J. CHEM. SOC. FARADAY TRANS., 1994, 90(23), 3521-3525

Determination of the NMR Monomer Shift and Dimerization Constant in a Self-associating System by Direct Application of the Least-squares Method

Henry K. S. Tan
Department of Applied Chemistry, National Chiao Tung University, Hsinchu, Taiwan, ROC

The monomer shift and dimerization constant of a self-associating system have been determined directly by applying the least-squares method to the NMR experimental data. Equations have been obtained in which two analytic methods can be applied. In the first method, a value of the dimerization constant is assumed and by linear regression the error sum of squares (ESS) is calculated. The minimum ESS is located by a five-point searching algorithm. In the second method, non-linear regression analysis of the model equation with the least-squares method results in a single equation of unknown K. A conventional root finding method is then applied from which the values of monomer shift and dimerization constant are determined. The numerical procedures proposed in this paper have been illustrated with an example using previously published experimental NMR data. Results of calculations are discussed.

NMR experiments have been widely used in the study of hydrogen bonding, kinetics and equilibrium behaviour of a variety of systems. The determination of the monomer shift and dimerization constant of a self-associating system has been investigated by Purcell *et al.*,[1] Chen and Shirts[2] and Chen and Rosenberger.[3] As pointed out by these authors the determination of the monomer shift by extrapolating experimental data to infinite dilution is not very reliable. Therefore, Purcell *et al.* applied a direct search procedure, Chen and co-workers proposed both iterative and graphical methods for more accurate determination of the monomer shift and dimerization constant. In this paper we first briefly review the equations and methods used by Chen and co-workers in their studies.[2,3] We then present alternative methods which are based on the direct application of the least-squares method for the curve fitting of non-linear equations. We also derive the necessary expressions for the direct determination of monomer shift and dimer dissociation constant from the experimental data obtained. An example, using previously published experimental results, is used to illustrate the methods proposed in this paper. In addition, we also include in this work an analytic way of determining monomer shift and dimer constant which is based on Chen and Rosenberger's graphical procedure.

Previous Methods used in the Treatment of Experimental Data

The observed chemical shift, δ, for a monomer–dimer equilibrium system can be expressed as the weighted average of monomer and dimer shifts[2–4] with the form of

$$\delta = \frac{[\mathrm{A}]}{[\mathrm{A}]_0}\,\delta_\mathrm{m} + \frac{[\mathrm{A}]_0 - [\mathrm{A}]}{[\mathrm{A}]_0}\,\delta_\mathrm{d} \qquad (1)$$

where $[\mathrm{A}]$ is the equilibrium monomer concentration, $[\mathrm{A}]_0$ is the total concentration, while δ_m and δ_d are the monomer and dimer shifts, respectively.

The equilibrium constant, K, for the self-association of A to the dimer A_2 is defined by

$$K = \frac{[\mathrm{A}]_0 - [\mathrm{A}]}{2[\mathrm{A}]^2} \qquad (2)$$

Combining eqn. (1) and (2) and after simplification, the following equation is obtained[2,3,5]

$$\delta = \delta_\mathrm{d} - \left(\frac{\delta_\mathrm{d} - \delta_\mathrm{m}}{2K}\right)^{1/2}\left(\frac{\delta - \delta_\mathrm{m}}{[\mathrm{A}]_0}\right)^{1/2} \qquad (3)$$

This form of equation has been used by several authors[6,7] to determine δ_m and K with the value of δ_m obtained from extrapolation at dilute concentration. As shown by Chen and Shirts, eqn. (3) can also be transformed to

$$\delta = \delta_\mathrm{m} + (\delta_\mathrm{d} - \delta_\mathrm{m}) \times \{\surd(1 + 8K[\mathrm{A}]_0) - 1\}/\{\surd(1 + 8K[\mathrm{A}]_0) + 1\} \qquad (4)$$

Note that eqn. (4) is an explicit form for δ in contrast to the implicit form given in eqn. (3).

The iterative method proposed by Chen and Shirts starts with an assumed value of δ_m and by applying a quadratic polynomial regression to eqn. (3). The intercept and the limiting slope (corresponding to the first and second coefficient, respectively, in the quadratic polynomial expression) provide an estimate of the value of K. This estimate of K is then substituted into eqn. (4) and again a quadratic polynomial regression is carried out for the determination of a more accurate estimate of δ_m. The process continues until the convergence in δ_m to the desired accuracy is achieved. Although this is a very useful method for determining monomer shift, in view of the inherent iterative calculations required, the use of some other methods may be warranted. Therefore, Chen and Rosenberger proposed a graphical approach for dealing with eqn. (3) or (4). The main argument in this graphical method is that if the correct value of δ_m is located then linear or quadratic polynomial regression with respect to either eqn. (3) or (4) will lead to an identical quality of curve fitting. In other words, the coefficient of the quadratic term of the polynomial equation approaches zero. Using eqn. (3) and by assuming various values of δ_m, a series of δ_m *vs.* K values are generated from both linear and quadratic regressional analysis. Similarly, using eqn. (4) and assumed values of K, a series of K *vs.* δ_m values can also be obtained with both linear and quadratic regressional analysis. The intersection of K *vs.* δ_m plots from both linear and quadratic treatment yields the correct values of δ_m and K. This graphical method is very efficient and is an improvement of the iterative scheme used previously by Chen and Shirts. Applying this graphical procedure together with the final numerical refinement can lead to a precise determination of values of monomer shift and dimerization constant within the accuracy of the measured spectroscopic data.

Proposed Methods

To facilitate the derivation of a general expression for directly determining the dimerization equilibrium constant and monomer shift we convert eqn. (3) to the following form

$$\delta = \delta_d + (\delta_m - \delta_d)\{\sqrt{(1 + 8K[A]_0)} - 1\}/(4K[A]_0) \quad (5)$$

Let a denote δ_d, b denote $(\delta_m - \delta_d)$ and let the measured δ be denoted by Y, and the concentration $[A]_0$ be denoted by X. The error sum of squares from a set values of δ_d, δ_m and K is

$$\text{ESS} = \sum_{i=1}^{N} (Y_i - \{a + b[\sqrt{(1 + 8KX_i)} - 1]/(4KX_i)\})^2 \quad (6)$$

where N is the number of data points.

It will be assumed that the three parameters of δ_d, δ_m and K are to be determined from eqn. (6) when the ESS is a minimum. By partial differentiation of eqn. (6) with respect to a, b and K, and equating the derivatives to zero, we obtain

$$a = \frac{\sum Y_i \sum (G_i)^2 - \sum G_i \sum G_i Y_i}{N \sum (G_i)^2 - \sum G_i \sum G_i} \quad (7)$$

$$b = \frac{N \sum G_i Y_i - \sum G_i \sum Y_i}{N \sum (G_i)^2 - \sum G_i \sum G_i} \quad (8)$$

where

$$G_i = [\sqrt{(1 + 8KX_i)} - 1]/(4KX_i) \quad (9)$$

The value of K corresponding to the minimum ESS can then be determined from the following equation

$$\begin{aligned} f(K) = a \sum \left[\frac{1}{\sqrt{(1 + 8KX_i)}} - G_i\right] \\ + b \sum \left[\frac{1}{\sqrt{(1 + 8KX_i)}} - G_i\right] \\ - \sum \frac{Y_i}{\sqrt{(1 + 8KX_i)}} + \sum G_i Y_i = 0 \end{aligned} \quad (10)$$

Note that in eqn. (10), a, b and G are all functions of K and are defined in eqn. (7)–(9), respectively.

Two methods can be used for determining the values of K, a and b which give the minimum ESS. The first method is by relying on eqn. (6)–(9). The second method corresponds to the determination of K which satisfies eqn. (10). In the first method, a value of K is assumed, G_i is calculated from eqn. (9), and a and b are then calculated from eqn. (7) and (8). Upon substituting the calculated values of a, b and the assumed value of K into eqn. (6), the corresponding ESS is obtained. The process can then be repeated with other assumed values of K. The minimum value of the ESS is located by successive iteration. One systematic way of locating this minimum value of the ESS is by the following procedures. If the minimum ESS is known to lie between K_1 and K_5, then if K_3 denotes the midpoint between K_1 and K_5 the calculated values of ESS are such that for $K_1 < K_3 < K_5$, ESS(1) > ESS(3) < ESS(5).

Corresponding values of the ESS at K_2 and K_4 are then calculated, where K_2 and K_4 are the midpoints between K_1 and K_3 and between K_3 and K_5, respectively. There are three possible outcomes after the calculation of ESS(2) and ESS(4): (i) ESS(2) > ESS(3) < ESS(4), (ii) ESS(1) > ESS(2) < ESS(3), (iii) ESS(3) > ESS(4) < ESS(5). For case (i), the old ESS(2) and ESS(4) become the new ESS(1) and ESS(5) in the next stage of iterative calculation. For cases (ii) and (iii), the old ESS(1), ESS(2), ESS(3), or ESS(3), ESS(4), ESS(5) becomes the new ESS(1), ESS(3), ESS(5) in the next iteration. In this manner the interval between K_1 and K_5 is reduced by half in each cycle of repetitive calculation.

The second method is in essence the determination of the real and physical meaningful value of a root that satisfies a non-linear algebraic equation. In this case, Newton–Raphson, secant, bisection or some other conventional numerical methods for root finding can be applied. The secant method is preferred if the derivative expression is too complicated to evaluate with the Newton–Raphson method.

In this paper, our primary aim is to suggest alternative methods to the previous ones used by Chen and Shirt[2] and Chen and Rosenberger.[3] However, we must point out that the graphical method proposed by Chen and Rosenberger is a very ingenious way of effectively determining K and δ_m without resorting to a complete regressional analysis which leads to the derivation of eqn. (10). Utilizing Chen and Rosenberger's idea we formulate an analytic expression which can aid us in the determination of δ_m and K. In doing so, plots of K *vs.* δ_m from both linear and quadratic regression can be omitted. The analytic expression is obtained based on the fact that when the correct K or δ_m is located, the quadratic term coefficient for the quadratic polynomial equation should approach zero. This criterion differs from that of Chen and Rosenberger, since in their method the quadratic term does not have to approach zero as long as the value of δ_m or K from both linear and quadratic regression converges.

Under the criterion that the quadratic term coefficient should equal zero when the correct δ_m or K is located, we have, after solving for the quadratic term coefficient, C, from quadratic polynomial regression by the method of determinants,

$$\begin{aligned} C = f(Z_i) = [\sum Z_i \sum Z_i^3 - \sum (Z_i)^2] \sum Y_i \\ + [\sum Z_i \sum Z_i^2 - N \sum Z_i^3] \sum Z_i Y_i \\ + [N \sum Z_i^2 - (\sum Z_i)^2] \sum Y_i Z_i^2 = 0 \end{aligned} \quad (11)$$

where

$$Z_i = f(\delta_m) = \left(\frac{Y_i - \delta_m}{X_i}\right)^{1/2}$$

if eqn. (3) is used,

$$Z_i = f(K) = \frac{\sqrt{(1 + 8KX_i)} - 1}{\sqrt{(1 + 8KX_i)} + 1}$$

if eqn. (4) is used. The validity of eqn. (11) is, of course, subject to the condition that

$$\begin{aligned} N \sum Z_i^2 \sum Z_i^4 + 2 \sum Z_i \sum Z_i^2 \sum Z_i^3 \\ \neq (\sum Z_i^2)^3 + N (\sum Z_i^3)^2 + \sum Z_i^4 (\sum Z_i)^2 \end{aligned}$$

As in the case of eqn. (10), any standard numerical methods for root finding can be applied in eqn. (11) to solve for K or δ_m. Again the secant method which avoids the need for obtaining the derivative expression is more convenient to apply than if the Newton–Raphson method is used.

Note that in eqn. (10), the root, K, obtained corresponds to the minimum ESS attainable. On the other hand, eqn. (11) indicates that both linear and quadratic polynomial regression lead to the convergence of a set of values of K, δ_m and δ_d. It should also be pointed out that if the implicit form of eqn. (3) is used, the values of K, δ_m and δ_d and the corresponding ESS will be different from those determined from using the explicit form of eqn. (4).

As pointed out by Chen and Shirts,[2] not all the self-associating systems are so simple that their equilibrium relationship can be easily depicted by eqn. (2). They have made references to previous studies[8,9] and believe that for fairly dilute solutions and for systems such as hindered phenols and lactams, the monomer–dimer equilibrium is dominant because of geometric preferences. Thus for these systems, the methods proposed in this paper may be useful. In carrying out the numerical treatment of the experimental data as outlined in this work, we must make the prerequisite assumptions that the value of K is unique and that the criteria of $\delta_d > \delta > \delta_m$ or $\delta_m > \delta > \delta_d$ must be met.

Results and Discussion

To illustrate the use of the proposed method we employ the experimental NMR data for solutions of δ-valerolactam in $CDCl_3$ as reported by Purcell *et al.*[1] These spectral data were obtained with a Varian DA-60 IL spectrometer and with the use of tetramethylsilane as internal reference. Table 1 gives the original data in which the solute concentration is expressed in terms of mole fraction. Following the work of Chen and Shirts,[2] we have also converted the concentration to mol kg^{-1} and the resulting conversions are also given in Table 1. To carry out linear regression analysis, the value of K is assumed so that a and b can be calculated from eqn. (7) and (8). From the calculated values of a and b the ESS corresponding to the assumed value of K is then determined from eqn. (6). With a series of assumed values of K, a plot of ESS *vs.* K can be obtained and is shown in Fig. 1. As can be seen from Fig. 1, the presence of a minimum ESS is evident although the exact location cannot be easily determined by visual inspection. The ESS as shown in Fig. 1 is fairly constant between K values of 5.2 and 5.6 kg mol^{-1}. By applying partial regression analysis to eqn. (3), that is assuming a value of δ_m and carrying out the linear regression calculation, plots of ESS *vs.* δ_m and ESS *vs.* δ_d were also obtained as shown in Fig. 2 and 3. Comparing Fig. 2 and 3, it can be seen that for the range of ESS between 220 and 250 Hz^2 the dimer shift does not vary as much as the monomer shift. This is in agreement with Purcell *et al.*'s finding, which used the direct search procedure,[10] that the calculated dimer shift is nearly independent of the preset monomer shift. We should also point out that the results of Fig. 1 and 2 do not give the same value for the minimum ESS. This is to be expected since two different forms of the equations, one explicit the other implicit, are used. The explicit type of equation is to be preferred if transformation to a linear form can be accomplished. Graphical output, as shown in Fig. 1, indicates the region in which the minimum ESS and the corresponding K are situated. The exact location, however, must be determined by analytical means. Applying eqn. (6)–(8) and employing the five-point searching algorithm outlined in the previous section, the value of K corresponding to the minimum ESS is calculated. Table 2 gives the iterative calculated results starting with three different initial K ranges. The first is for K between 3 and 7, the second for K between 1 and 9 and the third for K between 2 and 6. In the same table the corresponding values of δ_m and δ_d at K_3 for each iterative calculation are also

Table 1 Experimental NMR data of valerolactam in $CDCl_3$, at 306 K[a]

mole fraction	concentration /mol kg^{-1}	δ/Hz[b]
0.010	0.0839	382.0
0.020	0.1695	408.8
0.030	0.2569	420.7
0.041	0.3551	426.6
0.052	0.4556	442.7
0.061	0.5396	451.0
0.073	0.6541	444.8
0.083	0.7519	453.3
0.091	0.8316	459.5
0.097	0.8923	458.4
0.150	1.4659	469.6
0.203	2.1157	476.9
0.255	2.8432	484.6
0.300	3.5600	484.7
0.359	4.6522	488.3
0.408	5.7248	491.7
0.449	6.7689	491.9
0.490	7.9809	494.3

[a] From ref. 1. [b] From TMS.

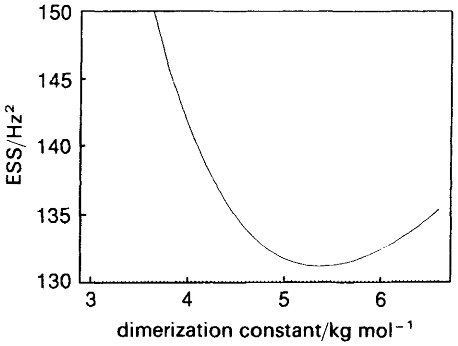

Fig. 1 ESS *vs.* dimerization constant

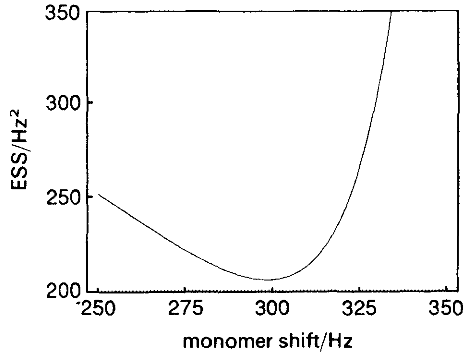

Fig. 2 ESS *vs.* monomer shift

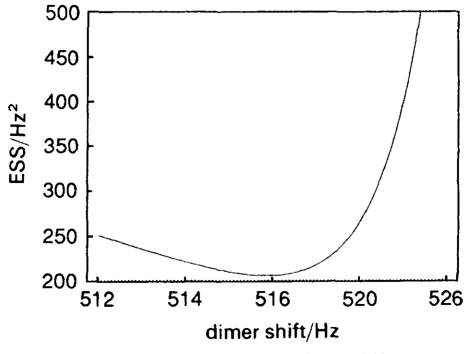

Fig. 3 ESS *vs.* dimer shift

Table 2 Calculated results using a five-point searching algorithm for locating the minimum ESS

iteration no.	K_1/kg mol^{-1}	ESS(1)/Hz^2	K_3/kg mol^{-1}	ESS(3)/Hz^2	K_5/kg mol^{-1}	ESS(5)/Hz^2	δ_{m3}/Hz	δ_{d3}/Hz
1	3.000	176.485	5.000	131.718	7.000	137.828	308.99	517.31
2	4.000	141.557	5.000	131.718	6.000	132.398	308.99	517.31
3	5.000	131.718	5.500	131.214	6.000	132.398	303.03	516.61
4	5.000	131.718	5.250	131.212	5.500	131.214	305.98	516.94
5	5.250	131.212	5.375	131.155	5.500	131.214	304.50	516.77
6	5.313	131.168	5.375	131.155	5.438	131.171	304.50	516.77
7	5.344	131.158	5.375	131.155	5.406	131.160	304.50	516.77
8	5.359	131.156	5.375	131.155	5.391	131.157	304.50	516.77
1	1.000	661.437	5.000	131.718	9.000	153.942	308.99	517.31
2	3.000	176.485	5.000	131.718	7.000	137.828	308.99	517.31
3	4.000	141.557	5.000	131.718	6.000	132.398	308.99	517.31
4	5.000	131.718	5.500	131.214	6.000	132.398	303.03	516.61
5	5.000	131.718	5.250	131.212	5.500	131.214	305.98	516.94
6	5.250	131.212	5.375	131.155	5.500	131.214	304.50	516.77
7	5.313	131.168	5.375	131.155	5.438	131.171	304.50	516.77
1	2.000	281.500	4.000	141.557	6.000	132.398	321.92	519.08
2	4.000	141.557	5.000	131.718	6.000	132.398	308.99	517.31
3	5.000	131.718	5.500	131.214	6.000	132.398	303.03	516.61
4	5.000	131.718	5.250	131.212	5.500	131.214	305.98	516.94
5	5.250	131.212	5.375	131.155	5.500	131.214	304.50	516.77
6	5.313	131.168	5.375	131.155	5.438	131.171	304.50	516.77

listed. Note that in these three sample calculations with different initial estimates of K values, convergence to minimum ESS is achieved within six to eight iterations.

In the above procedure, the determination of K is by using partial regression expressions. A complete regression analysis leads to eqn. (10) which can be used for the direct determination of K. To find the root of the non-linear equation $f(K) = 0$, the secant method is used. Although the Newton–Raphson method can also be applied, evaluation of the derivative expression for $f(K)$ can be complicated and tedious. By the use of the secant method the following expression is employed in the successive iteration

$$K_3 = K_2 - \frac{(K_2 - K_1)f(K_2)}{f(K_2) - f(K_1)}$$

where K_1 and K_2 are the two initial guesses of K to start the root-finding calculation. After each calculation, the value of K_3 determined becomes the new K_2 and the old K_2 is the new K_1 in the next iterative calculation. The process is repeated until K converges to the desired accuracy. Table 3 gives the calculated results with three different initial sets of K_1 and K_2 values. In all three cases, K is converged to three significant figures within five to seven iterations.

With either five-point minimum ESS searching or direct application of eqn. (10), the value of K is found to be 5.37 kg mol^{-1}. The corresponding δ_m and δ_d are calculated to be 304.50 and 516.77 Hz, respectively. The minimum ESS is determined to be 131.16 Hz.^{2}

Table 3 Convergence of the root of $f(K)$ using the secant method

initial values of K_1 and K_2					
K_1 3.00	K_2 4.00	K_1 7.00	K_2 8.00	K_1 1.00	K_2 2.00
K	$f(K)$	K	$f(K)$	K	$f(K)$
4.68414	0.077193	3.92623	0.20432	2.75318	0.56828
5.15510	0.021375	6.35708	−0.07247	3.67854	0.25986
5.33544	0.003424	5.72065	−0.02979	4.45818	0.10953
5.36984	0.000186	5.27647	0.00911	5.02612	0.03529
5.37182	0.000001	5.38050	−0.00081	5.29617	0.00719
5.37184	0.000000	5.37205	−0.00002	5.36527	0.00061
5.37184	0.000000	5.37183	0.00000	5.37171	0.00001
				5.37184	0.00000

By using eqn. (3) with various assumed values of δ_m and applying quadratic polynomial regression, a series of values for the quadratic term coefficient, C, *vs.* δ_m is obtained and these are plotted in Fig. 4. Similarly, using eqn. (4) with various assumed values of K and after applying the quadratic polynomial regression, a plot of quadratic term coefficient *vs.* K can be obtained, see Fig. 5. Analytically, the value of either K or δ_m can be obtained by solving eqn. (11) with a conven-

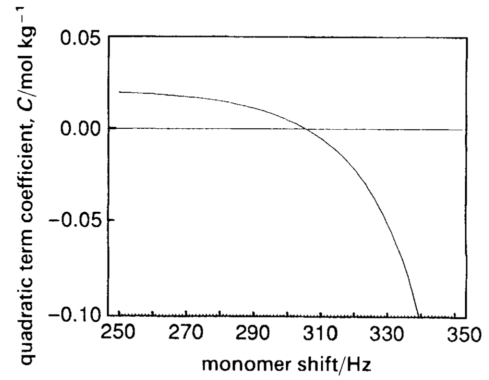

Fig. 4 Quadratic term coefficient *vs.* monomer shift using eqn. (3) and (11)

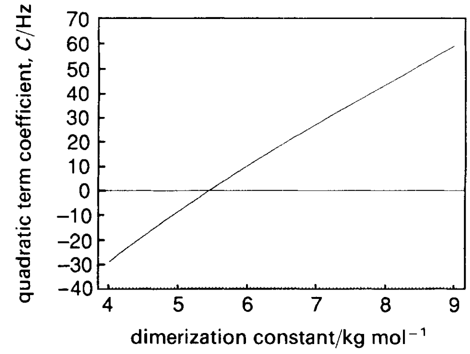

Fig. 5 Quadratic term coefficient *vs.* dimerization constant using eqn. (4) and (11)

tional root-finding method. For this example, the calculated K value is 5.44 kg mol^{-1} when the regression is applied to eqn. (4), and δ_m is calculated to be 305.32 Hz when regression is applied to eqn. (3).

Conclusions

In this study we have examined the previous methods used in the determination of monomer shift and dimerization constant for self-associating systems. We have derived a general expression from which the values of K, δ_m and δ_d corresponding to the minimum ESS can be determined. By the direct application of a least-squares method, the determination of K reduces to the problem of root finding for a non-linear algebraic equation. In addition, we have introduced in this paper an analytic method which modifies a previous graphical procedure for the determination of monomer shift, δ_m. In contrast to the previous iterative method of Chen and Shirts[2] and the graphical method employed by Chen and Rosenberger,[3] we have applied direct minimization of the ESS in the determination of the dimerization constant and chemical shifts for both the monomer and the dimer. We feel that the direct minimization of ESS is a better and more reasonable approach to data treatment. We have proposed an analytical method which is an efficient way of replacing the graphical procedure. Note that the numerical value read from graphs cannot be very accurate and numerical interpolation by analytic means is usually warranted. Therefore, analytical determination is often preferable to graphical means. We must emphasize that the methods proposed in this work may not be applicable for all self-associating systems. It is suggested that more experimental data should be used to examine the utility of the methods and the validity of the model equations used in the data treatment.

The author is indebted to Dr. L. G. Walker of Ryerson Polytechnic University, Toronto, Canada, for his help and assistance in the preparation of this article.

References

1 J. M. Purcell, H. Susi and J. R. Cavanaugh, *Can. J. Chem.*, 1969, **47**, 3655.
2 J-S. Chen and R. B. Shirts, *J. Phys. Chem.*, 1985, **89**, 1643.
3 J-S. Chen and F. Rosenberger, *Tetrahedron Lett.*, 1990, **31**, 1975.
4 H. S. Gutowsky and A. Saika, *J. Chem. Phys.*, 1953, **21**, 1688.
5 J-L. Dimicoli and C. Helene, *J. Am. Chem. Soc.*, 1973, **95**, 1036.
6 J. R. Barrio, F-T. Liu, G. E. Keyser, P. Van Der Lijn and N. J. Leonard, *J. Am. Chem. Soc.*, 1979, **101**, 1564.
7 M. Borzo, C. Detellier, P. Laszlo and A. Paris, *J. Am. Chem. Soc.*, 1980, **102**, 1124.
8 J. A. Walmseley, *J. Phys. Chem.*, 1978, **82**, 2031.
9 B. G. Somers and H. S. Gutowsky, *J. Am. Chem. Soc.*, 1963, **85**, 3065.
10 R. Hookes and J. A. Jeeves, *J. Assoc. Comput. Mach.*, 1961, **8**, 212.

Paper 4/02996H; *Received* 19*th May*, 1994