The results given in this work show, that coordination compounds with 'frozen' ligand conformation may serve as examples to illustrate the *Auwers-Skita* rule concerning the stereochemistry of heterogeneous catalytic hydrogenation [15].

We thank the *Swiss National Foundation for Scientific Research* for financial assistance as well as Dr. *K. Noack* from *Hoffmann-La Roche* for CD.-spectra.

REFERENCES

[1] *G. R. Brubaker*, Chem. Educ. *51*, 608 (1974) and references therein.
[2] *M. Saburi, T. Sawai & S. Yoshikawa*, Bull. chem. Soc. Japan *45*, 1086 (1972).
[3] *M. Saburi & S. Yoshikawa*, ibid. *45*, 1443 (1972).
[4] *H. G. Hamilton, Jr., & M. D. Alexander*, J. Amer. chem. Soc. *89*, 5065 (1967).
[5] *P. C. Harrington, S. Linke & M. D. Alexander*, Inorg. Chemistry *12*, 168 (1973).
[6] *L. J. DeHayes, M. Parris & D. H. Busch*, Chem. Commun. *1971*, 1398.
[7] *M. Wermeille, E. Sledziewska & K. Bernauer*, Helv. *57*, 180 (1974).
[8] *H. Ito & J. Fujita*, Bull. chem. Soc. Jap. *44*, 741 (1971).
[9] *J. L. Love & H. K. Powell*, Inorg. nuclear Chemistry Letters *3*, 113 (1967).
[10] *N. F. Curtis, D. A. Swann, T. N. Waters & I. E. Maxwell*, J. Amer. chem. Soc. *91*, 4588 (1969).
[11] *L. G. Warner & D. H. Busch*, J. Amer chem. Soc. *91*, 4092 (1969).
[12] *M. Goto, H. Matsushita, M. Saburi & S. Yoshikawa*, Inorg. Chemistry *12*, 1498 (1973).
[13] Inorg. Chemistry *9*, 1 (1970).
[14] *H. Stöckli-Evans*, to publish.
[15] *E. Brown & M. Ragault*, Bull. chem. Soc. Japan *47*, 1727 (1974) and references therein.

---

90. Considerations on the Interpretation of Mass Spectra via Learning Machines

by **Philippe Kent** and **Tino Gäumann**

Department of Physical-Chemistry of the Ecole Polytechnique Fédérale of Lausanne

(16. I. 75)

Summary. The application of learning machine techniques to the interpretation of mass spectra is investigated. An attempt to improve the characteristic ratio of the number of training points to the number of adjustable parameters is made by the use of reduced spectra. The ensuing results are corrected for the unequal representation of the different substances in the data bank. Reasonable performance is obtained for spectra of simple (monofunctional) substances, but predictive abilities are poor for more complicated substances.

Introduction. – The automated interpretation of mass spectra has been attempted by a variety of methods including library matching [1], artificial intelligence [2], nearest neighbour techniques [3] and learning machines [4]. The association of the first two methods has also been advanced in [5] and a review of the latter two methods, as well as some related techniques appears in [6].

Library matching requires a large bank of spectra if there is to be a reasonable probability of correct structure elucidation. Substantial memory requirements and access time are disadvantages here.

Computer times remain an important factor with artificial intelligence methods and, as with library matching, moderate size computers are necessary. Nearest neighbour techniques and analysis *via* learning machine vectors are related and escape both constraints through the reduced dimensionality of the problem and the possible use of a desk calculator.

Contradictory opinions have been voiced concerning the validity of learning machines as an interpretive tool for mass spectra [4] [7] [8]. The object of this paper is to present the results which we have obtained with this technique, emphasis being placed on the ease of subsequent interpretation.

Data Bank. – A bank of 1601 mass spectra from the *Aldermaston* tapes, with appended *Wiswesser Line Notation* formulae, was used in this study. A summary of the substances represented is shown in Table 1. The elements of structure specified in the first column are considered functions. Secondary and tertiary carbon atoms have also been listed, but are not considered functions here. The functions chosen correspond to simple character groups in the *Wiswesser Line Notation* [9].

The numbers in the second column, headed '1', indicate the number of spectra of substances containing the corresponding structural element when the molecule contains one or more of only one type of function. The line 'Total' gives the total number of spectra of such monofunctions. Similarly, the column headed '2' indicates the number of spectra with two types of function present in the molecule (difunctions),

Table 1. *Data Bank*

No	Function	1	2	3	$\geqslant 4$	0	$\leqslant 2$	All
1	Primary alcohol	45	179	54	25	–	224	303
2	Secondary alcohol	25	48	5	2	–	73	80
3	Tertiary alcohol	8	17	0	0	–	25	25
4	Ether	61	134	69	43	–	195	307
5	Ketone	34	52	40	43	–	86	169
6	Aldehyde	29	54	17	20	–	83	120
7	Acid	24	41	12	7	–	65	84
8	Ester	155	145	15	14	–	300	329
9	Dioxo	0	1	10	2	–	1	13
10	Primary amide	4	5	1	0	–	9	10
11	Secondary amide	0	1	1	0	–	1	2
12	Tertiary amide	2	1	0	0	–	3	3
13	Primary amine	25	17	4	5	–	42	51
14	Secondary amine	14	22	16	13	–	36	65
15	Tertiary amine	16	41	55	26	–	57	138
16	Phenyl	68	196	44	11	–	264	319
17	Insaturation	42	288	115	75	–	330	520
18	Cycles	23	180	130	79	–	203	412
	Tertiary carbon	58	46	1	5	3	107	113
	Secondary carbon	206	169	26	2	22	397	425
	Total	575	711	196	89	30	1286	1601

'Phenyl' corresponds to unfused benzene rings, fused rings appear under 'Cycles'.

'Insaturation' will not appear with unfused benzene rings.

Heterocycles with nitrogen and oxygen will appear as 'Secondary' or 'Tertiary amine' and 'Ether' respectively, as well as under 'Cycles'.

E.g. CH_3—CO—CH_2—CO—CH_3 is a monofunction.

and the total number of such spectra. Columns '3' and '4' correspond to spectra of substances containing three types of functions and to those containing four or more types of functions respectively. Paraffins appear under the heading '0'. The columns headed '2' and 'All' furnish the number of spectra with two types of functions or less, and the total number of spectra containing the corresponding function, respectively.

Substances containing atoms other than carbon, hydrogen, oxygen and nitrogen have been excluded.

General. – An introduction to the theory of learning machines may be found in [10]. Briefly, the method consists of representing patterns (spectra) as points in (n-dimensional) hyperspace and seeking (n-1-dimensional) hypersurfaces which will divide that hyperspace into regions containing all and only points representing a specified class of pattern. The surface is represented by a 'weight vector', $\boldsymbol{W}$, whose components or 'weights' are the coefficients of the terms of the polynomial expression of the surface. The components of the 'pattern vector', $\boldsymbol{X}$, representing a point, correspond to the terms of the polynomial. The scalar product of the weight vector and the pattern vector will give the distance from the point to the surface; the sign of such product will define a positive and a negative side of the surface.

Normally, one surface is used to distinguish between two complementary classes, the presence or absence of a specific feature for example. Although it is possible to distinguish between more than two, not necessarily complementary classes with one surface, the treatment is more complicated; it is easier to create a series of 'yes-no' surfaces to implement higher logic.

The quest of the required surfaces is the central problem. For this the data bank is separated into two parts: training sets for seeking the required surfaces, and prediction sets for evaluating the surfaces' performance as classifiers.

Training proceeds by first defining an initial, arbitrary weight vector. Each pattern point of the training set is then presented, one by one, for a response. A positive scalar product of the pattern vector and the weight vector is considered a positive response, indicating, for example, the presence of a certain feature. A negative scalar product is considered a negative response and indicates the absence of the feature in this case. If the response is correct, the next point is presented. If the response is incorrect, the weight vector is modified before the next point is presented. The process is repeated until all the points of the training set are correctly classified by the weight vector (100% recognition), or some arbitrary limit attained.

The negative feedback thus imposed on the weight vector upon erroneous response is of the form:

$$\boldsymbol{W}' = \boldsymbol{W} + K \cdot \mathrm{Sn} \cdot \boldsymbol{X}$$

where $\boldsymbol{W}'$ is the new weight vector
$\boldsymbol{W}$ the previous weight vector
Sn the sign of the correct response
$\boldsymbol{X}$ the pattern vector incorrectly classified

and K is either a constant, the smallest integer greater than $|(\boldsymbol{W} \cdot \boldsymbol{X})|/(\boldsymbol{X} \cdot \boldsymbol{X})$, or an expression of the type $K = \lambda \cdot |(\boldsymbol{W} \cdot \boldsymbol{X})|/(\boldsymbol{X} \cdot \boldsymbol{X})$, where $0 < \lambda \leqslant 2$.

The three possibilities for K correspond, respectively, to the 'fixed increment', the 'absolute' and the 'fractional correction rule'. The flow diagram in Fig. 1 summarizes the training sequence.

The value of the ratio T/N, where T is the number of points in the training set and N is the number of adjustable weights necessary for one classification, is of interest. For high values of T/N, a surface trained to 100% recognition should have a good predictive ability. For a low value of T/N, training will be faster but predictive ability questionable, and the influence of the initial weight vector more pronounced. A value of T/N greater than two is considered to be essential, values greater than five are recommended [11].

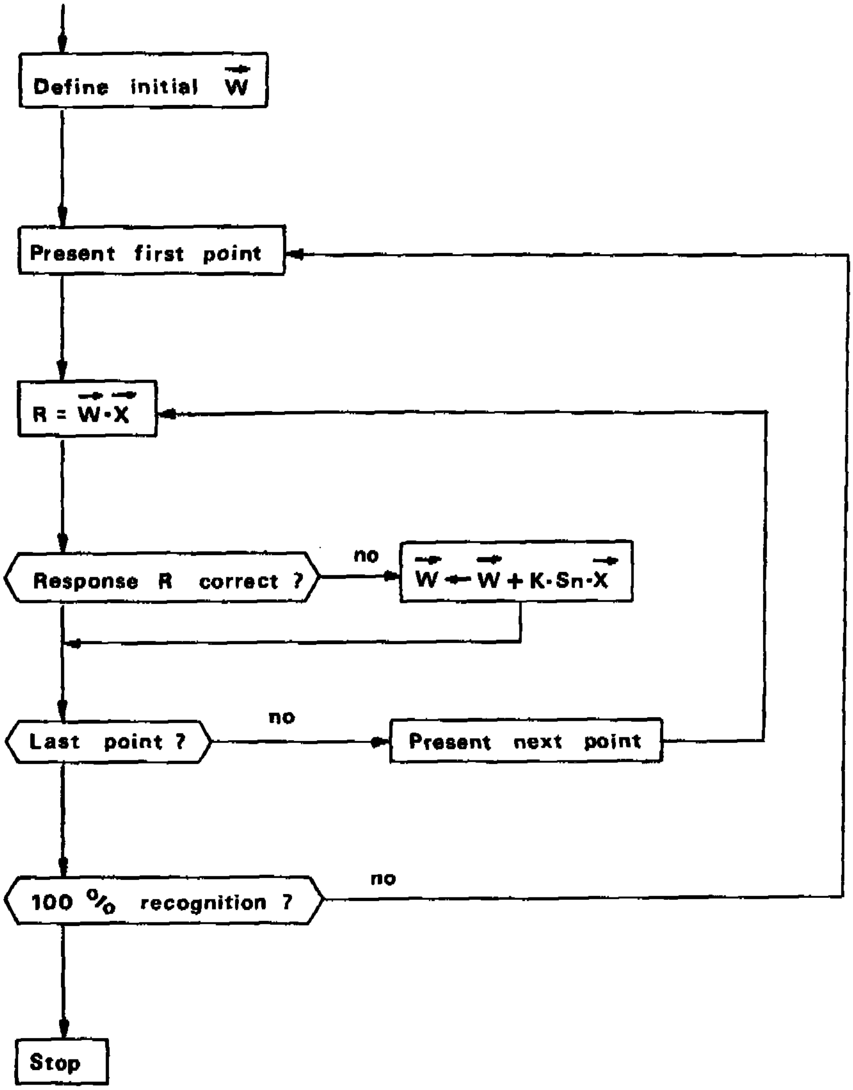

Fig. 1. *Training Flow Chart*

The use of a large number of parameters to characterize a pattern or the use of complicated decision surfaces is unfavourable, as is the use of 'committees' or 'layers' of surfaces, for all increase the value of N. An increase in T may be difficult to manipulate, may prevent 100% recognition from being attained and will adversely affect training times.

The predictive ability is, however, the final test as to the validity of a decision surface. To evaluate it, the points of the prediction set are presented for response by the weight vector and a tally is kept of correct responses. No corrections are made on the weight vector.

Experimental Part

In the present study the aim was to identify the presence of characteristic groups of atoms within the molecule *via* its mass spectra. The criteria may be a function as cited in Table 1, or any one of several of those functions taken together (the presence of oxygen for example). The first column of Table 2 lists the criteria used.

In order to reduce the number of parameters characterizing the spectra, the corresponding reduced spectra between masses 21 and 118 are used. When I_m is the intensity of the peak at $M/e = m$ of a mass spectrum, the intensities, R_j, of the fourteen peaks of the reduced mass spectrum are derived from the following formula:

$$R_j = \sum_{n=0}^{6} \mathrm{I}_{(20+14\cdot n+j)} \quad \text{for j} = 1,2, \ldots 14$$

The spectra are normalized for total ionisation equal to one.

As suggested in [12], use is made of an average reduced spectrum for each of the 16 functions of table 1 present as monofunctions. The intensities, A_{p_j}, of the average reduced spectrum of function p are given by:

$$A_{p_j} = (\Sigma R_j)/N_p \quad \text{for j} = 1,2, \ldots 14$$

where N_p is the number of reduced spectra of monofunctions containing function p used to obtain the average reduced spectrum of that function, and the summation is over those spectra.

To create a buffer zone around each of the 16 main points, the 120 (= 16 · 15/2) simple linear combinations between the 16 average reduced spectra obtained above are employed. The intensities, B_{pq_j}, of such a combination between the average reduced spectra of function p and function q are calculated using the relation:

$$B_{pq_j} = (A_{p_j} + A_{q_j})/2 \quad \text{for } j = 1,2, \ldots 14 \text{ and } p \neq q$$

where A_{p_j} and A_{q_j} are the intensities of the average reduced spectra of function p and q respectively.

The training set consisted of the 16 average reduced spectra and the 120 linear combinations between them.

The decision surfaces used here are either hyperplanes (HP) with $N = 15$ or hyperspheres (HS) with $N = 16$. With 136 pattern points in the training set we obtain a T/N ratio of 9.1 for hyperplanes and 8.5 for hyperspheres. The hypersphere has the advantage of being able to completely enclose a region of hyperspace with only one supplementary adjustable weight. A committee of hyperplanes (one more than there are dimensions) may also achieve this, but with catastrophic consequences on the T/N ratio ($T/N = 0.6$). Tests have been conducted using generalized quadratic surfaces (hyperparaboloid, hyperellipsoid or hyperhyperboloid) with $N = 120$ and $T/N = 1.1$, and quadratic surfaces with axes parallel to the reference axes with $N = 29$ and $T/N = 4.7$, yielding, however, mediocre results ($\langle P \rangle \leqslant 60\%$).

A null initial weight vector is chosen to minimise the effects of the choice when short training times are involved. Negative feedback is applied, as mentioned above, using the fixed increement correction rule with $K = 1$. An unmodified fractional correction rule does not permit the use of a

Table 2. *Predictive Ability*

Criteria	Monofunctions				Difunctions			
	HP		HS		HP		HS	
presence of	$\langle P\rangle$ (P_+, P_-)	$\langle C\rangle$ (C_+, C_-)	$\langle P\rangle$ (P_+, P_-)	$\langle C\rangle$ (C_+, C_-)	$\langle P\rangle$ (P_+, P_-)	$\langle C\rangle$ (C_+, C_-)	$\langle P\rangle$ (P_+, P_-)	$\langle C\rangle$ (C_+, C_-)
Oxygen	88 (93, 83)	89 (85, 93)	88 (93, 83)	89 (85, 93)	61 (53, 69)	61 (63, 59)	56 (41, 72)	57 (59, 55)
Nitrogen	88 (100, 77)	91 (81, 100)	93 (93, 93)	93 (93, 93)	58 (66, 50)	58 (57, 59)	73 (56, 91)	77 (86, 67)
Cycles or insaturation	87 (77, 97)	88 (96, 81)	77 (53, 100)	84 (100, 68)	63 (31, 94)	71 (83, 58)	63 (28, 97)	74 (90, 57)
Phenyl	98 (97, 100)	98 (100, 97)	98 (97, 100)	98 (100, 97)	67 (38, 97)	77 (92, 61)	69 (38, 100)	81 (100, 62)
Carbonyl	80 (87, 73)*	81 (76, 85)	75 (80, 70)*	75 (73, 78)	56 (59, 53)*	56 (56, 57)	58 (59, 56)*	58 (58, 58)
Carbonyl, non nitrogen bonded	85 (83, 87)	85 (86, 84)	87 (83, 90)	87 (89, 84)	63 (41, 84)	65 (72, 59)	63 (44, 81)	65 (70, 59)
Acid or ester	92 (90, 93)	92 (93, 90)	70 (43, 97)	78 (93, 63)	63 (50, 75)	63 (67, 60)	52 (16, 88)	53 (56, 51)
Ketone or aldehyde	82 (97, 67)	85 (74, 95)	82 (97, 67)	85 (74, 95)	56 (28, 84)	59 (64, 54)	56 (28, 84)	59 (64, 54)
Amine	83 (87, 80)*	83 (81, 86)	87 (80, 93)*	87 (92, 82)	58 (53, 63)*	58 (59, 57)	48 (31, 66)*	48 (48, 49)
Amide	68 (83, 53)*	60 (26, 94)	93 (100, 87)*	80 (60, 100)	40 (43, 38)*	44 (13, 75)	56 (43, 69)*	54 (23, 85)
Alcohol	87 (80, 93)*	87 (92, 82)	83 (83, 83)*	83 (83, 83)	52 (19, 84)*	53 (55, 51)	48 (25, 72)*	48 (47, 49)
Ester	–	–	90 (97, 83)*	91 (85, 96)	–	–	52 (56, 47)*	52 (51, 52)
Ether	88 (77, 100)	91 (100, 81)	85 (70, 100)	88 (100, 77)	53 (22, 84)	55 (58, 52)	56 (22, 91)	62 (70, 54)
Primary or secondary amine	92 (93, 90)	92 (90, 93)	93 (93, 93)*	93 (93, 93)	53 (22, 84)	55 (58, 52)	55 (22, 88)*	58 (64, 53)
Tertiary amine	72 (81, 63)	70 (54, 86)	72 (44, 100)	88 (100, 77)	59 (53, 66)	60 (61, 58)	50 (19, 81)	50 (50, 50)
Primary alcohol	78 (67, 90)	80 (87, 73)	75 (53, 97)	81 (94, 67)	–	–	–	–
Secondary alcohol	57 (20, 93)*	65 (71, 58)	53 (12, 93)*	58 (60, 56)	52 (13, 91)*	54 (57, 51)	48 (3, 94)*	41 (33, 49)
Tertiary alcohol	98 (100, 97)*	94 (89, 100)	98 (100, 97)*	94 (89, 100)	73 (59, 88)*	76 (71, 80)	70 (47, 94)*	78 (80, 77)

'Carbonyl, non nitrogen bonded' corresponds to aldehydes or ketones or acids or esters. For 'Phenyl', 'Cycles' and 'Insaturation' see notes to table 1. HP: Hyperplane; HS: Hypersphere; P: Predictive ability; C: Confidence level; * incomplete training; see text.

null initial weight vector and, as with the absolute correction rule, leads, in our experience, to lower predictive abilities, despite shorter training times, than the fixed increment rule chosen.

Pattern vector components 1 to 14 are the intensities of the peaks of the reduced mass spectra; a fifteenth component is used when a hyperspherical decision surface is desired and is the sum of the squares of the intensities of the peaks of the reduced spectra. A last component is equal to one.

Weight vector components 1 to 14 are put under a special constraint impeding negative values which would have little physical meaning in the case of mass spectra. The remaining components are free of constraints.

Training is stopped when 100% recognition is attained or when 1000 passes through the training set have taken place, whichever first occurs.

Results. – When evaluating predictive abilities, prediction subsets are taken sequentially from the bank summarised in Table 1, for each criterion.

Prediction subsets for monofunctions nominally contain 30 spectra of positive class and 30 spectra of negative class. The subsets for difunctions nominally contain 32 spectra of each class.

The percentage of correct classifications is calculated separately for the positive (P_+) and the negative (P_-) class:

$$P_+(P_-) = \frac{\text{number of points correctly classified positive (negative)}}{\text{number of points of positive (negative) class tested}} \cdot 100$$

and an average prediction, $\langle P \rangle$, calculated:

$$\langle P \rangle = (P_+ + P_-)/2$$

A confidence level is also defined for each class:

$$C_+(C_-) = \frac{\text{number of points correctly classified positive (negative)}}{\text{number of points classified positive (negative)}} \cdot 100$$

and an average confidence calculated:

$$\langle C \rangle = (C_+ + C_-)/2.$$

In Table 2, the results under the heading 'HP' are the predictions and confidence levels obtained when hyperplanar decision surfaces were used. Results obtained with hyperspherical decision surfaces appear under the heading 'HS'. An asterisk after the figures for $\langle P \rangle$ indicates that training was stopped without 100% recognition being attained after 1000 passes through the training set.

The first group of results pertains to spectra of monofunctional substances. Results pertaining to spectra of difunctional molecules appear in the second group.

Discussion and conclusion. – When considering the results in Table 2, it should be noted that the mode of presentation eliminates effects due to a possible inequality in the number of spectra present in each class. Accordingly, an average prediction or confidence level of fifty percent signifies a useless 'decision'. One should also take into account a statistical error of $1/\sqrt{M}$, where M is the total number of spectra (dependent on the data bank) in the prediction subset.

If one is concerned with simple, monofunctional substances, the results seem sufficient to justify the use of our decision surfaces as an aid to the (inexperienced) human interpreter seeking confirmation of the presence or absence of a suspected group of atoms.

Difunctional substances are the first step to more complicated molecules, and the decision surfaces obtained are inadequate for their interpretation.

If a larger number of parameters than those employed here is used, the ease of subsequent interpretation will suffer and problems will arise in order to maintain a reasonable T/N ratio so that insignificant information is not used for decision making. The separation of significant from insignificant information is essential and explains the use we have made of average reduced mass spectra for training; indeed, the risk of irrelevant information (in the form of a singular case) influencing a decision surface during training is greatly diminished. As with library searches and nearest neighbour techniques, the learning machine approach is valid for substances contained in the space spanned by the training set. Artificial intelligence is not thus limited.

In conclusion, better predictive abilities than those obtained here are necessary before the proposed interpretation of mass spectra can compete with library searches or artificial intelligence systems when complex molecules are involved. We believe that an improvement of predictive abilities in this case lies not in the use of a large number of parameters, but in the use of a restricted number of sophisticated parameters and on decision surfaces based not on simple 'chemical' structural groups but, again, on more sophisticated criteria. The problem is to determine such parameters and criteria systematically.

Thanks are extended to Dr. *Clerc* for the spectra and the corresponding *Wiswesser* formulae.

REFERENCES

[1] *H. S. Hertz, R. A. Hites & K. Biemann*, Analyt. Chemistry *43*, 681 (1971).

[2] *D. H. Smith, B. G. Buchanan, R. S. Engelmore, A. M. Duffield, A. Yeo, E. A. Feigenbaum, J. Lederberg & C. Djerassi*, J. Amer. chem. Soc. *94*, 5962 (1972).

[3] *C. F. Bender, H. D. Shepherd & B. R. Kowalski*, Analyt. Chemistry *45*, 617 (1973).

[4] *T. L. Isenhour & P. C. Jurs*, Analyt. Chemistry *43*, 20A (1971).

[5] *K.-S. Kwok, R. Venkataraghavan & F. W. McLafferty*, J. Amer. chem. Soc. *95*, 4185 (1973).

[6] *T. L. Isenhour, B. R. Kowalski, P. C. Jurs*, Critical Reviews in Analytical Chemistry *4*, 1, (July 1974).

[7] *K. Varmuza & P. Krenmayr*, Z. analyt. Chem. *266*, 274 (1973).

[8] *J. T. Clerc, P. Naegeli & J. Seibl*, Chimia *27*, 639 (1973).

[9] *E. G. Smith*, 'The Wiswesser Line Formula Chemical Notation', McGraw Hill, N.Y. 1968.

[10] *N. J. Nilsson*, 'Learning Machines', McGraw Hill, NY 1965.

[11] *W. S. Meisel*, 'Computer-Oriented Approaches to Pattern Recognition', Academic Press, N.Y., London 1972.

[12] *L. R. Crawford & J. D. Morrison*, Analyt. Chemistry *40*, 1469 (1968).