Trilinear PARAFAC decomposition of synchronous fluorescence spectra of mixtures of the major metabolites of acetylsalicylic acid

The Analyst

Joaquim C. G. Esteves da Silva* and Sandra A. G. Novais

LAQUIPAI, Chemistry Department, Faculdade de Ciências, R. Campo Alegre 687, P4150 Porto, Portugal. E-mail: jcsilva@fc.up.pt

Received 8th April 1998, Accepted 30th July 1998

Mixtures of the three major metabolites of acetylsalicylic acid (salicylic, gentisic and salicyluric acid) were analyzed by synchronous molecular fluorescence spectroscopy. Three-way data matrices were generated by acquisition of spectra as a function of the pH (between 2 and 11) and of different relative concentrations of the three components. The PARAFAC trilinear model, without restrictions and using one factor per metabolite, was used in the data analysis. A full decomposition of the data matrices into the spectra, concentration and pH profiles was obtained. This result shows that molecular fluorescence spectroscopy can be used for the development of robust analytical methods for the simultaneous determination of the three major metabolites of acetylsalicylic acid in complex background samples.

Introduction

Molecular fluorescence (MF) spectroscopy is a well known highly sensitive analytical technique.[1] The synchronous fluorescence (SyF) mode of MF spectral acquisition is of particular interest for the analysis of complex samples because higher spectral resolution and background reduction are achieved, increasing the potential of MF for the simultaneous determination of multiple analytes in multicomponent samples.[2–10]

One characteristic of MF methods is the ease of developing procedures to generate three-way spectral data matrices (several sets of two-way matrices) from one sample.[11–17] These matrices can be obtained by the variation of the intrinsic instrumental factors, such as the excitation and emission wavelengths, and/or by the controlled modification of experimental factors, *e.g.*, pH in the case of fluorophores with acid–base properties. If the two-way data matrices correspond to samples with the same qualitative composition, but where the relative contributions of the components to the overall spectra are different, the three-way data matrices are candidates for a unique decomposition process by chemometric factor analysis techniques.[18–39] If a unique decomposition is achieved, then the so-called 'second-order advantage' is obtained, allowing the robust estimation of analyte concentrations in mixtures that contain unknown interferences.[32]

The decomposition analysis of three-way data structures was first investigated in the field of psychometrics.[18–21] The first application to chemical analysis of the 'unique decomposition property' was in the multicomponent analysis of mixtures of aromatic hydrocarbons by MF spectroscopy coupled to liquid chromatography.[22] Since this first application, the potential of performing analytical chemistry with three-way data matrices has been increasingly recognized and great developments are being observed in chemometric methodologies to deal with these types of data structures.[26–39] However, the number of applications of these methodologies to experimental data is still very limited.[22,24,33–39]

This paper present the results of the chemometric decomposition by PARAFAC (parallel factor analysis)[20] of three-way SyF spectra sets, corresponding to mixtures of the three major acetylsalicylic acid metabolites, salicylic acid (SA), gentisic acid (GE) and salicyluric acid (SU). PARAFAC is a three-way analysis method in which the basic trilinear model is particularly suitable for the analysis of spectroscopic data.[31] In this work, experimental data were generated by varying the following three factors: (i) wavelength (between 250 to 450 nm); (ii) pH (between about 2 and 11); and (iii) concentration of the three metabolites (SA and GE, 1×10^{-6}–1×10^{-5} M; SU, 1×10^{-7}–1×10^{-6} M). The direct analysis of drugs and their metabolites in complex and not completed characterized matrices, such as biological fluids, is important in clinical analytical chemistry and usually constitutes a difficult problem.[40,41]

The objectives of this work were the following: (i) to study the decomposition of the three-way spectral data matrices corresponding to the three major metabolites of acetylsalicylic acid; (ii) to establish the basis to the development of robust methods for the direct and simultaneous robust quantification of the three acetylsalicylic acid metabolites in complex matrices; and (iii) to assess the usefulness of the PARAFAC model for the analysis of real experimental data sets.

Theory

The acquisition of a SyF spectrum for one sample constituted by the mixture of *nc* components with acid–base properties originates a vector $\boldsymbol{x}$ with *nw* elements (*nw* is the number of wavelengths of the SyF spectrum) of responses (fluorescence intensities). If the pH of the sample is changed, its SyF spectrum will be affected because the fluorescence properties of the three components will change upon protonation/deprotonation reactions. The acquisition of a SyF spectrum as function of a selected number of pH values (*nph*) generates a two-way data matrix $\boldsymbol{D}$ for the sample under analysis, which can be expressed by the following bilinear model:

$$d_{ij} = \sum_{s=1}^{nc} p_{is} w_{js} + e_{ij} \qquad i = 1,\ldots,nph;\ \ j = 1,\ldots,nw \qquad (1)$$

where d_{ij} are the elements of matrix $\boldsymbol{D}$, p_{is} and w_{js} are the elements of the matrices $\boldsymbol{P}$ and $\boldsymbol{W}$ which contain the pH concentration profiles and pure spectra of the *nc* components,

respectively, and e_{ij} are the elements of the matrix of the spectroscopic error.

Several data analysis strategies have been developed for the decomposition of a data matrix such as ***D*** without any *a priori* information about the sample.[42–45] These methodologies usually achieve the decomposition of the data matrix into the two basic quantities, in the case under discussion the pure spectra of the components and their pH–concentration profiles, by imposing restrictions, selected by physico-chemical criteria, in the mathematical decomposition process.[43–45] These restrictions are necessary because the direct decomposition of a two-way data matrix has a factor analysis limitation, which is called 'the rotational problem'.[18–21] Indeed, this decomposition is not unique and there are an infinite number of possible solutions consistent with any given data set.

A strategy to overcome 'the rotational problem' is by obtaining several two-way data matrices (three-way data matrix) for the same sample by varying another variable of the system under analysis. In chemical analysis the relative concentration of the components is the next variable that needs to be changed because the final objective of the trilinear decomposition is the development of quantitative analytical chemistry methodologies. The resultant three-way data structure can be expressed by the following trilinear model, which is an extension of eqn. (1):

$$t_{ijk} = \sum_{s=1}^{nf} p_{is} w_{js} c_{ks} + e_{ijk} \qquad (2)$$

$$i = 1,...,nph;\ \ j = 1,...,nw;\ \ k = 1,...,nc$$

where t_{ijk} are the elements of the experimental matrix ***T***, c_{ks} are the elements of the new matrix ***C*** which contain the relative concentration of the components and *nf* is the number of factors of the model. Eqn. (2) also represents the basic three-way PARAFAC model.[18–21] After selecting the number of factors for the model, the three basic unknown matrices ***P***, ***W*** and ***C*** are calculated by an iterative alternating least squares method without any restrictions, taking as a first approximation for the three matrices a random generated set.

The quality of the fit of experimental data to the tested model is assessed by the value of the loss function after convergence is achieved, which is defined by

$$\sum_{i=1}^{nph}\sum_{j=1}^{nw}\sum_{k=1}^{nc} {e_{ijk}}^2 = \sum_{i=1}^{nph}\sum_{j=1}^{nw}\sum_{k=1}^{nc} (t_{ijk} - t_{ijk}')^2 \qquad (3)$$

where t_{ijk}' are the predicted elements of the experimental matrix.

Experimental

Reagents

Stock standard solutions of SA (Sigma, St. Louis, MO, USA), GE (Aldrich, Gillingham, Dorset, UK) and SU (Sigma) were prepared in 0.10 M potassium nitrate (Merck, Darmstadt, Germany). To the mixtures of these three acids, obtained by rigorous dilution in 0.10 M potassium nitrate, a standard solution of 0.100 M nitric acid (Titrisol, Merck) was added to a total concentration of 0.0100 M. A solution of 0.02 M decarbonated potassium hydroxide (Merck) was used as a titrant.[46]

Experimental design of the metabolite solutions

A linear relationship between the SyF spectral intensity and the concentration of the three metabolites is expected and consequently eight metabolite solutions, corresponding to a two-level, three-factor full factorial design, would be sufficient to characterize the relationship. However, to check for non-linear behavior, another set of eight solutions of the metabolites were prepared using a similar design but with different levels. The overall 16 samples were randomly scheduled for analysis. Nevertheless, owing to experimental time limitations, only 12 randomly selected from the complete 16 set were analyzed. The concentrations corresponding to the levels were as folows: SA, 3.2, 6.0, 16 and 22 μM; GE, 3.1, 6.3, 16 and 22 μM; and, SU, 0.46, 0.92, 1.8 and 2.3 μM.

Instrumentation

Potentiometric pH measurements were made as described previously.[16–17] The experiments were carried out under nitrogen at 25.0 ± 0.2 °C. The potentiometric cell was calibrated with three buffer solutions (pH_1 = 9.043, pH_2 = 6.784 and pH_3 = 3.883) with the ionic strength adjusted to 0.1 M.[47]

SyF measurements were made with a Perkin-Elmer (Norwalk, CT, USA) LS-50 luminescence spectrometer with a flow cell. A peristaltic pump forced the displacement of the titrated solution into the flow cell after pH adjustment to the previously defined value. The following 24 pH values were used in all titrations (standard deviations in parentheses): 2.51(4); 2.73(4); 2.97(4); 3.18(8); 3.41(8); 3.85(8); 4.15(12); 4.71(8); 4.93(8); 5.37(8); 5.65(5); 5.92(8); 6.30(11); 6.55(12); 6.81(11); 7.16(15); 7.78(11); 8.51(11); 8.75(8); 9.01(12); 9.29(14); 9.58(15); 9.86(14); and 10.17(5). After stopping the peristaltic pump, SyF spectra were recorded with excitation between 250 and 450 nm, with the following settings: excitation and emission slit widths, 7.5 nm; wavelength difference between the excitation and emission monochromators, 60 nm; scan rate, 200 nm min^{-1}; and spectra digitized at every 0.5 nm throughout the spectral range.

Data analysis

The resolution of the raw SyF spectra was reduced to 5 nm (41 points per spectrum) to be used in the data analysis. The program TRILIN, obtained from P. M. Kroonenberg (Department of Education, Leiden University), was used for the PARAFAC decomposition. The three-way data structures (a total number of 11 808 points) were composed of 41 wavelengths, 24 pH values and 12 different relative concentrations of the three acids and were used without any pre-processing. A copy of a three-way data set corresponding to a mixture of the three acetylsalicylic acid metabolites can be obtained from the corresponding author.

Results and discussion

Preliminary analysis

Table 1 shows the characteristics of the SyF spectral band (with a Gaussian shape) of the three individual acetylsalicylic acid

Table 1 SyF spectral and acid–base characteristics of the three major acetylsalicylic acid metabolites

Metabolite	Maximum of the SyF band (nm)[a]	Acid–base characteristics[b]	
		pK_{a1}	pK_{a2}
SA	312	2.82	
SU	330	3.44	8.24
GE	340	2.98	[c]

[a] The width of the SyF band at 50% height is about 30 nm. [b] Literature values.[48,49] [c] Result not available.

metabolites observed experimentally under the settings used in this work. Fig.1 presents typical experimental spectral sets for the mixture of the three constituents under analysis. This figure shows that the SyF spectra of the individual metabolites are highly overlapped and the resultant spectra are characterized by one or two superimposed bands. This lack of spectral resolution was expected from the data in Table 1 because the maxima of the spectral bands of the three metabolites are separated by only 28 nm and the width (at 50% height) of each SyF band is about 30 nm.

The analysis of Fig. 1 also shows the existence of spectral variations at the beginning (low pH values) and end (higher pH values) of the titration. These variations are due to the deprotonation of the different species of the three metabolites, which are diprotic acids, and all have a pK_{a1} of about 3 and SU has a pK_{a2} of about 8. The second protolysis of SA is too small (< 13) to be detected in the pH range covered in this study and that of GE (protolysis of the phenolic group) was not found in the literature.[48,49]

Decomposition of the three component mixture

In addition to the data set, PARAFAC requires the definition of the number of factors for the model. The analysis of data sets corresponding to one and mixtures of two metabolites showed that only one factor is sufficient to describe the trilinear variation of each metabolite. Following this rule, the number of factors of the decomposition model for the three component mixture was set at three.

After setting a first approximation for the pH profiles, spectra and relative concentration for the three factors constituted by random numbers, PARAFAC convergence was achieved after about 500 iterations when the loss function became less than 1×10^{-8}. The calculated quantities of the three orders are shown in Figs. 2–3 and Table 2.

The calculated spectra are shown in Fig. 2 and, considering the information obtained from experiments with only an individual metabolite which are summarised in Table 1, these spectra match exactly those of the pure metabolites. Also, the analysis of Fig. 2 stresses the high overlap of the three pure SyF spectra.

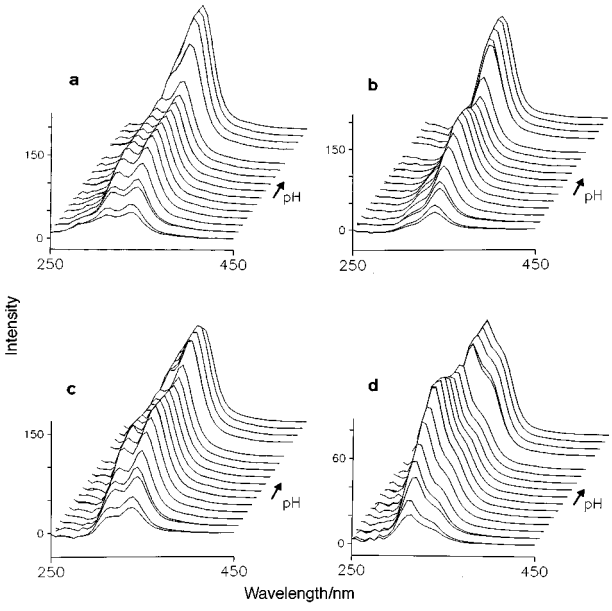

Fig. 1 Typical synchronous spectral data sets of mixtures of salicylic, gentisic and salicyluric acid with the following respective concentrations: (a) 22, 22 and 0.9 μM; (b) 22, 6.3 and 2.3 μM; (c) 22, 22 and 2.3 μM; and (d) 16, 3.2 and 0.5 μM.

Fig. 3 shows the pH profiles, *i.e.*, the SyF intensity variations due to each metabolite as the pH is changed. Examination of this figure shows the following: (i) the profile due to SU has a zero intensity value before about pH 7, and increases with an S-shaped curve until the end of the titration; (ii) the profiles due to the other two acids, SA and GE, never reach a zero intensity value; (iii) the profiles due to SA and GE have a similar shape until about pH 5, *i.e.*, an increase in intensity is observed; and (iv) the profile due to GE shows a decreasing trend beginning at about pH 8, but that due to SA remains approximately constant. These variations are a direct consequence of two factors. (a) Individual SyF characteristics of the different acid–base species of each acid. The following species have a SyF spectrum with the instrumentation settings used in this work and described above in the Instrumentation section: (i) basic species of SU (the acid and amphiprotic species are not fluorescent); and (ii) amphiprotic species of SA and GE (the acid and basic species are not fluorescent). (b) Acid–base characteristics of the three metabolites. The following properties can be obtained from the analysis of Fig. 3: (i) the second protolysis of SU is responsible for the S-shaped curve of the respective pH profile; (ii) the increase in intensity of the SA and GE profiles are due to the first protolysis; this variation does not begin at zero intensity because the acid species are already partially deprotonated at about pH 2; and (iii) the decrease in intensity of the GE pH profile corresponds to the second protolysis; it does not reach a zero intensity value because the deprotonation is not completed at pH ≈ 10. Because the observed variations of the SyF spectra variations are due to the protolysis reactions of the fluorescent species a detailed analysis of Fig. 3, *i.e.*, the calculation of the pH value at 50% of the maximum high for the three profiles, allows an estimation of the respective pK_a values, which are the following (the error of the graphical estimates is about 0.1 pK_a units): SA, pK_a = 2.9; GE, pK_{a1} = 2.9 and pK_{a2} = 10.2; and SU, pK_a = 8.2. These pK_a values are similar to others obtained from the literature and shown in Table 1.

In order to assess the quality of the calculated relative concentrations of the three metabolites, the plots of the experimental *versus* the predicted concentrations, both sets after scaling between 0 and 1, were obtained and the respective linear regression parameters are given in Table 2. The analysis of these

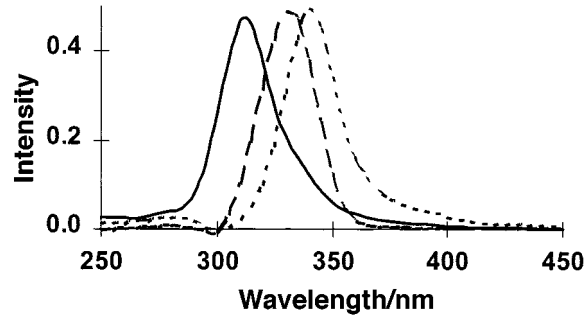

Fig. 2 Calculated SyF spectra of the three metabolites: ——, salicylic acid; ----, salyciluric acid; and ·····, gentisic acid.

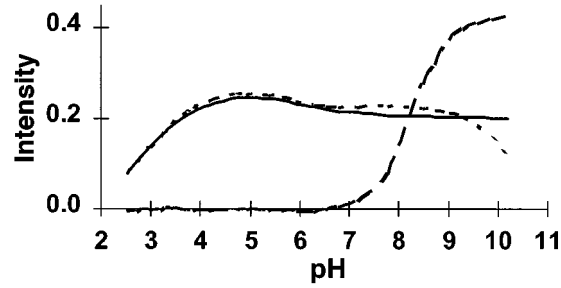

Fig. 3 Calculated pH profiles for the three metabolites: ——, salicylic acid; ----, salicyluric acid; and ·····, gentisic acid.

Table 2 Linear regression parameters for the plots of the real *versus* predicted concentrations (both sets of values scaled between 0 and 1).

		Confidence interval (95%)	
Metabolite	r^a	Intercept	Slope
SA	0.9939	[−0.004; 0.081]	[0.882; 1.030]
SU	0.9929	[−0.052; 0.051]	[0.878; 1.040]
GE	0.9972	[−0.024; 0.037]	[0.872; 0.973]

[a] Correlation coefficient.

regression parameters, and particularly the observation that the intercept is about zero and the slope about unity, show that the predicted concentrations are very good relative estimates of the experimental values.

Conclusion

It has been demonstrated that SyF spectroscopy is suitable for generating three-way data structures, following a trilinear model, suitable for the development of robust analytical methodologies for the determination of the three major acetylsalicylic acid metabolites in the micromolar concentration range. A full trilinear decomposition by PARAFAC of the experimental data matrices generated by the acquisition of SyF spectra as a function of the pH and different relative concentrations of the three metabolites was achieved.

Acknowledgments

Project grant PBIC C QUI 2177 95 from JNICT (Lisbon) is acknowledged.

References

1 G. G. Guilbault, in *Comprehensive Analytical Chemistry*, ed. G. Svehla, Elsevier, Amsterdam, 1977, vol. VII, ch. II.
2 J. B. F. Lloyd, *Nature (London)*, 1971, **232**, 64.
3 P. Jonh and I. Soutar, *Anal. Chem.*, 1976, **48**, 520.
4 S. G. Wakeman, *Environ. Sci. Technol.*, 1977, **11**, 272.
5 J. B. F. Lloyd, *Analyst*, 1980, **105**, 97.
6 T. Vo-Dinh, *Anal. Chem.*, 1978, **50**, 396.
7 T. Vo-Dinh and P. R. Martinez, *Anal. Chim. Acta*, 1981, **125**, 13.
8 T. Vo-Dinh, R. B. Gammage and P. R. Martinez, *Anal. Chem.*, 1981, **53**, 253.
9 E. L. Inman and J. K. Winefordner, *Anal. Chem.*, 1982, **54**, 2018.
10 T. Vo-Dinh, *Appl. Spectrosc.*, 1982, **36**, 576.
11 J. N. Miller, *Analyst*, 1984, **109**, 191.
12 I. M. Warner, G. Patonay and M. P. Thomas, *Anal. Chem.*, 1985, **57**, 463A.
13 T. T. Ndou and I. M. Warner, *Chem. Rev.*, 1991, **91**, 493.
14 C. L. Stevenson and T.Vo-Dinh, *Anal. Chim. Acta*, 1995, **303**, 247.
15 D. M. Hueber, C. L. Stevenson and T. Vo-Dinh, *Appl. Spectrosc.*, 1995, **49**, 1624.
16 J. C. G. Esteves da Silva and A. A. S. C. Machado, *Analyst*, 1995, **120**, 2553.
17 J. C. G. Esteves da Silva and A. A. S. C. Machado, *Analyst*, 1997, **122**, 1299.
18 J. B. Kruskal, Multilinear Models, in *Research Methods for Multimode Data Analysis*, ed. H. C. Law, C. W. Snider, Jr., J. A. Hattie and R. P. McDonald, Praeger, New York, 1984, pp. 36–62.
19 J. B. Kruskal, in *Multiway Data Analysis*, ed. R. Coppi and S. Bolasco, Elsevier, Amsterdam, 1989, pp. 7–18.
20 R. A. Harshman and M. E. Lundy, in *Research Methods for Multimode Data Analysis*, ed. H. C. Law, C. W. Snider, Jr., J. A. Hattie and R. P. McDonald, Praeger, New York, 1984, pp. 122–215.
21 J. K. Carrol and J. J. Chang, *Psychometrika*, 1970, **35**, 283.
22 C. J. Appellof and E. R. Davidson, *Anal. Chem.*, 1981, **53**, 2053.
23 M. D. Russel and M. Gouterman, *Spectrochim Acta, Part A*, 1988, **44**, 857.
24 M. D. Russel and M. Gouterman, *Spectrochim. Acta, Part A*, 1988, **44**, 863.
25 M. D. Russel, M. Gouterman and J. A van Zee, *Spectrochim. Acta, Part A*, 1988, **44**, 873.
26 D. S. Burdick, X. M. Tu, L. B. McGown and D. W. Millican, *J. Chemom.*, 1990, **4**, 15.
27 E. Sanchez and B. R. Kowalski, *J. Chemom.*, 1990, **4**, 29.
28 A. K. Smilde and D. A. Doornbos, *J. Chemom.*, 1991, **5**, 345.
29 A. K. Smilde, *Chemom. Intell. Lab. Syst.*, 1992, **15**, 143.
30 Y. Zeng and P. K. Hopke, *J. Chemom.*, 1992, **6**, 65.
31 S. Leurgans and R. T. Ross, *Stat. Sci.*, 1992, **7**, 289.
32 K. S. Booksh and B. R. Kowalski, *Anal. Chem.*, 1994, **66**, 782A.
33 K. S. Booksh, Z. Lin, Z. Wang and B. R. Kowalski, *Anal. Chem.*, 1994, **66**, 2561.
34 J. M. Henshaw, L. W. Burgess, K. S. Booksh and B. R. Kowalski, *Anal. Chem.*, 1994, **66**, 3328.
35 R. Tauler, A. K. Smilde, J. M. Henshaw, L. W. Burguess and B. R. Kowalski, *Anal. Chem.*, 1994, **66**, 3337.
36 A. K. Smilde, R. Tauler, J. M. Henshaw, L. W. Burguess and B. R. Kowalski, *Anal. Chem.*, 1994, **66**, 3345.
37 M. Gui, S. C. Rutan and A. Agbodjan, *Anal. Chem.*, 1995, **67**, 3293.
38 I. E. Bechmann, *Talanta*, 1997, **44**, 585.
39 P. Hindmarch, K. Kavianpour and R. G. Brereton, *Analyst*, 1997, **122**, 871.
40 F. Salinas, A. M. Peña, I. D. Merás and M. S. Durán, *Analyst*, 1990, **115**, 1007.
41 D. G. Konstantianos, and P. C. Ioannou, *Analyst*, 1992, **117**, 877.
42 H. Gampp, M. Maeder, C. J. Meyer and A. D. Zuberbuhler, *Talanta*, 1985, **32**, 1133.
43 H. Gampp, M. Maeder, C. J. Meyer and A. D. Zuberbuhler, *Talanta*, 1986, **33**, 943.
44 A. A. S. C. Machado and J. C. G. Esteves da Silva, *Chemom. Intell. Lab. Syst.*, 1993, **19**, 155.
45 J. C. G. Esteves da Silva, A. A. S. C. Machado and C. S. P. C. O. Silva, *Anal. Chim. Acta*, 1996, **318**, 365.
46 A. Albert and E. P. Serjeant, *The Determination of Ionization Constants*, Chapman and Hall, London, 1971, pp. 13–14.
47 M. T. S. Vasconcelos and A. A. S. C. Machado, *Talanta*, 1986, **33**, 919.
48 E. P. Serjeant and B. Dempsey, *Ionisation Constants of Organic Acids in Aqueous Solution*, Pergamon Press, Oxford, 1979.
49 E. B. Gonzalez, N. N. Daeid, K. B. Nlan and E. Farkas, *Polyhedron*, 1994, **13**, 1495.

Paper 8/02690D