *endo* group may introduce

TABLE 2

Principal components of ^{13}C-n.m.r. data for the 2-substituted norbornanes tested

Substituent	*Exo* substitution			*Endo* substitution		
	Comp 1	Comp 2	Comp 3	Comp 1	Comp 2	Comp 3
—	17.7	5.1	−1.8	—	—	—
CH_3	8.6	−5.9	−3.2	10.3	−6.2	4.7
NH_2	−10.4	−5.4	−3.5	−8.4	−3.0	4.9
OH	−29.0	−1.1	−2.8	−26.9	1.6	3.2
CN	14.9	−2.9	−1.6	15.8	−1.6	2.1
COOH	0.03	2.3	−3.2	0.7	4.7	2.0
$COOCH_3$	0.2	2.1	−3.1	0.9	4.5	1.9
CH_2OH	1.8	2.8	−3.8	3.7	3.0	4.2
% Variance accounted for	87.1	7.0	5.0			

additional steric interaction. The comparison of communality values in component 1 and component 3 indicates that electronic nature has less influence than the steric factor for the γ-carbons. The scores in component 3 (Table 2) indicate that norbornane (unsubstituted) can be grouped with the *exo* substituents; thus the *endo* groups appear to introduce additional steric interactions.

It is also obvious that the variances for C-4 and C-5 are not fully accounted for by the first three components. This emphasizes the need to examine minor components in detail; they may account for only a small fraction of the total variance of the data but for a very large fraction of the variances for individual variables.

Principal components analysis of mass spectral data

Qualitative and quantitative aspects of mass spectral studies involve: (a) identification of precursors that contribute to a given mass line, and the extent of their contributions; (b) in spectra of mixtures, identification of pure mass lines corresponding to interference-free fragments obtained from each constituent; (c) detection of any contributions from impurities in the system studied. The use of factor analysis of mass spectral data for such aspects has been studied by several authors. Ritter et al. [6] used the mass spectra of mixtures of hydrocarbons to demonstrate that the correct number of components could be identified by factor analysis. Knorr and Futrell [16] proposed methods involving non-orthogonal rotation of the factors to resolve the spectra into contributions from individual components. Their method required that each component in a mixture should have at least one pure mass line.

Here, p.c.a. is used to study associations between mass lines, and resolution of mixtures. The data on mixtures of hydrocarbons reported by Ritter et al. were used. The method of resolution proposed here does not require pure mass lines for each component, although they may be convenient; rather, the associations among the individual mass lines of the components are used for resolution of the mixtures, with and without the use of pure mass lines. It is shown that the eigenvector coefficients and communalities not only offer an elegant method of obtaining information regarding the pure mass lines of individual components in the mixtures, but also seem to provide information regarding fragmentation patterns.

The mass spectral data can be represented in two ways: the spectra as points (vectors) in the mass line space, or the mass lines as points (vectors) in the spectral space. The principal components procedure was applied to both these types of representations. Because the mass lines of fragments obtained from a single compound are highly correlated, it is often thought that their inclusion in the data set merely increases the dimensionality of the data; however, their inclusion helps to obtain a statistically better representation of the spectra. Moreover, there is normally no prior information available regarding the individual mass lines of each component, especially in

studies of mixtures. Hence, almost all the mass line intensities reported by Ritter et al. [6] are used here.

Mixtures of cyclohexane and cyclohexene. The data set on mixtures of cyclohexane and cyclohexene consisted of 20 mass intensities recorded for four samples (80%, 60%, 40% and 20% (v/v) cyclohexane). Initially, p.c.a. was done for all the 20 mass lines; five mass lines were found to contribute negligibly to the total variance and so were excluded from further processing. The resulting (4 × 15) data matrix of mass intensities was studied by using the covariance as well as the correlation matrix. The eigenvector coefficients obtained based on the correlation matrix and the communalities are given in Table 3. It is seen that two components are required to account for the total variance.

The plot of the eigenvector coefficients based on the covariance matrix was used to study the associations between the 15 mass lines (Fig. 1). It can be seen that mass lines 67, 54, 82, 81 and 79 are collinear from the origin and are contributed by cyclohexene. Similarly, mass lines 84, (56), 55, 69 and 43 are collinear from the origin. Hence, these two sets of line intensities can be identified as pure mass lines, i.e., contributed solely by cyclohexene and cyclohexane, respectively. It can also be inferred that the 41, 39, 27 and 28 lines are contributed to by both cyclohexane and cyclohexene. Figure 1 also indicates how the *m/z* 41 can be resolved into contributions from cyclohexane and cyclohexene, based on the projections of the vector $\overrightarrow{0, 41}$ onto the two vectors formed by the two sets of pure lines.

TABLE 3

Principal components analysis of cyclohexane/cyclohexene data set (autoscaled data matrix)

Mass	Comp 1	Comp 2	Communalities		
			Factor 1 (h_1)	Factor 2 (h_2)	$h_1 + h_2$
27	0.119	0.397	0.134	0.863	0.997
28	0.276	0.218	0.722	0.260	0.982
39	−0.044	0.423	0.018	0.980	0.998
41	0.230	0.302	0.501	0.499	1.000
42	0.313	0.114	0.928	0.071	0.999
43	0.313	0.114	0.928	0.071	0.999
54	−0.224	0.309	0.476	0.523	0.999
55	0.314	0.108	0.934	0.063	0.997
56	0.319	0.077	0.964	0.032	0.996
67	−0.241	0.286	0.550	0.448	0.998
69	0.319	0.083	0.964	0.038	1.000
79	−0.210	0.326	0.418	0.582	1.000
81	−0.233	0.296	0.514	0.480	0.994
82	−0.226	0.307	0.484	0.516	1.000
84	0.315	0.094	0.940	0.048	0.988

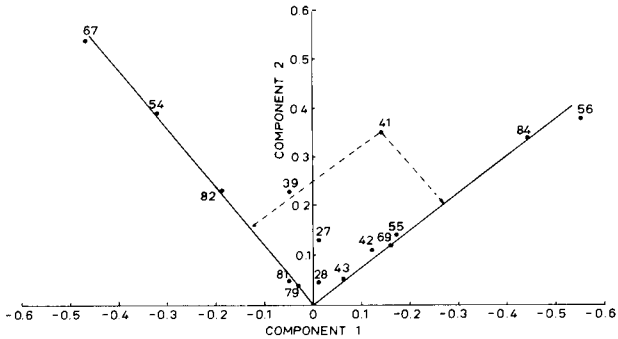
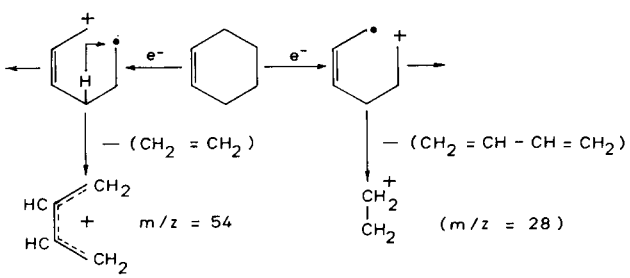


Fig. 1. Eigenvector coefficients of the cyclohexane/cyclohexene data (based on variance/covariance matrix).

Examination of the communalities in Table 3 shows that the pure lines corresponding to cyclohexane have a predominant contribution from only one factor. (The communalities indicate the number of components required to account for the variance for each variable; the variance of each variable is equal to unity in this case, as the data matrix has been autoscaled.) In the case of cyclohexene, the nearly equal communalities in the two factors imply equal contributions from two sources for lines 67, 54, 82, 81 and 79. However, the eigenvector coefficients of these mass lines indicate a single source. The nearly 50% contribution to each factor suggests that the molecular ion 82 of cyclohexane should have two modes of formation. This information is in conformity with the nature of cyclohexene, which is known [17] to undergo retro-Diels—Alder reaction; on primary ionization of the double bond, the cyclohexene molecule undergoes allylic cleavage



Thus, the eigenvector coefficients and communalities seem to provide information on the mechanism of fragmentation pattern as well.

Mixtures of cyclohexane and n-hexane. The data set on cyclohexane/n-hexane mixtures consisted of 18 mass intensities recorded for seven samples (pure cyclohexane, pure n-hexane and 5 mixtures with 90, 80, 50, 20 and 10% (v/v) cyclohexane). Three mass lines were found to contribute negligibly to the total variance and were excluded from further processing. The resulting (7 × 15) data matrix of line intensities was studied with both

covariance and correlation matrices. The eigenvector coefficients obtained based on the correlation matrix, along with the communalities for each variable (mass line) are given in Table 4.

When mass line 28 is included in the processing, three factors are required to account for the total variance of the data set. The communalities indicate that line 28 has contributions from three factors, whereas the other mass lines have contributions from at most two factors. Hence, the third factor indicates the presence of an impurity, contributing mainly to *m/z* 28. Ritter et al. suspected nitrogen to be another source of *m/z* 28, besides cyclohexane and cyclohexene.

The exclusion of *m/z* 28 from p.c.a. indicated that the variability in the mass spectra could be accounted for by only two factors. The eigenvector coefficients along with the communalities obtained based on the covariance matrix are given in Table 5. A plot of the eigenvector coefficients is given in Fig. 2. The collinear masses 57 and 86 arise from n-hexane; masses 84, 85 and 54 arising from cyclohexane form another set of collinear vectors.

Resolution of mass spectra of mixtures based on pure mass lines. If pure mass lines for each constituent in a mixture can be identified, resolution of the mixtures is relatively straightforward. The intensities of these lines in the spectra of pure constituents, recorded under the same conditions as that of the mixtures, can be used to obtain the relative contribution from each constituent. Alternatively, if a mass line is known to have contributions from

TABLE 4

Principal components analysis of cyclohexane/n-hexane data based on the correlation matrix

Mass	Components			Communalities			
	1	2	3	Comp 1 (h_1)	Comp 2 (h_2)	Comp 3 (h_3)	$h_1 + h_2$
27	0.249	0.341	0.008	0.193	0.193	—	0.953
28	0.068	0.427	−0.844	0.057	0.303	0.634	0.360
29	0.283	0.080	0.009	0.981	0.010	—	0.991
39	−0.198	0.450	0.405	0.480	0.337	0.146	0.817
41	0.221	0.461	0.197	0.598	0.354	0.034	0.952
42	0.262	0.269	0.172	0.841	0.120	0.026	0.961
43	0.282	0.101	0.101	0.974	0.017	0.009	0.991
54	−0.282	0.101	0.101	0.974	0.017	0.009	0.991
55	−0.272	0.201	−0.026	0.906	0.067	0.0006	0.973
56	−0.262	0.296	0.103	0.841	0.146	0.001	0.987
57	0.284	0.040	0.077	0.988	0.003	0.0007	0.991
69	−0.281	0.119	−0.007	0.967	0.024	—	0.991
84	−0.280	0.146	0.071	0.960	0.035	0.0006	0.995
85	−0.278	0.141	−0.131	0.947	0.033	0.002	0.980
86	0.283	0.052	0.038	0.981	0.004	—	0.985
Eigenvalues	12.253	1.664	0.890				
% Variance	81.68	11.10	5.93				

TABLE 5

Principal components analysis of cyclohexane/n-hexane data based on the variance/covariance matrix (excluding m/z 28)

Mass	Components		Communalities		$h_1 + h_2$	Variance	$(h_1 + h_2)$/Variance
	1	2	Comp 1 (h_1)	Comp 2 (h_2)			
27	0.055	0.136	0.168	0.047	0.215	0.241	0.892
29	0.201	0.142	2.25	0.052	2.302	2.325	0.990
39	−0.030	0.125	0.050	0.040	0.090	0.101	0.891
41	0.109	0.448	0.662	0.514	1.176	1.205	0.976
42	0.073	0.162	0.297	0.067	0.364	0.366	0.994
43	0.356	0.339	7.057	0.294	7.351	7.363	0.998
54	−0.033	0.014	0.061	0.0005	0.0615	0.063	0.976
55	−0.194	0.186	2.096	0.089	2.185	2.222	0.979
56	−0.309	0.547	5.317	0.767	6.084	6.103	0.997
57	0.593	0.326	19.582	0.272	19.854	19.895	0.998
69	−0.183	0.089	1.865	0.020	1.885	1.903	0.991
84	−0.517	0.383	14.88	0.376	15.256	15.271	0.999
85	−0.038	0.016	0.080	0.0006	0.0806	0.0857	0.940
86	0.152	0.089	1.287	0.020	1.307	1.313	0.995
Eigenvalues	55.686	2.562					

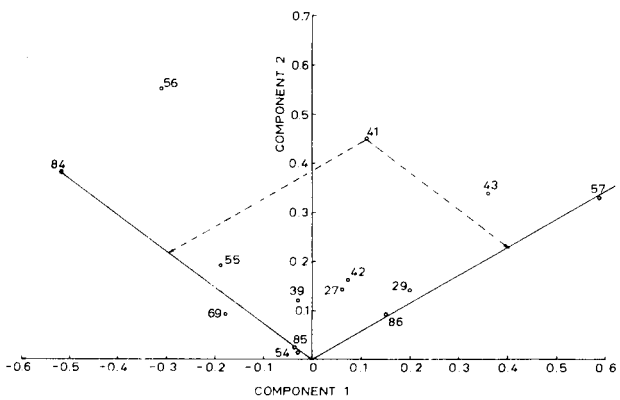


Fig. 2. Eigenvector coefficients of the cyclohexane/n-hexane data (based on variance/covariance matrix).

each constituent, it can be resolved to obtain the relative contributions of the constituents based on the intensities recorded for the pure mass lines in the same spectrum. The latter situation is of practical interest. For example, in the cyclohexane/n-hexane spectra, the line at m/z 41 has contributions from each component. The collinear vectors of the pure mass lines corresponding to cyclohexane and n-hexane in Fig. 2, form an oblique coordination system, and the projections of the vector $\overrightarrow{0, 41}$ onto the oblique axes

$\overrightarrow{0, 57}$ and $\overrightarrow{0, 84}$ would yield the relative contributions of n-hexane and cyclohexane to mass line 41.

The procedure for obtaining the projections in an oblique axes system is given in Fig. 3. In this diagram, C is a point the coordinates of which are known in the orthogonal system OX and OY; Θ ($\neq 90°$) is the angle between the two nonorthogonal vectors, OA and OB. Point C has to be expressed in terms of the two oblique axes. Lines CA_1 and CB_1 are drawn parallel to OB and OA, respectively, and CA_2 and CB_2 are drawn perpendicular to OA and OB from point C. If the lengths OA_2 and OB_2 are S_1 and S_2, and OA_1 and OB_1 are d_1 and d_2, respectively, then

$$S_1 = d_1 + d_2 \cos\Theta \text{ and } S_2 = d_2 + d_1 \cos\Theta$$

As d_1 and d_2 are the coordinates of C in the oblique axes system, they can be easily computed if the coordinates of points A, B and C are known in the orthogonal coordinate system formed by OX and OY. From p.c.a., the orthogonal coordinates are obtained as the eigenvector coefficients which can be used to derive the new nonorthogonal coordinate reference frame, in this case, the axes formed by the pure mass lines corresponding to cyclohexane and n-hexane.

This procedure was used to resolve the mass line 41 relative to lines 57 (n-hexane) and 84 (cyclohexane); the values are given in Table 6. These values are proportional [18] to the mole fractions of n-hexane and cyclohexane. The relative concentrations and so the mole fractions of the constituents in the different mixtures are readily obtained by using the values for the two pure constituents. It should be noted that though line 57 was chosen to represent n-hexane, cyclohexane also makes a small, but definite contribution to this line. The relative concentration ratios and the computed mole fractions are given in Table 7, after correction for the contribution to mass 57 by cyclohexane, in each spectrum. The mole fraction of cyclohexane (originally) taken by Ritter et al. is also given in this Table. The computed mole fraction values are in good agreement, except for spectrum 5.

TABLE 6

Resolution of mass line 41 based on lines 57 and 84

Spectrum No.	Reported intensity	Resolved values based on		Sum
		57	84	
1	7.1	0.8	6.2	7.0
2	6.4	1.6	4.8	6.4
3	8.0	2.2	5.0	7.2
4	9.3	5.0	4.2	9.2
5	9.1	6.6	2.1	8.7
6	8.8	7.6	1.2	8.8
7	8.8	8.4	0.06	8.46

TABLE 7

Resolution of the spectra of cyclohexane/n-hexane mixtures based on lines 57 and 84

Spectrum No.	Resolved values normalized to		Computed mole fraction (cyclohexane)	Reported mole fraction
	100% cyclohexane	100% n-hexane		
2	0.764	0.116	0.87	0.92
3	0.801	0.184	0.81	0.83
4	0.676	0.532	0.56	0.55
5	0.329	0.755	0.30	0.23
6	0.178	0.884	0.17	0.12

Resolution of mass spectra of mixtures without pure mass lines. The procedure described above for resolution of mass spectra can be applied even if there are no pure mass lines in the spectrum. This analysis showed that both cyclohexane and n-hexane contribute to lines 27, 29, 39, 41, 42, 43, 55, 56. The p.c.a. was applied by using these mass lines in the seven spectra. The contributions to each line in the spectra of the mixtures were computed relative to lines 29 and 55. Table 8 gives the resolved values for each spectra, the relative concentrations and the computed mole fractions for cyclohexane. These values agree very well with those obtained by using the pure mass lines. This demonstrates the salient feature of the procedure reported here: it does not require pure mass lines for each constituent in the spectrum of a mixture, as has been observed by Sedgwick [19]. However, to compute the mole fractions, either a well characterized spectrum of a pure constituent or the spectrum of a known mixture is necessary.

Resolution of mass spectra of mixtures based on spectra of pure constituents. The 15 mass lines in the seven spectra can be viewed as vectors in 7-dimensional space; the mass spectra are treated as variables. The (15 × 7) matrix of mass line intensities was used for p.c.a.; the results are given in Table 9. It can be seen that three factors are needed to account for the total variability in the data set. The removal of line 28 from the treatment showed that the variability in the data set could be accounted for by only two factors.

TABLE 8

Resolution of mass spectra of cyclohexane/n-hexane mixtures based on lines 29 and 55

Spectrum No.	Resolved values normalized to		Computed mole fraction (cyclohexane)
	100% cyclohexane	100% n-hexane	
2	0.786	0.094	0.89
3	0.903	0.162	0.85
4	0.667	0.545	0.55
5	0.371	0.859	0.30
6	0.082	0.890	0.08

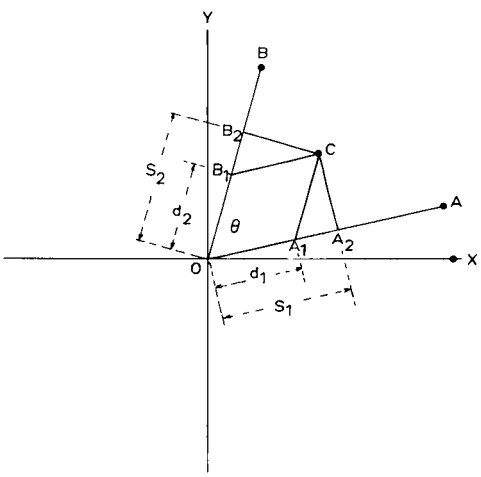
TABLE 9

Results of p.c.a. for cyclohexane/n-hexane data (mass spectra as variables)

Spectrum No.	Eigenvector coefficients			Percentage of variance accounted for in		
	Comp 1	Comp 2	Comp 3	Comp 1	Comp 2	Comp 3
1	0.390	−0.494	−0.115	60.9	38.9	0.1
2	0.344	−0.332	−0.135	72.7	26.9	0.2
3	0.385	−0.303	0.318	79.2	19.5	1.1
4	0.441	−0.032	−0.388	98.2	0.2	1.6
5	0.374	0.266	0.779	77.5	15.6	6.9
6	0.364	0.406	−0.132	66.7	33.0	0.2
7	0.338	0.563	−0.305	47.0	52.1	0.8
Eigenvalues	59.7	23.7	1.2			

These observations are in agreement with the inferences drawn from the earlier representation of mass lines space. An examination of Table 9 also shows that the impurity contributing mainly to line 28, affects spectra 3, 4 and 5, spectrum 5 being the most affected.

The plot of the eigenvector coefficients obtained using the (14 × 7) data matrix (after excluding *m/z* 28), is given in Fig. 4. Points 1 and 7 represent the spectra of 100% cyclohexane and 100% n-hexane, respectively; points 2—6 represent the spectra of mixtures of these two constituents, corresponding to 90, 80, 50, 20 and 10% (v/v) of cyclohexane. It can be seen that the (nonorthogonal) projections of points 2—6 onto the vectors 1 and 7 would yield the relative contributions. The resolved values (coordinates), the relative



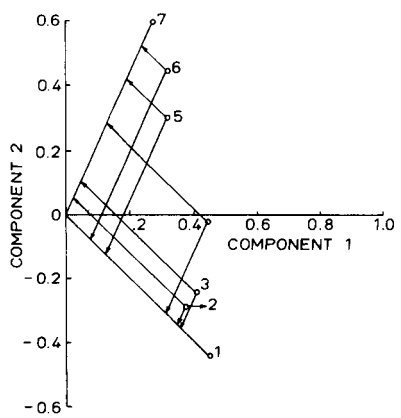


Fig. 3. The pure component projections in an oblique coordinate system.

Fig. 4. Eigenvector coefficients of mass spectra in mass line space.

concentration and the calculated mole fraction of cyclohexane in spectra 2—6 are given in Table 10. From the sequence of projections of these points, it appears that spectrum 2 was recorded under conditions slightly different from those for the other spectra.

Tables 7, 8 and 10 show that the resolved values for the mole fractions of cyclohexane and n-hexane are in good agreement with the actual values taken, except for spectrum 5.

The procedure described above for two-component resolution can be readily extended to multicomponent spectra and the resolved values, R, can be obtained by the relation, $R = X\ \Phi'^{-1}$, where X is the vector of the orthogonal eigenvector coefficients with respect to the orthogonal space formed by the eigenvectors and Φ' is the transpose of the matrix of direction cosines of the base vectors.

This resolution procedure can also be applied to other types of spectral data, where digital intensities are available. The eigenvector coefficients and communalities offer a powerful approach to studies of the association among the variables and to evaluating the number of factors required to account for each variable. For example, in g.c.-m.s., the mass fragments of co-eluting fractions might be checked for the number of components.

DISCRIMINANT FUNCTION METHOD

Discriminant functions are used to derive new variables which are linear or non-linear combinations of the original variables, and are more helpful in discriminating between "individuals" from different groups. Usually, the function is derived on the basis that it should provide a maximum value for the ratio of intergroup dispersion to intragroup dispersion. Fisher's discriminant is an example. The object is to arrive at a set of coefficients based on samples from known groups (the training/prototype set), using which a new sample can be classified as a possible member of one of the

TABLE 10

Resolution of mass spectra of cyclohexane/n-hexane mixtures based on mass spectra of pure constituents

Spectrum No.	Coordinates in		Normalized w.r.t.		Computed mole fraction of cyclohexane
	axis 1	axis 2	100% cyclohexane	100% n-hexane	
1	0.628	0.0	1.0	0.0	1.0
2	0.497	0.069	0.793	0.105	0.88
3	0.526	0.125	0.838	0.191	0.81
4	0.419	0.348	0.668	0.525	0.56
5	0.213	0.497	0.340	0.757	0.31
6	0.113	0.575	0.113	0.575	0.17
7	0.0	0.657	0.0	1.0	0.0

groups. Coomans et al. [20], have described in detail the methods and applications of the discriminant method in chemistry.

Discriminant analysis of hair data

A small-scale study with hair samples from the general population of metropolitan Bombay was done [21] as part of a programme to investigate the role of hair as an indicator of environmental exposure. The elemental data were processed in an attempt to discriminate between 21 samples from an area of high industrial activity and 21 samples, selected randomly, from the rest of the metropolis. Ten elements (Cr, Mn, Fe, Co, Cu, Zn, As, Se, Br, Cd) were selected. The p.c.a. plot of component 1 vs. component 2 (Fig. 5) provided no discrimination, therefore the Fisher discriminant technique was applied for these ten elements. It was observed that Cr, Mn, As and Cd had higher values for the discriminant coefficients and the method was repeated for these four elements. The discriminant coefficients are given in Table 11. The discriminant scores on the samples are plotted in Fig. 6. Principal components analysis of the same data set for Cr, Mn, As and Cd gave a plot (Fig. 7) of the first two components. It can be seen that the samples drawn

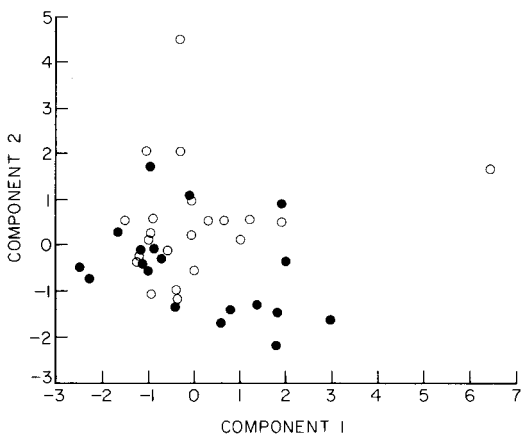


Fig. 5. Principal components results for data on 10 elements from 21 samples: (○) Chembur; (●) other areas.

TABLE 11

Discriminant coefficients obtained for hair samples by the Fisher discriminant method

Cr	Mn	Fe	Co	Cu	Zn	As	Se	Br	Cd
0.84	−2.42	0.36	0.33	0.41	0.44	−0.71	0.02	0.28	−0.82
1.03[a]	−1.80[a]	—	—	—	—	−0.46[a]	—	—	−0.80[a]

[a]Method applied to Cr, Mn, As and Cd only.

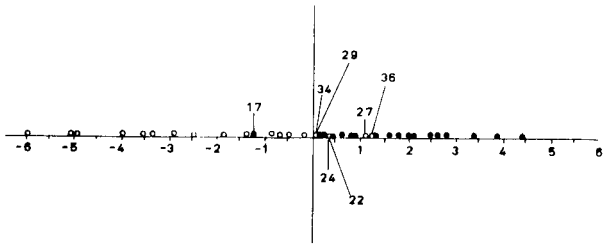


Fig. 6. Discriminant scores for data on 4 elements (Cr, Mn, As and Cd): (○) Chembur; (●) other areas.

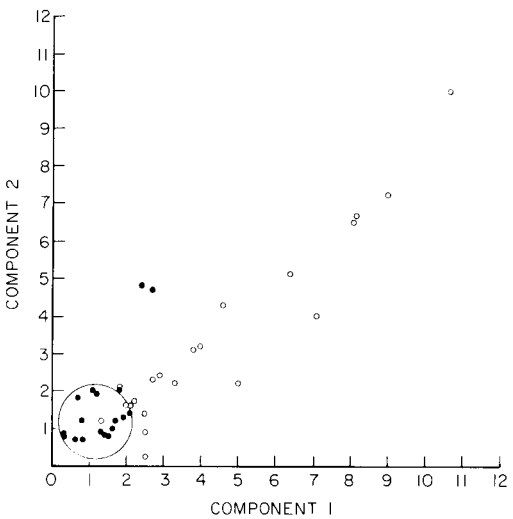


Fig. 7. Principal components results for data on the 4 elements: (○) Chembur; (●) other areas.

randomly from the nonindustrial areas are tightly clustered, whereas the samples drawn from the relatively small Chembur area, which is highly industrialized, are widely scattered, possibly because of higher levels of individual exposure.

Thus, a combined approach with the Fisher discriminant technique to choose the discriminating variables and p.c.a. to provide a better visual representation, offers a powerful method of feature generation and selection and aids interpretation.

Conclusions

The principal component and discriminant techniques have been widely used in feature extraction and selection. The former method has been generally used for representation, with consequent emphasis on variables with large variance. The discriminant method concentrates on variables that are

characteristic to a group and thus enhance class separation. A judicious combination of these approaches is shown to result in a better interpretation of the data. The major difficulty in factor analysis is to relate the derived components (factors) with the actual physical variables operative in the data set. The eigenvector coefficients and the communality patterns used in this work provide insight into an understanding of these relationships. The principal components method is surely more than a means of reducing dimensionality (to provide a visual picture), as is illustrated by the mass spectral resolution studies, where it enabled even the mode of fragmentation and the presence of extraneous constituents to be inferred.

The authors thank Dr. M. Sankar Das, Head of the Analytical Chemistry Division, for his interest and support.

REFERENCES

1 P. S. Shoenfeld and J. R. DeVoe, Anal. Chem., 48 (1976) 403 R.
2 B. R. Kowalski, Chemometrics, Anal. Chem., 52 (1980) 112 R; Anal. Chem., 54 (1982) 232 R.
3 K. Varmuza, Pattern Recognition in Chemistry, Springer Verlag, New York, 1980.
4 P. C. Jurs and T. L. Isenhour, Chemical Applications of Pattern Recognition, Wiley-Interscience, New York, 1975.
5 S. Wold, Pattern Recognition, 8 (1976) 127.
6 G. L. Ritter, S. R. Lowry, T. L. Isenhour and C. L. Wilkins, Anal. Chem., 48 (1976) 591.
7 J. B. Grutzner, M. Jautelat, J. B. Dence, R. A. Smith and J. D. Roberts, J. Am. Chem. Soc., 92 (1970) 7107.
8 G. Hangac, R. C. Wieboldt, R. B. Lam and T. L. Isenhour, Appl. Spectros., 36 (1982) 40.
9 E. R. Malinowski, Anal. Chem., 49 (1977) 612.
10 D. N. Lawley and A. E. Maxwell, Factor Analysis as a Statistical Method, Butterworths, London, 1963.
11 T. W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley-Eastern, New Delhi, 1972.
12 E. R. Malinowski and D. G. Howery, Factor Analysis in Chemistry, Wiley-Interscience, New York, 1980.
13 H. H. Harman, Modern Factor Analysis, University of Chicago Press, Chicago, IL, 1980.
14 R. Gnanadesikan and M. B. Wilk, in P. R. Krishnaiah (Ed.), Multivariate Analysis, Vol. 2, Academic Press, New York, 1969.
15 J. B. Stothers, Carbon-13 Nuclear Magnetic Resonance Spectroscopy, Academic Press, New York, 1972, p. 179.
16 F. J. Knorr and J. H. Futrell, Anal. Chem., 51 (1979) 1236.
17 H. Budzikiewicz, C. Djerassi and D. H. William, Mass Spectrometry of Organic Compounds, Holden-Day, San Francisco, CA, 1967.
18 E. R. Malinowski and M. McCue, Anal. Chem., 49 (1977) 284.
19 R. D. Sedgwick, in R. A. W. Johnston (Senior Reporter), Mass Spectrometry, Vol. 6, Royal Society of Chemistry, London, 1981, p. 194.
20 D. Coomans, D. L. Massart and L. Kaufman, Anal. Chim. Acta, 112 (1979) 97.
21 K. R. Bhat, J. Arunachalam, S. Yegnasubramanian and S. Gangadharan, Science Total Environ., 22 (1982) 169.

Analytica Chimica Acta, 157 (1984) 261—272
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

COMPUTERIZATION OF QUALITY CONTROL TESTS OF LIQUID CHROMATOGRAPHY MEDIA

PER BAECKLUND and ROLF DANIELSSON*

Department of Analytical Chemistry, Institute of Chemistry, University of Uppsala, P.O. Box 531, S-751 21 Uppsala (Sweden)

LARS HAGEL

Pharmacia Fine Chemicals, P.O. Box 175, S-751 04 Uppsala (Sweden)

(Received 14th July 1983)

SUMMARY

A dual microcomputer system is described for quality control tests of chromatography media at medium speed (time resolution 1 s). One dedicated microcomputer handles real-time control and measurement in a multi-experiment situation. The other computer is a personal computer that handles setting-up of experiments and evaluations. A special, simple high-level language is implemented through an interpreter in the dedicated computer. Precautions against loss of data and the design of reliable and easy-to-use operating procedures are discussed. The results in terms of performance, economy and computer familiarity from daily use of the system during three years are also given.

Quality control in the production of liquid chromatography media involves various chemical and physical analyses as well as a performance test. In this test, well known sample mixtures are eluted through a packed column and the elution volume, yield (i.e., area), resolution and zone broadening are evaluated from the recorded detector signals. These experiments are rather tedious to perform on traditional gel media such as Sephadex and Sepharose (Pharmacia Fine Chemicals), and several manual test stations are needed to obtain the necessary capacity. With steadily increasing production of established and new products, the capacity of the manual test systems becomes a time-limiting factor. Some years ago, a project was started to exploit the full capacity of each test station, and to reduce the time required for evaluation. Additional requirements were that the different tests were to be done and evaluated independently of each other, and that the system could be handled with minimal further training. Further, the system had to be flexible enough to allow future reconstruction and also had to provide a high degree of security against loss of experimental data.

At the commencement of this study in 1978, the commercially available systems for the above purpose were various chromatography integrators (Perkin-Elmer, 1974), a microcomputer specially designed for measurement and control (MY-16; Mydata, Sweden, 1978), a combination of a data

0003-2670/84/$03.00

logger and a desk-top computer (Hewlett-Packard, 1976), and an expensive minicomputer-based system. Only the last system fulfilled all the above requirements. Because of the high cost of this system, a dedicated multi-task microcomputer system was developed. This is used for control of valves and power relays and for data acquisition at moderate speed, typically one data point per second per channel. Routines for starting and evaluating experiments are highly operator-orientated, with dialogue procedures for interactive control, and with a high security margin against errors. The purpose of this paper is to describe the implementation of the system and to discuss the results obtained in its daily use over a three-year period.

THE DUAL MICROCOMPUTER APPROACH

Various ways of implementing laboratory routines with computers have been described. Since the early work dealing with minicomputer systems [1], the advent of microcomputers has drastically changed the situation. Microcomputers have been used as small stand-alone systems [2] as well as intelligent interfaces to a larger host computer [3]. The control and measurement tasks for the quality control procedures described above are simple enough to be handled by a dedicated microcomputer system. However, to run a number of independent experiments simultaneously, the system must provide some sort of multi-task operation. To set up the experiments and to collect and evaluate the results are tasks involving interactive data handling, which calls for a high-level programming language. For this purpose, an advanced desk-top computer (personal computer) seems appropriate, but very few, if any, systems of this kind are well suited to handle real-time measurement and control tasks during interactive work.

The costly step to a minicomputer or a main-frame system can be avoided by using two microcomputers, one dedicated system connected to the instruments, and one personal computer for data handling and operator dialogue procedures. In this way, the dual role of the computer regarding the two different interfaces towards the instruments and the human operator [4] is eliminated, and the two computers can be matched to the different demands in a cost-efficient way. The two microcomputer systems, for signal processing and data processing, will be described briefly with special attention to the separation of tasks and the implementation of non-standard microcomputer operations. A more detailed report on the signal-processing part is available on request. Figure 1 outlines the system; a similar approach has been used for voltammetric instrumentation [5].

The signal-processing computer

The basic experiment can be described as a sequence of events following a predefined time schedule. The timing of events (e.g., turn on, switch off,

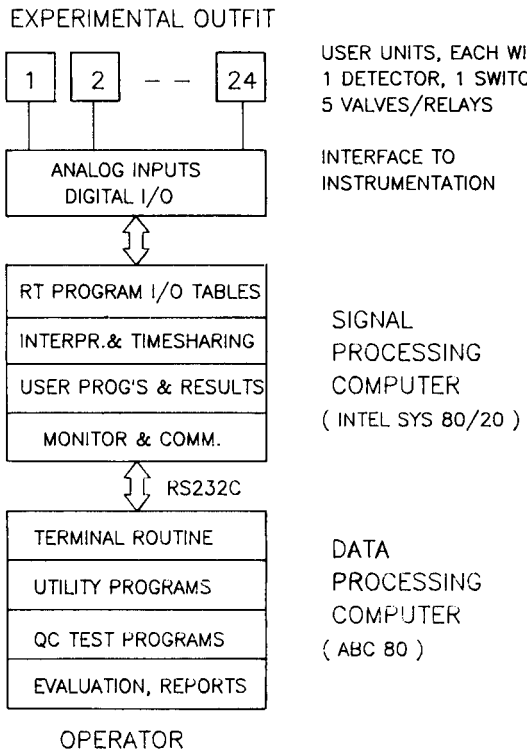


Fig. 1. The dual microcomputer system.

sample signal) can generally be expressed in terms such as wait N seconds, wait for some signal, repeat a block of events N times. Such tasks can readily be done using a minimum configuration system with sufficient input/output and memory capabilities. There are, however, considerable software limitations for such cheap systems; usually they are supplied with a simple monitor program for loading and displaying data, etc., via a console device (e.g., a CRT). For program development a larger system is preferable, e.g., a host computer with a cross-assembler or a cross-compiler. The resulting machine-code program, properly modified for each experiment, is then loaded into and executed by the signal-processing microcomputer. Such procedures have been described in several reports [3, 6, 7]. With the dual microcomputer system, the machine code preparation could be done by the data-processing microcomputer, but this approach was considered to be less efficient because of the uneven task division. In experimental work, the programs for the signal-processing computer have to be extended or modified rather frequently. Thus, powerful assembling and debugging facilities are needed possibly with another type of computer, which put heavy demands on the data-processing computer. An alternative procedure was adopted, which provided more independent operation for the two computers. To

facilitate the design of experiments and still retain flexibility, a set of instructions for the basic events and the timing procedures was formulated. The corresponding machine-code routines were assembled and stored in resident memory (PROM) in the signal-processing computer. These routines are activated at run-time according to a list of statements previously loaded by the data-processing computer. The statements are denoted by a two-character label, with additional parameters if required (Table 1). The list of statements constitutes a measurement and control program for the experiment.

Hardware. A dedicated microcomputer system INTEL SYS-80/20 was chosen for the signal-processing tasks. It was equipped with an analog input card (Data Translation 1742) and two expansion cards (INTEL iSBC-116) for input/output and data storage (RAM). The main features are an eight-bit processor (8080), 36 kbytes of RAM, 12 kbytes of PROM, 18 parallel I/O ports with 8 TTL signals each, and 32 analog input channels for ±10 V with 12-bit resolution. There are also three serial I/O ports for data communication and three timer/counter circuits. The input/output signals are arranged for connection of 24 user units. One analog input channel for detector signal, five digital outputs for ON-OFF control of magnetic valves and one digital input for an external switch are assigned to each unit. This division was designed so that an experiment could occupy one or two users units, depending on complexity.

Software. The INTEL computer was delivered with a resident monitor with a number of one-letter commands for loading and displaying data, etc. To serve as a convenient operating system in this application, the monitor was extended with a number of special commands. The monitor extensions were assembled and made resident in non-volatile memory (about 3 kbytes of PROM).

TABLE 1

Control and measurement statements
(All statements are terminated by (CR). Spaces are ignored)

Operator		Meaning
Flag		One of eight data bits denoted A through H
/usnr/		Optional user number 1—24, current user if deleted
Variable		Analog input (X), timer value (T) or flag (F followed by A—H)
Name		Two characters, the first a letter except F, X, T
Data		Decimal integer or hh' mm" ss for hours, minutes, seconds
Constant		Variable by name or immediate data
Mnemonic	*Operator(s)*	*Comment*
AW	flag/usnr/	Wait until flag is set
CT	/usnr/	Clear timer
DA, DB, DC	constant	Set up loop A, B, C for specified number of counts
EA, EB, EC		End of loop A, B, C
EN		End of program
LE ... = ...	name, data	Assign data to named variable
PR	var/usnr/	Store current value of variable
RA, ..., RH	/usnr/	Clear flag A, ..., H
SA, ..., SH	/usnr/	Set flag A, ..., H
WA	constant	Wait the specified time
*	—	Followed by text for remarks only

Multiple experiments are handled by running 24 independent user programs in a time-sharing mode. A fixed block of RAM memory is allocated to each user for storage of program and results. With a normal program, about 600 data points can be stored. Space is also reserved for a short message concerning the current experiments. In addition to the user programs, there is a separate executive program, the real-time program, which performs the actual input and output of instrumental signals. One of the timers is programmed to give an interrupt signal at 40-ms intervals, i.e., 25 times per second. At each signal, the control is passed in a circular manner to one of the 24 users, or to the real-time program. A flow chart of this time-sharing operation is shown in Fig. 2.

The user program statements, as recognized by an interpreter program, work on data tables for analog input signals and digital input/output signals. These tables are updated by the real-time program at 1-s intervals. In the meantime, all the user programs are checked once. When a wait statement is in operation, a user clock variable is decremented. Otherwise, the next statement is extracted and handled by the interpreter until a new wait

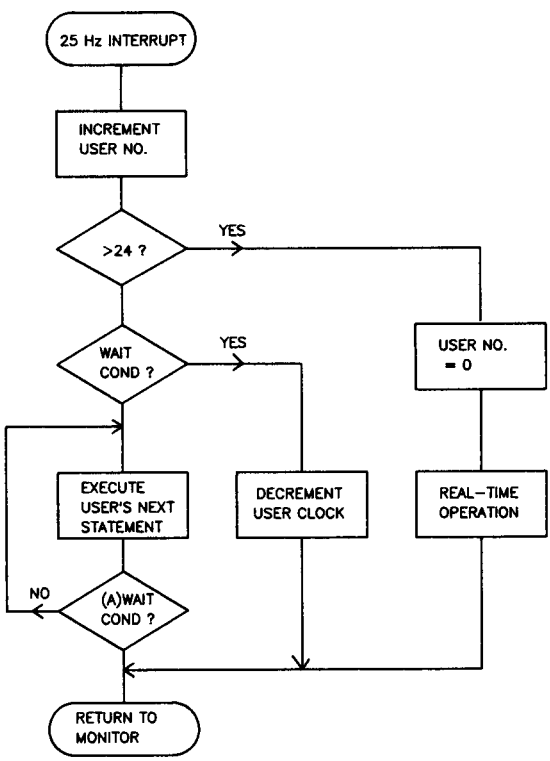


Fig. 2. Flow diagram for the interrupt-controlled time-sharing operation of the signal-processing computer.

(or await) condition is met. Input signals are sampled by moving data from the input table to the user's result area, while output signals are controlled by changing flag bits in the output table. A set of pointers is used to control program execution and data storage.

When the statement (or statements) of the current user has been executed or if the execution time exceeds 20 ms the control is returned to the monitor. As the execution time is rather short compared to the 40-ms interrupt interval, the signal-processing computer is mostly running under monitor control. In this way, a dual task system is achieved, with the measurement and control activities as the primary task and the monitor operations as the background task. All communication with the data-processing computer takes place via the monitor, and there are no real-time aspects to take into consideration for the data-processing computer.

An example of the control and measurement program. An example showing the level of the monitor operation and the use of measurement and control statements is provided by a performance test on Sephacryl S-300SF. A Pharmacia K26/50 column (height 30 cm) was used. The sample was 2 ml of cytochrome C (2 mg ml^{-1}) eluted with phosphate buffer (pH 7, ionic strength 0.05 M). The Pharmacia P-500 pump was operated at 20 ml h^{-1}. A u.v. detector (Pharmacia UV1) with a 3-mm flow cell was used at 254 nm. A 100X amplifier (a non-inverting operational amplifier circuit) was built in to give a 10-V full scale output signal. With a nominal absorbance setting of 0.05 (for the original 10-mV recorder) this signal corresponds to an absorbance of 0.5. A pen recorder (Pharmacia REC2) on the 10-V range was also connected to the output. The test station was assigned to user number 4, and the data input to the monitor is listed in Table 2.

The first three monitor commands are to clear the user RAM area. Next,

TABLE 2

Input to monitor for running the performance test on Sephacryl S-300SF. Monitor commands and the statements loaded are to the left; to the right are explanatory comments

K4	Delete old program for user No. 4	PRX4	Store analog value of unit No. 4
Z4	Clear user No. 4	WA59	Wait 59 s
A4	Append the following program to user No. 4	EA	End of loop A, wait 1 s
		PRT4	Store current timer value for user No. 4
AWA4	Await ON condition for flag A, unit No. 4	SE4	Set flag E, unit No. 4
		WA1	Wait 1 s
CT4	Clear timer for user No. 4	RE4	Clear flag E, unit No. 4
SH4	Set flag H, unit No. 4	EN	End of program
SE4	Set flag E, unit No. 4	(Esc)	Terminate program input
WA1	Wait 1 s	FM4	Fill message line for user No. 4
RE4	Clear flag E, unit No. 4	...	(Text to be entered)
SF4	Set flag F, unit No. 4	(Esc)	End of message line input
WA6″	Wait 6 min	B4	Begin program execution for user No. 4
RF4	Clear flag F, unit No. 4		
WA300″	Wait 300 min	...	Other activities, e.g., checking status
PRT4	Store current timer value for user No. 4	V4	List results for user No. 4, here 362 two-byte values
DA360	Do loop A 360 times		

the new program statements are entered. At this stage the program is just loaded, but the events taking place during performance are as follows. The experiment is started by closing a switch connected to the external flag input (flag A). The user timer (T) is cleared and will then show the number of seconds from start. The pen recorder is turned on via flag H, and an event mark is made on the recorder chart by closing a relay, operated via flag E, for 1 s. The sample solution is then injected during 6 min by operating a valve controlled by flag F. The peak to be detected appears after several hours, and a wait period of 300 min is inserted. After that, the current timer value is stored, followed by 360 data points (X) at 1-min intervals. The 59-s wait period inside the loop is due to an extra wait cycle caused by the end-of-loop statement. Finally, the timer value is stored, and an event mark is made on the recorder chart, which then is switched off.

The programming sequence is terminated by an escape character. Then a message line with arbitrary text, also terminated by an escape character, can be assigned to the user. The next monitor command initiates the program execution, and the monitor is available for other activities. The current experiments can be followed by monitor commands for listing pointers and table contents. When the user program is ready, the stored values are listed by the last monitor command. The values are given, as 2-byte words in hexadecimal notation. Decoding is not done by the monitor, because the data structure is inherent in the program and the interpretation is a task for the program-generating unit. For illustration of the given example, the data points were transferred to a computer with plotting facilities (HP-9836 with HP-7470), and the corresponding curve is shown in Fig. 3.

As can be seen from the example, the commands for the extended monitor and the measurement and control statements form a comprehensible basis for the operation of the signal-processing computer. This feature facilitates coordination with the data-processing computer and adds to the security of the system. All memory-addressing is inherent in the specific command

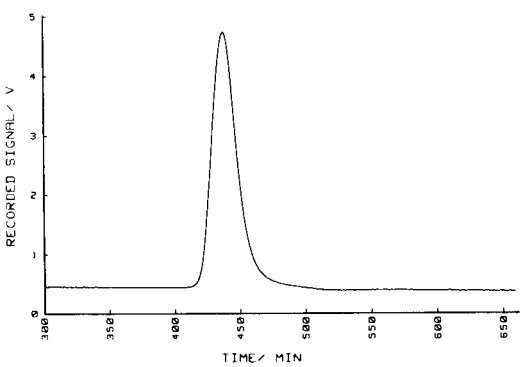


Fig. 3. Performance test on Sephacryl S-300SF: 2 mg of cytochrome C eluted with phosphate buffer, pH 7. Detection at 254 nm; absorbance of 0.05 V^{-1}; flow rate 20 ml h^{-1}.

and the user numbering. This eliminates the risks of unintentional program interference or possible loss of control over the system.

The data-processing computer

To conduct the various experiments, a proper INTEL monitor sequence must be prepared and executed, followed by evaluation of the results obtained. While the normal data-handling capabilities of a personal computer can be utilized for the latter purpose, the data communication with the signal-processing computer needs some consideration. One of the requirements for the total system was to make it possible for laboratory chemists to participate in program development. Only then can the benefits of laboratory computerization be fully exploited. To facilitate this process, some utility programs common to all experiments were developed. Only the utility programs operate in direct connection with the signal-processing computer. For the rest, the data-processing computer operates off-line, which means that mistakes in programming and maintenance are unlikely to affect the performance of running experiments.

Personal computer. The personal computer chosen was the ABC-80 (Luxor AB, Sweden), which is similar to PET, Apple or TRS-80. It is programmable in BASIC and supplied with 32-kbyte RAM, a dual floppy disk drive and a printer. The connection to the signal-processing computer is made via a built-in serial interface (RS232C) and a general terminal routine (T80PRT) operating at 1200 baud. The basic data communication is conducted by a BASIC subroutine, which sends the contents of an output buffer to the INTEL monitor and returns the received sequence in an input buffer. The buffers are ordinary string variables to the calling BASIC program.

Quality control test programs. The programs used for running an ordinary quality control test are summarized in Fig. 4. For each test, a corresponding test program is stored under an appropriate name ("X" and "Y" in Fig. 4). In the initial step, the selected test program is run, if required, via a menu program. A dialogue then takes place about the specific conditions (test identifier, user unit, instrumental parameters, etc.). As a result, two disk files are prepared. The information to be passed to the signal-processing computer (i.e., the measurement and control program and the message line) are stored on a fixed data-transfer file. Arbitrary information concerning the current experiment to be used for later evaluation and test reports are stored on a text file, which is given a suitable identifier by the operator. A hard copy of the same information is also obtained, which may be attached to the test station in question.

Utility programs. So far, the work of the data-processing computer has dealt with preparation of the data needed for the current experiment. As mentioned, all communications with the signal-processing computer takes place via the utility programs. Two such programs will be discussed; one for the initiation and one for the termination of an experiment. The initiating

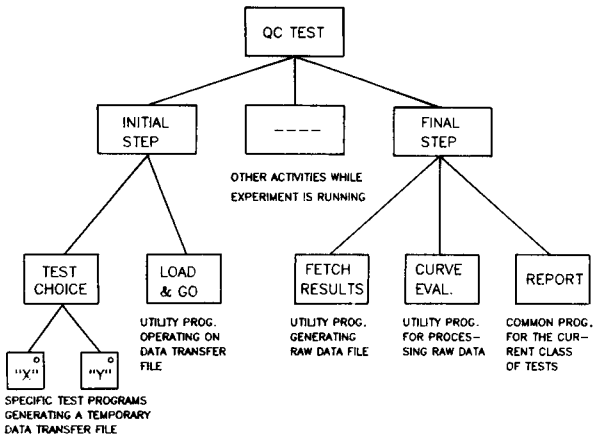


Fig. 4. The structure of the operations of the data-processing computer for a typical quality-control test. To show the connection between the programs, the JSP notation is used. The tasks included in a block are organized as a sequence of sub-tasks, shown as a row of blocks at a lower level. These blocks in turn may be split up, etc. A small circle within the sub-blocks indicates selection of one of the blocks.

utility program handles the "load and go" procedures (cf. Table 2). The purpose of this program is to take the control and measurement program and the message line from the data-transfer file, and add proper monitor commands for the start procedures. First, the user number for the program is requested and the status of the selected user is checked. An uncleared condition, indicating a running or unterminated experiment, is notified as "inaccessible user" and followed by a new request. Next, the measurement and control program and the message line are loaded. By a monitor command the same information is listed back and a comparison is made to guarantee proper loading. If no errors are detected a "begin" command is given, which completes the "load and go" operations.

When an experiment is completed, the terminating utility program is run ("fetch result" in Fig. 4). The status of the requested user program is checked and an error message is given if the experiment is still running. Otherwise, the contents of the user's result area is read twice and compared to ensure correct reading. The results are written in a data file named by the same identifier as the previously mentioned text file. The data file is read back and checked, and only then is the user RAM area cleared.

Evaluation and reports. Most quality control tests follow the same scheme of evaluation, based on peak data for characterization. A utility program for peak evaluation using interactive graphical control was developed. With this program the main part of the screen displays the curve to be evaluated. In graphical mode, a 3*2 point matrix for each character position is utilized, and with 20 lines a resolution of 60*80 points is obtained. This seems to be rather poor, but when fast routines are used for redrawing the curve, the

screen can act as a "moving window" for a selected part of the curve. The bottom line shows the x-position of a cursor and a numerical display of the corresponding (x, y) values for the curve. The point on the curve is also blinking. The keyboard is re-defined, so that single key-strokes initiate graphical commands for moving cursor and display window, changing scale parameters, and finding maximum and minimum.

The evaluation includes the following steps, prompted by a text message on the top line. (a) Display the curve and make corrections of erroneous points; the cursor point is replaced by interpolated value from the adjacent points. (b) Select the portion of the curve to be evaluated; move the cursor to first and last point, respectively. (c) Select points for baseline correction by moving cursor. The background is taken as the straight lines connecting the selected points on the curve. After subtraction the curve is redrawn. (d) Select peak to be characterized by moving the cursor to the first and the last point, respectively. The area is calculated by Simpson's rule. (e) The peak maximum is taken as the maximum y value within the x interval of the peak. The corresponding x value is taken as the peak position. By moving the cursor another point may be selected. (f) Steps (d) and (e) may now be repeated until all peak parameters have been obtained.

When the evaluation program is run, the test identifier (part of the file names) is requested, and the contents of the corresponding text file are presented on the screen. The operator verifies the selected experiment, and the data file is loaded. The curve is evaluated as described, and for each peak a calibration factor for the test substance is requested. Using the instrumental parameters contained in the text file, the peak characteristics are presented as results.

For the example shown in Fig. 3, the results obtained are shown in Table 3. Usually the original recorder chart is attached to the report. In doubtful cases, or for computer failure, results can be recalculated on data obtained with a digitizer. By the interactive procedures described, the operator controls the calculations. Thereby experience and knowledge of the operator on judgement of experimental raw data are utilized and maintained.

System maintenance

An important point in laboratory computerization is to establish the level of system knowledge for users with different assignments. This process is

TABLE 3

Results from a performance test on Sephacryl S-300SF

Peak time (h)	7.39
Elution volume (ml)	147.8
Peak interval (h)	6.91—8.49
Peak area (Vh)	1.677
Calibration factor (Vh)	1.860
Yield (%)	90.2

facilitated by the computer separation and the hierarchal program structure. As much as possible of the fundamental configuration and operation of the computers has been maintained, making the original computer manuals practicable for training and documentation. The minimum requirement for the laboratory workers is to be able to run a BASIC program with built-in dialogue procedures and to act on error messages. The demands on people planning the experiments are fundamental knowledge of BASIC programming including personal computer operation and familiarity with the special measurement and control instructions. Those who are responsible for the system operation should be acquainted with the monitor commands and with all aspects of the personal computer. Detailed information on the software and the hardware of the signal-processing computer is documented for service purposes only.

RESULTS AND DISCUSSION

The results of this work and the experience gained from running the system in routine quality control for three years may be described in terms of performance, economy and computer familiarity.

The data sampling and numerical integration of the peak areas have reduced the variation of the evaluations to 2% r.s.d., compared to 7% when the peaks were digitized manually, and a variation of 10—15% when a disc integrator trace was evaluated. The use of a BASIC programmable personal computer facilitates the development of operator-orientated software routines for running the system. No special knowledge of computers or programming is necessary; the system is regarded solely as dedicated analytical instrumentation. The design of the evaluation program makes the evaluation resemble the manual procedure and allows the operator (not the computer) to take decisions on the results. The software routines also generate check protocols, quality-control parameters and finally a report of the results. These features reduce manual typing and experimental errors and also generate a more structured and standardized documentation.

Security against loss of experimental data is essential for a quality control laboratory. The system prevents accidental loss of data in the following ways. The communication program between the data-processing computer and the signal-processing computer is constructed in such a way that no test can be halted by mistake. The system can be operated manually (e.g., sample injection) in case of a breakdown of the signal-processing computer. The detector signal is continuously traced by a recorder, and the resulting chromatogram may be evaluated separately by a digitizer and a desk-top computer to generate the experimental results. The data-processing unit is, in case of a failure, easily replaced by another identical computer which is otherwise used for work of less priority. The system configuration, however, has proved to have remarkable reliability, and during a three-year period essentially no hardware-related breakdown has been recorded, except for one

failure in a floppy disk drive. The incidents that occurred were powerline failures on some occasions. Because of the very low frequency of such failures, no power back-up facilities have yet been included in the system. Even though the system was developed for use with chromatographic experiments, it is basically an all-purpose system for control and data acquisition of medium-speed processes. The major limitation is the time resolution of one second which is necessary for the establishment of the time-sharing feature.

The favourable economy of the system is based on reductions in evaluation time, increased utilization of each test station and the low investment cost. Among the results that were achieved was a reduction in the time of evaluations from 45 to 5 min. The former time included the re-calculations that were necessary when disc integrators were used. The capacity of each station increased by 20—40% because of the possibility of automatic column switching and sample injection, e.g., over weekends. The hardware cost of the system per test station depends on how many users are allocated to each station, but even when the cost of a recorder is included, the cost of the whole equipment is less than that of a typical recorder with integration assembly.

The time spent on developing the system corresponded to approximately one man-year, which makes the pay-off time of the original system two years, without regard to the value of the knowledge gained in interfacing computers. The successful fulfilment of this work has made the laboratory personnel more computer-orientated. This computer familiarity may in the long run turn out to be the most valuable result of this work. Meanwhile, the benefit of the work may also be illustrated by the fact that another similar system is now to be installed.

REFERENCES

1 S. P. Perone, Anal. Chem., 43 (1971) 1288.
2 D. Betteridge and T. B. Goad, Analyst, 106 (1981) 257.
3 H. C. Smit, Anal. Chim. Acta, 122 (1980) 201.
4 E. Ziegler, Anal. Chim. Acta, 122 (1980) 315.
5 P. Baecklund and R. Danielsson, Anal. Chim. Acta, 154 (1983) 61.
6 R. P. J. Duursma, H. Steigstra, R. G. Logchies and H. C. Smit, Anal. Chim. Acta, 144 (1982) 13.
7 H. J. Skov and L. Kryger, Anal. Chim. Acta, 122 (1980) 179.

Analytica Chimica Acta, 157 (1984) 273—290
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

A SOFTWARE PACKAGE FOR THE GENERATION OF NOISE WITH WIDELY DIVERGENT SPECTRAL PROPERTIES
The Simulation of Realistic Stationary Detector Noise

J. M. LAEVEN and H. C. SMIT*

Laboratory for Analytical Chemistry, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam (The Netherlands)

J. V. LANKELMA

Stichting Mathematisch Centrum, Department of Applied Mathematics, Kruislaan 413, 1098 SJ Amsterdam (The Netherlands)

(Received 2nd August 1983)

SUMMARY

The computer program discussed is capable of generating stationary noise with optional statistical parameters. Five standard noise types can be generated: white, 1/f, first order, Gaussian, and damped cosine noise. The input to the program can be a measured or arbitrarily chosen autocovariance function (or power spectrum). The autocovariance function (power spectrum) computed from the generated noise will be in perfect agreement with the input autocovariance function, if an infinite number of noise data is generated. Some theoretical work on noise generation on the basis of the autocovariance function is described. The autocovariance function of 1/f noise is theoretically derived from the power spectrum, based on a model described earlier. A few examples of the use of the program and an example of a possible application are given. A peak-finding procedure is tested with a simulated chromatogram contaminated with different types of noise. Applications are possible in data processing, information extraction, simulation and automation.

The use of computers in data analysis has created possibilities of developing information extraction methods (i.e., extraction of relevant information from analytical signals) in a systematic manner. Consequently, determining and reducing the influence of noise is gaining importance. However, the relative complexity of the mathematics involved in data analysis to yield optimal results has been a limiting factor. The development of chemometrics can be partly explained from this background. To date, information extraction methods in analytical chemistry were usually based on, and tested (if at all) with, a white or first-order noise model. It is known, however, that many detectors exhibit noise characteristics different from the white or first-order model [1]. This means that the algorithm originating from this approach will be far from optimal in practice and even may give erroneous results.

The fields in which reliable noise models are necessary are very diverse. In data processing, filter procedures have to be developed and tested with use

0003-2670/84/$03.00

of realistic noise models to yield optimal performance. In information extraction methods, these models are necessary for qualitative analysis (e.g., peak-finding procedures in chromatography; interpretation of spectra) and quantitative analysis (e.g., determination of peak area and retention times in chromatography). Further, in simulation, when processes are simulated using systems theory, the need of reliable noise models is obvious. Also in the field of automation, noise models play an important role; programs to interpret spectra automatically have to be developed with use of different noise types to minimize "misses" and "false alarms".

In this paper a program capable of generating stationary noise with selectable and realistic characteristics is presented. This program can be of much help in all the research fields mentioned above.

THEORETICAL BACKGROUND

A set of independent numbers u_j, representing white noise with statistical properties, is defined as follows

$$E[u_j] = 0 \tag{1}$$

$$E[u_i u_j] \equiv R_{uu}(i-j) = \delta_{i-j} \tag{2}$$

where $E[u_j]$ represents the expected value of u_j (mean value), δ_{i-j} is the delta Dirac distribution, and $R_{uu}(i - j)$ is the autocovariance function (ACVF) of u. (Symbols are defined in Table 1.) The general definition of the ACVF of a stationary stochastic variable y is given by

$$R_{yy}(\tau) \equiv \lim_{T\to\infty} (1/T) \int_{-T/2}^{T/2} y_t y_{t+\tau} \mathrm{d}t \tag{3}$$

A (discrete) system with impulse response h_k, where k represents (discrete) time, is also defined. This means that if a Dirac pulse δ is applied to the input of the system, then the output of the system equals the impulse response h. It is well known from signal theory that any signal can be represented by a series of Dirac pulses [2]. If such a (discrete) signal is applied at the input of the described system, then the output of the system can be found by summing the responses on each distinct pulse. If this procedure is followed for white noise at the input of the system, the result at the output will be the convolution summation

$$x_l = \sum_{k=-\infty}^{\infty} h_k u_{l-k} = \sum_{k=1-M}^{M-1} h_k u_{l-k} \tag{4}$$

In practice, the impulse response of a system is finite in time, thus infinite summation can be replaced by finite summation over $2M-1$ values, where M is sufficiently large, as is done in Eqn. 4. The convolution means that the responses of the system to all input values in the past are summed to obtain the output of the system.

TABLE 1

List of symbols and definitions

u_j	White noise number
$E[\ ]$	Expected value of the expression within the brackets
$R_{xx}(\tau)$	Autocovariance function of the function x_t with time lag τ
$R_{uu}(i-j)$	Discrete autocovariance function of the (discrete) stochastic variable u
δ_i	Delta Dirac distribution, for the discrete case: $\delta_i = \begin{cases} 1 & i = 0 \\ 0 & i \neq 0 \end{cases}$
x_l	Discrete filter output
h_k	Filter impulse response
$2M-1$	Number of significant impulse response coefficients
FT	Fourier transform
FT^{-1}	Inverse Fourier transform
$H(\omega_l)$	Fourier-transformed impulse response
$G(\omega)$	Spectral power density function
$E_1(\xi)$	Exponential integral
t_0	Expected time duration of pulses making up the current in a flame ionization detector
K	Detection threshold parameter
ϕ_t	Function describing an asymmetric peak
H	Peak top amplitude
A	Asymmetry factor
t_R	Retention time
σ_p	Gaussian standard deviation of a peak
σ_n^2	Variance of baseline noise
S/N	Signal-to-noise ratio, defined as amplitude max/σ_n
$\bar{x}$	Mean value of a random sample of the distribution x
μ	Mean value of the distribution x
$s_{\bar{x}}$	Standard deviation of the mean value of a random sample of distribution x
T	Observation time
T_x	Time constant

The statistical properties of the system output (given by Eqn. 4) can be evaluated as follows

$$E[x_l] = \sum_{k=1-M}^{M-1} h_k E[u_{l-k}] \overset{(1)}{=} 0 \tag{5}$$

$$R_{xx}(i) = E[x_l x_{l+i}] \overset{(4)}{=} \sum_{j=1-M}^{M-1} \sum_{k=1-M}^{M-1} h_j h_k E[u_{l-k}\, u_{l+i-j}]$$

$$\overset{(2)}{=} \sum_{j=1-M}^{M-1} \sum_{k=1-M}^{M-1} h_j h_k \delta_{l-j+k} = \sum_{k=1-M}^{M-1} h_k h_{k+i} \tag{6}$$

(Numbers placed on the equal signs refer to eqns. derived before)

The last part of Eqn. 6 is known as the (discrete) system covariance function. Thus, the result is that the ACVF of the signal x_l (the output of the system) is equal to the system covariance function.

Equation 6 can be used to compute a shaping filter with impulse response h, which filters white noise in such a manner that the ACVF of signal x at the output of the shaping filter is in the desired form. So, if noise with a

certain ACVF is to be generated, the impulse response h of the shaping filter must be computed using Eqn. 6. This impulse response can be obtained by Fourier transformation of Eqn. 6

$$\mathrm{FT}[R_{xx}(i)] = \mathrm{FT}\left[\sum_{k=1-M}^{M-1} h_k h_{k+i}\right] \tag{7}$$

Because $R_{xx}(i)$ is even in i, Eqn. 7 will hold if i is replaced by $-i$. Thus the convolution theorem, i.e., convolution in the time domain is multiplication in the frequency domain, can be applied to Eqn. 7 and results in

$$\mathrm{FT}[R_{xx}(i)] = H^2(\omega_l) \tag{8}$$

where $H(\omega_l)$, the complex frequency response, represents the Fourier-transformed impulse response h_k. From this expression the impulse response of the shaping filter can be calculated

$$H(\omega_l) = \{\mathrm{FT}[R_{xx}(i)]\}^{1/2} \tag{9}$$

$$h_k = \mathrm{FT}^{-1}\{\mathrm{FT}[R_{xx}(i)]\}^{1/2} \tag{10}$$

Thus, summarizing, the impulse response of the shaping filter, generating noise with the required ACVF properties, can be found by Fourier transformation of the ACVF, taking the square root, and Fourier back-transformation. Because of the symmetry of the functions involved, a fast Fourier cosine transformation will do. After evaluation of the impulse response in the described manner, the signal x_l with the desired ACVF characteristics can be generated according to Eqn. 4.

Flicker noise

Some detectors exhibit noise characteristics that may hamper the signal analysis and interpretation to a high extent. Flicker or 1/f noise, caused by electrical disturbances of an incompletely understood nature, is an example of a noise difficult to handle. The problem of this type of noise is that the power of the noise is inversely proportional to the frequency down to very low frequencies. This implies that the highest noise power occurs in the same frequency area where generally analytical signals exhibit their maximum power. An example of a detector with this noise characteristic is the flame ionization detector (FID) [1].

A realistic approximation model of the 1/f noise power spectrum, in which the (actually non-existent) singularity at $\omega = 0$ is avoided, has been given by Smit and Walg [3]

$$G(\omega) = c \int_{t_m}^{t_n} [1/(1 + \omega^2 t_0^2)]\, \mathrm{d}t_0 = c\, \{[\arctan(\omega t_n) - \arctan(\omega t_m)]/\omega\} \tag{11}$$

where t_n and t_m are the time constants, $t_n > t_m > 0$, t_0 is the expected value of time duration of the pulses making up the current in a FID, and c is some constant.

Fourier back-transformation of the 1/f power spectrum to obtain the autocovariance function of 1/f noise can be done with a cosine transformation, because $G(\omega)$ is even

$$R_{xx}(\tau) = \mathrm{FT}^{-1}[G(\omega)] = 2c \int_0^{\infty} \int_{t_m}^{t_n} [\cos(\omega\tau)/(1 + \omega^2 t_0^2)]\, \mathrm{d}\omega\, \mathrm{d}t_0 \tag{12}$$

Introducing a new variable $z = \omega t_0$ with $\mathrm{d}z = t_0\, \mathrm{d}\omega$, and using the identity

$$(2/\pi) \int_0^{\infty} [\cos(ax)/(1 + x^2)]\, \mathrm{d}x = \mathrm{e}^{-a}$$

rearrangement of Eqn. 12 yields

$$R_{xx}(\tau) = 2c \int_{t_m}^{t_n} \frac{1}{t_0} \int_0^{\infty} [(\cos z\tau/t_0)/(1 + z^2)]\, \mathrm{d}z\mathrm{d}t_0$$

$$= c\pi \int_{t_m}^{t_n} \exp(-\tau/t_0)\, \mathrm{d}t_0/t_0 = c\pi\, [E_1(\tau/t_n) - E_1(\tau/t_m)] \tag{13}$$

where E_1 is an exponential integral, the general definition of which is given by Abramowitz [4]

$$E_1(\xi) = \int_{\xi}^{\infty} (e^{-s}/s)\, \mathrm{d}s$$

The constant $c\pi$ in Eqn. 13 can be rewritten to obtain the ACVF in the appropriate form

$$R_{xx}(\tau) = \sigma^2\, [E_1(\tau/t_n) - E_1(\tau/t_m)]/\ln(t_n/t_m) \tag{14}$$

Numerically, the obtained ACVF can be evaluated by using polynomial and rational approximations given by Abramowitz [4].

Theoretical approximations of the power spectrum and the ACVF of 1/f noise are available now, so that it becomes possible to compare the characteristics of measured noise sources with those of 1/f noise in both the Fourier domain and the time domain. Furthermore, the ACVF of measured noise can be used to classify the noise type as 1/f without the necessity of Fourier transformation, where problems arise because a reliable power spectrum can mostly only be found if optimal windows are used.

As the ACVF and power spectrum characteristics of 1/f noise are now known approximately, it is useful to compare 1/f noise with first-order noise. If the two power spectra are compared, one can see that if the dominating time constants t_n and T_x (see Eqns. 14 and 15) are approximately equal, the power of 1/f noise in the very low frequency area and the high frequency area is larger than the power of first-order noise in those regions, whereas for the middle region the reverse is true.

DESCRIPTION OF THE SAMSON PROGRAM

The purpose of the FORTRAN program SAMSON (Spectrum/ACVF Matching Synthesis Of Noise) is to generate zero mean, stationary noise based on a predefined ACVF/power spectrum given as input to the program; this ACVF/power spectrum can originate from measured noise (e.g., detector noise), computed externally and subsequently read by SAMSON, or can be a standard ACVF generated by the program. Four common standard types of ACVF are possible

(1) first-order noise $\sigma^2 \exp(-|\tau|/T_x)$ (15)

(2) Gaussian noise $\sigma^2 \exp(-\tau^2/T_x^2)$ (16)

(3) damped cosine first-order noise $\sigma^2 \exp(-|\tau|/T_x) \cos \omega_0 \tau$ (17)

(4) "1/f noise" $\sigma^2 [E_1(\tau/t_n) - E_1(\tau/t_m)]/\ln(t_n/t_m)$ (18)

First-order noise can originate from shaping filters (or processes) described by a first-order differential equation (low-pass filters). Many amplifiers can be considered as first-order systems. Gaussian filters cause the second type of noise. The third type originates from a bandpass filter as used in lock-in amplifiers. Generally it can originate from filters described by negative discriminant second-order differential equations with very high $\omega_0 T_x$, allowing the sine term to be ignored in the solution. The fourth type, 1/f noise, is observed in many detectors and other noise sources. In all cases, the described types of noise are observed as the output of the shaping filter, if white noise is applied as input.

To extend the use of standard ACVFs, an option to change individual ACVF values is inserted in the program, e.g., to allow study of the noise characteristics if spikes periodically occur in the ACVF, as may be observed with detectors that are zero-adjusted during a series of measurements. Of course, this option is only useful if the desired ACVF does not deviate too much from one of the standard types. If the deviation is too large, the ACVF can better be generated externally, and subsequently read by SAMSON.

Summarizing, any ACVF/power spectrum can be used as input to SAMSON. The program will generate noise with properties according to this ACVF/power spectrum and store the noise on a disc file. The ACVF/power spectrum, used as input, and the output ACVF/power spectrum, calculated from the generated noise, can optionally be stored on disc too, permitting their use for other purposes. Furthermore, the input and output ACVFs/power spectra are plotted simultaneously in one graph, allowing careful inspection and comparison. Of course, the similarity of input and output ACVFs/power spectra is a statistical matter, thus depending on the number of noise data generated. In the computer system used (HP 1000, model 45-F, operating under RTE IVB) the configuration permits 16 384 noise data to be generated at one time as a maximum. The exact number depends on the number of weighting coefficients necessary to describe the filter,

because these coefficients and the noise data are stored in the same array for efficiency reasons. Because of the boundary conditions, imposed by the structure of the computer system, SAMSON is written as an independent program communicating with the user interactively. The number of possible options and the need for visual inspection and interpretation of the chosen input ACVF/power spectrum and the resulting noise justifies this choice. Furthermore, the interactive mode allows use of SAMSON for educational purposes.

In Fig. 1 a flow chart is given of the main parts of the program (written in FORTRAN IV). The required memory is 124 kbyte. SAMSON is built up of five main parts

(1) Main program: I/O control, overlay calling and library routines (fast Fourier transform, white noise generation); ACVF/spectrum computation; computation of the shaping filter;

(2) Overlay FNCGN: generation of standard ACVFs;

(3) Overlay DSCIO: reading of input ACVF/spectrum from file; filtering of the white noise using the computed shaping filter; storage of the generated noise on disc;

(4) Overlay DSCO2: storage input/output ACVFs/power spectra on disc;

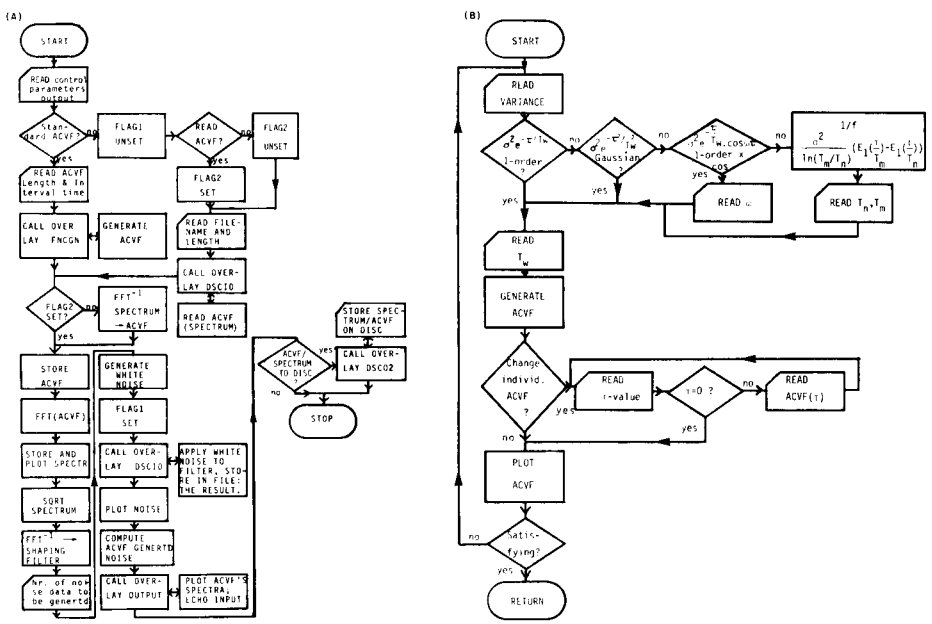


Fig. 1. (A) Flow chart of the main program of SAMSON; FFT stands for Fast Fourier Transform. (B) Flow chart of overlay FNCGN; four standard ACVF types are available for noise generation.

(5) Overlay OUTPT: Supply of plot facilities; output control (hard copy of input parameters, filter coefficients, used filter names etc.).

The time required for the noise generation should be considered. It can be seen in Eqn. 4 that each noise number has to be generated from $2M-1$ white noise numbers, where M depends on the ACVF length and on the number of significant filter coefficients (of course, these two criteria are related). For simplicity, it can be assumed that the filter is sufficiently described by 100 (M = 100) coefficients and that 10 000 noise data have to be generated. It follows that $\sim 2 \times 10^6$ multiplications/additions are needed. With conventional hardware/software this could result in rather time-consuming computations. However, in the computer system used, a vector instruction set is available, allowing considerable reduction in computation time. In the example, the total computation time is only 135 s. If RTE-VI is available, this time can be reduced further. In fact, RTE-VI permits some useful adaptation of the program; these are not described in detail here because their effect is rather system-dependent. The main adaptation is the extension of the number of data that can be generated up to 200 000 at the cost of some speed.

Confidence interval

To determine whether the ACVF of the noise generated by SAMSON fits the model used as input ACVF, the confidence intervals for the input ACVF and for the given number of noise data generated have to be estimated. This can be done using the Bartlett formula [5]

$$\sigma^2 [R_{xx}(\tau)] = [1/(T-\tau)^2] \int_{-T+\tau}^{T-\tau} (T-|r|-\tau) \{R_{xx}^2(r) + R_{xx}(r-\tau)R_{xx}(r+\tau)\} \, dr \qquad (19)$$

where r is an auxiliary variable (time).

In SAMSON optionally the confidence intervals (99.72%) are depicted: the ACVF of the generated noise should be within the interval $R_{xx}(\tau) \pm 3\sigma [R_{xx}(\tau)]$, where $R_{xx}(\tau)$ is the input ACVF.

RESULTS

In order to illustrate the range of possibilities of SAMSON, noise of all the mentioned standard types was generated. In addition, the ACVF of measured noise of different origins was used as input to the program, and the noise according to this ACVF was generated. The results are shown in Fig. 2. It can be seen that the agreement of the input ACVF with the ACVF of the generated noise is good. This is also true for the power spectra. The deviations are all within the Bartlett confidence intervals (Eqn. 19), which are not depicted in Fig. 2 to avoid confusion. In addition, for different numbers of noise data of the four standard noise types, the ACVFs obtained were never

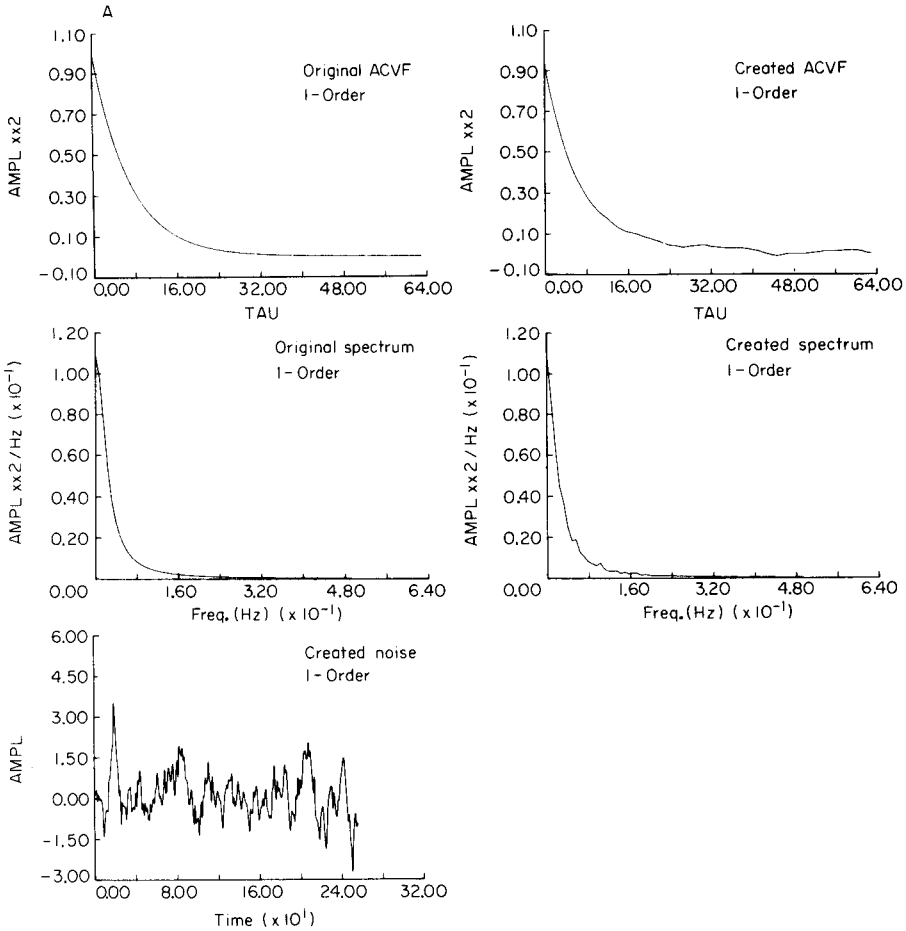


Fig. 2. Results of noise generation by SAMSON. At the left of each part (A—F) are shown the desired ACVF and the corresponding power spectrum, which are used as input to SAMSON; for ACVFs of measured noise, a record of the noise is also shown. At the right of each part, the resulting noise, its ACVF and power spectrum are depicted. Thus the similarity of input and output functions can be compared easily. (A) First-order noise (expression 15). (B) Gaussian noise (16). (C) Damped cosine noise (17). (D) 1/f noise (18). (E) Noise from the light signal leaving the monochromator of an inductively-coupled plasma emission spectrometer. (F) Noise originating from a flame ionization detector.

significantly different from the input ACVF used. The figures of the ACVFs and power spectra are based on 5000 noise data generated. Only a part of these 5000 numbers is shown in the figures. Table 2 lists the parameters for the standard ACVFs of Fig. 2.

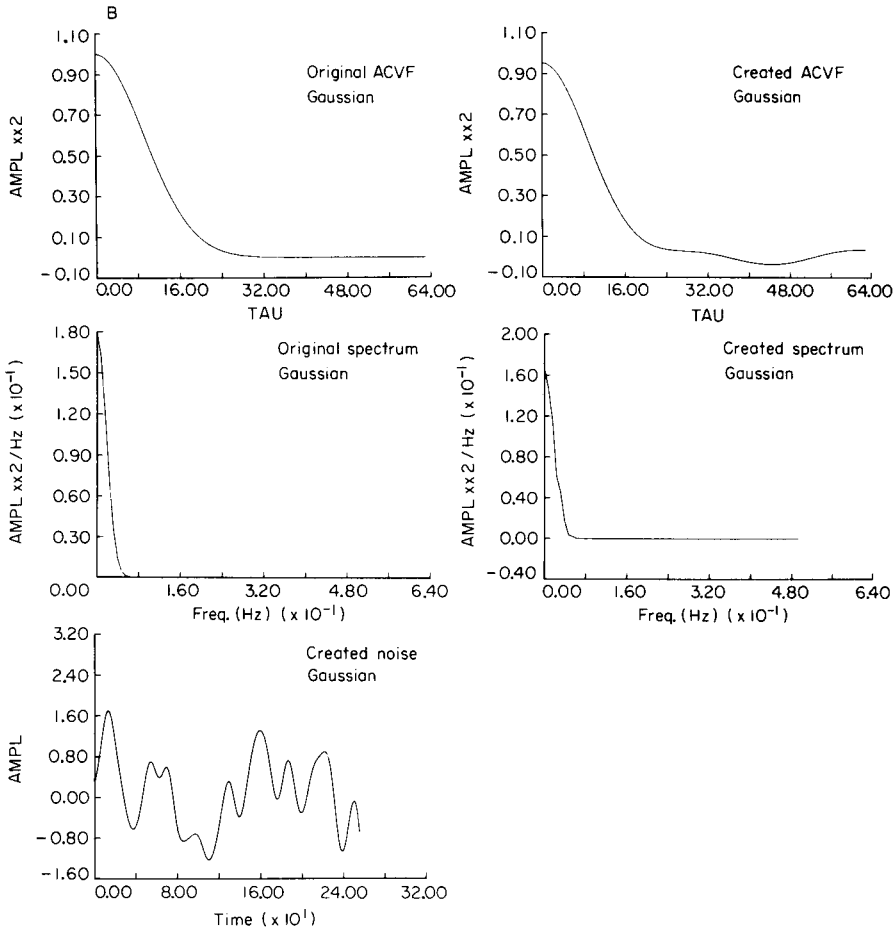


Fig. 2B.

APPLICATION TO TESTING OF PEAK-FINDING PROCEDURES

To demonstrate the possible applications of SAMSON, a realistic example is used. In chromatography, various attempts are being made to replace the usual visual interpretation of the chromatogram by automatic interpretation. Of course, to do so, the algorithm designer must formulate the problem of rather complex interpretation in a systematic manner. Many steadily improving peak-finding procedures have been developed. The main drawback of even the most sophisticated of these procedures is that the assumed noise models are generally not compatible with the noise observed in practice. For this reason, it is necessary to test the performance of the peak-finding procedure in use with different types of noise. SAMSON can readily be used for this purpose, and as a demonstration a first screening test of a peak-finding procedure is given below.

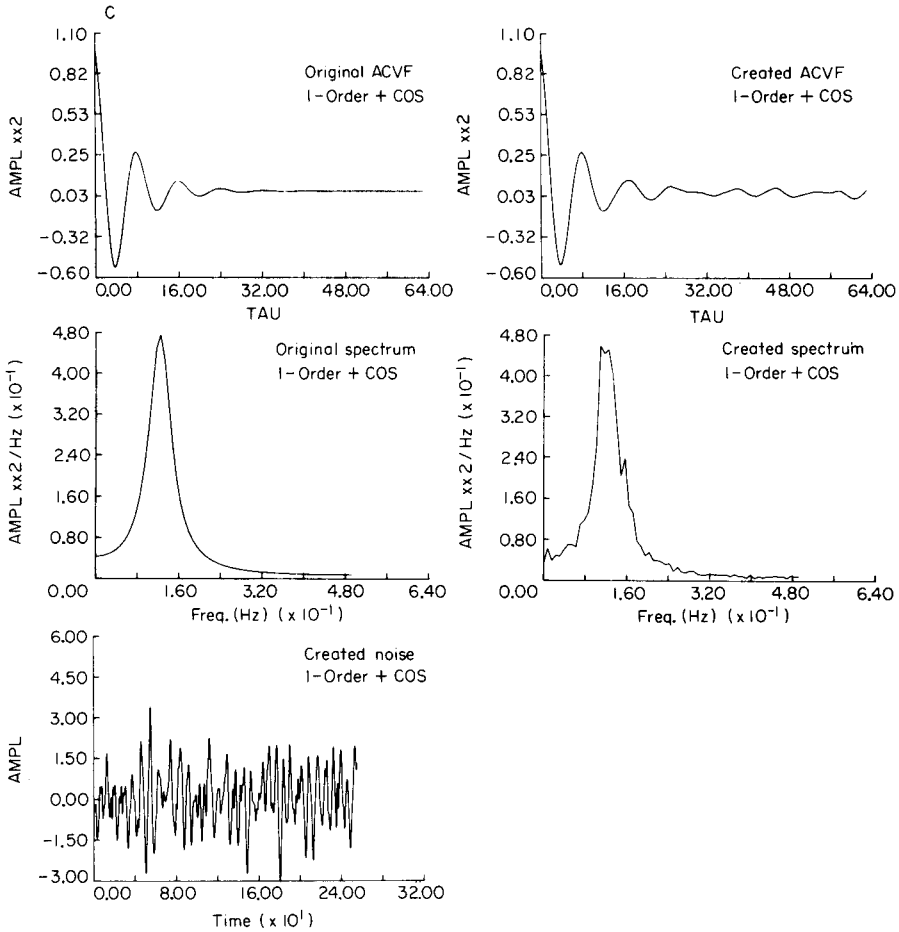


Fig. 2C.

A slightly modified version of the simple peak-finding procedure developed by Bobba and Donaghey [6] was used; the procedure was originally intended to handle chromatographic data with a microcomputer. The purpose of this test procedure is not to present an exhaustive statistical study on peak-finding procedure testing, but mainly to demonstrate the use of SAMSON.

Algorithm

The peak-finding program starts reading a certain preset number of data points, and evaluates the standard deviation and the approximated derivative ($\Delta x/\Delta t$) to determine linear drift. If these values do not exceed the preset values, the chromatographic system is considered to be stable. If they exceed the preset values, a new series of data is read and the procedure is

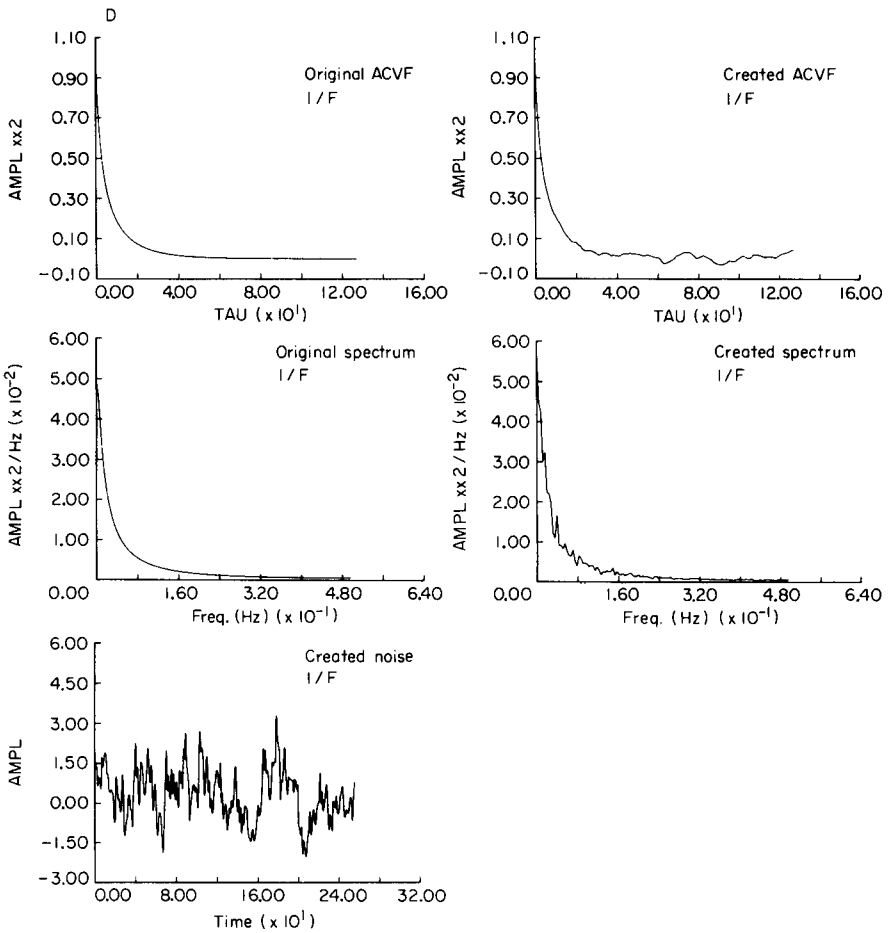


Fig. 2D.

repeated. After stabilization, detection of the first peak-start begins. A new preset number of data (equal to the preset parameter window length) is read and the standard deviation (σ_n) is evaluated. Peak start is detected if the last read value exceeds $K\sigma_n$, where K is a threshold parameter which can be chosen freely by the user (e.g., a read value > baseline + $3\sigma_n$ means that the probability a peak is present is >99.72%). If the last value does not exceed the threshold, another value is read until it does. After peak start detection, new values are read; the peak maximum is detected if a data point is read, which is $3\sigma_n$ smaller than the preceding value, which is considered to be the peak maximum. Peak end is detected if the approximated derivative of two successive values exceeds the slope determined at the peak start. The approximated derivative is defined as $x_j - x_{j-1} + K\sigma_n$. The peak-finding procedure is then repeated to the end of the data.

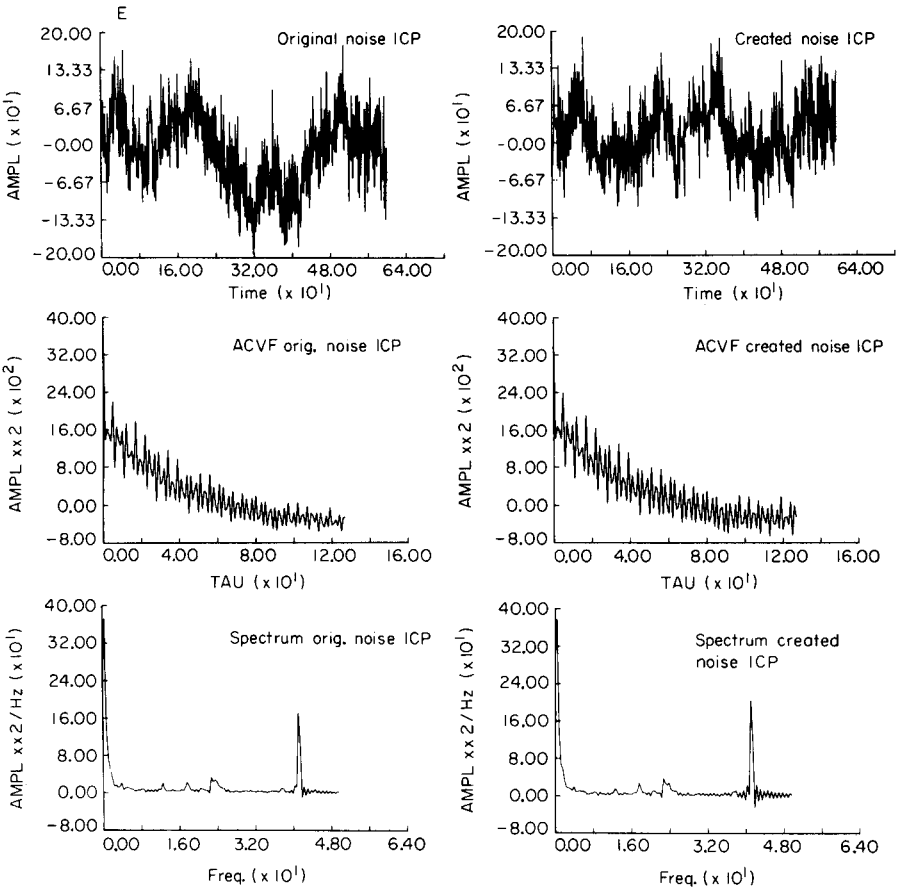


Fig. 2E.

Test conditions

A test chromatogram (400 data points) with 6 peaks was simulated (Fig. 3; Table 3) according to the model given by Fraser and Suzuki [7]

$$\phi(t) = H \exp\left([-\ln 2/A^2]\ \{\ln\left[1 + A\ (t - t_R)/\{\sigma_p\ (2 \ln 2)^{1/2}\}\right]\}^2\right) \qquad (20)$$

Peak starts, peak ends and the retention times were stored to be compared with the parameters detected by the peak-finding procedure. Peak end and peak start for the unresolved peaks were located at the minimum between the respective peak maxima.

Subsequently, five noise files of different types, containing 40 000 data each, were generated with SAMSON on the basis of an ACVF, length 64, and $\sigma_n^2 = 1.0$ (Table 4). The following test procedure was then chosen: for each of the five noise types, 400 data were added to the test chromatogram:

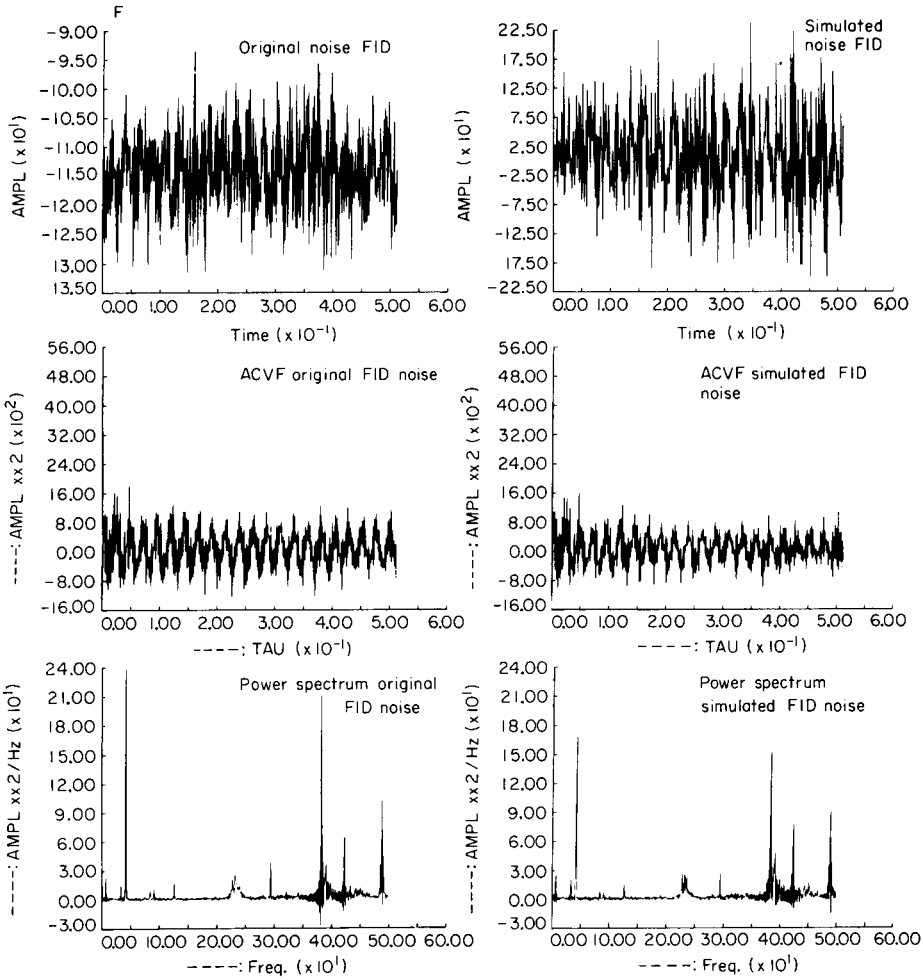


Fig. 2F.

noisy chromatogram = chromatogram + $p \times$ noise, where p is 1.0, 0.3 or 0.05. Thus the S/N ratios (maximum amplitude/σ_n) for the highest peak were about 10, 33 and 200. This test procedure was repeated 100 times with new noise data. Then, each noisy chromatogram was subjected to the peak-finding procedure six times, with input parameters: window length (4, 8 and 12), and K factor (threshold parameter, 1.0 and 2.0). If K was 3.0, bad results were always obtained in a preliminary study. The performance of the peak-finding procedure was quantified by awarding scores 1 to 10 for each peak parameter (peak start, peak maximum and peak end) correctly detected when a peak was in fact present. If the absolute difference ΔQ (time units) between the detected peak parameter and the real value was <10, then the detection was awarded $10-\Delta Q$, yielding a maximum score of 180 (six peaks, three parameters each), which was expressed as 100% performance.

TABLE 2

Parameters of the ACVFs shown in Fig. 2

Noise type	σ^2	T_x	ω_0
First order	1.0	8.0	
Damped cosine	1.0	6.0	0.77
Gaussian	1.0	12.0	
1/f	1.0	30 and 0.1	

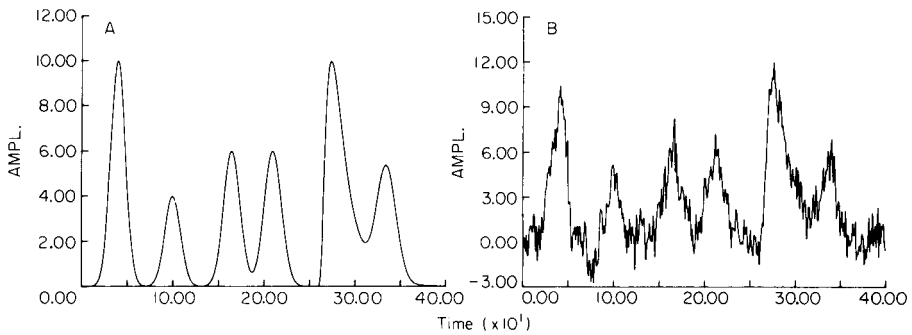


Fig. 3. (A) Simulated chromatogram to test the peak-finding procedure. (B) Test chromatogram with noise added (S/N for the highest peak is 10).

TABLE 3

Peak parameters of the chromatogram in Fig. 3

Peak No.	t_R	σ_{Peak}	A	Ampl. max.
1	40	8.0	0.0	10
2	100	8.5	0.0	4.0
3	165	9.0	0.0	6.0
4	210	9.5	0.0	6.0
5	275	10.0	0.7	10.0
6	335	10.5	0.0	5.0

Summarizing, the peak-finding procedure performance was tested with a chromatogram contaminated with five different noise types at three S/N levels. Further, the peak-finding procedure parameters, window length and threshold value, were varied. For each noise type, at each level, and for each individual K factor and window length, the test was repeated 100 times to obtain reliable results. Thus, altogether the peak-finding procedure was run 9000 times.

TABLE 4

Parameters of the ACVFs of the noise types used in the test chromatogram

Type of noise	σ_N	Time const.	ω_0	Spikes	
				τ	ACVF(τ)
First-order	1.0	8.0	—	—	—
Damped cosine first-order	1.0	8.0	0.555	—	—
Gaussian	1.0	3.0	—	—	—
1/f	1.0	0.1			
		10.0	—	—	—
First order + spikes	1.0	8.0	—	10	0.5
				20	0.2
				30	0.07
				40	0.02

Results

The scores of the peak-finding procedure under the different test conditions are summarized in Table 5. The mean score $\bar{x}$ (%) of the 100 runs is given. The 95% confidence interval on μ is $\bar{x} - 1.98\ s_{\bar{x}} < \mu < \bar{x} + 1.98\ s_{\bar{x}}$, where $0.4 < s_{\bar{x}} < 1.4$. The mean scores are significantly different at the 90% confidence interval if they differ (between noise types) more than 1.5, according to the Smith—Satterthwaite method [8].

A critical look at the table reveals some features of the peak-finding procedure under test. The deviation of the detected peak parameters averaged over the three S/N levels for K factor and window length where the procedure has optimal performance, is 5.5 (time units), which is about 0.6 σ_p. This implies a rather low detection performance. The deviation is almost independent of the S/N ratio: 6.1, 5.6 and 4.8 for S/N ratios of 10, 33 and 200, respectively.

TABLE 5

Performance of the peak-finding procedure applied to the test chromatogram contaminated with noise of different types. The score of the procedure is 100 if all peak parameters are detected without errors

σ_{Noise}	1.0						0.3						0.05					
K factor	1			2			1			2			1			2		
Window	4	8	12	4	8	12	4	8	12	4	8	12	4	8	12	4	8	12
First order	21	35	35	29	36	36	24	35	39	34	42	41	23	36	43	35	48	51
Spiked first order	19	35	37	23	38	39	21	37	42	28	43	43	18	37	43	32	50	51
Gaussian	25	34	39	32	34	33	27	36	38	34	39	38	23	36	39	39	51	49
Damped cosine	39	44	44	37	29	32	35	46	48	39	41	42	29	47	50	44	53	53
1/f	21	39	37	31	35	35	23	45	45	38	46	44	24	42	47	37	52	52

If the performance of the peak-finding procedure is considered in relation to the type of noise, some different features can be observed. First, it can be seen that the performance is approximately equal if the chromatogram is contaminated with first-order or spiked first-order noise; the optimal procedure parameters, K factor and window length, are about equal in both cases. The same is true for a chromatogram contaminated with Gaussian or 1/f noise, although the performance with 1/f noise is better. Table 6 gives an overview of optimal parameter combinations for the different types of noise.

The test procedure described shows some features of the peak-finding procedure under different noise circumstances. The procedure yields the parameters that have to be chosen to obtain the best performance for each type of noise and S/N ratio. An overall conclusion is that the algorithm for the peak-finding procedure needs improvement, because its performance is poor. Of course, simple addition of noise to simulated chromatograms is no guarantee of a realistic simulation of real chromatograms: care has to be taken that the peak profiles are not affected significantly by the filter. Thus, the chosen noise characteristics should be in accordance with the profile of the chromatogram.

Of course, the test procedure is too simple to obtain an exact picture of the behaviour of the peak-finding procedure; for example, "misses" and "false alarms" are not detected, nor is the performance as a function of peak resolution. Yet, the test procedure appears to be satisfactory as a preliminary screening method.

TABLE 6

Recommended window length and threshold value for different types of noise and S/N ratios

Type of noise	S/N	K-factor		Window	
		1	2	8	12
First-order	10		x	x or x	
	33		x	x or x	
	200		x		x
Spiked	10		x	x or x	
first-order	33		x	x or x	
	200		x	x or x	
Gaussian	10	x			x
	33		x	x or x	
	200		x	x or x	
Damped cosine	10	x		x or x	
	33	x		x or x	
	200		x	x or x	
1/f	10	x		x or x	
	33	x or x		x or x	
	200		x	x or x	

CONCLUSIONS

The computer program SAMSON is a useful tool for many purposes. The property of SAMSON to generate stationary noise with zero mean on the basis of an input (or standard) autocovariance function allows significant use of prior knowledge about analytical problems. The results presented above indicate that the agreement of the input ACVF with the resulting ACVF of the generated noise is good. This implies that it is possible to simulate noise with the same statistical properties as the noise of some source measured in practice. Consequently, software packages for many purposes can be tested exhaustively in circumstances giving a reliable picture of analytical practice. The example of testing a peak-finding procedure under different noise conditions indicates that the simulation of different kinds of noise to test the procedure is worth the extra effort: significant differences in performance with respect to different types of noise are observed. It would be useful to develop a more sophisticated test procedure for peak-finding procedures, because such a test procedure allows careful examination of the existing procedures and permits an optimal choice by the user for particular laboratory purposes.

Applications of SAMSON are possible in several fields. In information extraction methods, it can be used for simulation of detector signals or spectra interpretation. In data processing, it serves in filter procedures (e.g., Kalman filtering, forecasting). In systems theory, it may be useful in studies of the influence of different noise types on the output of a system and on the possibility of controlling a process, if the noise is applied at the input. In quality control, it may help in the development of test procedures to measure objectively the performance of software packages for information extraction. The steadily increasing computing and memory capacity of microcomputers allows development of more sophisticated methods for extraction of analytical information at low cost. The development of many of these methods may be aided by SAMSON.

REFERENCES

1 H. C. Smit and H. L. Walg, Chromatographia, 8 (1975) 311.
2 R. Deutsch, System Analysis Techniques, Prentice Hall, Englewood Cliffs, NJ, 1969, Ch. 5.
3 H. C. Smit and H. L. Walg, Chromatographia, 9 (1976) 483.
4 M. Abramowitz, Handbook of Mathematical Functions, Dover, New York, 1970, pp. 228, 231.
5 M. S. Bartlett, J. R. Stat. Soc., 138 (1946) 27.
6 G. M. Bobba and L. F. Donaghey, J. Chromatogr. Sci., 15 (1977) 47.
7 R. D. B. Fraser and E. Suzuki, Anal. Chem., 41 (1979) 37.
8 J. S. Milton and J. J. Corbet, Applied Statistics with Probability, Van Nostrand, London, 1979, p. 308.

Analytica Chimica Acta, 157 (1984) 291—302
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

ANALYTICAL EVALUATION OF TOTALLY PYROLYTIC GRAPHITE CUVETTES FOR ELECTROTHERMAL ATOMIC ABSORPTION SPECTROMETRY

D. LITTLEJOHN, I. DUNCAN, J. MARSHALL and J. M. OTTAWAY*

Department of Pure and Applied Chemistry, University of Strathclyde, Cathedral Street, Glasgow G1 1XL (Great Britain)

(Received 7th September 1983)

SUMMARY

The performance characteristics of electrothermal atomiser cuvettes made of totally pyrolytic graphite (TPG) are compared to those of coated and uncoated electrographite. An analytical programme was devised to determine the useful operational lifetime of each cuvette and assess the effect of cuvette age on the sensitivity and precision of the determinations of lead, manganese and vanadium by atomic absorption spectrometry. The main advantages of TPG are enhanced cuvette durability and improved sensitivity and precision, especially for involatile elements. The characteristics of uncoated electrographite were generally unsatisfactory compared to the pyro-coated and TPG cuvettes studied.

Although atomisation from metal surfaces has been investigated by various workers, it is widely accepted that graphite is at present the most suitable material for atomiser cuvette fabrication for electrothermal atomisation. Standard cuvettes supplied by all manufacturers of atomic absorption spectrometers are normally machined from electrographite and this traditional form of analytical-grade graphite has proved satisfactory for the volatilisation and atomisation of many analyte species. However, it has long been recognised that the porous structure of electrographite allows loss of atoms by gaseous diffusion [1], and the soaking of solutions into the substrate tends to promote carbide formation by elements such as molybdenum and vanadium, which leads to inefficient atomisation.

The coating of electrographite with a layer of pyrolytic graphite was suggested by L'vov in 1967 [2] to overcome these detrimental effects. Since then numerous procedures for the coating of graphite have appeared. Some workers have described the benefit of atomisation from a metal carbide surface such as tantalum carbide [3, 4], particularly for the control of certain interference effects, but pyrolytic graphite has received universal acceptance as the most useful coating. Accordingly, most manufacturers of electrothermal atomisers supply pyro-coated electrographite cuvettes as an alternative to the standard uncoated version. The problem with coated tubes,

0003-2670/84/$03.00

however, is that with use the pyrolytic graphite layer is removed, and hence variable sensitivity and precision can be encountered for many elements during the lifetime of the cuvette.

There is clearly a need for an atomiser material which gives extended cuvette lifetime, improved sensitivity and more consistent atom formation than current forms of graphite used in electrothermal atomisation. In this respect, the apparently attractive features of glassy carbon have been investigated [5, 6]. Unfortunately, measurements indicate that with present designs of atomiser power supply, slower heating rates and lower maximum temperatures are encountered with glassy carbon cuvettes compared to those of electrographite [5, 6]. In the past few years, Lersmacher and co-workers have investigated various procedures for the fabrication of electrothermal cuvettes. The forms of graphite studied have included microporous glassy carbon, glassy carbon [7] and totally pyrolytic graphite [8]. Studies of the heating characteristics of cuvettes prepared from these materials [9] involved measurements of the power and energy required to heat cuvettes of different types to typical temperatures required for atomisation. Of the materials studied, totally pyrolytic graphite (TPG) was found to be the most beneficial alternative to electrographite for cuvette fabrication. In particular, the heating rate and maximum temperature that could be achieved with TPG, at a given applied voltage, was greater than for other forms of graphite. This was explained by the development of a power equation [9] relating the heating rate (dT/dt) of a cuvette to the applied voltage (V), resistance (R), mass (m) and specific heat (σ) of the graphite tube, i.e., $dT/dt = V^2/Rm\sigma$. Because the specific heat is similar for different forms of graphite at any given temperature, the heating rate is inversely proportional to Rm.

It was shown [9] that $Rm = L^2\rho D$, where L is the length, and ρ and D are the resistivity and density, respectively, of the graphite cuvette. For a constant tube length, the heating rate at any applied voltage will be greatest for the least dense graphite. The resistivity and density of different forms of carbon and graphite are given in Table 1. The values quoted for electrographite and the glassy carbons are similar to those published by de Galan et al. [6]. Of the various forms of carbon tabulated, TPG has the lowest density. Clearly, TPG should be attractive as an atomiser material because

TABLE 1

Density and resistivity of different forms of graphite [10]

Graphite	Density (D) ($g\ cm^{-3}$)	Resistivity (ρ) ($10^{-4}\ \Omega\ cm$)	ρD ($10^{-4}\ \Omega g\ cm^{-2}$)
Electrographite	1.78	8.05	14.29
Pyro-coated electrographite	1.77	8.18	14.46
Glassy carbon	1.48	40.09	59.21
Microporous glassy carbon	1.41	30.91	43.64
Total pyrolytic graphite	2.02	2.50	5.05

Erratum

D. Littlejohn, I. Duncan, J. Marshall and J. M. Ottaway, Analytical Evaluation of Totally Pyrolytic Graphite Cuvettes for Electrothermal Atomic Absorption Spectrometry.

Anal. Chim. Acta, 157 (1984) 291—302.

p. 292, last line before Table 1 should read:

*"TPG has the lowest value of ρD"

* Analytica Chemica Acta 162, 198 : 457 280118

rapid heating rates are necessary for efficient atom formation especially with platform atomisation. In addition, TPG is impervious and fairly robust and has low porosity; although gradual evaporation of graphite is expected during use, the totally pyrolytic nature of the atomisation surface will be consistent throughout the lifetime of the cuvette. Therefore, TPG should give improved performance over pyro-coated electrographite, especially for involatile elements.

It was considered desirable to compare the analytical characteristics of electrographite, pyro-coated electrographite and totally pyrolytic graphite cuvettes for the same atomiser under similar operating conditions. This would provide a reliable indication of the practical advantages of TPG for atomic absorption spectrometry (a.a.s.). An evaluation programme was devised to establish the useful lifetime of these three types of graphite cuvette and to measure any changes in the sensitivity and precision of absorbance signals for lead, manganese and vanadium caused by a deterioration of the atomiser surface during use.

EXPERIMENTAL

Instrumentation

A Pye-Unicam SP-9 spectrometer and atomiser were used with a Pye-Unicam 9095 video programmer and FAS-I autosampler. The commercial, one-piece electrographite and pyro-coated cuvettes were 30 mm in length, with an internal diameter of 5 mm and wall thickness in the range 790—860 μm. The total pyrolytic graphite cuvettes were prepared by Lersmacher by pyrolytic deposition [8]. The TPG cuvettes were of the same length and internal diameter as the commercial tubes, but of 320-μm wall thickness to give a similar electrical resistance ($R = L\rho/X$, where X is the cross sectional area of the graphite [9]). The TPG cuvettes were prepared as cylinders without end rings. Rings of electrographite were fitted over each end of the cylinder to allow electrical contact via the standard design of contact flange used with the SP-9 atomiser. The cuvette was subjected to a second short deposition period to give good electrical contact between the rings and the tube.

The SP-9 furnace power supply is normally calibrated so that the set temperature is achieved when electrographite tubes (coated or uncoated) are used. Because of a slight difference in the conductivity and heating characteristics of TPG, temperatures obtained with the TPG cuvettes for voltage-controlled heating were lower than the programmed values for settings less that 1800—2100°C. Temperatures measured for settings greater than 1800—2100°C were higher than the programmed value. The temperature discrepancy at both ends of the temperature scale was normally not more than 50—150°C and occurred only in the voltage-control mode. In the 1800—2200°C range, the set and measured temperatures for the TPG cuvettes were the same. Under temperature (emissivity)-controlled heating, the set and measured temperatures were similar to within 50°C for all cuvettes studied.

Evaluation programme

The test experiments were based on the format of a repeated cycle which was continued throughout the lifetime of a tube. The cycle involved the measurement of seven absorbance signals for lead at an atomisation temperature of 2300°C, then seven signals for manganese at 2500°C, and finally seven signals for vanadium at 2800°C. In addition to this sequence, temperature measurements with the pyrometer, blank runs and occasional check signals with a second concentration of each element, brought the average number of atomisation firings per cycle to about 32. These cycles were repeated for each tube until it broke, or until the relative standard deviation reached 20% or more. The same spectrometer conditions were set each time that signals for a particular analyte were measured. The temperature programmes applied for each tube were the same, and are given in Table 2.

Performance characteristics were evaluated for each cuvette from the following data.

Total atomisation firings. The number of atomisation stages at 2300°C, 2500°C and 2800°C were counted. This did not include cleaning stages, which involved "gas flow" operation.

Total atomisation lifetime. This is probably a more important characteristic than the number of atomisation stages because the duration of heating is taken into account. The total length of time of all heating stages to 2300°C, 2500°C and 2800°C, plus the time of the cleaning stages, was calculated for each cuvette.

Weight loss. Each cuvette was weighed before and after use. The weight loss was calculated and expressed as mg s^{-1} of atomisation lifetime.

A.a.s. sensitivity for Pb, Mn and V. The mean peak height and peak area signals obtained during each cycle were calculated (n = 7; 20-μl samples). This allowed a sensitivity comparison between the different cuvettes and indicated the effect of tube age on signal magnitude. Solution concentrations were chosen to give absorbances $\leqslant 0.4$. This was achieved with concentrations of 20 $\mu g\ l^{-1}$ lead, 2 and 4 $\mu g\ l^{-1}$ manganese, and 100, 200 and 500 $\mu g\ l^{-1}$

TABLE 2

Conditions for electrothermal atomisation

Phase	Lead (283.3 nm)			Manganese (279.5 nm)			Vanadium (318.4 nm)		
	Temp. (°C)	Ramp No.	Hold (s)	Temp. (°C)	Ramp No.	Hold (s)	Temp. (°C)	Ramp No.	Hold (s)
Dry	130—280	6	45—60		as Pb			as Pb	
Atomise[a]	2300	0[b]	5[c]	2500	0[b]	5[c]	2800	0[b]	6 or 8
Clean[a]	2700	0[b]	3	2700	0[b]	3	2800	0[b]	5 or 3

[a]Temperature-controlled heating, 20-μl injection volumes. [b]Zero ramp equivalent to heating rate of 2000°C s^{-1}. [c]Gas-stop conditions.

vanadium. All measurements were made using the Pye-Unicam 9095 computer and printer which were coupled to the SP-9 spectrometer.

Precision of signal measurement. The relative standard deviations of the peak height and peak area signals were calculated for each cycle, to illustrate the effect of the atomisation surface and tube age on precision.

Reagents and temperature measurements

Stock solutions were prepared by dissolving appropriate salts of the highest available purity in dilute nitric acid to give a final acid concentration of not less than 10^{-2} M. Working solutions were freshly prepared from these stock solutions as required by dilution with distilled water.

Temperatures were measured with an Ircon series 1100 optical pyrometer. Slight variations in the emissivity of the cuvettes were ignored and temperatures were calculated on the basis of an emissivity of 1.

RESULTS AND DISCUSSION

Cuvette lifetime

The number of atomisation firings and the total atomisation lifetime achieved with each cuvette are given in Table 3. The two TPG cuvettes studied gave 170—200 atomisation firings more than the standard electrographite cuvette and 60—100 more than the pyro-coated version. There is a difference between the number of cycles obtained with the two TPG cuvettes because an additional number of atomisation sequences was required with the first TPG cuvette to establish satisfactory operating conditions, measure tube temperatures and check the linearity of response. Consequently, the average number of atomisation stages per cycle was greater for the first TPG cuvette. However, calculation of the total duration of atomisation and cleaning stages showed that the total lifetimes of the TPG cuvettes were similar. This parameter is probably a better indicator of cuvette durability than the number of atomisation firings because the number of atomisation sequences

TABLE 3

Lifetime of SP-9 cuvette

Type	No. of cycles	Total no. of atomisation stages	Total atomisation lifetime (s)	Weight loss (mg s^{-1})
Electrographite	4	134	1186	0.0272
Pyro-coated electrographite	8.5	243	2181	0.0165
Total pyrolytic graphite 1	10	349	2869	—
Total pyrolytic graphite 2	11	308	2713	0.0158

obtained with a cuvette will depend on the temperature and atomisation period selected by the operator.

The results presented in Table 3 show that TPG increased the useful cuvette lifetime by 500 to 700 s compared to pyro-coated electrographite, and by 1700 s compared to the standard uncoated tube. The protocol deliberately included absorbance measurements at three atomisation temperatures to simulate a procedure for the determination of three elements of different volatility in aqueous samples. This general trend has been confirmed by similar tests done by other workers [10]. Although the data obtained are probably representative of the relative performance of each type of graphite cuvette studied, it is likely that corrosive or salt-containing matrices would decrease the useful lifetime of all the cuvettes.

The rate of weight loss incurred by the TPG cuvettes was substantially lower than that of the electrographite tube, but only slightly better than that of the pyro-coated tube. Given the much smaller wall-thickness of the TPG cuvettes (320 μm) compared to the others (800 μm) it is probable that a slightly thicker TPG wall would have given a longer lifetime than 2800 s. The weight loss rate for the pyro-coated SP-9 cuvette was less than for a pyro-coated Perkin-Elmer HGA-500 tube (0.038 mg s^{-1}) subjected to the same evaluation procedure. An HGA-500 tube coated by the Philips/Pye-Unicam procedure [8] gave a much lower loss rate (0.016 mg s^{-1}), similar to the coated SP-9 tube.

Comparison of sensitivity

The mean peak height and area signals (n = 7) for each cycle obtained from the SP-9 computer are presented in Figs. 1—3. Because the signals obtained with the two TPG cuvettes were similar, only one set of results is presented for each element.

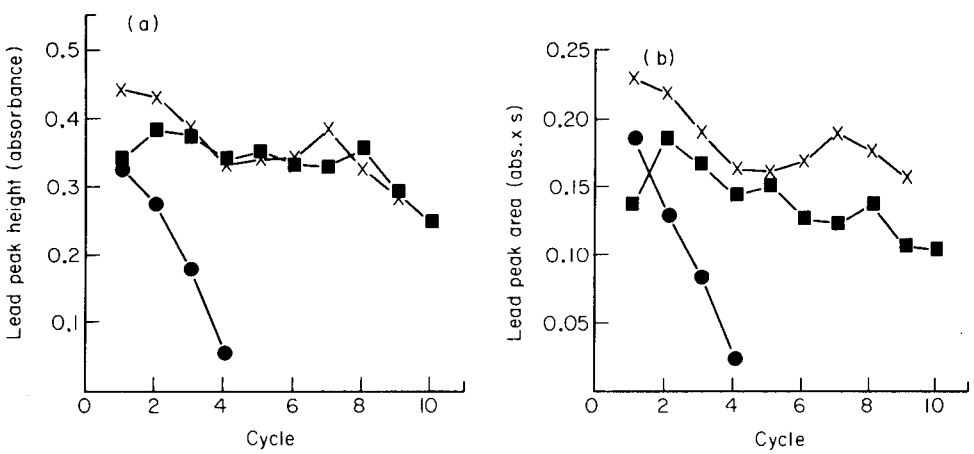


Fig. 1. Change in (a) mean peak height and (b) mean peak area signals for lead (20 μl of a 20 μg l^{-1} solution) during the lifetime of: (•) electrographite; (×) pyro-coated electrographite; (▪) TPG cuvettes.

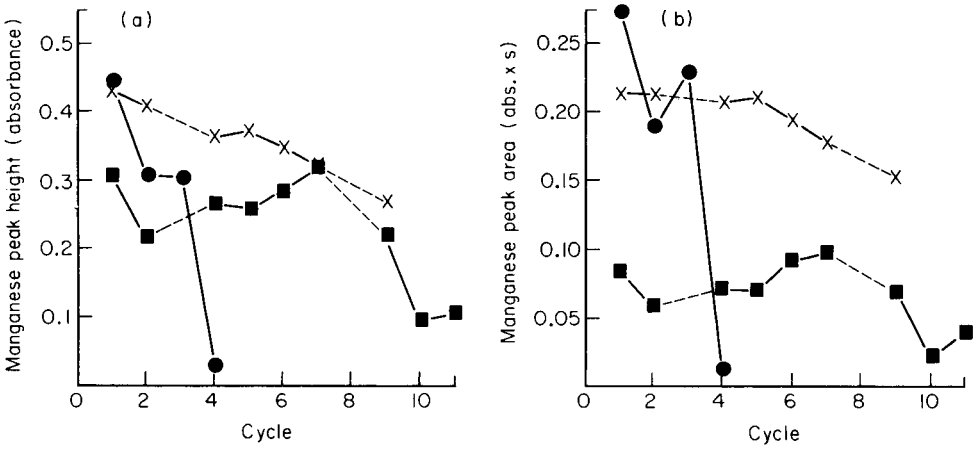


Fig. 2. Change in (a) mean peak height and (b) mean peak area signals for manganese during the lifetime of: (•) electrographite; (X) pyro-coated electrographite; (▪) TPG cuvettes. (Values for cycles 3 and 8 missing because of autosampler fault.) See text for manganese concentrations used.

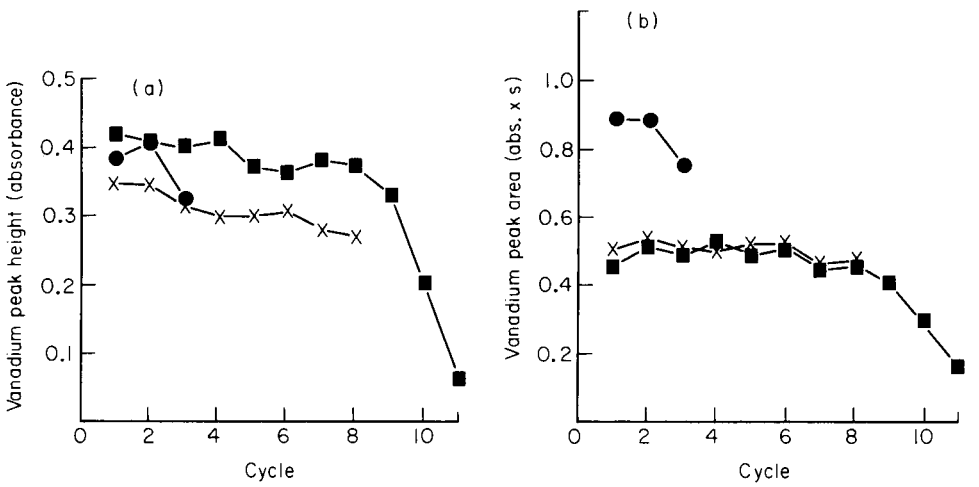


Fig. 3. Change in (a) mean peak height and (b) mean peak area signals for vanadium during the lifetime of: (•) electrographite; (X) pyro-coated electrographite; (▪) TPG cuvettes. See text for vanadium concentrations used.

Lead. Initially peak height sensitivity was best for the pyro-coated cuvette, but after cycle 2 the performances of the TPG and pyro-coated cuvettes became almost identical (Fig. 1a). A similar set of mean peak height signals was obtained for both the TPG cuvettes studied. The uncoated electrographite cuvette showed a surprising decrease in the mean peak height absorbance from cycle to cycle, although within-cycle precision was reasonable (see Table 5). Similar tube-age trends were observed for peak area signals (Fig. 1b).

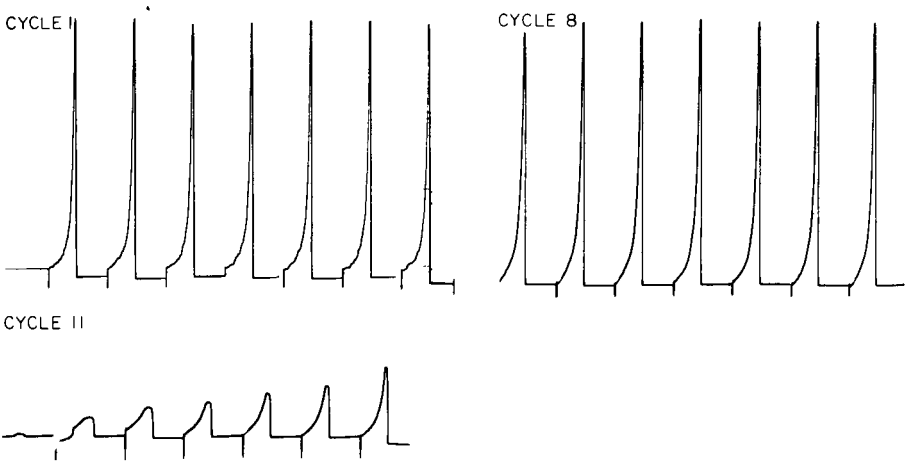


Fig. 4. Recorder tracings for repeated 20-μl injections of 100 μg l^{-1} vanadium obtained during cycles 1, 8 and 11 with a TPG cuvette. The first peak of each cycle is at the right hand side of each set of signals.

Manganese. The pyro-coated tube gave the greatest peak height sensitivity for manganese (Fig. 2a), but there was a general decrease in the mean absorbance from cycle to cycle. The change in sensitivity was not apparent within each cycle, and it probably reflects a long-term effect of tube age on signal magnitude. A more dramatic fall-off in the mean peak height absorbance was observed for the uncoated electrographite cuvette. The sensitivity achieved with TPG was slightly poorer than for pyro-coated electrographite, but the mean peak height signal was more consistent, decreasing only in the last two cycles. A 2 μg l^{-1} manganese solution was used with the TPG and pyro-coated cuvettes. A 4 μg l^{-1} solution was required with the uncoated electrographite tube to give a similar absorbance.

Similar trends were observed for the manganese mean peak areas (Fig. 2b), although a greater difference was observed in the sensitivity obtained with the pyro-coated and TPG cuvettes than was found for peak height. It was also apparent that peak area signals for the pyro-coated and TPG cuvettes were more consistent from cycle to cycle than the height values. This suggests that the difference in the mean manganese peak height between cycles was related more to changes in the rate of atom formation rather than to differences in the total number of atoms formed.

Vanadium. The TPG cuvette gave by far the best sensitivity for vanadium. Similar peak height absorbances were obtained for 20 μl of 100, 200 and 400 μg l^{-1} vanadium solutions in TPG, pyro-coated electrographite and uncoated electrographite cuvettes, respectively (Fig. 3a). Except for the last two cycles, the TPG signals were remarkably consistent throughout the lifetime of the tube. Specimen recorder tracings obtained for vanadium with the TPG cuvette during various cycles are shown in Fig. 4. A gradual decrease in the mean peak height was observed with the pyro-coated cuvette, but

Erratum

p. 299, lines 4 and 5 should read:

"... measurement*, the peak area sensitivity for the TPG cuvette is approximately twice that for the pyro coated cuvette".

* Analytica Chimica Acta, 162 (1984) 45

that the SP-9 atomiser gives slightly better sensitivity than the HGA-500, probably because of the smaller internal diameter and volume of the Pye-Unicam cuvette.

the change in sensitivity with age was not as great as might have been expected, which indicates the durable nature of the pyro-coating. The peak area signals (Fig. 3b) show a similar trend to the peak height values. Under the conditions of measurement, the peak area sensitivities for the pyro-coated and TPG cuvettes were almost identical.

Totally pyrolytic graphite is of most benefit for the atomisation of refractory, involatile analyte elements. A TPG cuvette gave a factor of 1.4 improvement over the pyro-coated cuvette in peak height sensitivity for molybdenum. A comparison of the various sensitivities for each of the cuvettes is given in Table 4. The values are representative of the general performance of each type of graphite as an atomiser material. The SP-9 values are compared in Table 4 to two sets of sensitivity data for the Perkin-Elmer HGA-500 furnace, one obtained in this laboratory for a pyro-coated cuvette, during a similar study with the HGA-500 to that reported here [11], the other taken from Perkin-Elmer literature [12]. These values indicate the benefit of TPG, especially for the determination of involatile elements. The data also show that the SP-9 atomiser gives slightly better sensitivity than the HGA-500, probably because of the smaller internal diameter and volume of the Pye-Unicam cuvette.

Comparison of precision

The relative standard deviation (r.s.d.) of peak height and peak area signals calculated by the SP-9 computer during each cycle are given in Tables 5—7. In general, better precision was obtained with the TPG cuvette, and peak height precision was better than peak area precision for the atomisation conditions chosen.

Lead. Peak height and area r.s.d. values for TPG were mainly less than 3% and often better than 2% (Table 5); only during the last two cycles did precision become less favourable. Although the peak height precision for the pyro-coated cuvette approached that of TPG, the peak area values were

TABLE 4

Sensitivities (concentration in $\mu g\ l^{-1}$ per 0.0044 absorbance) for lead, manganese, vanadium and molybdenum (20-μl injections)

Cuvette	Pb (273.3 nm)	Mn (279.5 nm)	V (318.4 nm)	Mo (313.3 nm)
SP-9 electrographite	0.32—0.50	0.06	5.5	—
SP-9 pyro-coated electrographite	0.25	0.02—0.03	2.9	0.35
SP-9 TPG	0.25	0.03—0.045	1.1	0.25
HGA-500 pyro-coated electrographite [11]	0.55	0.06—0.17	8.0—14.6	—
HGA-500 literature values [12]	0.65	0.14	1.5	0.45

TABLE 5

Precision of peak height and peak area signals for lead with different cuvettes (for results given in Fig. 1; $n = 7$)

Cycle	Relative standard deviation (%)					
	Electrographite		Pyro-coated		TPG	
	Height	Area	Height	Area	Height	Area
1	1.3	6.7	3.6	3.9	1.3	3.5
2	3.5	5.0	4.4	6.1	3.1	3.3
3	5.7	5.8	3.0	4.6	4.4	1.6
4	30.1	36.0	2.3	6.5	1.9	1.4
5			2.7	9.0	1.6	1.1
6			1.8	6.3	1.7	2.3
7			2.5	8.6	1.9	2.0
8			2.6	4.9	1.2	1.4
9			3.4	5.5	7.0	8.5
10			—	—	7.4	9.2

TABLE 6

Precision of peak height and peak area signals for manganese with different cuvettes (for results given in Fig. 2; $n = 7$)

Cycle	Relative standard deviation (%)					
	Electrographite		Pyro-coated		TPG	
	Height	Area	Height	Area	Height	Area
1	1.3	1.4	1.8	2.0	1.7	5.6
2	3.6	5.3	3.7	4.6	3.7	7.4
3	5.5	6.8	—	—	—	—
4	34.8	62.7	3.0	5.1	1.1	4.9
5			2.1	1.4	1.9	1.7
6			3.5	2.9	1.3	4.7
7			6.6	7.2	1.3	2.6
8			—	—	—	—
9			13.6	14.8	1.0	3.3
10			—	—	12.0	28.2

considerably worse, in the range 4—9%, similar to the uncoated cuvette. A consideration of the sensitivity and precision obtained suggests that the peak height mode will give better detection limits for the conditions used here.

Manganese. During the first two cycles, all the cuvettes gave similar peak height precision (Table 6). Thereafter precision was best for TPG (<2%). The r.s.d. values for peak areas were very variable for all the cuvettes studied (2—7%). From the sensitivity and precision data, it is apparent that the TPG and pyro-coated peak height signals will give the best detection limits for manganese.

TABLE 7

Precision of peak height and peak area signals for vanadium with different cuvettes (for results given in Fig. 3; $n = 7$)

Cycle	Relative standard deviation (%)					
	Electrographite		Pyro-coated		TPG	
	Height	Area	Height	Area	Height	Area
1	2.0	5.5	2.1	3.1	0.6	2.8
2	2.8	6.6	1.3	2.2	1.6	1.2
3	12.8	10.3	2.8	4.6	0.9	1.7
4			2.8	3.2	1.1	1.7
5			0.8	3.4	1.3	1.2
6			2.5	3.0	1.2	1.2
7			4.0	2.5	1.9	3.1
8			2.0	3.5	1.6	1.5
9					7.0	8.5
10					20.7	12.2

Vanadium. The r.s.d. values for peak height for each cuvette were similar to the respective peak area values. Again TPG gave the best performance, with r.s.d. values generally better than 2%. The values for pyro-coated electrographite were 2—4%. The uncoated version showed satisfactory precision for the first 2 cycles before a severe degradation of performance occurred. From the sensitivity and precision data, there is likely to be little difference in the peak height and area detection limits for vanadium obtained with TPG and pyro-coated cuvettes.

CONCLUSIONS

The evaluation protocol described was probably more severe than typical routines used for electrothermal atomisation in most laboratories. The absolute lifetimes achieved with different cuvettes cannot therefore be directly compared with practical lifetimes obtained, for example, when only volatile elements are determined. However, the results presented allow a comparison between the different types of SP-9 cuvettes available, which will be representative of their relative general performance, irrespective of the analyte determined.

The study clearly indicates that the uncoated electrographite SP-9 cuvette is less satisfactory than the other forms of carbon tube studied, especially for applications involving the use of high atomisation temperatures. In contrast, the lifetime, sensitivity and precision of the pyro-coated SP-9 cuvette is reasonable. The fabrication of cuvettes from totally pyrolytic graphite is undoubtedly an important development in electrothermal atomisation. The results presented indicate that TPG gives a welcome improvement in operational lifetime, sensitivity and precision compared to existing commercial

forms of SP-9 cuvette, especially for the atomisation of involatile elements such as vanadium and molybdenum. Minor refinements in the deposition and fabrication procedures are likely to increase the lifetime of TPG cuvettes further. Practical applications are currently being investigated, and the resistance of TPG to chemical attack, particularly in the presence of acid- and salt-containing matrices, will be reported in due course.

The preparation of totally pyrolytic cuvettes by Dr. B. Lersmacher, Philips Research Laboratories, Aachen, West Germany, is gratefully acknowledged, as is the assistance given by Mr. D. Widmer and Dr. M. P. Wassall of Pye-Unicam. This work was made possible by the loan of equipment by Dr. P. J. Whiteside, Pye-Unicam, and by the provision of financial support from the Pye Foundation (for DL) and SERC (for JM).

REFERENCES

1 B. V. L'vov, Atomic Absorption Spectrochemical Analysis, Hilger, London, 1970.
2 B. V. L'vov and G. G. Lebedev, Zh. Prikl. Spektrosk., 7 (1967) 264.
3 E. Norval, H. G. C. Human and L. R. P. Butler, Anal. Chem., 51 (1979) 2045.
4 T. M. Vickery, G. V. Harrison and G. S. Ramelow, Anal. Chem., 53 (1981) 1573.
5 J. Korečkova, W. Frech, E. Lundberg, J. A. Persson and A. Cedergren, Anal. Chim. Acta, 130 (1981) 267.
6 L. de Galan, M. T. C. de Loos-Vollebregt and R. A. M. Oosterling, Analyst, 108 (1983) 138.
7 W. F. Knippenberg, B. Lersmacher, H. Lydtin and K. Schelhas, Ger. Offen., NE 2702189, 1978; U.S. Pat., 4 204 769, 1980.
8 B. Lersmacher, Ger. Offen., 2 959 275 (Cl. GOINZI/03), 1981.
9 A. McLean, J. M. Ottaway, P. Connor and M. A. Wassall, in preparation.
10 P. Whiteside, T. Dymont and M. P. Wassall, to be published.
11 D. Littlejohn, J. Marshall, I. Duncan, J. Hendry and J. M. Ottaway, in preparation.
12 Analytical Methods for Furnace Atomic Absorption Spectroscopy, Perkin-Elmer, Norwalk, U.S.A., 1980.

Analytica Chimica Acta, 157 (1984) 303—311
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

THE DETERMINATION OF TRACE METALS IN HUMAN FLUIDS AND TISSUES
Part 3. Contamination in Sampling of Blood Serum and Liver Tissue and their Stability on Storage

V. HUDNIK

Boris Kidrič Institute of Chemistry, Ljubljana (Yugoslavia)

M. MAROLT-GOMIŠČEK

Faculty of Medicine, Edvard Kardelj University, Ljubljana (Yugoslavia)

S. GOMIŠČEK*

Faculty of Natural Sciences and Technology, Edvard Kardelj University, Ljubljana (Yugoslavia)

(Received 26th July 1983)

SUMMARY

Contamination in sampling human serum and liver tissue under hospital conditions is described. For the determination of copper, zinc and cadmium, the use of disposable syringe needles or Menghini needles is not critical, but the results for manganese in serum suffer from systematic errors. It is shown that contaminants introduced onto the surface during the manufacture of syringe needles contribute to this error. Storage of human serum at −20°C is beneficial for the determinations of copper, zinc, cadmium and manganese. The material of the storage vessel (glass or polystyrene) is then not critical.

Metals in human fluids and tissues are of increasing interest not only in medical investigations but also in everyday clinical practice. The sodium, potassium, calcium and magnesium contents are important but iron, copper, zinc, aluminium etc. should also be determined for diagnostic, prognostic and therapeutic purposes. These determinations are easily done in clinical laboratories, given suitable instrumentation. However, the collection of fluid and tissue samples must always remain in the hands of medical personnel for ethical reasons. The problems of contamination in sampling can, therefore, be serious and even critical when the concentrations of metals are low.

The storage of biological samples must be accepted as an integral part of the analysis because the concentration of a particular element can be changed by numerous interactions as well as by degradation of the material investigated. While these risks are less pronounced with lyophilised tissue samples, biological fluids must be stored with all precautions.

All these problems have been indicated and studied though their consequences are often neglected. Kurz et al. [1] reported that disposable syringes

0003-2670/84/$03.00

and needles were tested and found to be free of extractable zinc and copper. Speecke et al. [2] assumed that in every really clean analytical laboratory, samples can be treated for the determination of copper and zinc in blood serum, whereas the normal cleanliness or precautionary practices in a hospital ward or a chemical laboratory are completely inadequate for the determination of elements in the ng g^{-1} range (e.g., manganese). Reimold and Besch [3] established that Becton-Dickinson and Monoject plastic disposable syringes and glass or plastic tips from SMI micropipettes are zinc-free, whereas Vacutainer Tubes should not be used, if the sample is to be analysed for zinc. Versieck et al. [4] demonstrated that needle biopsies of the liver are heavily contaminated by numerous trace elements, while Meinel et al. [5] found that biopsies taken with a Menghini needle did not show any significant change that could be explained as contamination by the needle itself. Nooijen et al. [6] reported that the use of untreated disposable biopsy needles did not lead to a significant contamination of the biopsies with copper or zinc.

Omang and Vellar [7] examined the concentration gradients obtained for calcium, copper and sodium obtained in samples of serum, sweat and urine during storage at room temperature, at 4°C and at −20°C. They found out that a concentration gradient was produced in the sample tubes for copper and sodium even after 22 days at room temperature, whereas most of the calcium content precipitated from the serum and urine samples. Anand and Ducharme [8] reported that the chromium concentration of serum and plasma remained constant in polyethylene vessels, if stored frozen at below −10°C. Polycarbonate tubes were found adequate for storage of serum and plasma at 4°C for up to two weeks. Anand et al. [9] discussed changes in trace element concentrations in biological fluids and the factors to be considered in selecting a container that is made of an appropriate material. Fisher et al. [10] evaluated the effects of container composition, storage duration and temperature on serum calcium, magnesium, copper, zinc, sodium and potassium levels. Moore and Meredith [11] examined the adsorption of lead in blood and water samples on various container materials; they demonstrated that at lead concentrations of 2 μmol l^{-1} there was little adsorption of lead on the materials examined except for soda glass over one month. Majer and Khalil [12] found that calcium and magnesium ions at trace levels in solution were adsorbed more slowly on propylene containers than on glass containers.

The aims of the present study were to clarify the risks arising from contamination during sampling of blood and liver tissue for the determination of copper, zinc, cadmium and manganese, and to investigate the changes in concentrations of these elements in human serum during storage in glass and polystyrene vessels at normal and low temperatures.

EXPERIMENTAL

Blood samples were collected with disposable needles under hospital conditions. All glass vessels used were first cleaned with nitric acid and then rinsed thoroughly with redistilled water, with all precautions for extreme trace analysis. Polystyrene vessels (Spex) were used as received.

Liver tissue (pig) samples were taken by means of Menghini needles as well as by glass needles of the same shape which were prepared in the laboratory.

For storage of blood serum, 10-ml vessels made of borosilicate glass (Pyrex), soft soda glass (Vega) and polystyrene (Spex) were used; the vessels were of similar geometry. They were cleaned with nitric acid in a special apparatus made of quartz and rinsed with redistilled water. Aliquots (2 ml) of fresh pooled human serum were transferred to each of these vessels, which were then stored at room temperature and at −20°C for 1—30 days. Copper, zinc, cadmium and manganese were determined by a.a.s. For each temperature and each time, the determinations were done on two samples stored in separate containers of the same material, and from each sample two parallel measurements were made. All the determinations were conducted by atomic absorption spectrometry as described previously [13].

RESULTS AND DISCUSSION

In medical practice, blood samples are collected with disposable needles or plastic catheters, and tissues are taken with surgical instruments and Menghini needles. The samples can be contaminated by constituents of the plastic, glass or metal parts of these instruments. The risk of contamination can be especially critical insofar as the metal parts of the instruments are made from stainless steel which contains elements like chromium, nickel, manganese, etc. in high concentrations [14].

Concentration levels of copper and zinc in human serum and liver tissue are relatively high, so that the results are unlikely to be seriously affected by contamination. In a literature survey [14], agreement of copper and zinc levels in serum was shown to be satisfactory. Experiments done here, in which disposable needles were rinsed with redistilled water, did not show an observable increase of copper and zinc concentrations, all the results being below 0.002 $\mu g\ ml^{-1}$ for copper and 0.005 $\mu g\ ml^{-1}$ for zinc. However, normal precautions regarding the purity of reagents and vessels used, as well as the testing of needles, are necessary to avoid unexpected systematic errors.

The situation with regard to the determination of cadmium and manganese is different. Although the concentrations of these two elements in liver tissue are about an order of magnitude higher than the detection limit of the electrothermal a.a.s. procedures, their concentrations in human sera were in some cases below the detection limit of the procedures used. At such concentration levels, contamination can cause systematic errors, making the results entirely unreliable. Even a brief survey of the data given in the litera-

ture for cadmium and manganese levels in blood serum indicates that systematic errors for both elements can be appreciable [14, 15].

Contamination caused by the sampling technique

To obtain some initial information about the influence of the method of sampling of blood on the results obtained for cadmium and manganese, samples were collected by various techniques. The results are shown in Table 1. It can be seen that the application of metal needles and plastic catheters as well as glass and polystyrene vessels did not lead to contamination with cadmium whereas metal needles caused increased manganese concentrations in blood serum. It should be emphasized that the values for cadmium in Table 1 are within the "normal" range for human blood serum, while the manganese values are too high in comparison with the more recent data listed by Versieck et al. [14]. This suggests that an additional uncontrollable contamination occurred in the present work. For this reason, the "normal" values for manganese in serum were omitted from Part 1 of this series [13]; the values found were in the range 0.8—8 ng ml^{-1}.

The results for cadmium and manganese obtained after sampling of pig liver tissue by Menghini needles and glass needles are presented in Table 2. The use of steel or glass needles did not cause any significant changes in concentration for either element.

In order to establish the extent and sources of the contamination, water and diluted hydrochloric acid were used with disposable needles instead of the blood serum. This should show quantitatively any contamination caused by dissolution of the metal surface under moderate (water) and more corrosive conditions (HCl) in comparison with biological fluids. The results of these experiments for disposable needles from two different producers are shown in Table 3. In all cases, an increase in the manganese concentration was observed, whereas cadmium was not detected. The composition of the steel needles was reportedly the same in both cases (18% Cr, 1% Mn, 0.001%

TABLE 1

The influence of sampling techniques on the determination of cadmium and manganese in blood serum[a]

Sample	Cd found (ng ml^{-1})				Mn found (ng ml^{-1})			
	A	B	C	D	A	B	C	D
1	2.6	2.9	2.6	3.0	9.0	4.0	3.2	2.0
2	2.0	2.0	2.0	2.0	1.5	2.0	1.5	1.4
3	0.95	0.95	0.95	0.95	10.2	2.8	1.0	1.0
4	2.4	2.0	1.6	1.0	4.2	4.2	2.2	2.2

[a]A, metal needle/glass vessel; B, metal needle/polystyrene vessel; C, plastic catheter/polystyrene vessel; D, plastic catheter/glass vessel.

TABLE 2

The influence of sampling techniques on the determination of cadmium and manganese in liver tissue

	Cd found ($\mu g\ g^{-1}$)			Mn found ($\mu g\ g^{-1}$)		
	$\bar{x}$[a]	s	s_r (%)	$\bar{x}$[b]	s	s_r (%)
Menghini needles	0.082	0.003	4	12.3	1.4	11
Glass needles	0.080	0.004	5	12.6	0.92	7

[a]Range: 0.077—0.086 for Menghini needles and 0.074—0.083 for glass needles. $N = 5$.
[b]Range: 11.4—14.7 for Menghini needles and 11.6—13.6 for glass needles. $N = 5$.

TABLE 3

Contamination in sampling with disposable needles

Rinse solution		Metal found ($ng\ ml^{-1}$)	
		Cd	Mn
H_2O	1[a]	<0.05	2.0
	2[a]	<0.05	<0.5
0.005 M HCl	1	<0.05	1.0
	2	<0.05	<0.5

[a]1, Braunula, B. Braun Melsungen; 2, Instrumentaria, TIK Kobarid.

Cd). These experiments suggested that contamination by manganese is caused by impurities which are left on the surface of the metal after the manufacturing process and not by dissolution of surface layers of metal. In further tests, successive samples were taken with the same syringe and disposable needle; cadmium was not detected in the solution whereas the manganese concentration decreased on repeated sampling, as shown in Table 4. It is clear that the plastic parts of the syringe did not cause the manganese contamination. The above explanation of the contamination was supported by further experiments in which normal Menghini needles were used; the concentrations of cadmium and manganese were not higher even for the first sampling. These needles are used for a long period and therefore the surface layer is cleaned regularly.

The production process for disposable needles includes preparation of the tube from sheet metal, cutting to shorter pieces, grinding and electropolishing of the tips, and cleaning and fastening the tubes to plastic syringe bodies. Electropolishing is omitted by some manufacturers. During production, the metal needle comes in contact with various electrolytes, lubricants and wash solutions which can all contribute to the surface contamination. The results

TABLE 4

Contamination caused by disposable needles with plastic syringes on successive sampling[a]

Rinse solution	No. of successive sampling	Mn found (ng ml^{-1})	Rinse solution	No. of successive sampling	Mn found (ng ml^{-1})
H_2O	0[b]	<0.5	0.005 M HCl	0[b]	<0.5
	1	1.4		1	3.0
	2	1.4		2	<0.5
	3	0.6		3	<0.5
	4	0.6		4	<0.5

[a]In all cases, the cadmium found was <0.05 ng ml^{-1}. [b]Plastic syringe alone.

of experiments in which metal needles from different production stages were used are shown in Table 5. The values for cadmium, copper and zinc are below the detection limits of the procedures applied, and only the manganese concentrations are elevated. As the needles passed through the various production stages, the concentration of manganese decreased whether rinsing was done with water or dilute hydrochloric acid. It can be concluded that the critical stage as far as contamination with manganese is concerned is the preparation of the needles from sheet metal. Additionally, the acidic medium can increase the contamination by dissolving some metal from the bulk material.

Storage of blood serum samples

During storage of blood serum, various reactions and interactions can occur. The denaturation of proteins, bacterial growth, changes in pH, photochemical reactions, adsorption and desorption of elements can all affect the concentrations of particular trace elements. It is practically impossible to estimate the individual effects of particular phenomena, because of mutual interactions and overlapping. For practical purposes, it is only meaningful to

TABLE 5

Contamination in sampling caused by needles from different stages of production

Needle	Metal found (ng ml^{-1})[a]							
	Cd		Mn		Cu		Zn	
	A	B	A	B	A	B	A	B
Untreated	<0.05	<0.05	20	95	<2	<2	<10	<10
Ground	<0.05	<0.05	1	2	<2	<2	<10	<10
Polished	<0.05	0.05	<0.5	0.5	<2	<2	<10	<10

[a]A, Redistilled water; B, 0.005 M HCl.

observe the total effect of all reactions and interactions during storage, and then to establish proper storage conditions.

To study these effects, three sorts of vessel material, and two temperatures of storage were examined. Copper, zinc and manganese were observed at their normal concentration levels whereas cadmium-spiked serum (9 ng ml^{-1}) was used. The results obtained are shown in Figs. 1 and 2. These results show that the influence of the vessel material was not generally significant. The only exception was manganese, for which an increase of concentration was observed after ten days of storage in glass vessels. This increase was pronounced for storage at room temperature and can be ascribed to leaching of manganese from glass. The changes of manganese levels in serum stored in polystyrene vessels were small at room temperature and negligible at −20°C.

The differences in the behaviour of the copper, zinc and cadmium levels in serum in relation to the temperature and material used are not large but become more evident after 10—15 days of storage. Additionally, after the first day of storage, some irregularities were observed for zinc; such irregularities are often observed for standard aqueous solutions of metal ions at low concentrations. After several days of storage at room temperature, some flocculation was observed in serum samples. In these cases, accurate measurement of sample volume became difficult and the deviations of parallel

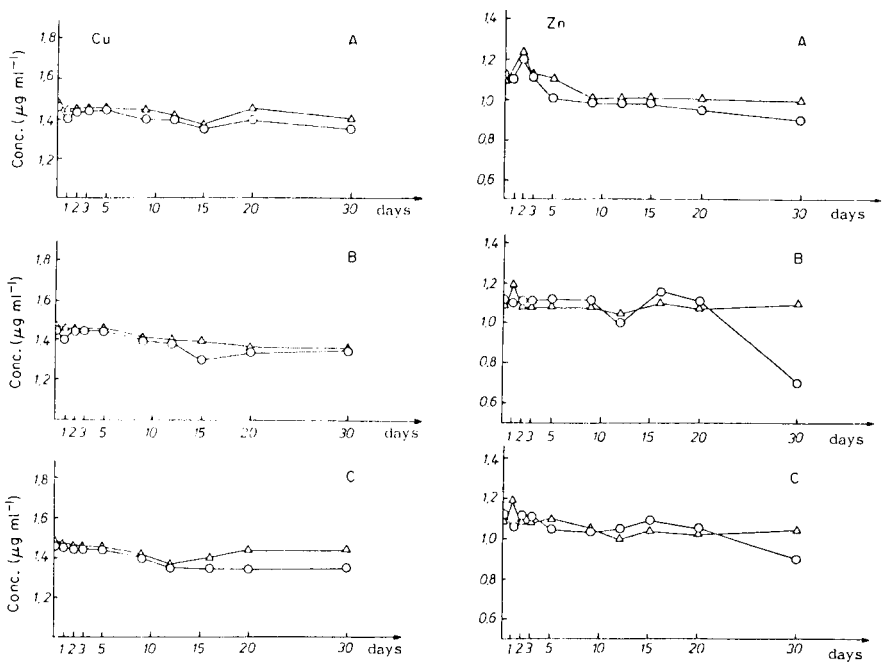


Fig. 1. Influence of vessel material, temperature and storage duration on copper and zinc levels in serum. Vessel: (A) soft soda glass; (B) pyrex glass; (C) polystyrene. Temperature: (○) room; (△) −20°C.

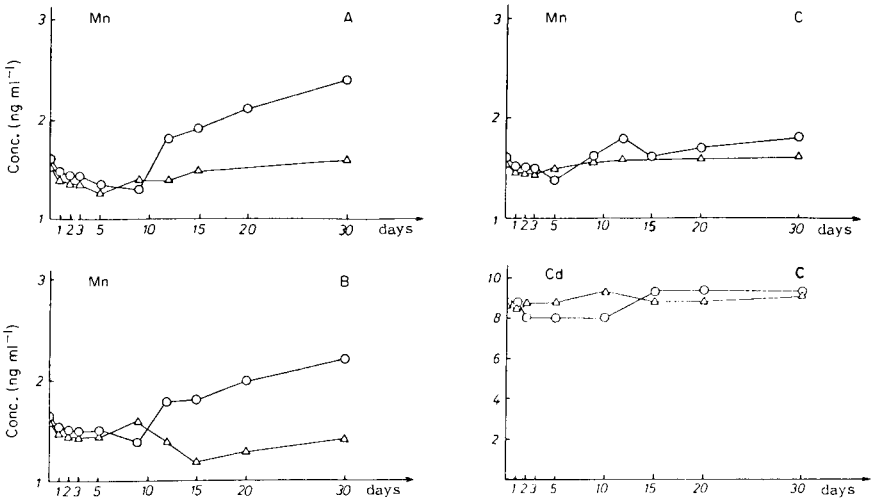


Fig. 2. Influence of vessel material, temperature and storage duration on manganese and cadmium levels in serum. Vessel: (A) soft soda glass; (B) pyrex glass; (C) polystyrene. Temperature: (○) room; (△) −20°C.

measurements were higher. This was more pronounced for zinc and can be ascribed to the different binding of copper and zinc in proteins.

Conclusions

On the basis of all these observations, it can be concluded that contamination in sampling caused by common surgical instruments can be critical for elements in the ng g^{-1} range. However, the main source of contamination is not the constituents of the bulk material, but rather contaminants introduced onto the surface during the manufacture of the instrument. Accordingly, it is advisable to check not only any new product used but also each batch of that product.

Storage of serum at −20°C has a beneficial effect on the stability of results for all the metals investigated, and is therefore preferable to storage at room temperature.

The authors are grateful to the Research Community of Slovenia for financial support. We also thank Dr. Jacques Versieck for his critical review of this manuscript.

REFERENCES

1 D. Kurz, J. Roach and E. J. Eyring, Biochem. Med., 6 (1972) 274.
2 A. Speecke, J. Hoste and J. Versieck, N.B.S. Spec. Publ., 422 (1976) 299.
3 E. W. Reimold and D. J. Besch, Clin. Chem., 24(4) (1978) 675.
4 J. Versieck, A. Speecke, J. Hoste and F. Barbier, Clin. Chem., 19(5) (1973) 472.

5 B. Meinel, J. Ch. Bode, W. Koenig and F. W. Richter, J. Clin. Chem. Clin. Biochem., 17 (1979) 15.
6 J. L. Nooijen, C. J. A. van den Hamer, J. P. W. Houtman and S. W. Schalm, Clin. Chim. Acta, 113 (1981) 335.
7 S. H. Omang and O. D. Vellar, Z. Anal. Chem., 269 (1974) 177.
8 V. D. Anand and D. M. Ducharme, N.B.S. Spec. Publ., 422 (1976) 611.
9 V. D. Anand, J. M. White and H. V. Nino, Clin. Chem., 21(4) (1975) 595.
10 G. L. Fisher, L. G. Davies and L. S. Rosenblatt, N.B.S. Spec. Publ., 422 (1976) 575.
11 M. R. Moore and P. A. Meredith, Clin. Chim. Acta, 75 (1977) 167.
12 J. R. Majer and S. E. A. Khalil, Anal. Chim. Acta, 126 (1981) 175.
13 V. Hudnik, M. Marolt-Gomišček and S. Gomišček, Anal. Chim. Acta, 157 (1984) 143.
14 J. Versieck and R. Cornelis, Anal. Chim. Acta, 116 (1980) 217.
15 G. V. Iyengar, W. E. Kollmer and H. J. M. Bowen, The Elemental Composition of Human Tissues and Body Fluids, Verlag Chemie, Weinheim, 1978, p. 22.

Analytica Chimica Acta, 157 (1984) 313—326
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

SIMULTANEOUS DETERMINATION OF BORON, PHOSPHORUS AND SULPHUR IN SOME BIOLOGICAL AND SOIL MATERIALS BY INDUCTIVELY-COUPLED PLASMA EMISSION SPECTROMETRY

M. W. PRITCHARD

Grasslands Division, Department of Scientific and Industrial Research, Private Bag, Palmerston North (New Zealand)

J. LEE*

Applied Biochemistry Division, Department of Scientific and Industrial Research, Private Bag, Palmerston North (New Zealand)

(Received 16th June 1983)

SUMMARY

Atomic emission lines at 249.678 nm, 214.911 nm and 180.735 nm are used simultaneously to determine total concentrations of boron, phosphorus and sulphur respectively, in plant and soil sample digests by means of inductively-coupled plasma emission spectrometry. The digestion procedures chosen enable these elements to be determined simultaneously on a single digest. Analytical data are compared after using various digestion procedures on standard reference materials. Actual and potential interelement interferences in the plasma are discussed together with attempts to overcome these interferences. Limitations for simultaneous determinations of boron, phosphorus and sulphur remain in the digestion and dissolution procedures.

Boron, phosphorus and sulphur in biological and geological samples are usually determined independently using separate spectrophotometric techniques such as the molybdate method for phosphorus (AOAC 7.095), and the curcumin procedure for boron [1], as well as the turbidimetric method for sulphur [2, 3]. These time-consuming procedures do not generally allow simultaneous determinations, often require separate digestions or extractions, and are prone to a range of interferences. While x-ray fluorescence can be used to analyse for phosphorus and sulphur in geological and biological materials, boron cannot be determined in this way.

Major advantages of inductively-coupled plasma emission spectrometry (i.c.p.e.s.) for measurements on sample digests or extracts are its multi-element capability, high sensitivity and relative freedom from chemical interference. Plasma conditions are suitable for the excitation and measurement of non-metals such as boron, phosphorus and sulphur. To exploit the multi-element capability of i.c.p.e.s., analyses for all elements would ideally be done on the same digests or extracts. Notwithstanding these advantages, spectral and matrix interference effects are a potential source

0003-2670/84/$03.00

of error for complex samples [4]. Errors from spectral interferences can be minimized by applying on-line or off-line corrections [4—6]. Matrix interference effects can be decreased by optimising a number of instrumental operational parameters such as the viewing height above the load coil, forward power delivered to the plasma, and argon gas flow rates. Matrix matching can be used to lessen any effects that the sample matrix may have on the accuracy of the determination.

The manner of sample preparation is critical for accurate results. Variable amounts of all three elements are lost when biological and geological samples are decomposed by dry ashing [7]. These losses cannot be prevented with certainty by using ashing aids. Losses of these elements may also be incurred during acid digestions in open vessels and care must be taken to ensure that the temperature during digestion is never allowed to become too high [3, 8]. Hydrofluoric acid is the decomposition reagent of first choice for geological materials containing large quantities of silicon, so that the silicon is removed as its volatile fluoride. When silica is volatilized, mannitol must be used to prevent loss of boron [7]. An alkaline oxidation using sodium hypobromite has been used to digest soil samples prior to the determination of total sulphur and phosphorus in soils [9, 10] and for the determination of boron in plant and soil materials [11].

Although a digestion procedure may be suitable for boron, phosphorus or sulphur, it is difficult to select a digestion procedure suitable for all three elements in a given sample matrix. This paper reports the simultaneous determination of boron, phosphorus and sulphur in several reference materials by i.c.p.e.s. Actual and potential inter-element interferences in the plasma are discussed together with procedures which overcome these interferences.

EXPERIMENTAL

Apparatus

An Applied Research Laboratories (ARL) 34000 emission spectrometer was used. The vacuum polychromator was equipped with a movable primary slit operated under computer control for scanning in the immediate vicinity of any analytical line. A standard ARL torch of the Fassel design was used in conjunction with a GMK nebulizer system (LabTest Equipment Co.) with forced sample delivery using a Gilson Minipuls II peristaltic pump (2.0 ml min^{-1}). Other instrumental conditions were as given elsewhere [12, 13]. The analytical lines selected for this study were B I 249.678 nm, P I 214.911 nm and S I 180.735 nm.

Observation height in the plasma

Solutions containing 5 μg ml^{-1} each of boron, phosphorus and sulphur in 2 M hydrochloric acid prepared from Specpure (Johnson Matthey) compounds were aspirated into the plasma and the signal response was measured at different viewing heights in the plasma. Koirtyohann et al. [6] have

recently pointed out deficiencies associated with using an external reference point and showed that an internal plasma reference point is more useful in describing the optimal observation zone in inductively-coupled plasmas. Although the viewing height was measured with respect to the height above the load coil in this plasma, no measurements were made within the "initial radiation zone" as described by Koirtyohann et al. In the plasma used this zone is <10 mm above the load coil, where inter-element effects may be severe [6]. These may be demonstrated by nebulizing large concentrations of sodium or yttrium salts into the plasma. Unless otherwise stated, all measurements were made at 18 ± 2 mm above the load coil.

Spectral interferences

A series of dilutions (in 2 M hydrochloric acid) from single element solutions, made from Specpure chemicals or analytical-grade chemicals for which relevant trace impurities had been identified, were nebulized into the plasma while the apparent response was recorded at the wavelength of interest for each of the three elements. The metals were prepared from their respective chlorides except for aluminium which was obtained from a Carius tube digestion of aluminium oxide (Johnson Matthey). In this way, a calibration graph of apparent signal vs. concentration for each potential interferent was constructed [14]. Critical concentration ratios were also calculated, and non-linear on-line corrections were derived and applied to any significant spectral interferent [4—6, 12, 15]. All of the major and some minor elements in the reference material (i.e., Al, Ca, Co, Cr, Cu, Fe, K, Mg, Na, Ti, and Zn) were screened for potential spectral interference at the boron, phosphorus and sulphur emission lines.

Calibration

Calibration graphs were obtained from direct intensity measurements (corrected for background contributions and spectral interferences by on-line correction) made on standard multi-element solutions. Standard solutions were prepared from Specpure potassium sulphate, potassium dihydrogen-phosphate and boric acid in 2 M hydrochloric acid (redistilled constant boiling).

Sample digestions and extractions

Biological samples. A general procedure [3, 16] using concentrated nitric acid (redistilled from analytical-grade reagent) was used for all digestions except where indicated. Approximately 0.3—0.4 g of each sample (dried over phosphorus pentoxide for 24 h in a vacuum desiccator immediately prior to weighing) was weighed directly into an acid-washed, 50-ml Erlenmeyer flask. The flasks had previously been cleaned by repeatedly heating those containing nitric acid until a boron blank below the limit of detection was obtained for subsequent digestions. A 5-ml aliquot of nitric acid was added to the flask and the sample was left to steep overnight at room tem-

perature. The sample was heated gently on a hot plate for 4 h at 80° C and after complete oxidation of all organic matter (the solution became clear) the temperature of the block was increased to 140° C and the remainder of the acid was volatilized. The residue was dissolved in 2 M hydrochloric acid with slight warming (≈40° C). Oxidation of NBS Bovine Liver and IAEA Milk Powder (which are particularly resistant to nitric acid oxidation) was further assisted by the initial addition of 1 ml of 70% perchloric acid, which was later removed by fuming; 0.5 ml of 100-volume hydrogen peroxide (AnalaR) was added to the nitric acid during the digestion of the IAEA Blood reference material.

Soil samples. Two methods were used. The sodium hypobromite oxidation procedure has been described elsewhere [9, 10]. Recently the method has been applied to boron determinations in plant and soil materials [11]. Approximately 0.4 g of soil was accurately weighed into 100-ml acid-washed silica beakers, 12 ml of sodium hydroxide/sodium hypobromite solution was added [9], mixed by swirling, allowed to stand for 10 min and heated on a hot-plate at 280° C until evaporated to dryness. The samples were allowed to stand for 1 h more at this temperature, then cooled, the residue was dissolved in 1 ml of deionized water, and 1 ml of formic acid was added. The samples were made up to volume (5 ml) with 2 M hydrochloric acid.

For the nitric acid/mannitol/hydrofluoric acid digestions, ca. 0.3 g of material was accurately weighed into 100-ml polyethylene beakers, 5 ml of concentrated nitric acid was added and the beakers were suspended over a boiling water bath for 2 h. Additional 5-ml aliquots of nitric acid were added if the oxidation of organic matter was incomplete. If necessary, oxidation was completed by adding 1 ml of 70% perchloric acid (Baker-analysed Ultrex-grade reagent) for particularly resistant samples (e.g., Soil 5). After excess of nitric acid had been boiled off, 1.25 ml of a 0.1% mannitol solution was added to each beaker and the solution was stirred while on the water bath. After 15 min, 6 ml of 40% hydrofluoric acid was added and the solutions were slowly evaporated to dryness. After further heating on the water bath for 1 h, the samples were taken up in 2 M hydrochloric acid and diluted to 10 ml in stoppered graduated polyethylene centrifuge tubes. The total acidity was maintained at 2.0 ± 0.2 M hydrochloric acid.

RESULTS AND DISCUSSION

Viewing height above the load coil

Profiles of signal intensities vs. the viewing height for 10 μg ml^{-1} solutions of boron, phosphorus and sulphur were measured. The line-to-background ratio profiles are given in Fig. 1. Ideally the signal maxima for all three elements should coincide. The profiles also indicate the relative sensitivities of the analytical lines. From these profiles it is evident that 16 ± 2 mm is the best viewing height for simultaneous determination of boron, phosphorus

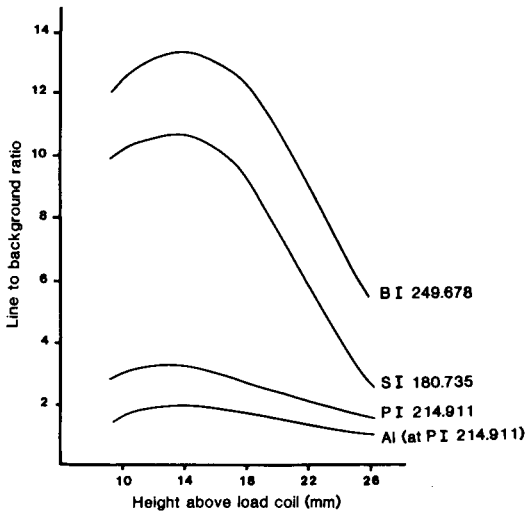


Fig. 1. Line-to-background signal ratios plotted as a function of height above the load coil while nebulizing 10 mg l^{-1} solutions of boron, phosphorus and sulphur into the plasma. A profile is also given for Al (4140 mg l^{-1}) viewed at the P I 214.911 nm line. (The 3σ-detection limits for B, P and S in 2 M HCl are 0.02, 0.10 and 0.03 μg ml^{-1}, respectively, at 16 mm. Aerosol gas flow 0.8 l min^{-1}, pump rate 2.0 ml min^{-1}.

and sulphur in this plasma. Potential interferents sometimes exhibit different signal/height profiles from the analyte allowing the interferent/analyte signal ratio to be minimized by the choice of viewing height [6, 17]. Profiles for several likely interferents (in terms of their high concentrations in the sample matrix) on the analytical lines of boron, phosphorus and sulphur were also measured. One of these (Al at P I 214.911 nm) is also shown in Fig. 1. This was characteristic of profiles for most of the interferent elements measured. Unfortunately, this profile maximum for the analyte elements is also the maximum for the signal response vs. plasma height profiles for several potential interferents in the viewing window for the boron, phosphorus and sulphur lines.

In this study the aerosol gas flow was optimized for the S I 180.735 nm line. This line showed the largest variation in signal-to-background ratios with changes in aerosol flow rate.

Sodium was expected to be present at high levels in many of the extracts. The effect of this element on the net intensity at the boron, phosphorus and sulphur lines was studied. The results are reported in Table 1. Only the boron line showed any significant reduction in net intensity in the presence of 1 mg ml^{-1} sodium in solution. At a viewing height of 18 mm, matrix effects caused by sodium in 2 M hydrochloric acid at concentrations up to 10 mg ml^{-1} are less than 10% when forced sample delivery is used. Changes to the spectral background under each of the emission lines for

TABLE 1

Effect of sodium on the line intensities of boron, phosphorus, and sulphur

Height above load coil (mm) (± 2 mm)	Change in signal-to-background ratio[a]		
	B I 249.678	P I 214.911	S I 180.735
14	−4.3	0	+0.5
18	−1.6	0	0
22	−4.6	−2.4	−3.6
26	−3.2	0	0

[a] 10 μg ml^{-1} analyte measured in the presence or absence of 1 mg Na ml^{-1} in 2 M HCl, expressed as the percentage difference in the line to background ratios measured with and without sodium present.

boron, phosphorus and sulphur, were negligible (<0.5%) at sodium concentrations up to 10 mg ml^{-1}.

Spectral interferences

Plant and animal tissue, rocks and soils contain sufficient Al, Ca, Cu, Fe, K, Mg and Na for these elements to be potential spectral interferents. Serially diluted solutions of these elements and of Co, Cr, Ti and Zn were nebulized into the plasma and the apparent concentrations of boron, phosphorus and sulphur were measured at the relevant analytical line to check for any spectral emission originating from elements other than the analyte. Table 2 lists

TABLE 2

Spectral interferences

Analyte line	Interferent	Effect	ng μg^{-1} [a]	CCR[b]
B I 249.678 nm	Co	Line overlap	2.0	500
	Fe	Broadened line wing overlap	0.19	5263
	Mn	Broadband?	0.15	6666
P I 214.911 nm	Al	Broadband/ recombination continuum	3.6	277
	Cu	Broadened line wing overlap	7.72	129
	Fe	Line overlap and/or line scatter	0.6	1666
S I 180.735 nm	Al	Recombination continuum	0.29	3448
	Ca	Line overlap	13.0	77
	Mn	Line overlap	2.8	357

[a] Apparent analyte concentration (ng ml^{-1}) for each μg ml^{-1} interferent. [b] Critical concentration ratio, the ratio of interferent to analyte concentration at which the intensity ratio of interferent to analyte within the spectral window is unity [15].

the nature of the spectral interference together with the critical concentration ratio for each interferent/analyte pair. It can be seen that some of the interferences are serious enough to warrant the use of on-line correction, particularly the interferences of Cu, Fe and Al on phosphorus, Ca and Mn on sulphur, and Fe on boron.

The source of each spectral interference was studied by scanning the spectrum in the region of the analytical line of interest while solutions containing the interfering element were nebulized into the plasma. Wavelength scans were also made in the region of the analyte line while extracts and digestions of reference materials were nebulized. Only the scans of reference material digests and extractant solutions showing the most detail are presented here.

Boron. In Fig. 2A it is apparent that the wings of the emission lines of the iron lines at 249.653 nm and 249.649 nm [18] overlap the boron viewing window. This necessitates the application of on-line correction for boron measurements when iron is present in concentrations exceeding 500 times that of boron by weight. The profile in the region of the B I 249.678-nm line obtained during nebulization of an alkaline extract of the IAEA reference Soil 5 containing ca. 200 μg Fe ml^{-1} shows that on-line

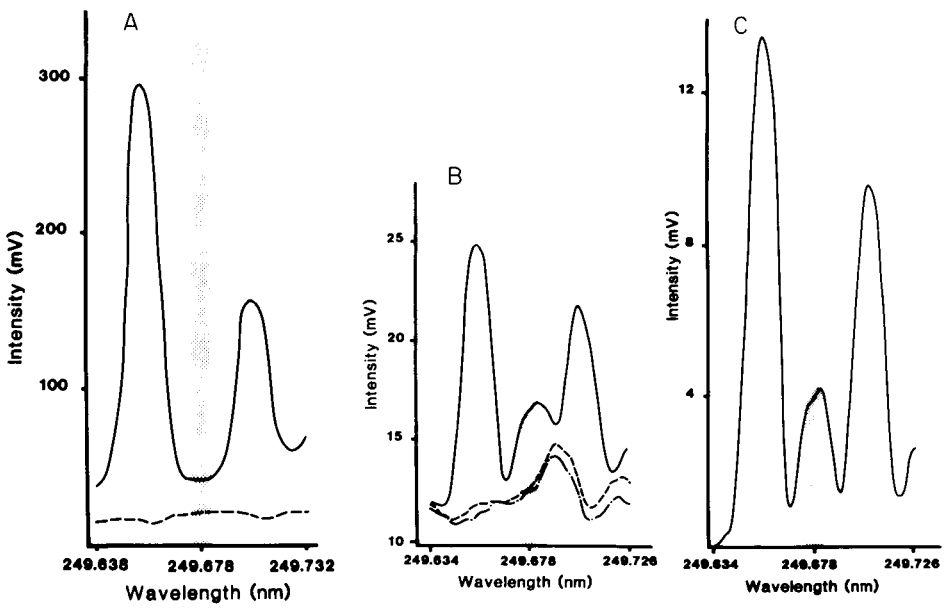


Fig. 2. Plots of signal intensity vs. wavelength in the region of the B I 249.678-nm line. (A) Solution nebulized: (——) 10 mg ml^{-1} iron solution in 2 M HCl; (— — —) 2 M HCl; (the apparent B concentration caused by the Fe wing overlap about 2 mg l^{-1}. (B) Solution nebulized: (——) an alkaline extract of Soil 5 in 2 M HCl; (- - -) 2 M HCl; (- - - -) alkaline oxidation blank. (C) The difference plot after removal of plasma background from (B). The hatched areas represent the spectral bandpass.

spectral correction is necessary. The contribution from the alkaline oxidation blank to the plasma background is negligible (Fig. 2B). A difference plot (Fig. 2C), with the plasma background structure removed, shows that the B I 249.678-nm line is clearly resolved and the only other contribution to the overall signal is derived solely from the overlapping iron lines (representing an apparent boron signal equivalent to 0.19 μg B l^{-1} for every mg Fe l^{-1} present).

Phosphorus. An iron line at 214.917 nm [19] overlaps the spectral window set to observe the phosphorus line at 214.911 nm. The degree of interference can be seen when wavelength scans obtained from nebulizing separate solutions containing 1% iron and 0.5 mg l^{-1} phosphorus into the plasma are studied (Fig. 3A). The signal from the iron solution is ca. 10 times the signal from 0.5 mg l^{-1} phosphorus. From the critical concentration ratios listed in Table 2 the contribution of iron to the phosphorus signal can be calculated to give an apparent concentration of 0.6 μg phosphorus for each mg of iron present. Thus for accurate determination of phosphorus in the presence of moderate levels of iron, on-line correction of the spectral interference is necessary. Background emission in the region of the P I 214.911-nm line is complicated by NO band emission [20]. A difference plot (Fig. 3B) of the scans obtained during nebulization of 1% iron in 2 M hydrochloric acid shows that the wing of the adjacent iron line at 214.917 nm lies within the spectral window for the P I 214.911-nm line.

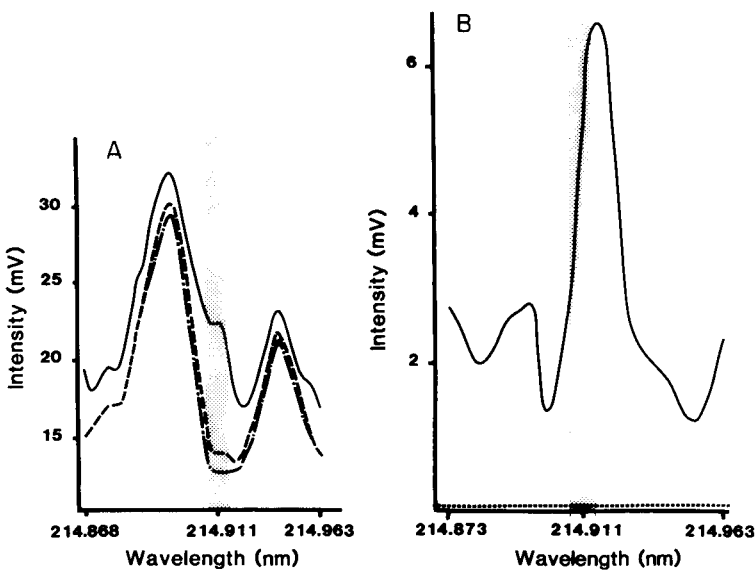


Fig. 3. (A) Spectral region of the P I 214.911-nm emission line, recorded while nebulizing (····) 2 M HCl; (- - -) 0.5 mg l^{-1} P in 2 M HCl; (—) 10 mg ml^{-1} Fe in 2 M HCl. (B) Difference plot between the 10 mg ml^{-1} Fe spectrum and the 2 M HCl blank spectrum. The hatched areas represent the spectral bandpass.

A copper emission line at 214.897 nm can also be detected but the concentration of copper in the reference materials considered here is too low for any correction to be necessary. Concentrations of copper >70 mg l^{-1} in the solution would be needed before on-line correction would be required.

Aluminium exhibits perhaps the most interesting spectral structure of all the interfering elements in the region of the phosphorus line. The gross signal at the phosphorus wavelength is increased in the presence of aluminium by two main effects, a broadband continuum and a wing overlap of an adjacent aluminium emission line. In Fig. 4A, the phosphorus line can be seen to lie on the side of a broad emission peak from aluminium. This is the fourth doublet in a series of $3p^2P^0$—nd^2D aluminium doublet transitions seen in the emission spectrum of aluminium in the plasma between 225 nm and 210 nm [21]. The size of this wing within the spectral window can be seen more clearly in the difference plot (Fig. 4B) where an increasing signal is obtained over the whole scan from 214.874 nm to 214.964 nm. When this wing is compared with the line-width at half maximum peak height of the P I 214.911-nm line in spectral scans between 214.895 nm and 214.927 nm, the special nature of this aluminium doublet is apparent. These doublets degrade into a broadband continuum below 210 nm [21]. An additional broadband continuum resulting from a recombination of aluminium ions with electrons also contributes to the signal obtained in this region of the spectrum when aluminium is nebulized.

The spectrum obtained when an alkaline oxidation extract of Soil 5 was nebulized shows a general background signal enhancement (Fig. 5A). This is attributed to the presence of aluminium in the extract which gives an

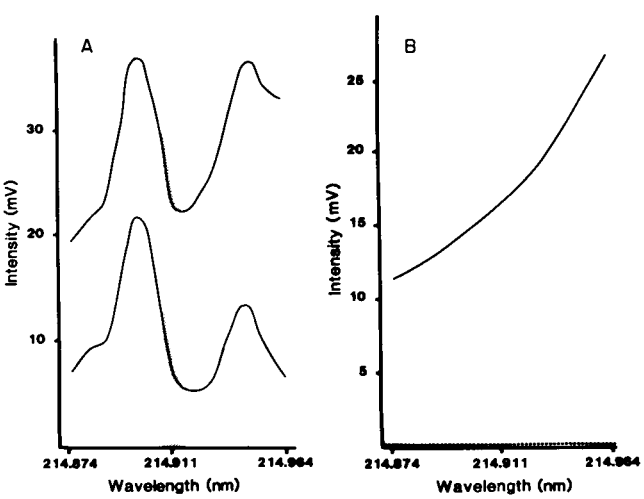


Fig. 4. (A) Spectrum obtained in the region of the P I 214.911-nm line when Al solution (4140 mg l^{-1}) was nebulized (upper trace); the lower trace is from 2 M HCl; the hatched area represents the spectral bandpass. (B) Difference plot of the spectra in (A).

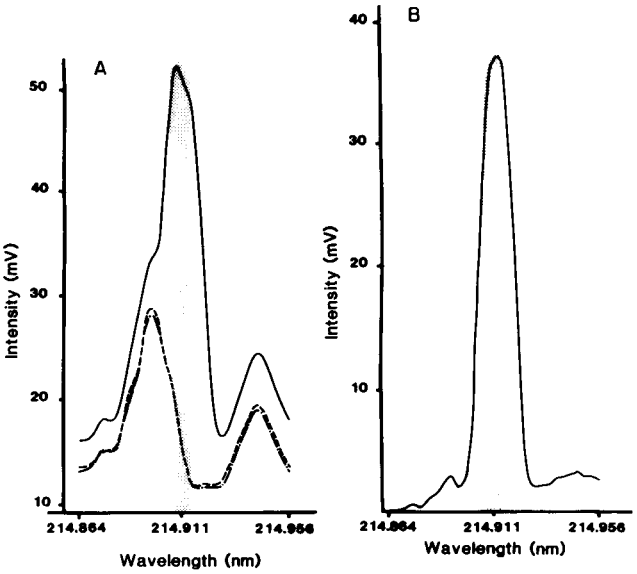


Fig. 5. (A) Spectrum between 214.868 nm and 214.963 nm produced by nebulizing: (—) alkaline oxidation extracts of soil 5; (- - -) an alkaline oxidation blank; (- · - ·) 2 M HCl. (B) Difference plot of soil extract and oxidation blank spectra in (A).

apparent phosphorus concentration of 3.6 μg for every mg of aluminium present. The slight shoulder on the edge of the phosphorus line is attributed to an iron emission line at 214.917 nm (iron is also present in the extract). However, a difference plot (Fig. 5B) shows a prominent phosphorus signal notwithstanding the presence of aluminium (seen as a background enhancement of the phosphorus line) and iron in the extract.

Sulphur. Scans in the region of the S I 180.735-nm line for the Soil 5 extracts showed a clearly defined emission line with no observed interferences. However, spectral scans in this region when plant acid digestion solutions were nebulized showed a more complex spectral structure. It was readily apparent that on-line correction was necessary for Al, Cu, Fe and Mn as indicated in Table 2. The nature of these spectral interferences has been discussed elsewhere [12].

Calcium seriously interferes with sulphur because of a calcium line at 180.739 nm. This gives (Table 2) a signal equivalent to an apparent sulphur concentration of 13 μg for every mg of calcium present in the solution. Aluminium and manganese interferences yield an apparent sulphur concentration of 0.29 and 2.8 μg, respectively, per mg of interferent. Consequently, on-line corrections for spectral interferences of Al, Ca and Mn on sulphur were applied in all plant analyses. Accurate determinations of the high concentrations of calcium are needed as any error in the value of calcium is reflected in the correction factor applied to the determination of sulphur.

TABLE 3

Analyses of biological samples by i.c.p.e.s. ($\mu g\ g^{-1}$ dry weight)

Sample	B			P			S		
	Assigned values	This work	Literature values	Assigned values	This work	Literature values	Assigned values	This work	Literature values
NBS 1573 (Tomato leaves)	(30)[a]	38 ± 0.1	37[c] 33[d]	3400	3459 ± 8	3400[c] 3340[d] 3500[f]	—	5848 ± 58	5900[f]
NBS 1571 (Orchard leaves)	33	31 ± 3	33[c] 32[d] 41[c]	2100	1883 ± 100	1980[d] 1800[j] 2200[e]	(1900)	1890 ± 100	2080[d] 2000[c] 2300[i]
NBS 1570 (Spinach)	(30)	27 ± 3.5	27.8[g]	5500	4814 ± 300	5000[d] 5400[f] 2983[g]	—	3834 ± 58	3700[f] 4470[h]
NBS 1568 (Rice flour)	—	n.d.[b]	0.72[g]	—	1420 ± 2	1443[g]	—	1059 ± 5	—
NBS 1577 (Bovine liver)	—	n.d.	3.2[d]	11 000	11 521 ± 35	10 500[d] 11 000[j]	—	7353 ± 81	7100[d] 7200[i]
NBS 1572 (Citrus leaves)	—	69 ± 0.3	—	1300	1332 ± 11	—	4070	3822 ± 58	—
IAEA V-8 (Rye flour)	—	n.d.	—	592	500 ± 3.7	—	(620)	475 ± 4	—
IAEA V-9 (Cotton cellulose)	—	0.7 ± 0.06	—	—	n.d.	—	—	60 ± 1.3	—
IAEA A-11 (Milk powder)	(1.7)	1.9 ± 0.2	—	9100	9942 ± 620		(18 000)	3286 ± 258	—
IAEA H5 (Animal bone)	—			(95 780)	104 342 ± 580	—	(2350)	2337 ± 116	—
IAEA H-4 (Pig muscle)	—	14.9	—	(6830)	9015	—	(8300)	9424	
IAEA A-13 (Animal blood)	—	n.d.	—	—	693 ± 19	—	—	5992 ± 96	—

[a] ( ) indicate non-certified value for information only. [b] n.d. = not detected. [c] [22]. [d] [23]. [e] [24]. [f] [25]. [g] [26]. [h] [2]. [i] [27]. [j] [14].

Analytical performance

Biological samples. Standard reference materials from the National Bureau of Standards and the International Atomic Energy Agency (IAEA), were used to assess the accuracy of the overall method for the simultaneous determination of boron, phosphorus and sulphur. The results, reported in Table 3 are based on a minimum of three replicates. There is a paucity of published results for boron and sulphur in standard reference materials apart from NBS 1571 (orchard leaves) and NBS 1573 (tomato leaves).

A prime consideration in the choice of the digestion vessels used was that the digestions should be kept simple so that a large number could be done daily in order to utilize the high sample throughput of the instrument. Closed vessels were therefore not considered. Losses of specific chemical forms of the analytes (in particular, sulphur-containing amino acids, thiols etc.) may occur from open vessels but all recoveries were observed to be satisfactory when the rate of heating of the nitric acid digest was kept sufficiently low to prevent vigorous fuming. X-ray fluorescence and neutron activation data in Table 3 are in good agreement with the present data, indicating that losses during digestion were minimal [25, 27].

Precision data for boron, phosphorus and sulphur in two different standard reference materials with nitric acid digestion are presented in Table 4. Very acceptable precision is obtained both for consecutively repeated measurements on the same solution and for measurements on a composite solution repeated every week for 5 consecutive weeks.

Standard soil materials. Boron, phosphorus and sulphur were determined in two soil standard reference materials. One (Soil 5) had a high organic content while the other was mainly inorganic. Results for alkaline oxidation extractions and for nitric/hydrofluoric acid "total" digestions, with and without the presence of mannitol, are given in Table 5. A shortage of certified values hinders full evaluation of total recovery and accuracy particularly for boron.

Although all three digestion procedures gave similar sulphur values, the

TABLE 4

Precision of results for boron, phosphorus and sulphur by i.c.p.e.s.

	NBS 1571 Orchard Leaves		IAEA A-13 Animal Blood		
	Mean ($\mu g\ ml^{-1}$)	r.s.d.[a] (%)	Mean ($\mu g\ g^{-1}$)	r.s.d.[b] (%)	CE[c] (%)
B	0.306	0.74	—	—	—
P	18.6	2.3	693	2.7	0.2
S	18.1	2.6	5992	1.6	0.2

[a]Relative standard deviation for a composite solution of 7 individual 0.5-g digestions measured once a week over a period of five weeks (dilution 1:100 w/v). [b]Overall error including digestion ($n = 6$). [c]Counting error (r.s.d.) for three integrations on one sample.

TABLE 5

Concentration data for boron, phosphorus and sulphur in geological reference materials ($\mu g\ g^{-1}$, dry weight)

Standard	B		P		S
	Assigned	This work	Assigned	This work	This work
IAEA Soil 5	(63)[a]		1100		
NaOH/OBr extraction		32 ± 1		839 ± 16	489 ± 12
$HF/HNO_3/HClO_4$		n.d.[b]		1056 ± 53	483 ± 9
HNO_3/HF/mannitol		42 ± 3		944 ± 8	498 ± 15
NBS 1645					
$HF/HNO_3/HClO_4$		n.d.		478 ± 10	13 600 ± 500
HNO_3/HF/mannitol		n.d.		398 ± 13	12 219 ± 363

[a]One analysis only [28]. [b]n.d. = not detected.

recovery of phosphorus and boron from the alkaline oxidation method was low compared with that from the acid digests. There was good agreement between the certified value for total phosphorus in Soil 5 and the value given here for the total digestion procedure using nitric/hydrofluoric/perchloric acids. However, losses of boron may still occur during the digestion of Soil 5 even in the presence of mannitol. It appears that the alkaline oxidation method extracts only a portion of the total boron present although the literature value given is from one analysis only [28].

Conclusion

Accurate determinations of boron, phosphorus and sulphur can be made simultaneously on single digests of soils and biological samples by i.c.p.e.s. when the atomic emission lines at 249.678 nm, 214.911 nm and 180.735 nm, respectively, are used. However, less-common digestion procedures are required and on-line corrections for spectral interferences are needed for all three elements. Comparisons of assigned values and published data for standard reference materials with the results obtained after different digestion procedures gave good agreement for all elements with the exception of boron in Soil 5. Further studies will be directed towards developing better digestion procedures in open vessels, in order to obtain higher recoveries of boron from soil samples high in organic materials. Alkaline fusions may offer a better alternative [7].

REFERENCES

1 A. S. Baker, J. Agric. Food Chem., 12 (1964) 367.
2 L. L. Wall, C. W. Gehrke and J. Suzuki, Commun. Soil Sci. Plant Anal., 11 (1980) 1087.
3 R. W. Blancher, G. Rehm and A. C. Caldwell, Soil Sci. Soc. Am. Proc., 29 (1965) 71.
4 P. W. J. Boumans and F. J. de Boer, Spectrochim. Acta, Part B, 30 (1975) 309.

5 R. L. Dahlquist and J. W. Knoll, Appl. Spectrosc., 32 (1978) 1.
6 S. R. Koirtyohann, J. S. Jones, C. Jester, P. Yates and A. Dennis, Spectrochim. Acta, Part B, 36 (1981) 49.
7 R. Bock, A Handbook of Decomposition Methods in Analytical Chemistry, International Textbook Co., Glasgow, 1979.
8 P. J. Randall and K. Spencer, Commun. Soil Sci. Plant Anal., 11 (1980) 257.
9 M. A. Tabatabai and J. N. Bremner, Soil. Sci. Soc. Am. Proc., 34 (1970) 62.
10 W. A. Dick and M. A. Tabatabai, Soil Sci. Soc. Am. J., 41 (1977) 511.
11 C. C. Kowalenko, Commun. Soil Sci. Plant Anal., 10 (1979) 1421.
12 J. Lee and M. W. Pritchard, Spectrochim. Acta, Part B, 36 (1981) 591.
13 D. L. Miles and J. M. Cook, Anal. Chim. Acta, 141 (1982) 207.
14 N. R. McQuaker, P. D. Kluckner and G. N. Chong, Anal. Chem., 51 (1979) 888.
15 P. W. J. M. Boumans, Line Coincidence Tables for Inductively Coupled Plasma Atomic Emission Spectrometry, Pergamon Press, Oxford, 1980.
16 P. Schramel, A. Wolf and B. J. Klose, in P. Bratter and P. Schramel (Eds.), Trace Element Analytical Chemistry in Medicine and Biology, W. de Gruyter, Berlin, 1980.
17 F. J. M. J. Maessen, J. Balke and J. C. de Boer, Spectrochim. Acta, Part B, 37 (1982) 517.
18 A. N. Zaidel, V. K. Prokofjev, S. M. Raishli, V. A. Slavnyi and E. A. Shreider, Tables of Spectral Lines, Plenum, New York, 1970.
19 M. L. Parsons, A. Forster and D. Anderson, Atlas of Spectral Interferences in ICP Spectroscopy, Plenum, New York, 1980.
20 R. D. Reeves, S. Nikdel and J. D. Winefordner, Appl. Spectrosc., 34 (1980) 477.
21 G. F. Larson and V. A. Fassel, Appl. Spectrosc., 33 (1979) 592.
22 D. C. de Bolt, J. Assoc. Off. Anal. Chem., 63 (1980) 802.
23 E. S. Gladney, Anal. Chim. Acta, 118 (1980) 385.
24 J. Benton-Jones, Jr., Commun. Soil Sci. Plant Anal., 8 (1977) 349.
25 D. S. Winter, personal communication, 1981.
26 R. C. Munter, R. A. Grande and P. C. Ahn, I.C.P. Inf. Newsl., 5 (1979) 368.
27 E. T. Turney, D. B. Curtis and E. S. Gladney, Anal. Chem., 49 (1977) 1741.
28 R. Dybczynski, A. Tugsavul and O. Suschny, Geostand. Newsl., 3 (1979) 61.

Analytica Chimica Acta, 157 (1984) 327—332
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

PHASE-RESOLVED FLUORIMETRIC DETERMINATION OF TWO ALBUMIN-BOUND FLUORESCEIN SPECIES

LINDA B. McGOWN

Department of Chemistry, Oklahoma State University, Stillwater, OK 74078 (U.S.A.)

(Received 12th September 1983)

SUMMARY

A new approach is described for phase-resolved fluorescence spectroscopy, for use in resolving mixtures of two components with very similar fluorescence spectra and lifetimes. Results are given for application of the technique to solutions containing fluorescein physically bound to albumin and fluorescein isothiocyanate covalently bound to albumin. Because the two fluorescein species have essentially identical fluorescence spectra and a phase-angle difference of only 2°, the conventional phase-resolved method in which measurements are made at the two phase angles at which the fluorescence contribution from one or the other of the components is zero will not resolve the components. Solutions containing 25—50 nM of each component were successfully resolved by making measurements at two other phase angles and solving the pair of simultaneous equations that is generated. Accuracy is best (average relative error, 3%) using detector phase angles corresponding to a 45° shift from the phase angles of the components. Relative standard deviations of ±16% are obtained at these phase angles. Solutions containing 5—50 nM fluorescein and 50—500 nM fluorescein isothiocyanate conjugated to albumin could also be resolved, with an average relative error of 16% and ±2.4 r.s.d. The method could be used for simultaneous determination of a fluorophore in two different microenvironments, as in protein-ligand binding studies and in homogeneous immunoassay.

The use of phase and modulation measurements for determining fluorescence lifetimes has been well-described [1, 2]. Phase-resolved spectrofluorimetry, in which phase-resolved detection is used to eliminate the emission of a single fluorescence lifetime from a multicomponent mixture, should encourage the development of quantitative methods which exploit differences in fluorescence lifetimes to improve the selectivity of fluorimetric methods.

To measure a fluorescence lifetime, τ, by means of phase and modulation measurements, the sample is excited with sinusoidally modulated light having an angular frequency, ω, and intensity $E(t)$ at time t:

$$E(t) = 1 + m_{ex} \sin \omega t$$

where m_{ex} is the degree of modulation. The resulting fluorescence, $F_1(t)$, of a single component sample with exponential decay will be phase-shifted by angle ϕ and appear as

$$F_1(t) = 1 + m_{ex} m \sin(\omega t - \phi)$$

0003-2670/84/$03.00

The demodulation, m, of the light caused by the fluorophore is given as $m = [1 + (\omega\tau)^2]^{-1/2}$ and the phase shift, ϕ, as $\phi = \arctan(\omega\tau)$. The fluorescence lifetime, τ, can be found from these relationships.

For a solution containing two fluorophores, A and B,

$$F_2(t) = k_A \sin(\omega t - \phi_A) + k_B \sin(\omega t - \phi_B)$$

where k_A and k_B are constants which depend on the spectral characteristics of the species. If the fluorescence emission is observed with a phase-sensitive detector, the phase-sensitive signal at detector phase, ϕ_D, will be

$$F_2(\phi_D) = k_A \cos(\phi_D - \phi_A) + k_B \cos(\phi_D - \phi_B).$$

By choosing ϕ_D to be 90° out of phase with one of the components, e.g., component A ($\phi_D = \phi_A \pm 90°$), that component will be "nulled" and only component B will be observed. Conversely, if $\phi_D = \phi_B \pm 90°$, only component A will be observed. This is the basis of phase-resolved fluorescence determinations [3, 4].

Lakowicz and Cherek [3] demonstrated the use of phase-resolved fluorescence for simultaneous determination of indole and 2,3-dimethylindole ($\Delta\tau$ = 6.5 ns) and showed that lifetime differences of 1 ns were sufficient to resolve derivatives of tyrosine and tryptophan. They were also able to resolve mixtures of PRODAN and TNS ($\Delta\tau$ = 8 ns) and dibenzo[a,h]anthracene and dibenzo[c,g]carbazole ($\Delta\tau$ = 26.3 ns) [4]. In each of these determinations, the two species in the mixtures had significantly different fluorescence spectral characteristics in addition to differences in fluorescence lifetimes of at least 1 ns, and generally much larger. Emission spectra run at the two ϕ_D values corresponding to the phase angles of the components ±90° would therefore show reasonably large fluorescence intensities arising from the nonnulled components.

The purpose of the study described here was to demonstrate the application of phase-resolved fluorimetry to the simultaneous determination of two species with very similar fluorescence lifetimes and essentially identical fluorescence spectra, such as might occur in resolution of the distribution of a fluorophore between two different environments. In these cases, wavelength cannot be used as a parameter to increase the selectivity for one component over the other. Therefore, changes in fluorescence lifetime offer an attractive alternative, using phase-resolved detection. However, it will be difficult to resolve the mixtures using the null-phase-angle approach because, for both components, the nonnulled component is being measured very close to its own null phase angle, resulting in very low fluorescence intensities. It was found that these mixtures can be resolved by measuring the fluorescence at two other detector phase angles and solving the two simultaneous equations

$$F(\phi_1) = \bar{I}_{A,\phi_1}[A] + \bar{I}_{B,\phi_1}[B] \quad (1)$$

$$F(\phi_2) = \bar{I}_{A,\phi_2}[A] + \bar{I}_{B,\phi_2}[B] \quad (2)$$

where $\bar{I}_{A,\phi_1}$, $\bar{I}_{B,\phi_1}$ and $\bar{I}_{A,\phi_2}$, $\bar{I}_{B,\phi_2}$ are the molar fluorescence intensities of A and B measured at ϕ_1 and ϕ_2, respectively, using standard solutions of A and B.

Maximum values for $\bar{I}_A$ and $\bar{I}_B$, and therefore maximum total fluorescence intensities, will be observed at $\phi_1 = \phi_A$ and $\phi_2 = \phi_B$. However, these phase angles occur at the peaks of the phase angle curves (see Fig. 1), resulting in very similar values for the molar fluorescence intensities for a component at the two phase angles. In other words, $\bar{I}_{A,\phi_A} \simeq \bar{I}_{A,\phi_B}$ and $\bar{I}_{B,\phi_A} \simeq \bar{I}_{B,\phi_B}$. Equations 1 and 2 become almost identical, making it difficult to solve accurately for [A] and [B].

It becomes apparent that in order to resolve mixtures of components A and B with similar lifetimes, selection of ϕ_1 and ϕ_2 must satisfactorily compromise between the two requirements that: (a) $\bar{I}_A$ and $\bar{I}_B$ at ϕ_1 and ϕ_2 must be large enough to allow reproducible measurement of fluorescence intensities; and (b) $|\bar{I}_{A,\phi_1} - \bar{I}_{A,\phi_2}|$ and $|\bar{I}_{B,\phi_1} - \bar{I}_{B,\phi_2}|$ must be large enough so that Eqns. 1 and 2 are sufficiently different that they can be accurately solved for components of interest. In other words, the compromise will be at phase angles somewhere between ϕ_A and ϕ_B, where the first requirement is best fulfilled, and $\phi_A \pm 90°$ and $\phi_B \pm 90°$, where the second requirement is best fulfilled.

We successfully resolved mixtures of fluorescein adsorbed to bovine serum albumin (BSA) and fluorescein isothiocyanate covalently conjugated to BSA (FITC-BSA), using the phase angles $\phi_1 = \phi_A - 45°$ and $\phi_2 = \phi_B - 45°$.

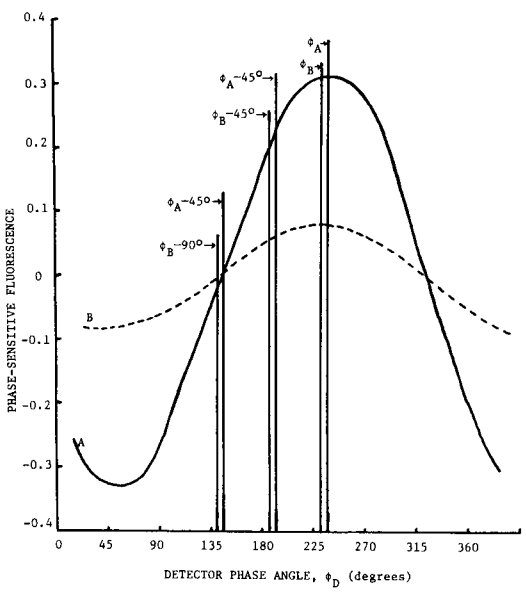


Fig. 1. Phase-sensitive fluorescence intensity as a function of detector phase angle: (—) 0.10 μM fluorescein; (---) 0.050 μM bound FITC. Modulation = 18 MHz, λ_{ex} = 490 nm, λ_{em} = 520 nm. ϕ_A and ϕ_B are the phase angles of fluorescein and bound FITC, respectively.

EXPERIMENTAL

Reagents

Fluorescein (sodium salt), bovine serum albumin (fatty acid-free, crystalline) and fluorescein isothiocyanate-conjugated bovine serum albumin were purchased from Sigma Chemical Co. Aqueous solutions were prepared with deionized water.

Molarities for the BSA-conjugated were calculated using a molecular weight of 66,500 for the BSA, and assuming 9.9 FITC molecules per BSA, as determined by the manufacturer. The molarities cited here refer to the bound FITC concentration, or 9.9 × [BSA].

Procedure

All measurements were made with an SLM 4800S spectrofluorimeter in the phase-resolved mode, with 18-MHz modulation. Constant wavelengths were used, corresponding to the fluorescein excitation and emission maxima (490 nm and 520 nm, respectively). Replicate (3—6) measurements were made for each solution at each phase angle, and each measurement was the electronic average of 100 samplings (requiring 30 s per measurement).

All fluorescein and FITC-BSA standard solutions and mixtures were prepared in 1×10^{-5} M BSA to insure complete uptake of the fluorescein by the BSA, as well as to provide an essentially constant scatter intensity.

Standard solutions of fluorescein and of FITC-BSA in 1×10^{-5} M BSA were used to find the molar fluorescence intensities of each component at each detector phase angle.

Fluorescence intensities of all solutions were measured first at one phase angle and then at the next, rather than going back and forth between phase settings, to minimize the effect of irreproducibility in setting the phase angles.

For the first set of mixtures (25—50 nM of each component), best results were obtained when data varying from the mean by more than 1 s.d. were rejected. For the second set of mixtures (5—50 nM fluorescein and 50—500 nM bound FITC), in which the signals were much larger and less sensitive to extraneous noise, the *Q*-test (90% confidence) was used to reject data.

RESULTS AND DISCUSSION

When 18-MHz modulation was used, the phase angles of the fluorescein and the FITC-BSA were found to differ by 1.7° corresponding to a $\Delta\tau$ of 0.3 ns. Actually, because there are numerous FITC labels attached to each protein in the FITC-BSA, it is likely that small differences exist between the lifetimes of the different FITC labels. However, because the proportional contributions from each separate lifetime component to the overall observed lifetime for FITC-BSA remain constant for the FITC-BSA standards and test solutions, it is possible to treat it as a single species with a constant lifetime.

The results for four minutes containing 25—50 nM each of fluorescein and FITC-BSA are shown in Table 1. Three sets of detector phase angles were used: ϕ_A and ϕ_B, $\phi_A - 45°$ and $\phi_B - 45°$, and $\phi_A - 90°$ and $\phi_B - 90°$. Accu-

TABLE 1

Phase-resolved fluorimetric determination of fluorescein (A) and fluorescein isothiocyanate (B)
(λ_{ex} = 490 nm, λ_{em} = 520 nm, modulation = 18 MHz, ϕ_A = 237.0°, ϕ_B = 235.3°)

Added (nM)		Concentration found (nM) and relative error (%)[a]											
[A]	[B]	(ϕ_A, ϕ_B)				($\phi_A - 45°$, $\phi_B - 45°$)				($\phi_A - 90°$, $\phi_B - 90°$)			
		[A]	Error	[B]	Error	[A]	Error	[B]	Error	[A]	Error	[B]	Error
50.0	50.0	45.1	−9.8	67.4	34.8	55.0	10.0	53.1	6.2	—	—	42.2	−15.6
50.0	25.0	53.7	7.4	15.0	−40.0	54.4	8.8	24.9	−0.4	22.9	−54.2	35.9	43.6
25.0	50.0	15.8	−36.8	51.3	2.6	23.5	−6.0	45.6	−8.8	—	—	50.5	1.0
25.0	25.0	36.2	44.8	12.5	−50.0	24.8	−0.8	23.0	−8.0	—	—	5.4	−78.5
Average rel. error (%)		1.4		−13.1			3.0		−2.7		—		−12.4
Standard deviations[b]		11%					16%				1740%		

[a](—) indicates that negative concentrations were found. [b]These are the standard deviations for the fluorescence intensity measurements. Average relative values are shown.

racy was best for measurements made at the 45° phase angles, and worst at the null phase angles. Negative FITC concentrations were calculated for three of the mixtures when the null phase angles were used. Other phase angle combinations were also tried (ϕ_A − 45°, ϕ_B; ϕ_A, ϕ_B − 45°), but the results were less accurate than those made at both phase angles shifted by 45° (set 2).

The standard deviations for the fluorescence intensity measurements for each method are also shown in Table 1. Precision was best for the measurements made at the phase angles of the components. This is probably due to minimal changes in intensity with fluctuations in phase angle settings at these curve maxima, as well as to the fact that the intensity magnitudes are greatest at these angles. The standard deviation for measurements at angles 45° out of phase with the component phase angles was not much larger than those at the phase angles, and the slightly decreased precision is compensated by the increased accuracy. Precision at the null phase angles was much worse than at the other angles.

Another disadvantage of the use of null phase-angle measurements is the difficulty in locating the exact null angle. Excessive noise at and near the null angle makes this a tedious and time-consuming task. An alternative approach is to find the phase-angle maximum and then shift by 90° to the null phase angle, but this can also be difficult due to the small change in intensity with phase angle in the peak area. The use of two equations at two detector phase angles as described here does not require the exact location of any particular phase angle, as long as two constant phase angles are used for all measurements in a particular determination. The results for the fluorescein/FITC mixtures indicate that the chosen phase angles should be shifted 45° from the component phase angles, but the actual angles used need not be exactly these.

TABLE 2

Phase-resolved determination of fluorescein (A) and FITC (B)
(Conditions as in Table 1)

Added, nM		Found, nM ($\phi_A - 45°$, $\phi_B - 45°$) and relative error (%)			
[A]	[B]	[A]	Error	[B]	Error
50.0	500.0	50.2	00.4	554	11
50.0	250.0	43.8	−12	309	24
25.0	500.0	32.4	30	550	10
25.0	250.0	27.8	11	302	21
5.0	500.0	5.9	18	579	16
Average rel. error			+9.5%		+16%
R.s.d. = ±2.4% for fluorescence intensity measurements					

Another set of mixtures was run in which the FITC concentrations were 5—100 fold greater than the fluorescein concentration. Again, best results (shown in Table 2) were obtained at $\phi_A - 45°$, $\phi_B - 45°$. Good results (average relative error of 4%) were obtained for FITC concentration at the null phase angle of fluorescein, but fluorescein concentration could not be determined at all at the null phase angle of FITC. Precision for this set of mixtures was much better than for the first set (shown in Table 1), because of the much larger signals being measured.

Conclusions

There are several types of experiments in which the distribution of a single species between two different environments must be determined. In protein-ligand binding studies, the distribution of a small molecule between a particular binding site in a protein and free aqueous solution is determined. Similarly, homogeneous (nonseparation) fluoroimmunoassay involves the distribution of a fluorescent-labelled antigen between an antibody and aqueous solution. Fluorescein is the most commonly used label for fluoroimmunoassay, which is why it was chosen for this study.

In both binding studies and fluoroimmunoassay, phase-resolved fluorescence could be used to determine the percents of free and protein-bound fluorophore without requiring any separations. Even very small differences in fluorescence lifetimes of species with essentially identical spectral characteristics can be successfully exploited by using the simultaneous equation approach, as demonstrated in this study.

REFERENCES

1 J. B. Birks and T. H. Monroe, Prog. React. Kinet., 4 (1967) 239.
2 R. D. Spencer and G. Weber, Ann. N.Y. Acad. Sci., 158 (1969) 361.
3 J. R. Lakowicz and H. Cherek, J. Biol. Chem., 256 (1981) 6348.
4 J. R. Lakowicz and H. Cherek, J. Biochem. Biophys. Methods, 5 (1981) 19.

Analytica Chimica Acta, 157 (1984) 333—342
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CATALYTIC EFFECT OF IRON(III) ON THE OXIDATION OF 2-HYDROXYBENZALDEHYDE THIOSEMICARBAZONE BY HYDROGEN PEROXIDE
Kinetic Fluorimetric Determination of Iron

A. MORENO, M. SILVA, D. PEREZ-BENDITO and M. VALCARCEL*
Department of Analytical Chemistry, Faculty of Science, University of Cordoba (Spain)
(Received 15th March 1983)

SUMMARY

A kinetic method is described for the determination of nanogram amounts of iron(III) based on its catalysis of the oxidation of 2-hydroxybenzaldehyde thiosemicarbazone by hydrogen peroxide in an ammoniacal medium. In order to monitor the reaction, the appearance rate of fluorescence of the oxidation product (λ_{ex} = 365 nm, λ_{em} = 440 nm) is measured. The calibration graph is linear in the range 10—60 ng ml^{-1} iron(III) with an r.s.d. of ±1.3%. The proposed method has few interferences and has been applied satisfactorily to the determination of iron in several alloys and minerals. A mechanism for the catalyzed reaction is proposed.

In a continuation of studies on the catalytic effects of metal ions on the oxidation of azomethine compounds [1], it was observed that trace amounts of iron(III) catalyze the oxidation of 2-hydroxybenzaldehyde thiosemicarbazone (HBTS) by hydrogen peroxide in ammoniacal medium. This redox-catalyzed reaction is used for the kinetic fluorimetric determination of iron(III). The oxidation of HBTS by hydrogen peroxide is catalyzed by manganese(II) [2] and a kinetic fluorimetric determination of zinc(II) based on its activating effect on this catalyzed reaction has been described [3]. The proposed method for the determination of iron(III) is slightly less sensitive than that for manganese [2], and the precision is similar. The determination of mixtures of both catalysts by differential kinetic analysis is possible [4].

The kinetic fluorimetric determination of iron(III) has seldom been described. One method is based on the 4,4′-bis(dicarboxymethylamino)-stilbene-2,2′-disulphonic acid/hydrogen peroxide indicator reaction in an acetic acid medium [5—7]. Iron can be determined in the range 0.02—0.1 μg in a final volume of 5 ml by its catalytic action on this reaction which is followed by measuring the rate of quenching of the fluorescence of the organic compound. Recently, Salinas et al. [8] have proposed a kinetic fluorimetric determination of iron(III) based on its effect on the hydrolysis product of 1,4-diamino-2,3-dihydroanthraquinone. The method proposed here is very

0003-2670/84/$03.00

sensitive and selective, permitting the determination of iron concentrations at the 10 ng ml^{-1} level.

EXPERIMENTAL

Reagents and equipment

All solvents and reagents used were of analytical grade. A standard iron solution (1.000 mg ml^{-1}) was prepared by dissolving 1.000 g of pure iron in 50 ml of (1 + 1) nitric acid and diluting to exactly 1 l with distilled water. Solutions of lower concentrations were prepared by appropriate dilution. A 0.03% (w/v) 2-hydroxybenzaldehyde thiosemicarbazone solution was prepared in ethanol. This solution was stable for at least a month. The HBTS was prepared by the method of Sah and Daniels [9], by condensing 2-hydroxybenzaldehyde with thiosemicarbazide in an ethanolic medium. The white crystals were isolated and recrystallized from ethanol.

Fluorescence was measured with a Perkin-Elmer MPF-43A spectrofluorimeter fitted with a device that permits direct recording of fluorescence/time graphs at fixed excitation and emission wavelengths. The instrumental conditions used were: sensitivity $\times 0.1$; emission and excitation slits, 6 nm. Under these conditions, a 8×10^{-6} M solution of quinine sulphate showed a fluorescence intensity of 70% on a scale adjusted to zero in the absence of quinine.

The pH values were measured with a Beckman 3500 pH meter equipped with a combined glass/calomel electrode.

Kinetic—fluorimetric determination of iron

To a solution containing up to 600 ng of iron(III) in a 10-ml standard flask, add 0.5 ml of 0.03% HBTS solution in ethanol, 1 ml of 1% (v/v) hydrogen peroxide solution, 1.5 ml of 4.3 M ammonia solution, 0.4 ml of 0.1 M hydrochloric acid and 2.5 ml of ethanol, and start the stopwatch. Make up to the mark with distilled water, mix and, after 1 min, transfer a portion to a 1-cm quartz fluorimeter cell maintained at 50 ± 0.1°C. Wait for 2 min before starting to record the fluorescence intensity (λ_{ex} = 365 nm, λ_{em} = 440 nm) as a function of time. Follow the uncatalyzed reaction under similar conditions without the addition of iron. The reaction rate for the catalyzed reaction is calculated from the difference in the slopes between the two fluorescence/time plots.

Preparation of samples

Nickel—chrome alloy. Add 5 ml of aqua regia to an exactly weighed amount of alloy (50 mg) in a glass vessel. Warm gently to complete dissolution. Cool and dilute to 100 ml in a volumetric flask with distilled water. In order to remove the interference of nickel, add 2 ml of concentrated hydrochloric acid to 10 ml of this solution. Percolate the solution through a Dowex 1-X8 anion-exchange column (2 cm diameter, 10 cm high).

Elute the nickel with 2 M hydrochloric acid, then elute the iron and other ions with distilled water and collect in a 100-ml volumetric flask. Make up to volume, take an aliquot of 2.0 ml and proceed as above for the determination of iron.

Aluminium bronze. Dissolve a quantity of bronze (50 mg) in the minimum volume of (1 + 1) nitric acid and dilute to 100 ml with distilled water in a volumetric flask. Take an aliquot of 20 μl for the iron determination.

Zinc concentrate. To a sample of the mineral (50 mg) in a 100-ml Erlenmeyer flask, add 10 ml of (1 + 1) hydrochloric acid dropwise and warm gently until hydrogen sulphide ceases to be evolved. Digest for 30 min, add 5 ml of concentrated nitric acid and heat the mixture until dissolution is complete. Cool, transfer the solution to a 100-ml volumetric flask and dilute with distilled water to the mark. The concentration of iron is very high in this solution. Therefore, dilute 1 ml of the sample to 100 ml with distilled water and take an aliquot of 0.5 ml from this solution for the measurement of iron.

Portland cement. To a glass vessel, add a known amount of cement (50 mg), 5 ml of water and 5 ml of concentrated hydrochloric acid. Place the vessel on a sand bath until dissolution is complete, and evaporate to dryness at 100—105°C. Cool, and add 2.5 ml of (1 + 1) hydrochloric acid. Digest for 10 min, dilute with 2.5 ml of hot water, and filter quickly. Wash the precipitate of silicic acid with hydrochloric acid (1 + 100) and dilute to 100 ml with distilled water in a volumetric flask. In order to eliminate the interference of calcium, add 10 ml of concentrated hydrochloric acid to 50 ml of this solution. Percolate the solution through a Dowex 1-X8 anion-exchange column (as above), and proceed as for the nickel—chrome alloy. Collect in a 250-ml volumetric flask. Take an aliquot of 0.25 ml for the determination of iron.

Dolomite and low silica magnesite—chrome. Follow the same procedure as for Portland cement. In the case of dolomite, in order to remove the interference of the calcium and magnesium ions, take 50 ml of the solution, add 10 ml of concentrated hydrochloric acid and percolate this solution through the Dowex 1-X8 anion-exchange column; then proceed as for Portland cement. Collect in a 100-ml volumetric flask, and use a 0.7-ml aliquot. For the determination of iron in low-silica magnesite—chrome, dilute the solution 1:100 with distilled water and take 0.6 ml for the determination of iron.

RESULTS AND DISCUSSION

Catalytic effect of iron(III)

2-Hydroxybenzaldehyde thiosemicarbazone in an ammoniacal medium is only slowly oxidized by hydrogen peroxide, sodium periodate and potassium peroxodisulphate. In the presence of iron(III), the rate of oxidation with hydrogen peroxide is increased and this change can be measured at the

excitation (365 nm) and emission (440 nm) wavelengths of the oxidation product. Figure 1 shows the effect of iron(III) on fluorescence/time curves for this reaction.

The influence of temperature on the reaction rate was studied in the range 15—60°C. The slopes (tan α) of the fluorescence/time curves at the different temperatures (Fig. 2a) show that the initial rate increases with temperature, but is constant between 50 and 60°C. Therefore, 50°C was selected for use. The initial temperature of all solutions was room temperature (ca. 25°C); under such conditions, for 15 ng ml^{-1} iron, the initial rate was 0.052 min^{-1}

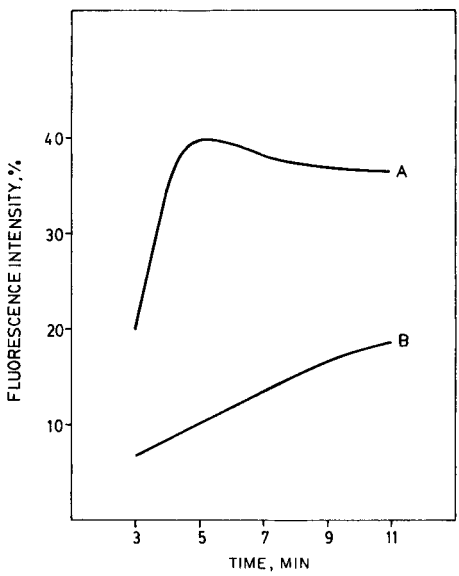


Fig. 1. Catalytic effect of iron(III) on the oxidation of HBTS by hydrogen peroxide in an ammoniacal medium: (A) reaction catalyzed by 50 ng ml^{-1} iron; (B) uncatalyzed reaction. (50°C, 7.7 × 10^{-5} M HBTS, 3.2 × 10^{-2} M H_2O_2, 30% ethanol, 0.65 M ammonia, pH = 11.3).

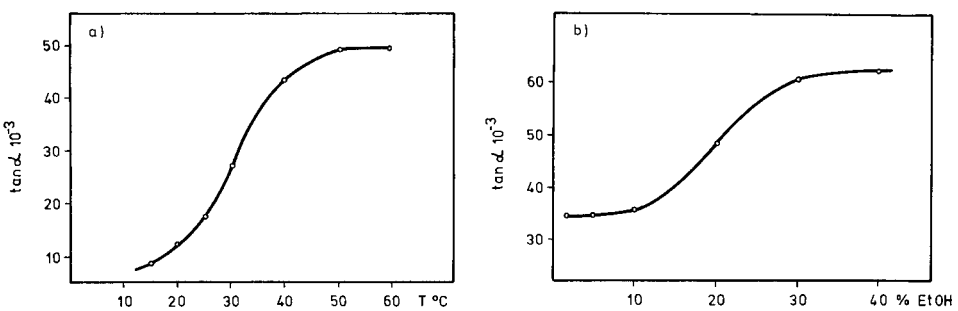


Fig. 2. (a) Effect of temperature on the initial rate (100 ng Fe ml^{-1}, 1.45 × 10^{-2} M H_2O_2, 1.3 × 10^{-4} M HBTS, 0.4 M ammonia); (b) variation of the initial rate with ethanol concentration (50 ng Fe ml^{-1}, 1.45 × 10^{-2} M H_2O_2, 7.7 × 10^{-5} M HBTS, 0.4 M ammonia, 50°C).

and the relative standard deviation was 0.64% (6 results). If the reagents were preheated to 50°C before mixing, the initial rate was 0.047 min^{-1} ± 0.95%. Therefore, the former method is superior because it has higher sensitivity and greater precision for the catalytic determination of iron. This method is also simpler.

The order of addition of reagents and adjustment of the ionic strength to 1 M did not appreciably affect the performance of the method. Nor did variations in the ethanol concentration of up to 5%, but for 5—30% ethanol, the initial rate increased as shown in Fig. 2(b). At higher ethanol concentrations, the rate was little affected. The enhancing effect of the ethanol could be attributed to its interaction because of its weak reducing properties, in one of the intermediate reaction stages.

The effect of HBTS concentration on the reaction rate was studied in the range $0.25—2.5 \times 10^{-4}$ M. The rate of the catalytic reaction increased linearly with HBTS concentration up to 7.7×10^{-5} M. At higher concentrations, the initial rate decreased linearly with HBTS concentration. Thus, 0.5 ml of 0.03% reagent solution in ethanol was selected for use. The initial rate increased slightly with concentration of hydrogen peroxide up to 1.8×10^{-2} M (partial order 1/4). Between 1.8×10^{-2} M to 2.6×10^{-2} M the rate was proportional to the hydrogen peroxide concentration and at higher values, this initial rate was constant. Therefore, 1 ml of 1% hydrogen peroxide (3.2×10^{-2} M) is recommended.

The rate of the catalytic reaction increased with increasing ammonia concentration up to 0.54 M; in the range 0.54—0.87 M, the initial rate was not affected by ammonia concentration and at higher values the rate decreased. In order to keep precision as high as possible, the ammonia concentration selected was 0.65 M (1.5 ml of 4.33 M ammonia solution). The pH—ammonium ion effect on the system was studied for mixtures of ammonium chloride and ammonia giving a total "ammonia" concentration of 0.65 M. In the experimental procedure, some 10^{-2} M hydrochloric acid was added to each sample in order to obtain a suitable pH value. In this case, the variation of the initial rate was affected by two variables: hydrogen and ammonium ions. The results of this study are shown in Table 1. The partial orders of hydrogen and ammonium ions were calculated simultaneously because both species are closely related. In fact, if the hydrogen ion concentration remains constant and if in all experiments the total ammonia concentration is 0.65 M, the ammonium ion concentration must be constant. At constant concentrations of other variables, therefore, $\tan\alpha/[H^+]^x[NH_4^+]^y = K_1$, a constant. Thus, the x and y values (partial orders in hydrogen and ammonium ions, respectively) were calculated by testing pairs of values which gave values of K_1 in different experiments with a minimum standard deviation between one another. These partial orders were calculated by using a computer program to be $x = y = 1$.

The influence of ammonia is decisive in this catalyzed reaction. Several experiments were done at optimum pH in the absence of ammonia and in

TABLE 1

Dependence of the rate of the catalyzed reaction on hydrogen and ammonium ion concentrations

Initial HCl conc.[a] ($\times 10^{-3}$ mol l^{-1})	pH[b]	Actual composition		tan α[e]	tan α
		$[H^+]$[c] ($\times 10^{-12}$ mol l^{-1})	$[NH_4^+]$[d] ($\times 10^{-3}$ mol l^{-1})		$[H^+][NH_4^+]$
0.00	11.46	3.47	4.05	0.049	3.488×10^{12}
0.98	11.41	3.89	4.55	0.056	3.166×10^{12}
1.47	11.39	4.07	4.76	0.065	3.352×10^{12}
2.45	11.35	4.47	5.22	0.081	3.474×10^{12}

[a]Initial ammonia concentration was 0.65 ml l^{-1}. [b]Measured with a pH meter. [c]Calculated from pH value. [d]Calculated from pH, known concentrations and pK_b for NH_3. [e]After subtraction of rate of uncatalyzed reaction.

the presence of other amine bases. In an alkali metal hydroxide medium, the catalytic effect of iron(III) decreased to only 10% of the value observed when ammonia was present in the reaction medium. With other nitrogen bases, such as butylamine and triethanolamine, the reaction rate was similar to that with ammonia.

The fluorescence intensity/time curves were recorded for solutions containing different quantities of iron(III), under the optimum experimental conditions. The iron-catalyzed oxidation of HBTS was first order with respect to iron.

The partial orders of the several variables obtained from the log—log plots of rate data for the catalytic oxidation of HBTS are summarized in Table 2. For the recommended procedure for the kinetic fluorimetric determination of iron(III), the following kinetic equation is proposed

$$d[HBTS]_{ox}/dt = K[HBTS][H^+][NH_4^+][Fe^{3+}] + K'$$

where $[HBTS]_{ox}$ is the concentration of the oxidized reagent, K is the conditional rate constant for the catalyzed reaction and K' is a constant for the uncatalyzed reaction.

TABLE 2

Summary of the kinetic data for the HBTS/H_2O_2/Fe(III) system

Variable and concentration range	Partial order	Variable and concentration range	Partial order
$[HBTS] < 7.7 \times 10^{-5}$ M	1	$[NH_3] > 0.87$ M	−2
$[HBTS] > 7.7 \times 10^{-5}$ M	−1	pH < 11.3	1
$[H_2O_2] < 1.8 \times 10^{-2}$ M	1/4	pH > 11.3	−1
1.8×10^{-2} M < $[H_2O_2] < 2.6 \times 10^{-2}$ M	1	$[NH_4^+] < 5.83 \times 10^{-3}$ M	1
$[H_2O_2] > 2.6 \times 10^{-2}$ M	0	$[NH_4^+] > 5.83 \times 10^{-3}$ M	0
$[NH_3] < 0.54$ M	3/4	10 ng ml^{-1} < $[Fe^{3+}]$ < 60 ng ml^{-1}	1
0.54 M < $[NH_3] < 0.87$ M	0		

Kinetic determination of iron

Calibration graphs. The intensity of the fluorescence/time curves were obtained for solutions containing different amounts of iron(III), under optimum conditions. To these curves, various well-established methods of preparing calibration graphs were applied. Table 3 shows the concentration range over which the calibration graphs are linear and the relative standard deviations for the determination of iron by initial rate, fixed-time and fixed-intensity methods. In the initial rate method, only an initial portion of the reaction is utilized. For the fixed-time method, a time interval of 3.5 min was chosen for the measurements of the fluorescence. For the fixed-intensity approach, the inverse of the time required to reach a fluorescence of 25% was plotted against the concentration of iron. The initial rate method was selected on the basis of the greater precision and the wider linear range than the other methods. The detection limit for this method was calculated [10] to be 0.11 ng ml^{-1}.

Interferences. For the determination of 10 ng ml^{-1} iron by the initial rate method, other ions can be tolerated up to the levels given in Table 4. In order

TABLE 3

Linear calibration ranges and relative standard deviations for the kinetic—fluorimetric determination of iron(III)

Method	Linear range (ng ml^{-1})	R.s.d.[a] (%)
Initial rate	10—60	1.33
Fixed time	10—60	1.50
Fixed intensity	15—60	1.76

[a]Obtained by 11 measurements on 20 ng Fe ml^{-1}.

TABLE 4

Tolerance for foreign ions in the determination of 10 ng ml^{-1} iron by the initial rate method

Ion added[a]	Amount tolerated (ng ml^{-1})
Ag(I), La(III), CN^-	10
Ni(II), Ca(II), $C_2O_4^{2-}$, tartrate	50
Os(VIII), Hg(II), IO_4^-	100
Hf(IV), Pb(II), UO_2(II)	250
Pt(IV), Mg(II), Pd(II), Au(III), Ce(IV), CrO_4^{2-}	500
Ba(II), Tl(I), Zn(II), Cr(III), Sb(III), NO_2^-	1000
Co(II), Al(III), Bi(III), Th(IV), Ti(IV)	2500
Sr(II), Zr(IV), Sn(IV), Cd(II), Sn(II), $P_2O_7^{4-}$, $S_2O_3^{2-}$, ClO_4^-, AsO_3^{3-}, PO_4^{3-}, S^{2-}	5000
K(I), Se(IV), Na(I), Be(II), NO_3^-, Br^-, CO_3^{2-}, Cl^-, citrate, SO_4^{2-}, I^-, SCN^-, $S_2O_8^{2-}$, F^-, ClO_3^-, IO_3^-, MoO_4^{2-}, VO_3^-, WO_4^{2-}, BrO_3^-, acetate,	10 000[b]

[a]Cations added as chlorides, nitrates or acetates, anions as sodium or potassium salts.
[b]Maximum tested.

to develop the application of this kinetic method for the determination of iron in nickel–chrome alloys and aluminium bronze, the tolerance limits for manganese and copper ions, which are also effective catalysts, were calculated more accurately. They were found to be 1:1.5 and 1:35 (w/w) for iron to foreign ion, respectively. These levels are higher than those present in the analyzed samples.

Applications

Results for the determination of iron in alloys and mineral samples from the Bureau of Analysed Samples (U.K.) and British Chemical Standards (U.K.) indicate the precision and accuracy of this method. Table 5 shows the results obtained for the various samples analyzed.

Nature of the fluorescence product, and mechanism of the catalyzed reaction

The oxidation products were isolated and characterized in order to suggest the nature of the fluorescent product of the catalytic reaction. A solution of the isolated product under the recommended experimental conditions for the determination of iron exhibited a very intense blue fluorescence (λ_{ex} = 365 nm, λ_{em} = 440 nm). Its elemental analysis revealed a smaller carbon content than HBTS, which suggested an oxidative hydrolytic breakage of the C=N— bond. The i.r. spectrum of this compound, compared with that of HBTS, revealed bands corresponding to N=N—O (1480 cm^{-1}), NO (1400 cm^{-1}) and SO_2 (1130 cm^{-1}) while bands corresponding to $\rangle$C=N— (1685 cm^{-1}), $\rangle$C=S (1225 cm^{-1}) and the aromatic ring had disappeared. This is in agreement with several studies by Walter and Rohloff [11] of the transformation of $\rangle$C=S to C—SO_3^- for numerous thiosemicarbazones in the presence of various oxidants. Also, the dianion of 2-hydroxybenzoic acid was detected in the filtrate, by qualitative organic analysis and u.v. spectrometry.

These data suggest the hydrolytic oxidation of 2-hydroxybenzaldehyde thiosemicarbazone catalyzed by iron(III) yielding the 2-*N*-oxide of 4-amine-5,5′-dioxo-1,5-oxathia-4-hydro-2,3-diazole:

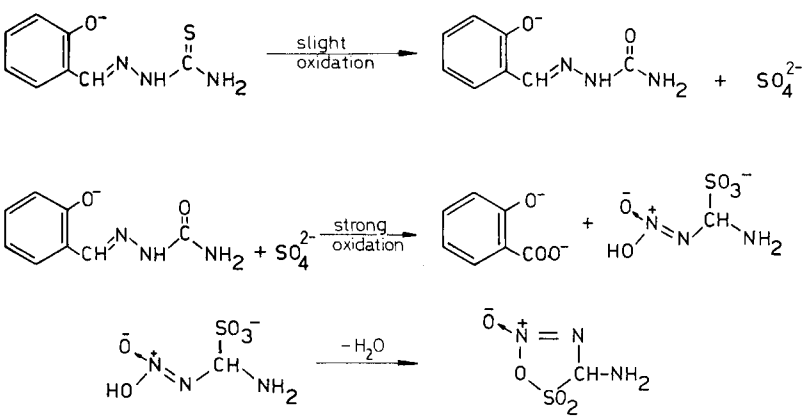


TABLE 5

Determination of iron in alloys and mineral samples

Sample	Iron content (%)	
	Reported	Found[a]
Nickel—chrome alloy (B.C.S. No. 310/1)	0.25	0.25 ± 0.01
Aluminium bronze (B.A.S. No. 32a)	2.67	2.68 ± 0.02
Zinc concentrates (B.A.S. No. 41dG)	10.0	10.07 ± 0.07
Portland cement (B.C.S. No. 372)	2.49[b]	2.50 ± 0.01[b]
Dolomite (B.A.S. No. 9h)	0.21[b]	0.207 ± 0.002[b]
Low silica magnesite—chrome (B.C.S. No. 396)	10.9	10.8 ± 0.1

[a]Mean ±r.s.d. for 6 separate determinations. [b]$AsFe_2O_3$.

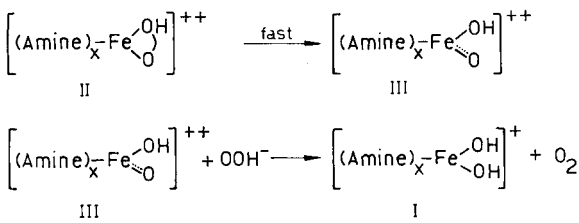
The elemental results for the heterocycle agreed reasonably well with the formula proposed. The presence of the azoxy group, a lactone ring and the rigidity and planarity in the proposed oxidation product could explain its intense fluorescence.

It is possible to detect an intermediate step in the above reaction when it proceeds in the presence of a slight excess of hydrogen peroxide (giving less sensitivity for the kinetic determination of iron). In this case, 2-hydroxybenzaldehyde semicarbazone (λ_{ex} = 365 nm, λ_{em} = 440 nm) and sulphate are formed. The azomethine compound was confirmed by elemental analysis and i.r., u.v. and fluorescence spectroscopy and the sulphate ion was confirmed qualitatively (formation of $BaSO_4$). Furthermore, with a large excess of hydrogen peroxide, these compounds yielded the indicated heterocyclic oxidation product and the dianion of 2-hydroxybenzoic acid. This latter reaction only occurs with 2-hydroxybenzaldehyde semicarbazone when sulphate is present. According to Wang [12], the catalytic action of iron(III) in this reaction may be due to its effect on the decomposition of hydrogen peroxide yielding oxygen, which acts as the true oxidant. Several data suggest this to be true, e.g., the presence of nitrogen bases in the reaction medium, strong catalytic activity only by iron(III) and manganese(II) [2], the pH value (11.3) at which the iron(III) shows its highest catalytic effect, and the low concentration range ($\approx 10^{-7}$ M) for the determination of iron. Therefore, the following mechanism for the decomposition of hydrogen peroxide by iron(III) under the present conditions is suggested.

$$x\ \text{Amine} + Fe^{3+} + 2\,OH^- \longrightarrow \underset{\text{I}}{\left[(\text{Amine})_x\text{-Fe}^{OH}_{OH}\right]^+}$$

$$H_2O_2 \overset{OH^-}{\rightleftharpoons} OOH^- + H^+$$

$$\underset{\text{I}}{\left[(\text{Amine})_x\text{-Fe}^{OH}_{OH}\right]^+} + OOH^- \longrightarrow \underset{\text{II (unstable)}}{\left[(\text{Amine})_x\text{-Fe}^{OH}_{O}\right]^{++}} + 2\,OH^-$$



In the presence of hydrogen peroxide, the two hydroxide ions in compound I may be replaced by a hydroperoxide (OOH^-) ion, yielding compound II. But, because the O—O bond length in OOH^- is only about 1.3 Å, compound II is unstable and the O—O bond tends to break to form compound III. This compound can readily react with a second hydroperoxide ion to yield oxygen and regenerate compound I.

REFERENCES

1 M. Valcarcel and F. Grases, Talanta, 30 (1983) 139.
2 A. Moreno, M. Silva, D. Perez-Bendito and M. Valcarcel, Talanta, 30 (1983) 107.
3 A. Moreno, M. Silva, D. Perez-Bendito and M. Valcarcel, Analyst, 108 (1983) 85.
4 D. Perez-Bendito, M. Silva and A. Moreno, Anal. Chim. Acta, (1984) in press.
5 S. U. Kreingold, E. A. Bozhevolnov, R. P. Lastovskii and V. V. Sidorenko, Zh. Anal. Khim., 18 (1963) 1356.
6 M. Laanmaa, M. L. Allsalu and H. Kokk, Tartu Miikliku Ulticodi, 219 (1963) 199.
7 A. A. Abraztov and V. G. Bocharova, Tr. Voronezh. Gos. Univ., 82 (1971) 182.
8 F. Salinas, C. Genestar and F. Grases, Anal. Chim. Acta, 130 (1981) 337.
9 P. P. T. Sah and T. C. Daniels, Recl. Trav. Chim. Pays-Bas, 69 (1959) 1545.
10 K. B. Yatsimirskii, Kinetic Methods of Analysis, Pergamon Press, Oxford, 1966, p. 19.
11 W. Walter and Ch. Rohloff, Justus Liebigs Ann. Chem., (1977) 447, 463, 485.
12 J. H. Wang, J. Am. Chem. Soc., 77 (1955) 4715.

Analytica Chimica Acta, 157 (1984) 343—348
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

DISTRIBUTION EQUILIBRIA OF ALUMINUM—PYROCATECHOL VIOLET—QUATERNARY ONIUM SALT ION-PAIRS IN MICELLAR SYSTEMS

Spectrophotometric Determination of Aluminum

SHOJI TAGASHIRA

Faculty of Science, Yamaguchi University, Yamaguchi 753 (Japan)

(Received 11th April 1983)

SUMMARY

The distribution equilibria of the ion-pairs of the aluminum—pyrocatechol violet complex with zephiramine, hexadecyltrimethylammonium (HTA) bromide or tetraphenylphosphonium (TPP) chloride were examined spectrophotometrically in surfactant micellar solutions. Hypsochromic shifts are attributed to dissolution of the ion-pair in the micelles. The TPP system was selected for aluminum determinations because the distribution of TPP ion-pair between water and Triton X-100 was less than that of the HTA or zephiramine ion-pair. A molar absorptivity of 64 000 l mol^{-1} cm^{-1} at 710 nm was obtained in 0.1% Triton X-100 solution. The proposed method was applied to the determination of aluminum in seaweed.

The color changes of certain metal complexes with acidic dyes such as pyrocatechol violet (pyrocatechol sulfonaphthalein) in cationic surfactant solutions have been applied in the spectrophotometric determination of such metals in order to improve sensitivity and selectivity [1—3]. Pyrocatechol violet is commonly used as a spectrophotometric reagent for aluminum [4], and forms water-soluble, colored chelates with many metals [5]. Chester et al. [6] reported that the addition of cationic surfactants such as hexadecyltrimethylammonium (HTA) bromide to aqueous solutions of aluminum and pyrocatechol violet caused a bathochromic shift. In these systems, the cationic surfactant not only forms an ion-pair with the anionic aluminum—pyrocatechol violet chelate but also dissolves or disperses the complex. However, the reported methods involve problems concerning spectral stability. When a large excess of the surfactant is present, the absorbance decreases with increasing surfactant concentration and there is a hypsochromic shift.

In aqueous solution, the absorption peaks of the aluminum—pyrocatechol violet chelate and its ion-pair with cationic surfactants are at 580 nm [4] and 670 nm [6], respectively. The peaks of the ion-pairs are at 590 nm in trichloromethane [7] and 597 nm in n-butyl acetate [8]. This suggests that at higher surfactant concentrations, the ion-pair is distributed inside the micelle. Tetraphenylphosphonium (TPP) chloride, which is considered not to

0003-2670/84/$03.00

form micelles, formed an ion-pair with an absorption peak at 710 nm in aqueous solutions. Despite addition of a large amount of this salt, there was no hypsochromic shift of the spectrum of the aqueous solution.

In the present paper, the distribution equilibria of the ion-pairs between water and surfactant micellar phases are examined, and the most suitable system is used for the spectrophotometric determination of aluminum.

EXPERIMENTAL

Reagents and apparatus

Pyrocatechol violet, tetraphenylphosphonium chloride, zephiramine (all from Dozin Chemical Laboratory), Triton X-100 and hexadecyltrimethylammonium bromide (Wako Pure Chemicals) were used. A 1 M ammonium sulfate—ammonia buffer (pH 9.5) and a 1% (w/v) ascorbic acid solution were prepared. A standard aluminum solution was prepared from $KAl(SO_4)_2 \cdot 12H_2O$ and standardized gravimetrically with ammonia (5.149×10^{-2} M). Other reagents were of analytical grade.

A Shimadzu model UV-200 recording spectrophotometer was used with 10-mm silica cells. A Toa HM-5A type pH meter was used with a combined glass electrode.

General procedures

An aliquot of aluminum solution was placed in a 10-ml volumetric flask. Ascorbic acid, pyrocatechol violet, TPP chloride and Triton X-100 were added successively (recommended concentrations are given in Fig. 4, except for 3.0×10^{-2} M TPP chloride). The flask was shaken slowly, the pH of the solution was adjusted with buffer to 9.1—9.5, and the absorbance was measured.

For liquid-liquid extraction, 15 ml of sample solution containing Cu^{2+}, Ni^{2+} and Co^{2+} was transferred to a 100-ml separatory funnel, 2 ml of aqueous 2% potassium diethyldithiophosphate [9] was added and the pH was adjusted to 1 with sulfuric acid. The solution was shaken with 15 ml of trichloromethane for 5 min. After phase separation, the aqueous phase was treated as above for the determination of aluminum.

RESULTS AND DISCUSSION

Distribution equilibria

Figure 1 shows the spectra of aqueous solutions containing aluminum, pyrocatechol violet and various concentrations of HTA bromide. The spectra of the ion-pair (spectra 2—5) have an isosbestic point at 640 nm. The spectra of the aqueous ion-pair system were slightly shifted to a shorter wavelength by additions of a large concentration of a nonionic surfactant such as Triton X-100; the peak was at 590 nm in trichloromethane. These facts support the distribution of ion-pairs between water and a micellar phase.

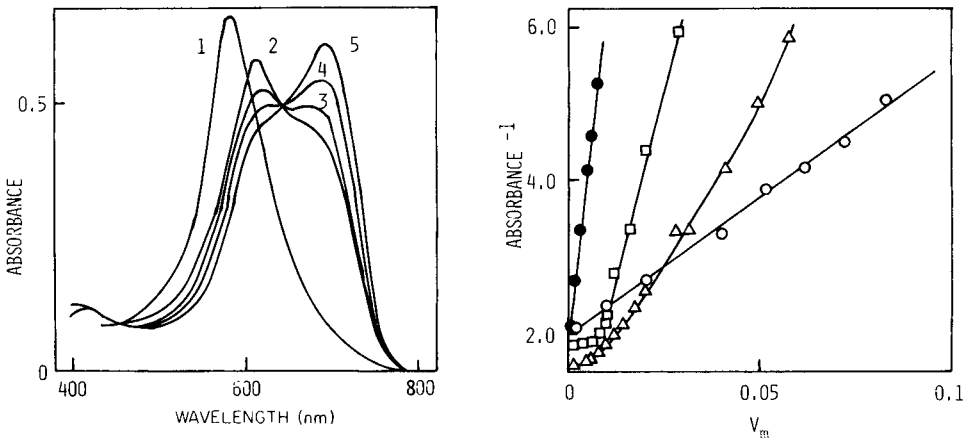


Fig. 1. Absorption spectra of the Al—pyrocatechol violet chelate and of the ion-pair with HTA (1.0×10^{-5} M Al^{3+}, 2.0×10^{-5} M pyrocatechol violet, pH 9.5) in the presence of HTA bromide concentrations of: (1) 0; (2) 0.05; (3) 0.11; (4) 0.22; (5) 0.33%. Spectra measured against pure water.

Fig. 2. Plots of reciprocal absorbance as a function of volume fractions of: (○, △, □) Triton X-100 and (●) HTA micelles (1.0×10^{-5} M Al^{3+}, 2.0×10^{-5} M pyrocatechol violet, 0.2% of ascorbic acid, pH 9.5), measured in the presence of: (○) 2.0×10^{-2} M TPP chloride at 700 nm; (□) 1.0×10^{-3} M HTA bromide at 688 nm; (△) 1.0×10^{-3} M zephiramine at 695 nm; (●) at 688 nm against pure water.

The distribution ratio of an ion-pair is given by $D = [M]_m/[M]_w$, where $[M]_w$ and $[M]_m$ are the ion-pair concentrations in water and the micellar phase, respectively. By selecting a suitable wavelength at which absorption by the ion-pair in the micelles is negligible, the absorbance of the solution is given by $A = \epsilon[M]_w l$, where ϵ is the molar absorptivity of the ion-pair in the water phase and l is the pathlength. The apparent concentration of the ion-pair in the solution is given by $[M]_a = [M]_m V_m + [M]_w V_w$, where V_w and V_m are the volume fractions of the water and micellar phases, respectively; V_m can be calculated by employing a molar volume ϕ_v, of the surfactant forming the micelle: $V_m = \phi_v(C - \text{c.m.c.})$, where C and c.m.c. are the concentration and critical micelle concentration [10], respectively, of the surfactant. For Triton X-100, $\phi_v = 1.29\,M^{-1}$ [11], and c.m.c. = 2.4×10^{-4} M; for HTA bromide, $\phi_v = 0.28\ M^{-1}$ [12], and c.m.c. = 8.0×10^{-4} M. From the above equations, the following working equation is derived

$$1/A = 1/(\epsilon[M]_a) + V_m(D - 1)/(\epsilon[M]_a)$$

Figure 2 shows plots of $1/A$ against V_m for Triton X-100 and HTA micelles. Plots for the HTA and zephiramine ion-pairs as a function of V_m for Triton X-100 were not linear in dilute Triton X-100 solutions, so that values of D could not be obtained in these systems. The distribution ratio (D = 17) obtained for the Al—pyrocatechol violet ion-pair with TPP between water

and Triton X-100 micelles is smaller than that (D = 250) for the ion-pair with HTA between water and HTA micelles. The latter ion-pair may be adsorbed on the cationic micelle surface and easily dissolved inside the micelle. The solubility of the Al—pyrocatechol violet ion-pairs in Triton X-100 micelles is qualitatively in the order HTA > zephiramine > TPP.

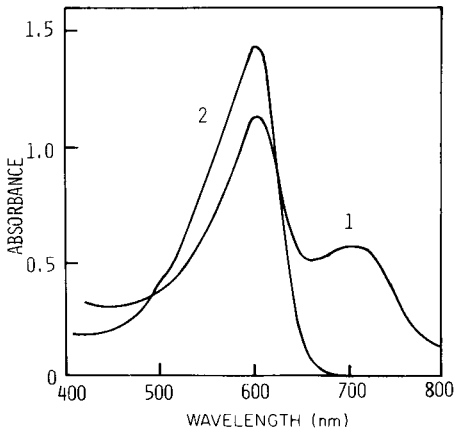
The bathochromic shift of solutions involving the aluminum complex and cationic surfactants has been explained by the deprotonation of the ligand on the micelle surface [13]. However, this explanation is inadequate, because the shift for the Al—pyrocatechol violet ion-pair with TPP occurred in the absence of surfactant micelles. Furthermore, Mukerjee and Cardinal [14] indicated that the effective dielectric constants at the micelle surface are lower than that of pure water; that for the HTA micelle is 43. The decrease in the dielectric constant increases the protonation of the ligand. A bathochromic shift was observed for tetraphenylarsonium chloride, but not for tetraethylammonium and tetrabutylammonium salts. The bathochromic shift of the spectrum may be attributed to charge transfer owing to π—π interaction between the phenyl groups of the complex and those of TPP chloride.

The hypsochromic shift resulting from the dissolution of the complex inside the micelles decreases the sensitivity and reliability for the determination of aluminum, and small amounts of precipitated ion-pair with TPP were observed after a few minutes. Therefore, Triton X-100 was used as a dispersant for the ion-pair.

Application to the determination of aluminum

Addition of TPP chloride to an aqueous solution containing aluminum and pyrocatechol violet caused a new absorption peak at 710 nm, at which wavelength the absorbance of the dye itself was negligible (Fig. 3). The effect of TPP chloride concentrations was examined in the range 1.0×10^{-3}—4.0×10^{-2} M; maximum absorbance was obtained with more than 2.0×10^{-2} M (Fig. 4). In the pH range 9.1—10.8, maximum constant absorbance was obtained. However, pyrocatechol violet in alkaline solutions was readily oxidized and the absorbance decreased with time, therefore 0.2% ascorbic acid was added to stabilize the absorbance. Under these conditions, constant absorbance was obtained with 5.0×10^{-5}—2.0×10^{-4} M pyrocatechol violet. In the presence of 0.1% Triton X-100 the solution was stable for 30 min. As discussed above, the quantity of the ion-pair with TPP in the Triton X-100 micelles is negligible because of the small distribution ratio and the low concentration of the surfactant.

The calibration graph was linear over the range 1.0×10^{-6}—1.5×10^{-5} M aluminum and the molar absorptivity was 64 000 l mol^{-1} cm^{-1}. This value is slightly larger than values reported when cationic surfactants were used (53 000 [6] and 58 000 [13] at 670 nm).



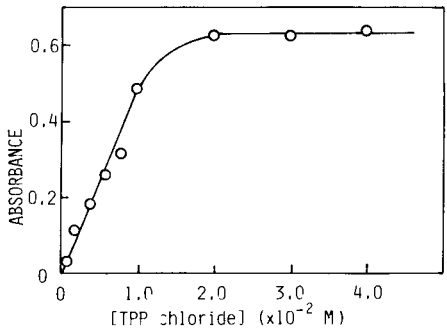


Fig. 3. Absorption spectrum of the Al—pyrocatechol violet ion-pair with TPP measured at pH 9.5 against pure water: (1) 1.0 × 10^{-5} M Al^{3+}, 5.0 × 10^{-5} M pyrocatechol violet, 3.0 × 10^{-2} M TPP chloride, 0.2% ascorbic acid, 0.1% Triton X-100; (2) reagent blank.

Fig. 4. Effect of TPP chloride concentration on absorbance at 710 nm (1.0 × 10^{-5} M Al^{3+}, 1.0 × 10^{-4} M pyrocatechol violet, 0.2% ascorbic acid, 0.1% Triton X-100, measured at pH 9.1—9.5 against reagent blank).

Removal of interfering ions

For a 1.0 × 10^{-5} M aluminum solution, no interference was found from a 1000-fold molar quantity of fluoride, 100-fold amounts of Na^+, K^+, Ca^{2+}, acetate and HPO_4^{2-}, or a 10-fold amount of Mg^{2+}. However, 10-fold amounts of Co^{2+}, Ni^{2+} or Cu^{2+} slightly increased the absorbance, and 10-fold amounts of Fe^{2+}, Fe^{3+} or CrO_4^{2-} gave turbid solutions.

In order to eliminate interferences by heavy metals, pretreatments by liquid-liquid extraction were developed. After diethyldithiophosphate extraction, 100-fold amounts of Cu^{2+}, Ni^{2+} and Co^{2+} caused no interference, but Fe and Cr were not completely extracted. However, the use of ammonium pyrrolidinedithiocarbamate (1 ml, 2%) and sodium fluoride (1 ml, 0.5%), in place of the diethyldithiophosphate, and extraction at pH 3.5—4.5, prevented interference from 50-fold amounts of Fe^{2+}, Fe^{3+} or CrO_4^{2-}.

Analysis of dried seaweeds

Three commercially available dried seaweeds were analysed which were found to contain about 120 mg kg^{-1} iron and 50 mg kg^{-1} copper. The samples were decomposed with hydrogen peroxide and concentrated sulfuric acid [15] and were presented for aluminum determination as 0.1 M sulfuric acid solutions. After extraction of the heavy metals, aluminum in the aqueous phase was determined by the procedure given. Three samples were found to contain 61.6, 45.6 and 7.47 mg kg^{-1} aluminum, respectively, in agreement with the values (64.8, 44.9 and 7.12 mg kg^{-1}) obtained by atomic absorption spectrometry.

The author is grateful to Prof. Y. Yamamoto, Prof. K. Hayashi and Dr. Iwamoto for valuable discussions and suggestions, and to Mr. A. Maeda for technical assistance.

REFERENCES

1 See, e.g., W. L. Hinze, in K. L. Mittal (Ed.), Solution Chemistry of Surfactants, Vol. 1, Plenum Press, New York, 1979, p. 79.
2 V. N. Tikthonov and T. Y. Stepanova, Zh. Anal. Khim., 34 (1979) 426.
3 He Xi-Wen and D. P. Poe, Anal. Chim. Acta, 131 (1981) 195.
4 W. K. Dougan and A. L. Wilson, Analyst, 99 (1974) 413.
5 G. Schwarzenbach, Die Komplexometrische Titration, Ferdinand Enke, Stuttgart, 1956, Ch. E.
6 J. E. Chester, R. M. Dagnall and T. S. West, Talanta, 17 (1970) 13.
7 T. Korenaga, S. Motomizu and K. Toei, Analyst, 105 (1980) 328.
8 Y. Shijo, Nippon Kagaku Kaishi, (1974) 1912.
9 G. K. Bundnikov, N. K. Shakurova, N. A. Ulakhovich, R. A. Cherkasov, V. V. Ovchinnikov, G. A. Kutyrev and V. F. Toropova, Zh. Anal. Khim., 32 (1977) 1326.
10 K. Kalyanasundaram and J. K. Tomas, J. Am. Chem. Soc., 99 (1977) 2039.
11 R. J. Robson and E. A. Dennis, J. Phys. Chem., 81 (1977) 1075.
12 H. V. Tartar, J. Phys. Chem., 59 (1955) 1195.
13 T. Matsushita, M. Kaneda and T. Shono, Anal. Chim. Acta, 104 (1979) 145.
14 P. Mukerjee and J. R. Cardinal, J. Phys. Chem., 82 (1978) 1620.
15 K. Hayashi, Y. Sasaki, S. Tagashira and K. Ito, Bunseki Kagaku, 29 (1981) T61.

Analytica Chimica Acta, 157 (1984) 349—354
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

THE DETERMINATION OF A CARBAMATE INSECTICIDE IN SOIL SAMPLES BY DIFFERENTIAL PULSE POLAROGRAPHY

MICHAEL L. HITCHMAN* and SUBRAMANIAM RAMANATHAN

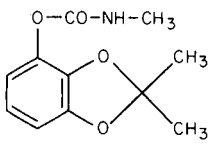
Department of Chemistry and Applied Chemistry and Thin Film and Surface Research Centre, University of Salford, Salford M5 4WT (Great Britain)

(Received 20th September 1983)

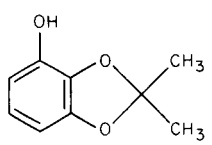
SUMMARY

Differential pulse polarography is shown to be a simple and potentially useful method for monitoring the degradation of a carbamate insecticide (Bendiocarb; 2,3-isopropylidenedioxyphenyl-*N*-methylcarbamate) in soil samples. Only extraction from the soil sample with dichloromethane and evaporation of the extract is needed prior to polarography. Calibration plots were linear over the range 10—50 mg l^{-1} at —0.94 V vs. Ag/AgCl in an acetate buffer of pH 5.0. There is no apparent interference from hydrolysis products or soil components.

Bendiocarb (I; 2,3-isopropylidenedioxyphenyl-*N*-methylcarbamate) is an established carbamate insecticide. Its physical and chemical characteristics



(I)



(II)

have been described [1] and its insecticidal activity has been summarised [2]. It is available for agricultural use under the tradenames Ficam, Garvox and Multamat. Bendiocarb is toxic to birds, fish and bees, although studies on mammals and fish indicate that there is no bio-accumulation. It is degraded in soil under aerobic conditions to carbon dioxide and a soil-bound residue which is not available to following crops. Clearly, in order to investigate the toxicology, biodegradability and environmental chemistry in detail, it would be desirable to have a simple, rapid and reliable method of determining the substance. Several analytical techniques have been suggested for Bendiocarb formulations [1]. For example, the substance shows an absorption in the ultraviolet region of the spectrum, but many other commercial formulations have similar spectra and interpretation of the results becomes too complex for routine analysis. The pyrolysis of Bendiocarb to a phenol (II) followed by gas chromatography (g.c.) with a flame ionisation detector

0003-2670/84/$03.00

has been suggested; the bendiocarb content of the sample is calculated from the peak area and a correction is made for the phenol (II) which is initially present in the sample [1]. High-performance liquid chromatography can be used, but no details have been published [3]. However, for residue analysis, g.c. is still used, with derivatization prior to the chromatographic separation to provide adequate sensitivity [3].

An alternative method based on differential pulse polarography (d.p.p.) is described in this paper. The method does not require a chromatographic separation or derivatization, nor is there any sign of interference in preparing the calibration plots from any of the phenol (II) which may be initially present in the sample. All that is required is a simple extraction from the soil sample, evaporation and straight polarography of the extract.

EXPERIMENTAL

Equipment and solutions

An E.G. & G.PAR Model 364 polarographic analyser was used with a Model 303 mercury drop electrode, including a Ag/AgCl reference electrode and a platinum auxiliary electrode. A glass cell of 10-ml volume was used and polarograms were recorded on a strip-chart recorder. All measurements were done at 22°C.

Bendiocarb samples (FBC, Ltd.) were in the form of a concentrated suspension (Bendiocarb 50-SC; 500 g l^{-1}). Stock solutions from the Bendiocarb 50-SC were prepared in deionised water and were not stored for more than two days. Appropriate aliquots were diluted with aqueous buffer solutions as required. The following solutions were tried as supporting electrolytes for polarography: ammonia/ammonium chloride buffer (pH 9.4), phosphate buffer (pH 6.8), Britton-Robinson buffers (pH 3—12), Tris buffer (pH 7—9), borax buffer (pH 8—10), and acetic acid/potassium chloride buffer (pH 3.5—8). In the case of acetic acid, the effect of the inert electrolyte was also investigated; sodium chloride, lithium chloride and potassium nitrate were tested. All buffers and other chemicals used were of analytical-reagent grade.

Samples

The soil used for studies of Bendiocarb treatment was John Innes Compost No. 1. This was used without any further grinding or sieving, in order to try and simulate practical soil conditions. Several samples of the soil (each 50 g) were treated with 1% (v/v) and 2% (v/v) bendiocarb stock solutions (24 ml) by shaking together in a closed bottle for a period of about 20 min. Control soil samples were prepared at the same time by treatment in the same way with deionised water. The quantity of Bendiocarb in the soil was monitored over a period of about a week by extraction and differential pulse polarography as described below. Whenever a Bendiocarb-treated sample was subjected to analysis, a parallel run was done on a control sample. Because degradation of Bendiocarb leads to, among other things, the phenol

(II), and because the rate of degradation depends on pH [1], then in order to compare experimental quantities of Bendiocarb in soil with those expected from known degradation rates, it was necessary also to monitor the soil pH as a function of time. For this purpose, a third set of soil samples was prepared by identical treatment with Bendiocarb and the pH of this set of samples was measured at the same time as each determination of Bendiocarb was done.

Procedures

Solutions for polarography were always first deaerated for at least 4 min with nitrogen which had been scrubbed through vanadium(II) chloride solution. For initial investigations, the potential was scanned cathodically from about −0.1 V (vs. Ag/AgCl) to locate the peaks of interest; routine scans were restricted to around this peak potential. For all the Bendiocarb determinations, an initial potential of −0.6 V and a scan rate of −5 mV s^{-1} were found to be adequate.

Bendiocarb was extracted from soil samples by refluxing the sample with dichloromethane (100 ml) for 30 min. The extract was filtered and evaporated to near dryness on a water bath. The residue was dissolved in a measured volume (usually 1—2 l) of aqueous electrolyte solution and then submitted to polarography as described above. For some samples, the residue was redissolved in dichloromethane and applied to a chromatographic column containing silica gel made up in dichloromethane and topped with anhydrous sodium sulphate. After initial elution with dichloromethane (100 ml), the fraction eluted with 60 ml of ethyl acetate/dichloromethane mixture (1 + 9) at a flow rate of 0.4 ml min^{-1} was collected, evaporated to near dryness and dissolved in aqueous solution (usually 50 ml) as before for polarography.

The pH of the soil samples was measured potentiometrically after equilibration for 30 min of a fixed amount of soil (10 g) with 20 ml of deionised water followed by filtration.

RESULTS AND DISCUSSION

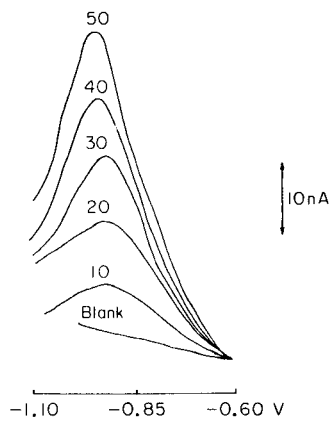
Of the various solutions tried as supporting electrolytes for the polarography, those containing acetic acid and potassium chloride yielded the most quantitatively useful current peaks for Bendiocarb. The dependence of peak current (i_p) and peak potential (E_p) on pH (over the range pH 3.5—8) and ionic strength (over the range 0.1—3 M) was systematically studied with the acetic acid/potassium chloride system. It was found that while, in general, i_p was independent of pH over the range, it increased about three-fold at constant pH 5 in going from ionic strength 0.1 M to 3 M. The E_p value also varied with ionic strength, becoming more negative with increasing ionic strength; at 3 M, $E_p \simeq -1.04$ V which approached the solvent decomposition potential. Thus, although high ionic strength would significantly increase the current sensitivity, the increased background current made it advisable to

determine bendiocarb at ionic strength 0.1 M where E_p = —0.94 ± 0.02 V. The use of NaCl, LiCl, or KNO_3 instead of KCl as the inert electrolyte had no significant effect on either i_p or E_p and so these salts could also be used in determinations. The studies reported here, however, were restricted to potassium chloride.

Figure 1 shows typical d.p. polarograms obtained with diluted stock solutions of bendiocarb. Least-squares treatment of the data gave the calibration equation

$$i_p = (7.96 \pm 0.11)\,10^{-1}\,[BC] - (1.79 \pm 0.38)\,nA \quad (1)$$

where [BC] is the bendiocarb concentration (mg l^{-1}) and the errors given are the standard errors. The correlation coefficient is 0.999 and the standard error of values of i_p about the regression line is 0.51 nA. Results falling within the limits of error given were obtained with other sets of diluted stock solutions, provided that the solutions were used within two days of being prepared; after that, peak heights decreased with time. All the results indicated that differential pulse polarography provides a reliable determination of Bendiocarb in the range 10—50 mg l^{-1}. For concentrations less than 10 mg l^{-1}, some improvement is clearly needed. When the ionic strength was increased to 2 M, it was found that the values of i_p were more than twice those obtained at ionic strength 0.1 M, yet E_p shifted only to —0.99 V from —0.94 V. Another possibility would be the formation of a derivative of Bendiocarb; a preliminary investigation was made of nitrosation of Bendiocarb using a procedure adapted from Gajan et al. [4]. The peak current for the reduction of the derivative from a Bendiocarb stock solution (10 mg l^{-1})



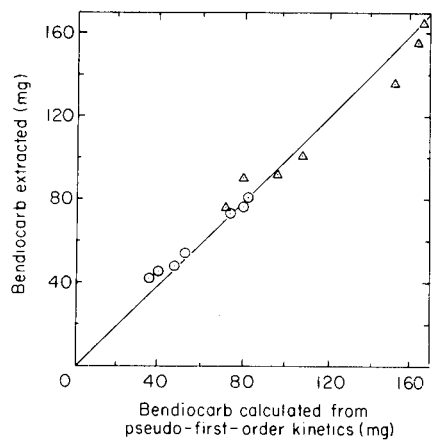


Fig. 1. Differential pulse peaks for diluted stock solutions covering the range 10—50 mg l^{-1} Bendiocarb: KCl/acetic acid buffer, pH 5, ionic strength 0.1 M.

Fig. 2. Comparison of experimental values of Bendiocarb found in soil samples with values calculated on the basis of pseudo-first-order kinetics. (○) Soil treated with 1% solution; (△) soil treated with 2% solution.

was about 30 times greater than i_p for the same solution without derivatisation. Also a solution of 1 mg l^{-1} Bendiocarb when derivatised gave a peak current which was clearly distinguishable from the background current.

Table 1 gives results found for the extraction of Bendiocarb from soil samples treated with 1% and 2% stock solutions. There was a steady decrease in the amount of Bendiocarb found, which is consistent with the decomposition with time of the insecticide, as discussed below. However, after one hour of treatment, decomposition should be negligible; comparing the Bendiocarb found after 1 h with the amount used to treat the soil sample initially (120 mg and 240 mg) shows that the extraction efficiencies were 67.3% and 69.2%, respectively. If it is assumed that the extraction efficiencies remain constant at these levels for subsequent extractions, then the amounts of Bendiocarb found can be compared with those expected on the basis of the reported kinetics for hydrolysis. Chromatographic separation of the extract did not result in any improvement of the analysis and, in fact, reduced the apparent extraction efficiency by another 10%. It is clearly preferable to apply polarography directly to the extract after dissolution in the buffer. Whiteoak et al. [1] gave half-life values ($t_{1/2}$) for the hydrolysis of Bendiocarb at 25°C; a ln $t_{1/2}$ vs. pH plot of their data provided a least-squares line for the pH range 5—9, which was adequate to allow interpolation at any intermediate pH value. As the Bendiocarb decomposition appears to be a base-catalysed hydrolysis, it seems reasonable to assume that for low concentrations of the insecticide in the soil, such as were used here, the rate of reaction will follow pseudo-first-order kinetics. Thus the concentration (C_t) of Bendiocarb at any time (t) will be related to the initial concentration (C_0) by

$$C_t = C_0 \exp[-(t/t_{1/2}) \ln 2] \quad (2)$$

where ln $2/t_{1/2}$ is equivalent to the pseudo-first-order rate constant for the decomposition. By measuring the pH of the soil samples, obtaining the corresponding $t_{1/2}$ values by interpolation from the ln $t_{1/2}$ vs. pH plot, and allowing for the extraction efficiencies, a comparison can be made between the amounts of Bendiocarb actually found at various times with those ex-

TABLE 1

Results for soil treated with 1 and 2% (v/v) Bendiocarb solutions

Treatment time		Treated with 1% Bendiocarb		Treated with 2% Bendiocarb	
days	h	pH of soil	Bendiocarb extracted (mg)	pH of soil	Bendiocarb extracted (mg)
	1	4.5	80.8	5.2	166.0
1		5.5	76.6	5.3	156.8
3		5.6	74.3	5.6	136.7
4		6.3	55.1	6.3	101.7
5		6.3	48.9	6.3	92.9
6	18	6.3	45.8	6.3	91.0
7	18	6.3	42.7	6.3	77.4

pected on the basis of first-order kinetics. Figure 2 makes this comparison; the correlation coefficient for the data is 0.991 and the standard error of the experimental points about the line of unit slope is 7.2 mg. In view of the limited data available on the half-lives for the hydrolysis [1] and of the fact that the pH varied between analyses, particularly soon after the initial treatment of the soil samples, the correlation is remarkably good.

From the results presented above, d.p.p. would appear to be a relatively simple but useful method for monitoring Bendiocarb in soil samples. The technique lends itself to routine work without the need for prior chromatographic purification. There appears to be no interference from any hydrolysis product of the Bendiocarb or from any other organic species which may be removed from the soil during the extraction. The method should also be valuable in following the kinetics of decomposition of Bendiocarb in soil samples and in studying the effect of environmental parameters, such as pH, on reaction rate. It would be useful for correlating Bendiocarb levels with the control of soil pests. In this context, it is worth commenting on the rates of application of Bendiocarb to soil samples. The treatment of 50-g soil samples with 24 ml of 1% and 2% (v/v) stock solutions of Bendiocarb 50-SC corresponds to application rates of the active ingredient of 2.4 g kg^{-1} and 4.8 g kg^{-1}, respectively. Typically recommended application rates in spray treatment are 500—1000 g $hectare^{-1}$, but the depth of penetration is obviously not predictable, nor is the weight of active ingredient per unit weight of soil. The manufacturer of Bendiocarb recommends treatment of seedling boxes before transplantation with 2.5—5.0 g of active ingredient per box (30 × 60 cm). For a soil depth of about 3 cm and a soil density of about 1.5 g cm^{-3}, this corresponds to application rates of 0.3—0.6 g kg^{-1}. The applications here were about ten times higher than this, but examination of the amounts extracted (Table 1) shows that Bendiocarb could be determined at the level of 0.3—0.6 g kg^{-1} active ingredient simply by dissolving the residue after extraction in, say, 100—200 ml of electrolyte to bring the solutions into the range of the calibration plot. Other possible applications would be for depth profiling of the insecticide in soil and for residues on foodstuffs and crops. The major limitation of the method would appear to be the lack of sensitivity at low Bendiocarb levels, which needs further investigation.

We thank FBC Ltd. for providing the Bendiocarb samples, Dr C. Tait-Bowman and Mr G. L. Sykes for helpful discussions and practical advice on the use of Bendiocarb for treatment of soil samples, and Dr R. J. Whiteoak for constructive comments and criticisms.

REFERENCES

1 R. J. Whiteoak, J. B. Reary and K. C. Overton, Anal. Methods Pest. Plant Growth Regulators, 10 (1978) 3.
2 R. W. Lemon, Proc. Br. Crop Protection Conf., 2 (1971) 570.
3 R. J. Whiteoak, personal communication.
4 R. J. Gajan, W. R. Benson and J. M. Finocchiaro, J. Assoc. Off. Agr. Chem., 48 (1965) 958.

Analytica Chimica Acta, 157 (1984) 355—357
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

Short Communication

A SIMPLE METHOD OF JUDGING THE ACCEPTABILITY OF ANALYTICAL METHODS

J. JOHN*, J. SEDLÁČEK and F. ŠEBESTA

Department of Nuclear Chemistry, Technical University of Prague, 115 19 Prague 1 (Czechoslovakia)

(Received 20th January 1983)

Summary. The method is based on calculation of the relative differences between values obtained for calibration standards by a sequential leave-one-out technique and accurate values obtained by a reference method. The method is suitable, namely, for comparing analytical procedures when a limited number of standards of accurately known composition is available.

Judging the acceptability of analytical methods is of great importance, particularly when it is necessary to choose the best way of evaluating results or comparing procedures. The usual recommended way is either calculation of the "total error" of the method [1] or the "divergence criterion" used in information theory [2]. The application of either of these procedures for judging the acceptability of a method is not common, mainly because of the sophistication needed to use the total error as the divergence criterion [3], and because fairly large sets of experimental data are needed for their application.

Because of the drawbacks of such accurate methods, an attempt has been made to find a method which could, at least in some cases, simplify the problem of comparing the acceptability of various analytical procedures. It should be applicable to the procedures most common in analytical laboratories, i.e., the construction of calibration graphs from up to ten points and the evaluation of results by means of programmable pocket calculators. The method chosen is a simplification of McFarren's method of total error [1] and the use of a method of "sequentially unknown samples" is proposed. The limiting factors for the use of this method are that the precision demanded of the standards used is at least one order of magnitude better than the expected precision of the tested analytical method, and that all the samples must fall within a concentration range in which they can be measured with comparable precision.

McFarren et al. [1] originally defined total error (T) as

$$T = (d + 2s)/\mu \qquad (1)$$

where d is the absolute value of the mean error, s is the standard deviation

0003-2670/84/$03.00

and μ is the true value for the sample. As the determination of the exact values of d and s is quite laborious, they are replaced by quantities which can be calculated from one calibration graph obtained from single measurements on standards. For this calculation, the method of sequentially unknown samples is proposed. The principle of this method is as follows. After the optimum type of calibration graph has been chosen, one of the standards is considered to be unknown. The experimental data measured for all the other calibration standards are used to construct the calibration graph and the concentration of the "unknown" is calculated from it. This calculation is sequentially repeated for all the calibration standards, which yields a number of results for "unknown" samples equal to the number of calibration standards. All the following calculations are based on the use of the relative errors (Δ) of the results obtained for "unknown" samples, calculated as $\Delta = 100(\xi - \mu)/\mu$, where ξ is the value calculated for the standard serving as the "unknown".

It was experimentally proved (Table 1) that for 6—10 calibration standards, the standard deviation (s) calculated from the same number of replicates is practically equal to the arithmetic mean of the absolute values of the relative errors ($\overline{|\Delta|}$). Accordingly, $\overline{|\Delta|}$ is referred to subsequently as precision and can be used as an estimate of the mean value of the standard deviation in the given concentration range. Under these conditions the precision can also replace $100\ s/\mu$ in Eqn. 1. In Eqn. 1, d/μ represents the relative bias of the method (i.e., the mean relative error of the results for the analyzed samples). Under the present conditions, $100\ d/\mu$ can be replaced by the absolute value of the arithmetic mean of the relative errors ($|\overline{\Delta}|$). From the above considerations, Eqn. 1 can be rewritten simply as $T = |\overline{\Delta}| + 2\overline{|\Delta|}$. This form is used below.

TABLE 1

Comparison of results obtained by the proposed method and statistical evaluation of multiple analyses of one sample

Method[a]	s (%)	$\overline{\lvert\Delta\rvert}$ (%)	$\overline{\Delta}$ (%)	T (%)
1	1.1	1.0	—0.1	2.1
2a	3.9	3.5	—1.9	8.9
2b	4.2	4.0	0.1	8.1
3a	2.2	2.3	—0.8	5.4
3b	2.0	2.4	—0.2	5.0
4	0.7	0.8	0.3	1.9

[a]Method 1: neutron activation analysis (n.a.a.) with γ-spectrometry and multiple peak evaluation (24-h cooling) and calculation of weighted mean without exclusion of outliers. Method 2: n.a.a. with counting of delayed neutrons and chemical determination of total uranium concentration for (a) 500 μg U per sample; (b) 100 μg U per sample. Method 3: n.a.a. with combined counting of delayed neutrons and γ-spectrometric determination of ^{239}U; (a, b) as in Method 2. Method 4: spectrometry of 185.7-keV γ-rays from ^{235}U with NaI (Tl) detector, 20 mg U per sample determined by a chemical method, 6—12 h counting.

This simplified method of judging the acceptability of analytical methods was tested on some radiochemical methods of isotopic analysis of uranium in small aqueous samples [4, 5]. The results obtained (Table 1) show good agreement between the precision calculated as $\overline{|\Delta|}$ and the standard deviation (s) calculated from multiple measurements on one sample.

The last problem was how to test simply the statistical significance of the value of $\overline{\Delta}$ as possible bias. The present small sets of data are most exactly tested by a non-parametric method, which needs further calculations. As $\overline{\Delta}$ is expected to be zero in a procedure based on a calibration graph, the results of the non-parametric test were compared with the results obtained by means of a t-test [6] in simplified form $t = |\overline{\Delta}|\sqrt{n}/\overline{|\Delta|}$, where n is the number of calibration standards used. The results obtained by both procedures for the six analytical procedures listed in Table 1 were the same. Thus, the use of Student's t-test in the above simplified form is recommended for establishing bias.

The authors thank Dr. B. G. Cooksey from the University of Strathclyde, Glasgow, for many fruitful comments.

REFERENCES

1 E. A. McFarren, R. J. Lishka and J. H. Parker, Anal. Chem., 42 (1970) 358.
2 K. Eckschlager and V. Štěpánek, Information Theory as Applied to Chemical Analysis, Wiley-Interscience, New York, 1979.
3 K. Eckschlager, I. Horsák and Z. Kodejš, Evaluation of Analytical Results and Methods, SNTL, Prague, 1980 (in Czech).
4 J. John, Ph.D. Thesis, Technical University of Prague, 1983.
5 J. John, J. Sedláček and J. P. Rambak, Radiochem. Radioanal. Lett., 53 (1982) 263.
6 K. Eckschlager, Errors, Measurements and Results in Chemical Analysis, Van Nostrand-Reinhold, London, 1969.

Analytica Chimica Acta, 157 (1984) 359—361
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

Short Communication

BASELINE COMPENSATION FOR GRADIENT ELUTION HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY WITH A DYNAMICALLY REFERENCED FLOW CELL

JONATHAN B. CROWTHER, DEREK D. DEZARO and RICHARD A. HARTWICK*

Department of Chemistry, Rutgers, The State University of New Jersey, New Brunswick, NJ 08903 (U.S.A.)

(Received 1st September 1983)

Summary. Stable baselines were obtained with a gradient test system of acetone in methanol by using a secondary pump circuit at atmospheric pressure to simulate the gradient formed by the primary (high pressure) circuit. The secondary gradient system is used dynamically to reference the flow cell, thus eliminating baseline drift when detector-active mobile phases are used in gradient elution.

Gradient elution is a powerful, versatile means of optimizing separations of complex mixtures [1, 2]. In high-performance liquid chromatography (h.p.l.c.), detector flow cells have customarily been of the static reference design, resulting in severe baseline drift when detector-active solvents are used in a gradient system [3]. For example, gradient elution used with refractive index detectors exhibits severe baseline drift except under the trivial conditions when the refractive indices of the low- and high-strength solvents are identical. With u.v.-visible detectors, the use of short wavelengths, ion-pairing agents, or weakly u.v.-absorbing solvents or buffer salts can all contribute to baseline drift.

A principle is outlined herein by which mildly detector-active mobile phases can be used during gradient elution, within the limits of detector linear range. A dynamic reference cell is created by generating a secondary gradient exactly simulating the primary high-pressure gradient. Volumes in the secondary system are matched to the flow rate of the primary column so that the time of arrival of the secondary gradient to the reference cell is exactly coincident with the primary gradient. The reference gradient is produced by using a second pumping system coupled either mechanically or electronically to the primary system. This principle is demonstrated in this communication by using two high-pressure pumps to generate gradients from a single low-pressure source.

Experimental

The apparatus consisted of two Waters M6000 pumps (Waters Associates, Milford, MA), both connected via short lengths of plastic tubing to a single

0003-2670/84/$03.00

stirred solvent reservoir. A third Waters M45 pump was used to form a gradient by pumping the high-concentration solvent into the stirred reservoir. Gradient shapes and slopes were varied by combinations of pump flow rates and high-strength solvent composition. Adjustments of arrival times of primary and secondary gradients at the flow cell was achieved by empirical calibration. Once calibrated, the system remained stable unless a new column was used.

Low-strength mobile phase was an aqueous solution that contained 10% (v/v) methanol and 0.010 mol dm^{-3} potassium hydrogenphosphate. A u.v.-active high-strength buffer was simulated by using the same mobile phase, but with 1 cm^3 of acetone per 100 cm^3 of solvent (0.140 mol dm^{-3}). The column was a Partisil 10-ODS-3 (Whatman, Clifton, NJ). The column was 4.6 mm i.d. × 250 mm. The flow rate for all separations reported was 1.5 cm^3 min^{-1}.

Results and discussion

Figure 1A shows the separation of 10 nmol each of the ribonucleosides, cytidine, uridine, guanosine, and adenosine under simulated gradient drift conditions, with acetone as a u.v. absorber in the high-strength solvent. The baseline drift starts at about 8 min, which was the delay time for the gradient to reach the head of the column. The severe baseline shift almost completely obscures the adenosine peak. Accurate quantitation and identification under these conditions would be difficult if not impossible. Figure 1B shows the same separation under the conditions of a dynamic reference cell. A nearly perfect subtraction of baseline drift is achieved, with no significant increase in noise. Integration of the small peak for adenosine was readily achieved. Continuous running of the instrument under isocratic conditions with the dynamic referencing solvent showed very stable and noise-free baselines.

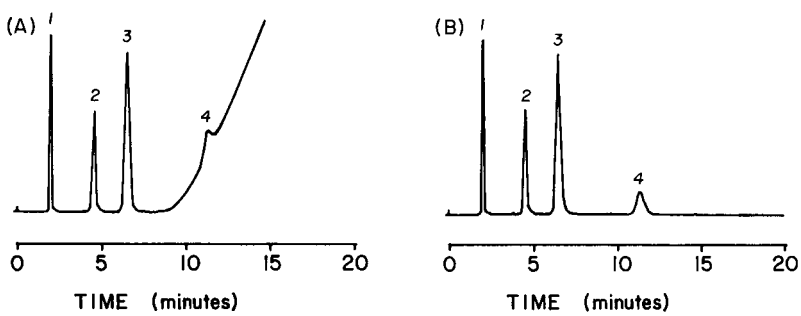


Fig. 1. Chromatograms without and with a reference stream: (A) the gradient separation of the four ribonucleosides cytidine, uridine, guanosine and adenosine, using 0.140 mol dm^{-3} acetone to simulate a u.v.-absorbing buffer, and a static (air) reference cell; (B) the same separation with a secondary gradient through the reference cell. Detection at 254 nm, 0.04 *A* full scale; integrator attenuation 7; flow rate 1.5 cm^3 min^{-1}.

In summary, these results demonstrate the practicality of compensating for baseline drift using a dynamic reference cell sensing a secondary flow stream produced by a duplicate pumping system. While demonstrated for u.v.-visible detectors, the concept can be readily applied to other detection methods. The primary limitation of this method is that the active solvent can only be varied within the linear dynamic range of the detector. Also, a mixing chamber of good design is needed on the secondary circuit to ensure low noise levels.

The authors acknowledge the National Science Foundation, Grant CHE 810024, for support of research from which this application was derived.

REFERENCES

1 C. Liteanu and S. Gocan, Gradient Elution Chromatography, Halsted Press (Wiley), New York, 1974.
2 P. Jandera, M. Janderova and J. Churacek, J. Chromatogr., 115 (1975) 9.
3 D. S. J. Atkin, R. J. Hamilton, S. F. Mitchell and P. A. Sewell, Chromatographia, 15 (1982) 97.

Analytica Chimica Acta, 157 (1984) 363—367
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

Short Communication

LIQUID-LIQUID EXTRACTION WITH 2-ACETYLPYRIDINEBENZOYLHYDRAZONE IN THE DETERMINATION OF TRACES OF COPPER, NICKEL, COBALT AND ZINC BY ATOMIC ABSORPTION SPECTROMETRY

M. GARCIA-VARGAS, M. P. HERNANDEZ-ARTIGA, and J. A. PEREZ-BUSTAMANTE*

Department of Analytical Chemistry, Faculty of Sciences, University of Cádiz, Puerto Real, Cádiz (Spain)

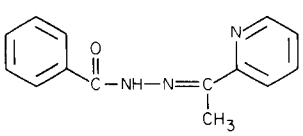
(Received 6th June 1983)

Summary. Trace amounts of Cu, Ni, Co and Zn in aqueous solutions are determined by their simultaneous extraction with 2-acetylpyridinebenzoylhydrazone followed by flame atomic absorption spectrometry. Isoamyl alcohol is the preferred solvent. Calibration graphs are linear, usually over the range 0—0.2 mg l^{-1} of aqueous phase. There are few interferences from other ions present in 500—1000-fold amounts.

Many atomic absorption spectrometric (a.a.s.) procedures, based on the separation and preconcentration of metal ions by liquid-liquid extraction have been reported [1]. Among the commonly used chelating reagents are sodium diethyldithiocarbamate [2], ammonium pyrrolidinedithiocarbamate [3], 8-quinolinol [4] and dithizone [5]. 4-Methylpentan-2-one is usually preferred as the solvent for a.a.s. but other solvents such as chloroform, ethyl acetate, butyl acetate and isoamyl alcohol have also been used [1].

Aroylhydrazones behave as polydentate ligands forming uncharged chelates with divalent transition metal ions [6], which can be extracted into organic solvents. These compounds have been widely employed in spectrophotometric and fluorimetric methods for the determination of metals [7] but have rarely been applied in a.a.s. methods [8].

This communication describes a sensitive method for the determination of trace amounts of copper, nickel, cobalt and zinc by flame a.a.s. after extraction with 2-acetylpyridinebenzoylhydrazone (I; APBH) and isoamyl alcohol. The metal—APBH complexes are reasonably stable in isoamyl alcohol. Optimal conditions for the extraction are described.



(I)

0003-2670/84/$03.00

Experimental

Instrumentation. The Pye-Unicam SP-9 atomic absorption spectrometer used was equipped with a standard air-acetylene burner head. Single-element (copper, zinc, cobalt and nickel) hollow-cathode lamps were used with deuterium background correction. The instrumental settings are summarized in Table 1. Other instrumentation included a Perkin-Elmer 575 spectrophotometer with 1.0-cm quartz or glass cells, and a Metrohm E-516 pH meter with a combined glass—calomel electrode.

Reagents. Stock standard solutions (1.000 mg ml^{-1}) were prepared by dissolving 1.000 g of pure metal in the minimum volume of an appropriate mineral acid. Single- and multi-element solutions of lower concentrations were prepared by appropriate dilution as required.

Solutions (0.05% and 0.1% w/v) of APBH in ethanol, isoamyl alcohol, 4-methylpentan-2-one ketone (MIBK) and n-butyl acetate were prepared. The solutions in MIBK are stable for a day, those in isoamyl alcohol for a month and those in the other solvents for a few days. The synthesis of PABH has been described elsewhere [9].

Potassium hydrogenphthalate/sodium hydroxide (0.1 M, pH 6.0) and other buffer solutions were prepared conventionally. All reagents were of analytical grade and all solutions were prepared with distilled deionized water.

Glassware was cleaned with dithizone dissolved into chloroform and all reagent solutions were stored in polyethylene bottles, previously soaked for several days in 6% (v/v) nitric acid. Special care was taken to avoid potential sources of contamination from the materials and chemicals used.

Procedure. An aliquot of the sample solution (containing up to 100 μg of each metal ion) was mixed with 5 ml of phthalate buffer solution (pH 6.0), diluted to 100 ml with water in a 250-ml separatory funnel and extracted with 10 ml of a 0.05% (w/v) solution of APBH in isoamyl alcohol for 2 min. After 5 min or more, to allow phase separation, the aqueous layer was discarded and the organic phase was dried with anhydrous sodium sulfate and transferred into stoppered glass tubes for a.a.s. (see Table 1).

TABLE 1

Conditions for atomic absorption spectrometry[a]

	Cu	Ni	Co	Zn
Analytical line (nm)	324.8	232.0	240.7	213.8
Spectral bandwidth (nm)	0.5	0.2	0.2	0.5
Lamp current (mA)	4	12	12	8
Integration time (s)	2	2	2	3
Zero set	IA	IA	IA	BR[b]

[a]Air and acetylene flows were 5.2 and 1.2 l min^{-1}, respectively. Background correction was used only for zinc. [b]Blank reagent.

Calibration graphs were prepared by using multielement standard solutions of appropriate volumes, covering the ranges 0.1—1 (Co), 0.1—2 (Ni and Zn) and 0.1—3 (Cu) $\mu g\ ml^{-1}$ in the organic phase, treated in exactly the same way as the samples.

Results and discussion

Spectrophotometric study of the metal—APBH chelates. When diluted solutions of metal ions in acidic or alkaline medium and a 0.1% (w/v) of APBH in ethanol are mixed, a yellow chelate is formed immediately. Spectrophotometric characteristics of the cobalt, nickel, copper and zinc complexes are detailed in Table 2. At the optimum pH values, many other metal—APBH complexes are also formed in aqueous ethanol. However, because a high proportion of ethanol is required to dissolve the metal—APBH complexes, liquid-liquid extraction is advisable. The extraction was used satisfactorily with flame a.a.s., providing increased sensitivity and selectivity.

Extraction/a.a.s. study. The effect of pH on the extraction of the metal chelates into different solvents was studied over the pH range 2—11 (Table 3). Isoamyl alcohol proved to be the most satisfactory, because the APBH reagent solution was more stable and the a.a.s. signals were more reproducible. To control the pH, 2.5 ml of the phthalate buffer was added when the phase/volume ratio was less than 5, and 5 ml when it exceeded 5.

TABLE 2

Spectrophotometric characteristics of the metal—APBH chelates in aqueous ethanol

Metal ion	λ_{max} (nm)	ϵ ($l\ mol^{-1}\ cm^{-1}$)	Optimum pH	Metal—APBH ratio
Cu(II)	365—375	3.1×10^4	2.5—3.5 (6.8—7.5)	n.d.[a]
Zn(II)	375	3.4×10^3	6.8—9.0	1:3
Ni(II)	376	2.5×10^4	5.3—8.3	1:3
Co(II)	362—370	1.5×10^4	6.3—7.0 (8.5—9.0)	1:3

[a]Not determined.

TABLE 3

Optimum pH ranges for the solvent extraction of the metal—APBH chelates

Metal ion	n-Butyl acetate[a]	Isoamyl alcohol	MIBK
Cu(II)	—	5—8	5.5—8
Zn(II)	5—8	4—7.5	5—11
Ni(II)	—	5—8	4.7—7.5
Co(II)	—	5—9	6—10

[a]Atomic signals of Cu, Ni and Co were not reproducible.

In other tests, the shaking time was varied from 0.5 to 5 min; 2 min was sufficient to provide the maximum a.a.s. signals. A period of 5 min was usually enough for phase separation. An aliquot of 100 ml of solution containing 5 μg of each metal ion was extracted with 10 ml of 0.01—0.1% (w/v) APBH solution in isoamyl alcohol. Changes in the reagent concentration over this range did not affect the a.a.s. signals; 10 ml of 0.05% (w/v) reagent solution was chosen, in order to overcome interferences.

A 10:1 V_w/V_o ratio was selected to provide reasonable sensitivity and precision. If this ratio were higher, insufficient organic phase would be obtained after extraction because of the solubility of isoamyl alcohol in water.

The extraction efficiency was calculated from four 100-ml aliquots, all of which contained 10 μg of each metal. These solutions were extracted with 10 ml of APBH in isoamyl alcohol; the dried organic layers were then carefully evaporated to dryness, and the residues were dissolved in 5 ml of hot 4 M nitric acid. The recoveries of copper, nickel and cobalt, calculated from their a.a.s. signals in these solutions, were higher than 92% in all cases. The zinc recovery was calculated from a spectrophotometric method with dithizone [10] and was found to exceed 90%.

The atomic absorption signals of the metal complexes remained constant for at least two days when the samples were stored between measurements at 4°C.

Atomic absorption determination of copper, nickel, cobalt and zinc. Under the recommended conditions, linear calibration graphs were obtained for the ranges 0—10 μg (Co), 0—20 μg (Ni and Zn) and 0—30 μg (Cu). The sensitivities for 1% absorption were 0.055, 0.063, 0.040 and 0.049 μg ml^{-1} related to the organic layer, for Co, Ni, Cu and Zn, respectively.

The precision of the method was checked for 5 aliquots containing 10 μg and 6 aliquots containing 5 μg of each metal ion. The relative standard deviations for 10-μg contents were ±0.74% (Co), ±0.62% (Ni), ±0.60% (Cu) and ±1.00% (Zn), and for 5-μg contents ±1.44% (Co), ±2.82% (Ni), ±1.0% (Cu) and 3.90% (Zn).

Influence of other ions. Copper, nickel, cobalt and zinc (5 μg of each metal ion per 100 ml of aqueous solution) were determined in the presence of 1000-fold amounts of 42 ionic species. There were no interferences from Na, K, Ca, Mg, Pb, Cd, Mn(II), Mo(VI), W(VI), V(V), As(V), NO_3^-, NO_2^-, Cl^-, F^-, Br^-, I^-, SCN^-, SO_4^{2-}, CO_3^{2-}, PO_4^{3-}, oxalate, tartrate and citrate. Thallium(I), Be, Al, Cr(III), La and borate did not interfere in the determinations of nickel, cobalt and copper. In the determination of zinc, 1000-fold amounts of La(III) (in the presence of tartrate) and Al(III) and Tl(I) (both in the presence of fluoride) were tolerated; Cr(III) and Be(II) in the presence of fluoride were tolerated in 500- and 100-fold amounts, respectively. Thorium(IV) and Zr(IV) in 1000-fold amounts interfered in the determinations of cobalt and zinc but could be masked with tartrate; they did not interfere in the determinations of nickel and copper.

When iron(II, III), U(VI), Sb(III) and Ti(IV), which react with the reagent,

were present in 1000-fold amounts, a higher concentration of reagent (0.1—0.2% w/v) was required in order to achieve the total extraction of copper, nickel, cobalt and zinc. Under analogous conditions bismuth may be present in 250-fold quantities.

The main advantage of the proposed method is its sensitivity and simplicity. In contrast to most other methods, one extraction is sufficient for the determination of Cu, Ni, Co and Zn in a sample. Furthermore, samples can be extracted immediately after they have been taken and the extracts stored for a couple of days, thereby alleviating transport problems.

REFERENCES

1 M. S. Cresser, Solvent Extraction in Flame Spectroscopic Analysis, Butterworths, London, 1978.
2 J. A. Dean and C. Cain, Anal. Chem., 29 (1957) 530.
3 K. M. Bone and W. H. Hibbert, Anal. Chim. Acta, 107 (1979) 219.
4 H. Goto and E. Sudo, Bunseki Kagaku, 10 (1961) 175.
5 H. V. Weiss, P. R. Kenis, J. Korkisch and I. Steffan, Anal. Chim. Acta, 104 (1979) 337.
6 M. Garcia-Vargas, M. Gallego and M. de la Guardia, Analyst, 105 (1980) 965.
7 M. Katyal and Y. Dutt, Talanta, 22 (1975) 151.
8 M. Gallego, M. Garcia-Vargas and M. Valcarcel, Microchem. J., 27 (1982) 328.
9 M. Garcia-Vargas, J. M. Bautista and P. de Toro, Microchem. J., 26 (1981) 557.
10 E. B. Sandell, Colorimetric Determination of Traces of Metals, Interscience, New York, 1959, p. 949.

Analytica Chimica Acta, 157 (1984) 369—372
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

Short Communication

SPECTROPHOTOMETRIC DETERMINATION OF NICKEL WITH CYCLOHEXYLIDINE-AMMONIUM 2-AMINOCYCLOHEXYLIDENE-1-CYCLOHEXENE-1-DITHIOCARBOXYLATE

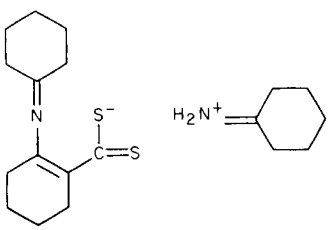
A. SAFAVI* and H. PARHAM

Chemistry Department, College of Arts and Sciences, Shiraz University, Shiraz (Iran)

(Received 20th September 1983)

Summary. An (extraction)-spectrophotometric method is reported for the determination of nickel(II) with cyclohexylidineammonium 2-aminocyclohexylidene-1-cyclohexene-1-dithiocarboxylate. The violet 1:2 chelate is soluble in aqueous ethanolic or acetonic media at pH 6—9, or can be extracted into methyl isobutyl ketone. The molar absorptivity of the complex is about 2.5×10^4 l mol^{-1} cm^{-1} at 550 nm.

Cyclohexylideneammonium 2-aminocyclohexylidene-1-cyclohexene-1-dithiocarboxylate (I) was first prepared by Takeshima et al. [1]. This ligand with the donating —CSS$^-$ group, forms an intense violet complex with nickel(II) ions. The complex is insoluble in water but is soluble in most organic solvents. The violet complex (λ_{max} = 550 nm) contrasts well with



(I)

the yellow reagent solution (λ_{max} = 410 nm), permitting the detection of very small amounts of nickel visually. The spectrophotometric procedure offers a simple and sensitive direct means of determining trace amounts of nickel.

Experimental

Reagents. Stock (1000 mg l^{-1} Ni^{2+}) solutions of nickel nitrate, $Ni(NO_3)_2 \cdot 6H_2O$ were prepared in distilled-deionized water. Working standards were prepared by appropriate dilution of the stock solution.

The reagent (I) was prepared and purified by the method of Takeshima et al. [1]. Reagent solution (1.708×10^{-3} M) was prepared by dissolving the yellow needle-like crystals in either ethanol or acetone. This solution was stable for several weeks.

0003-2670/84/$03.00

Apparatus. The spectra were recorded on a Beckman DK-2A u.v.-visible spectrophotometer with 1-cm glass cells. Absorbances at fixed wavelength were measured with a Spectronic 70 spectrophotometer.

Recommended procedures. The complex is soluble in aqueous ethanolic or acetonic solution. For this method, take 50 ml of sample or standard solution containing 1.2—180 μg of nickel(II), add 10 ml of the reagent solution in the appropriate solvent and dilute with ethanol or acetone to 100 ml. Measure the absorbance at 550 nm against a reagent blank.

For the extraction method, transfer 50 ml of solution containing 1.5—100 μg of nickel(II) to a 100-ml separating funnel, add 3 ml of the reagent solution in ethanol and extract the complex with 25 ml of methyl isobutyl ketone. Allow the phases to separate and measure the absorbance at 550 nm against a blank prepared in the same way with distilled water instead of the nickel solution.

Results and discussion

The nickel complex is insoluble in water but dissolves in organic solvents. Chloroform, carbon tetrachloride, dichloromethane, amyl alcohol and methyl isobutyl ketone (MIBK) were tested for extraction of the complex; MIBK was the most efficient. A single extraction sufficed for quantitative transfer of the nickel to the organic phase; a second extract gave no absorbance from any complex in the solvent.

The optimum pH range for complex formation was 6—9 (Fig. 1). Addition of buffers did not affect the result. The extraction of 100 μg of nickel(II) as the complex was also tested at various pH values; the optimum pH range in this case was 5.8—9, whether or not buffer was used. Accordingly, buffer solution was not used in further work, the pH was adjusted as required with ammonia solution or hydrochloric acid.

The process of complex formation seems to be very fast. No change in absorbance of the complex was observed between 1 min to several hours after complex formation. The complex was stable for days in aqueous solu-

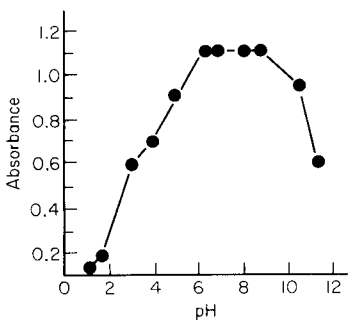


Fig. 1. Effect of pH on complex formation.

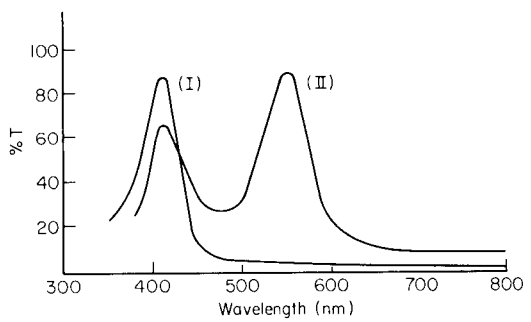


Fig. 2. Absorption spectra in aqueous ethanolic solutions: (I) the reagent; (II) the nickel complex in presence of excess of the reagent.

tions but less stable when dissolved in ethanol, acetone or MIBK.

The absorption spectra of the complex and the reagent are shown in Fig. 2. As neither Ni^{2+} nor the ligand shows absorption at 550 nm, this wavelength was chosen for absorbance measurements of the complex.

Calibration data. The calibration graphs were obtained for both methods described under Experimental. When an ethanolic aqueous solution was used, Beer's law was obeyed over the range 0.012—1.8 mg l^{-1} nickel in the final solution. When the extraction procedure was applied, Beer's law was obeyed for 0.012—2.0 mg l^{-1} nickel in the aqueous sample solution taken. The molar absorptivity was about 2.5×10^4 l mol^{-1} cm^{-1} in aqueous ethanolic solution. The Sandell sensitivity of the method is about 0.002 μg cm^{-2}, which is better than that obtained with dimethylglyoxime, even by the oxidizing method (0.004 μg cm^{-2}) [2].

Composition of the complex. The composition of the complex was established by the continuous variations method [3], the mole ratio method [4] and the slope ratio method [5]. All three methods showed that nickel ion and the ligand combine in a 1:2 ratio in a well defined complex.

Interferences. The effect of interfering ions on the determination of the nickel complex was investigated on solutions containing 100 μg of nickel in the final aqueous ethanolic solution. Aluminium(III), Bi(III), Zn(II), Cd(II), Pb(II), Mg(II), Ca(II) and Cr(III) in 500-fold amounts did not interfere with the formation of the nickel complex, but Cu(II), Co(II), Fe(III), Mn(II), Hg(II) and Ag(I) reacted with the reagent and thus interfered. The effect of 500-fold amounts of these interferences were easily removed by addition of saturated ammonium thiocyanate to the test solution prior to the addition of the reagent. Under the recommended conditions, thiocyanate did not react at all with the nickel. Thus the complexes of thiocyanate with the interfering ions were easily extracted into MIBK*. The reagent was then added to the aqueous solution and the procedure was completed as described above. No interferences resulted from the presence of 500-fold amounts of chloride, bromide, iodide, sulphate, nitrate, tartrate or citrate, but 10-fold amounts of carbonate and phosphate interfered. The latter ions were removed by addition of barium nitrate to the test solution and filtration of the precipitates prior to addition of the reagent to the test solution. EDTA and oxalate are serious interferences and must be absent.

In general, the proposed reagent provides a sensitive determination of nickel with a high degree of practicability. The formation and extraction of the complex is very fast at room temperature, and likely interferences are readily removed.

The authors are indebted to Prof. A. Massoumi for his kind advice and help.

*Ag(I) and Hg(II) were precipitated with NH_4SCN and removed by filtration.

REFERENCES

1 T. Takeshima, T. Hayashi, M. Muraoka and T. Matsuoka, J. Org. Chem., 32 (1967) 980.
2 E. B. Sandell, Colorimetric Determination of Trace Metals, Wiley-Interscience, New York, 1959.
3 P. O. Job, Ann. Chim., 9 (1928) 113.
4 J. H. Yoe and A. L. Jones, Ind. Eng. Chem. Anal. Ed., 16 (1944) 3.
5 A. E. Harvey and D. L. Manning, J. Am. Chem. Soc., 72 (1956) 4488.

ANALYTICA CHIMICA ACTA, VOL. 157 (1984)

AUTHOR INDEX

0003-2670/84/$03.00

Printed in The Netherlands

VOL. 157 No. 2 ANALYTICA CHIMICA ACTA MARCH 1, 1984

CONTENTS

(Abstracted, Indexed in: Anal. Abstr.; Biol. Abstr.; Chem. Abstr.; Curr. Contents Phys. Chem. Earth Sci.; Life Sci.; Index Med.; Mass Spectrom. Bull.; Sci. Citation Index; Excerpta Med.)