

Configurational Assignment of Diastereoisomeric 2-Hetero-substituted Sulphoxides by Application of Multivariate Analysis Methods to Their Nuclear Magnetic Resonance Parameters

Jose Luis Garcia Ruano*

Departamento de Quimica, Universidad Autonoma de Madrid, Cantoblanco, 28049-Madrid, Spain

Pedro Noheda and Jesus Sanz

Instituto de Quimica Organica, C.S.I.C. Juan de la Cierva 3, 28006-Madrid, Spain

Multivariate analysis methods (principal component analysis, discriminant analysis, and SIMCA) have been applied to the ^1H n.m.r. parameters of acyclic 2-hetero-substituted sulphoxides, in order to establish their relative configurations, with satisfactory results. These methods are quicker and probably also more accurate than previous methods based on a complex conformational study. The results indicate that chemical shifts can be accurately correlated with the relative configurations of the substrates; however this relationship is not overt and is impossible to derive directly.

The relationship between vicinal coupling constants and the stereochemistry of the protons involved can be a valuable tool for establishing the relative configurations of the chiral centres of diastereoisomeric sulphoxides. For rigid systems it has been possible to formulate a series of empirical rules¹ to relate the spatial arrangement of the sulphonyl group with the chemical shifts of the surrounding nuclei. Thus, the configurations of cyclic sulphoxides can be easily assigned from their ^1H n.m.r. parameters, as a result of the restricted conformational mobility of these systems, where only one, or two at maximum, conformations are involved in their conformational equilibria.

The situation is completely different with acyclic sulphoxides. Their conformational flexibility, determining the participation of many conformations in the equilibria, limits the application of n.m.r. parameters, as configurational criteria, to those cases where the different rotamer populations can be determined. In most cases this is not possible and thus several approximations are necessary, which decrease the accuracy of the conclusions. Thus, in the configurational assignments of many pairs of diastereoisomeric acyclic 2-oxygenated sulphoxides (see references in Table 1) it has been necessary to study the conformational behaviour of certain fragments of the molecules under various conditions, in order to infer the conformational preferences of other fragments, for which there is no direct information. The relative configurations deduced from such studies have had to be confirmed, in several cases, by X-ray diffraction methods.^{2,3} In addition, it was apparent that the role of chemical shifts in the solution of the problem was slight, in spite of the evident relationship between these parameters and the stereochemistry associated with each relative configuration. Thus the balance between the time required for these spectroscopic studies and the reliability of conclusions deduced from them is not very favourable. Therefore it seemed interesting to investigate other faster and more accurate methods of assigning the configurations.

We have therefore applied the methods of multivariate analysis to the ^1H n.m.r. parameters, and report here the application of such methods to the parameters collected in Table 2, corresponding to the compounds listed in Table 1. When this study was begun, there were enough data to affirm that the relative configurations of the substrates (1)–(21) were known. Compounds designated α have the same relative configuration at both chiral centres ($R,R/S,S$), whereas the β -epimers have the opposite ($R,S/S,R$) (see Figure 1). On the other hand, at that time we did not have enough data to assign the configurations of compounds (22)–(27). Taking into account that substrates (1)–(19) are 2-oxygenated and (20)–(27) 2-

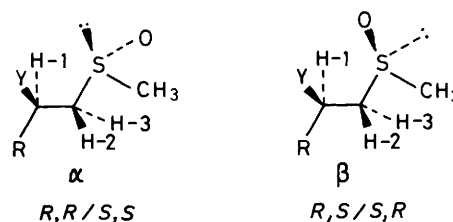


Figure 1. Configurations of compounds (1)–(27)

nitrogenated sulphoxides, we have used the n.m.r. parameters of the first set to define the α and β classes; those of the 2-amino sulphoxides have been used as a 'test set' [two compounds of this set, (20) and (21), had known configurations]. Recently, we have completed configurational studies on compounds (22)–(27), resulting in configurational assignments (see references in Table 1), confirming the validity of the conclusions established in the present paper.

Results and Discussion

Chemical shifts and coupling constants of compounds (1)–(27) (Table 1), obtained from the analysis of their spectra recorded with CDCl_3 and $(\text{CD}_3)_2\text{SO}$ as solvents, are collected in Table 2. Several β -epimers exhibited deceptively simple spectra in both solvents and they have not therefore been considered in Tables 1 and 2.

A detailed study of Table 2, indicates that the values of ΔJ ($J_{1,2} - J_{1,3}$) for the α -epimers are larger than those for the corresponding β -epimers. [especially in $(\text{CD}_3)_2\text{SO}$], suggesting a relationship between the relative configurations of the substrates and the values of their vicinal coupling constants. It is also evident, from the δ values of H-1, H-2, and H-3, that a similar relationship between these parameters and the relative configurations of the substrates cannot be established at first sight, although it must exist. Therefore, the problem of configurational assignments for this type of compound is usually solved on the basis of the values of the vicinal coupling constants, although the application of this criterion requires data corresponding to both diastereoisomers of each compound. In this paper we have applied several methods of multivariate analysis to the data in Table 2, in order to discover whether these values can be related to relative configuration.

The data matrix which includes the chemical shifts of compounds (1)–(27) in CDCl_3 and $(\text{CD}_3)_2\text{SO}$ (variables 1–3

and 7–9 in Table 2) will be called matrix D; matrix J includes the coupling constants for the same compounds and solvents (variables 4–6 and 10–12 in Table 2). Matrix T is formed by

Table 1. Configurations (see Figure 1) of the sulfoxides (1)–(27)

R-CHY-CH ₂ -SO-CH ₃			
α	R	Y	β
(1) ^a	Me	OH	(13) ^a
(2) ^b	Ph	OH	
(3) ^b	Ph	OMe	(14) ^b
(4) ^b	Ph	OAc	
(5) ^c	4-NO ₂ C ₆ H ₄	OH	
(6) ^c	4-NO ₂ C ₆ H ₄	OMe	
(7) ^c	4-MeOC ₆ H ₄	OH	(15) ^c
(8) ^c	4-MeOC ₆ H ₄	OMe	(16) ^c
(9) ^d	4-Pyridyl	OH	(17) ^d
(10) ^d	4-Pyridyl	OMe	
(11) ^e	Bu ^t	OH	(18) ^e
(12) ^e	Bu ^t	OAc	(19) ^e
(20) ^f	Ph	NHPh	(21) ^f
(22) ^g	Ph	NHMe	(25) ^g
(23) ^g	Ph	NMe ₂	(26) ^g
(24) ^g	Ph	NH ₂	(27) ^g

^a F. Alcudia, J. L. Garcia Ruano, J. H. Rodriguez, and F. Sanchez, *Can. J. Chem.*, 1979, **57**, 2426. ^b E. Brunet, J. L. Garcia Ruano, M. A. Hoyos, J. H. Rodriguez, P. Prados, and F. Alcudia, *Org. Magn. Reson.*, 1983, **21**, 643. ^c F. Alcudia, J. M. Llera, E. Brunet, L. Castillo, J. L. Garcia Ruano, and J. H. Rodriguez, *An. Quim., Ser. C*, 1986, **82**, 18. ^d J. L. Garcia Ruano, C. Pedregal, and J. H. Rodriguez, *Tetrahedron*, 1987, **43**, 1987. ^e E. Brunet, J. L. Garcia Ruano, J. H. Rodriguez, and F. Alcudia, *Tetrahedron*, 1984, **40**, 4433. ^f E. Brunet, M. C. Carreño, M. T. Gallego, J. L. Garcia Ruano, and F. Alcudia, *J. Chem. Soc., Perkin Trans. 2*, 1983, 937. ^g Ref. 6.

all the values of both chemical shifts and coupling constants shown in Table 2.

(a) *Principal Component Analysis.*—In principal component analysis (PCA) of a data matrix, variables (components) are calculated as linear combinations of the original ones, in such a way that each new variable accounts for the maximum amount of variance in the data.⁴

PCA is usually the first method selected for the study of a complex data matrix. It can be used to show, in a graphical form, the overall structure of the matrix. The amount of variance accounted for by each component can be used as a measure of the correlations among the variables.

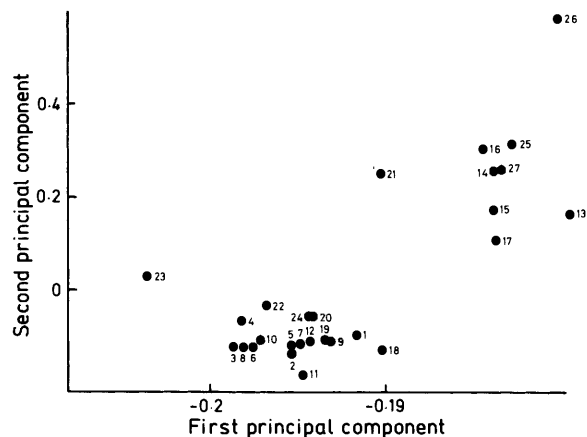


Figure 2. Principal component plot of compounds (1)–(27), using variables 4, 5, 6, 10, 11, and 12 (coupling constant data)

Table 2. N.m.r. parameters for the nuclei H(1), H(2), and H(3) in the sulfoxides (1)–(27)

Variable no.	CDCl ₃						(CD ₃) ₂ SO					
	$\delta(1)$	$\delta(2)$	$\delta(3)$	$J_{1,2}$	$J_{1,3}$	$J_{2,3}$	$\delta(1)$	$\delta(2)$	$\delta(3)$	$J_{1,2}$	$J_{1,3}$	$J_{2,3}$
Compd.	1	2	3	4	5	6	7	8	9	10	11	12
(1)	4.51	2.87	2.71	9.90	2.20	-13.20	4.13	2.83	2.80	10.40	2.50	-13.00
(2)	5.29	3.01	2.88	10.90	1.90	-13.00	4.98	3.03	2.85	11.00	2.60	-12.90
(3)	4.71	3.03	2.91	11.10	2.50	-13.10	4.59	3.20	2.84	11.10	2.60	-13.10
(4)	6.19	3.22	3.05	10.30	3.10	-13.30	6.04	3.44	3.07	10.80	2.90	-13.40
(5)	5.55	3.095	2.90	10.10	2.30	-13.30	5.11	3.04	2.97	11.00	2.20	-13.20
(6)	4.58	3.00	2.87	11.10	2.20	-13.00	4.76	3.20	2.935	10.90	2.70	-13.20
(7)	5.30	3.065	2.86	10.50	2.60	-13.00	4.91	3.02	2.81	11.10	2.40	-12.90
(8)	4.66	3.03	2.875	11.10	2.50	-13.00	4.50	3.20	2.81	11.00	2.50	-13.20
(9)	5.41	3.14	3.16	9.80	2.40	-13.20	4.97	3.00	3.08	11.10	2.40	-12.90
(10)	4.89	3.19	3.08	10.70	2.60	-13.10	4.62	3.19	2.95	11.00	2.60	-13.20
(11)	3.94	2.81	2.73	11.10	1.70	-12.90	3.50	2.64	2.73	11.30	1.80	-12.80
(12)	5.22	2.77	2.91	10.30	2.10	-13.30	5.02	2.91	2.93	10.20	2.20	-13.50
(13)	4.50	2.87	2.75	9.10	2.60	-13.00	4.13	2.92	2.89	6.30	6.10	-13.00
(14)	4.71	3.17	2.95	7.40	5.60	-13.30	4.59	3.20	3.07	7.50	6.10	-13.20
(15)	5.28	3.11	2.92	9.10	3.50	-13.00	4.91	3.075	3.03	6.9	6.00	-13.30
(16)	4.665	3.21	2.945	7.20	6.00	-13.30	4.51	3.20	3.045	7.20	6.80	-13.40
(17)	5.43	3.07	3.37	8.80	3.00	-13.30	5.04	3.09	3.29	7.60	5.00	-13.10
(18)	3.92	2.75	2.85	10.30	1.40	-13.10	3.51	2.71	2.95	9.90	2.20	-13.20
(19)	4.35	2.64	2.80	10.10	2.10	-12.90	4.92	2.84	3.11	10.30	2.30	-13.70
(20)	5.02	3.15	3.00	8.86	3.08	-13.34	4.84	3.16	2.96	11.50	3.04	-13.04
(21)	4.90	3.18	2.93	9.22	5.47	-13.27	4.79	3.23	3.10	7.81	7.10	-12.83
(22)	4.13	2.93	2.94	10.08	3.45	-13.24	3.97	3.08	2.82	10.74	3.48	-13.12
(23)	4.11	3.55	2.99	10.76	4.40	-13.06	4.09	3.66	2.84	11.15	4.90	-13.32
(24)	4.60	2.92	2.97	10.41	3.28	-12.80	4.21	2.99	2.81	10.83	3.29	-12.82
(25)	4.07	3.15	2.81	7.62	6.35	-12.97	3.95	3.16	3.00	7.26	7.12	-12.83
(26)	3.82	2.99	3.46	9.48	6.17	-12.75	4.00	3.24	3.41	10.23	5.85	-12.85
(27)	4.56	3.11	2.88	8.02	5.61	-12.83	4.25	3.01	2.95	7.55	6.54	-12.99

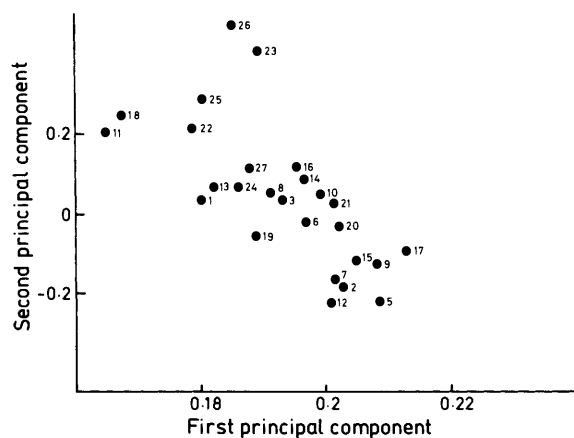


Figure 3. Principal component plot of compounds (1)–(27), using variables 1, 2, 3, 6, 7, and 8 (chemical shift data)

In our case, PCA has been applied separately to matrices D and J, since they differ in physical significance. This also avoids the mathematical problem of the different magnitudes of coupling constants and chemical shift values.

In Figure 2, all compounds studied are plotted with coordinates the first and second components calculated from data in matrix J. These two components account for 99.7% of the total variance. The compounds seem to form two groups. α -Compounds are plotted near the lower left corner, while β -compounds occupy the right-hand side. However, two β -compounds, (18) and (19), appear with the α -group. Similar results are obtained when the third component is used as a coordinate axis.

Figure 3 shows in a similar way the results obtained from the data in matrix D. No natural groups are observed in this plot, but many α - and β -compounds appear as close pairs.

Figures 2 and 3 show that the relationship between configuration and the data in matrix J (coupling constants) is more evident than that with the chemical shift values (matrix D). α , β -Pairs are closer in Figure 3, since their chemical shift values are more similar than their coupling constants.

These results justify the choice of coupling constants as a criterion for assigning configurations; and account for failures of the attempts to relate configuration with chemical shift values.

(b) *Discriminant Analysis.*—Discriminant analysis can be applied to a data matrix in which at least two groups of data have been previously defined. This method calculates one or more functions as linear combinations of the original variables, trying to minimize the ratio between the sum of variances of the compounds in each group and the difference between the group means. The results can be shown in graphical form, as in principal component analysis, by plotting the function values for each compound.

The calculations of the function values were carried out by means of a program written by us in BASIC, according to a method developed by Foley and Sammon.⁵

In order to improve the significance of our results, we selected compounds (20)–(27) (Table 3) as a test set; their values were not included in the calculations. These compounds are 2-nitrogenated sulphoxides and their configurations are now known. Their use as test set was expected to indicate whether the method could be applied to any 2-hetero-substituted sulphoxides regardless of the type of substituents at C-2.

From the data in matrix J, we can see that the classification is correct for compounds (1)–(20); however, compound (21) is erroneously assigned to the α -class (Figure 4). The co-ordinates

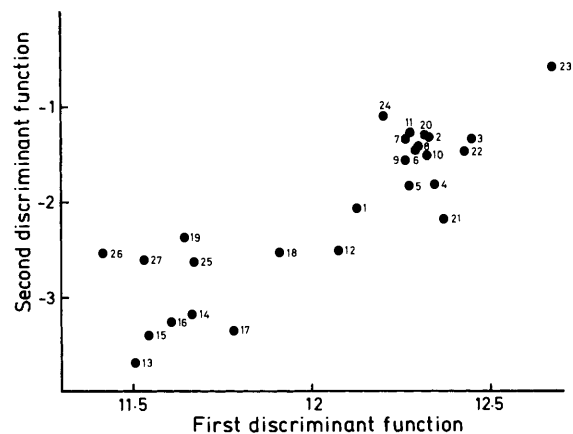


Figure 4. Discriminant analysis plot of compounds (1)–(27), using the function values (see text) calculated from variables 4, 5, 6, 10, 11, and 12 (coupling constant data)

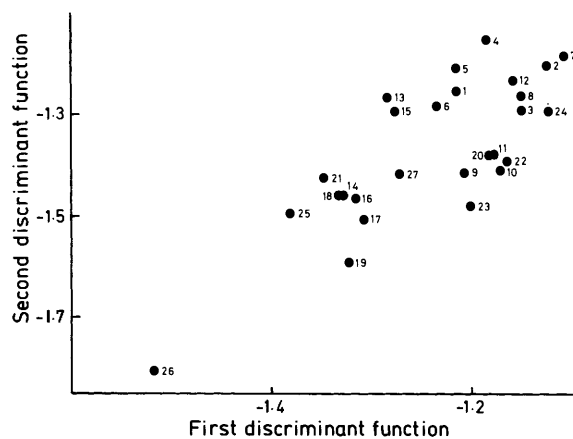


Figure 5. Discriminant analysis plot of compounds (1)–(27), using the function values (see text) calculated from variables 1, 2, 3, 6, 7, and 8 (chemical shift data)

used in the representation are the two first discriminant functions.

The results are better when matrix D is used: the assignment is 100% correct (Figure 5). This indicates that, although there is no natural grouping for α - and β -compounds from their chemical shift data (Figure 3) it is possible to relate these values to the configuration.

Compound (26) is correctly plotted in the β -class area, but separated from the other β -compounds. All β -compounds in Table 1 show a conformational preference for the rotamer which exhibits the sulphur function and the group R in an antiperiplanar arrangement. However, compound (26) is the only one in which the preferred conformation [in both CDCl_3 and $(\text{CD}_3)_2\text{SO}$] has the substituents R and Y (see Table 1) in an antiperiplanar arrangement as shown from protonation studies.⁶

Although α - and β -groups are differentiated when chemical shift data are used (Figure 5), a better separation can be obtained by choosing, among the 12 variables in matrix T, those showing highest discriminating power. Variables 1, 5, 6, 7, 9, 10, and 11 were selected according to the SIMCA results, and a better separation was then obtained for α - and β -compounds. The results of discriminant analysis are shown in Figure 6. The possibility of selecting variables in discriminant analysis is often used in order to achieve a good separation with a smaller number of variables.

Table 3. Residual values obtained for compounds (1)–(27) when fitted by the SIMCA method to classes 1 (α -compounds) and 2 (β -compounds)

Compd.	Class	Data Matrix					
		T		D		J	
		Cl. 1	Cl. 2	Cl. 1	Cl. 2	Cl. 1	Cl. 2
(1)	1	0.74	0.79	0.27	0.56	0.81	1.02
(2)	1	0.64	1.19	0.38	0.99	0.71	1.04
(3)	1	0.43	1.08	0.32	0.64	0.33	0.98
(4)	1	0.31	1.29	0.44	1.18	0.29	0.93
(5)	1	0.63	1.04	0.39	0.88	0.74	1.12
(6)	1	0.62	0.94	0.49	0.54	0.74	0.74
(7)	1	0.59	1.23	0.62	1.04	0.72	1.14
(8)	1	0.53	1.06	0.38	0.65	0.50	0.76
(9)	1	1.07	1.03	1.03	0.56	1.10	1.30
(10)	1	0.49	0.99	0.48	0.44	0.15	0.81
(11)	1	0.44	0.94	0.37	0.65	0.41	1.21
(12)	1	0.58	0.85	0.49	0.96	1.01	0.87
(13)	2	5.87	0.62	0.45	0.37	8.35	0.46
(14)	2	5.46	0.44	0.78	0.33	6.28	0.36
(15)	2	5.24	0.72	0.46	0.64	7.11	0.72
(16)	2	6.24	0.63	0.73	0.33	6.87	0.11
(17)	2	3.87	0.72	7.75	0.49	5.43	0.68
(18)	2	0.99	0.68	1.14	0.71	1.44	0.72
(19)	2	1.38	0.67	1.37	0.76	1.79	0.40
(20)	1	1.42	1.05	0.22	0.40	2.19	1.41
(21)	2	6.00	0.69	0.82	0.28	6.93	0.90
(22)	1	1.68	0.88	0.53	0.67	1.12	1.03
(23)	1	2.74	1.73	0.46	1.10	1.70	1.04
(24)	1	1.46	1.10	0.51	0.71	1.68	1.33
(25)	2	7.00	0.74	0.98	0.61	8.12	1.11
(26)	2	4.83	1.76	1.85	1.64	12.96	2.13
(27)	2	6.12	0.72	0.56	0.20	7.33	1.21

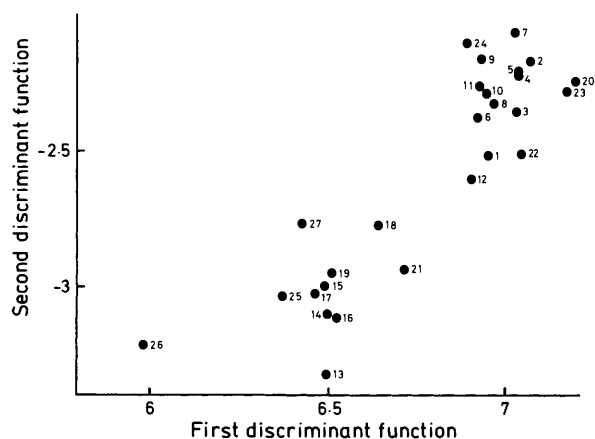


Figure 6. Discriminant analysis plot of compounds (1)–(27), using the function values calculated from selected variables data (see text)

(c) *SIMCA*.—A disadvantage of discriminant analysis is its lack of flexibility. In the *SIMCA* method, developed by Wold *et al.*,⁷ the data for each class are fitted, in an independent way, to a principal component (PC) model. Once the PC parameters

have been calculated, test set data are fitted to the different PC models; the quality of fit can be considered as a measure of the probability of pertinence of the compound to each class. The calculations were carried out by means of a program written in *BASIC* from the *SIMCA* description given by Wold.⁷

The first step in *SIMCA* is the determination of the number of components for each class of PC model. From our PC calculations, we have assumed that two components are enough to describe the n.m.r. data of compounds (1)–(19).

When all variables in matrix T are used, the *SIMCA* method assigns the β -configuration to all compounds in the test set. Similar results are obtained when matrix J is used. However, when the starting data are those in matrix D, all compounds in the test set are assigned correctly (Table 3).

These results appear to confirm that chemical shifts, rather than coupling constants, are the parameters that contain enough information to assign the configuration of the 2-hetero-substituted sulphoxides, regardless of the substituent in the 2-position. Nevertheless, this relationship is not evident, and multivariate methods are required to find it.

Multivariate analysis is a useful tool for extracting the relevant information from a complex data matrix. However, a careful study seems to be necessary in order to select both the mathematical method and the starting data. The results will depend on the different methods and techniques available. Our results with 2-hetero-substituted sulphoxides show that multivariate methods can be used for configurational determination from analytical data such as n.m.r.

Experimental

¹H N.m.r. spectra were recorded in the Fourier transform mode with a Varian XL-100-15 spectrometer coupled with a Varian 620/L 16 K computer, transforming 8 K data points, or with a Bruker WM-200-SY instrument (Aspec 2000 80 K computer) transforming 16 K data points, as indicated in the references in Table 1.

Calculations were carried out using an Olivetti M-20 microcomputer. Programs (principal component analysis, discriminant analysis and *SIMCA*) were written by us in *BASIC* according to the descriptions given in references 4, 5, and 7.

References

- 1 A. B. Foster, T. D. Inch, M. H. Quadir, and J. H. Webber, *J. Chem. Soc., Chem. Commun.*, 1968, 1086.
- 2 M. A. Hoyos, S. Martinez, and S. Garcia Blanco, *Acta Crystallogr., Sect. C*, 1983, **39**, 1018.
- 3 M. A. Hoyos, S. Martinez, and S. Garcia Blanco, *Acta Crystallogr., Sect. C*, 1983, **39**, 473.
- 4 E. R. Malinowski and D. G. Howery, 'Factor Analysis in Chemistry,' Wiley, New York, 1980.
- 5 D. H. Foley and J. W. Sammon, *IEEE Trans. Comput.*, 1975, **C-24**, 281.
- 6 E. Brunet, M. T. Gallego, and J. L. Garcia Ruano, *Tetrahedron*, 1986, **42**, 1423.
- 7 S. Wold and M. Sjostrom, in 'Chemometrics: Theory and Application,' ed. B. R. Kowalski, ACS Symposium Series no. 52, 1977, p. 243.

Received 13th May 1987; Paper 7/843