Towards stereochemical and conformational assignment in flexible molecules using NOEs and molecular modelling

Lech Kozerski,*[,a] Piotr Krajewski,[a] Krzysztof Pupek,[a] Paul G. Blackwell[b] and Michael P. Williamson*[,c]

[a] *Institute of Organic Chemistry, Polish Academy of Sciences, 00-961 Warszawa, Kasprzaka 44/52, Poland*
[b] *School of Mathematics and Statistics, University of Sheffield, Sheffield, UK S3 7RH*
[c] *Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, UK S10 2UH*

The title methods are used firstly to assign the relative stereochemistry of the four diastereomers of a 2,3-dihydrobenzofuran derivative, and secondly to define the preferred conformation of one diastereomer. The experimental constraints are obtained from steady-state NOE measurements and coupling constants, and are compared to calculated values using low-energy conformations calculated using the molecular mechanics program PCMODEL. The significance of the results is assessed by using a statistical analysis, based on an estimation of the magnitude of experimental errors, which produced the correct relative stereochemistries, showed that the experimental data are not fitted by any single conformation, and generated a preferred conformation which is very similar to that found in an X-ray structure. The method can give information on the relative stereochemistry of chiral centres several bonds apart.

Introduction

In this paper we propose a method for determining both the stereochemistry and the conformation of a flexible organic molecule in solution using NMR data. NMR data are inherently imprecise measures of structure. The relationship between the measured parameter [such as the NOE (nuclear Overhauser effect) or *J*-coupling] and conformation is not precisely defined. For the NOE, although the equations relating molecular geometry to observed NOE are well known,[1] they depend on a number of factors that are experimentally very hard to determine, such as correlation times and the rate of exchange between different conformations. Similarly, for *J*-couplings, the relationship between *J* and angle is usually defined by an equation of the form $J = A\cos^2\theta + B\cos\theta + C$ (the 'Karplus equation'[2]), although a relationship as simple as this does not adequately reproduce experimental values, and more sophisticated equations involving a greater number of adjustable parameters still have only moderate success.[3]

With this in mind, we propose to use statistical methods to judge the success of the conformational calculations. The use of statistical methods for structural calculations is not new. In general, these methods generate a set of conformers, often using molecular modelling, and then calculate the best fit between the set and the experimental data.[4–6] The approach adopted here is similar, and is essentially a linear least-squares fit, but handled in such a way that it (*a*) handles comparisons of large numbers of conformations in a statistically meaningful way, and (*b*) can handle ensembles of many conformations in a flexible and transparent manner.

The description of the conformation of a flexible molecule requires a very large number of parameters. In the presence of conformational averaging, which in most cases leads to an averaging of observed NMR parameters, NMR usually[7] does not provide enough data to define both conformations and populations.[8] All attempts at conformational analysis using NMR (including the assignment of stereochemical centres) must therefore use other additional techniques. Frequently this takes the form of molecular modelling. In small molecules with few degrees of freedom, it is often possible to calculate a complete list of all possible low energy conformations, either by a grid search or by molecular mechanics[4,9,10] or Monte Carlo methods.[11] In more complex cases, it is less easy to cover all possible conformations: either the description of a 'conformation' remains very vague,[12] or a limited number of conformations are produced by some constrained minimisation such as simulated annealing.[6,13]

Outline of the method

A key element of our approach is that we aim to use any pieces of information that may be available about the structures present in solution, weighted appropriately by our confidence in them. This is done by placing appropriate error margins on each piece of information, and carrying out a statistical analysis of the results. In the case of NOEs, the error value is estimated directly by making repeated measurements of the same NOE, while for *J*-couplings it is determined by the apparent error of fitting Karplus equations to known structures. We also use information from molecular modelling (*i.e.* the geometries of local minima and their enthalpies). Although such information is not directly experimental in the same sense as NOEs and *J*-couplings are experimental, it is originally based on experimental observations, since the force fields used are parametrised to give results that agree with experiment. Our treatment of error in the results from molecular modelling is based on the assumption that the geometries of local minima are likely to be reasonably good (for example, within 15° for each dihedral angle) but that the energies of these structures may not be very accurate. Energies are notoriously difficult to calculate, especially where electrostatics are concerned, and in any case, molecular mechanics generally calculates enthalpies rather than free energies. Therefore, we calculate all local minima that are likely to be significantly populated, which we have taken to mean anything within 3 kcal mol^{-1} of the lowest minimum, but thereafter ignore the predicted energies or derived populations. We then use the calculated conformations, and combinations of them, as an initial basis set from which to calculate expected NOEs and *J*-couplings for comparison to the experimental results. We use standard statistical methods to

compare the results and to determine, not just the single best structure, but all structures that are consistent with the data. The use of statistics in combination with the error values placed on the data allows us to state whether one solution is significantly better than all others, and thus to include all 'good' solutions.

The restriction of conformational space is done in stages. The first stage is to determine the relative stereochemistry of each chiral centre in the molecule, or alternatively to demonstrate that the data are not sufficient to determine the stereochemistry. To do this, we calculate all the low energy conformations for each possible relative set of chiralities. Following the outline method above, we then ask what combination of low energy conformations within each set of chiralities gives the best fit to the experimental data, and carry out a statistical analysis to determine whether one set of chiralities gives a significantly better fit than all the others. If it does, then we accept this set as the (most probable) relative chirality of the molecule. It is important to note the difference between this approach and the more conventional approach, in which individual conformers are either compared with the experimental results, or generated from experimental restraints, and the best fitting conformer is accepted as having the true relative stereochemistry. In this conventional approach, one's confidence in having the correct stereochemical assignment derives from the observation that all the best structures have the same stereochemistry.[6d] However, we stress (as discussed further below, in the section 'Determination of best conformational ensemble'), that if the molecule is flexible and is exchanging between several different conformers, such an approach may give an incorrect stereochemical assignment (*cf.* Fig. 3), and only by simultaneously considering an ensemble of structures with the same stereochemistry can one reliably distinguish the correct stereochemistry.

The analysis above will also have produced the 'best' set of low energy conformations. Typically, this set will use only a rather small number of the complete set of low energy conformations within the set of chiralities. We must then, as our second stage, use similar statistical methods to determine which conformations make a significant contribution to the conformational ensemble. In the case considered here, only two conformations are significant.

Finally, we seek to refine the structures produced. As noted above, we assume that the local geometry calculated by the molecular mechanics will be approximately correct, but not necessarily precisely correct. Therefore, we ask if there is a conformation close to the 'good' conformations that produces a significantly better fit to the experimental data than the original. In this way, we are taking into account the uncertainty or 'error' in the molecular mechanics calculations, and trying to get away from the circularity of seeking an answer only from amongst the small set of initially calculated conformers. If the best conformation or set of conformations also happens to be the lowest energy conformation(s), then our confidence in the molecular mechanics is further increased. If there is such a structure or set of structures, we accept this as the most likely solution.

Results

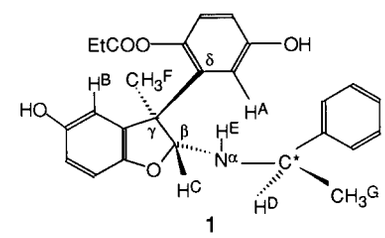
Conformational analysis

Reaction of *p*-quinone with the enaminone C_2H_5–CO–$C(CH_3)$=CH–NH–$CH(CH_3)C_6H_5$ produces a mixture of the four diastereomers **1**–**4**, from which the only isomer that can be purified and crystallised is **2**. The pairs **2** and **3**, **1** and **4** are in equilibrium (Scheme 1). This is evidenced by the observation of a transfer of saturation between the H^C signals in the pairs **1**–**4** and **2**–**3**, and also by the observation that purified **2,** when dissolved, forms an equilibrium mixture of **2** and **3** (for the equilibrium populations see Table 1). This process, similar to mutarotation of the anomeric position of sugars,[14] was first recognised in this class of compounds in our earlier study.[15] In ^{1}H NMR spectra at 500 MHz of mixtures of the four isomers, only protons H^C and H^D give resolved signals. In particular, the CH_3^F signals were coincident in [$^2H_5$]pyridine solution. Irradiation of this combined signal gave rise to the NOEs at the

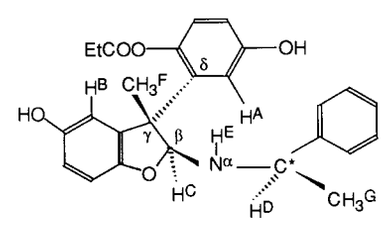
Table 1 Experimental NMR data for the diastereomeric mixture

	Compound			
	1	**2**	**3**	**4**
$^3J(H^C, NH^E)$/Hz	13.4	13.3	13.3	13.8[a]
$^3J(H^D, NH^E)$/Hz	2.9	2.2	2.9	1.0[a]
NOE $H^C\{CH_3^F\}$[b]	1.3	1.0	*c*	18.2
Configuration of C^β–C^γ bond	*trans*	*trans*	*cis*	*cis*
Equilibrium population (%) at temperature *T*/°C[d]				
	1	**2**	**3**	**4**
[$^2H_5$]Pyridine				
−30	67			33
−10	61			39
27	59	95	5	41
60	60			40
[$^2H_6$]DMSO				
27	65	97	3	35

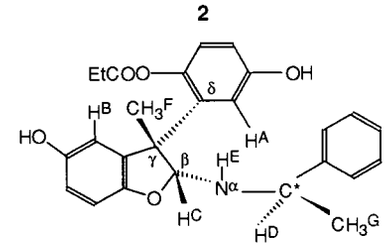
[a] [$^2H_5$]Pyridine. [b] Percentage NOE observed on H^C protons on saturation of overlapped CH_3^F signals from all diastereomers in [$^2H_5$]pyridine solution. [c] Not measured. [d] In each solvent, isomers **1** and **4** form an equilibrating pair, as do **2** and **3** (see Scheme 1).



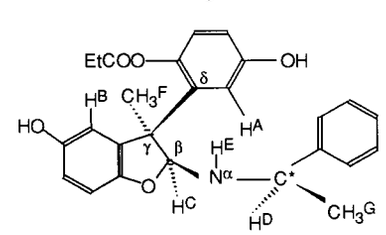
1



2



3



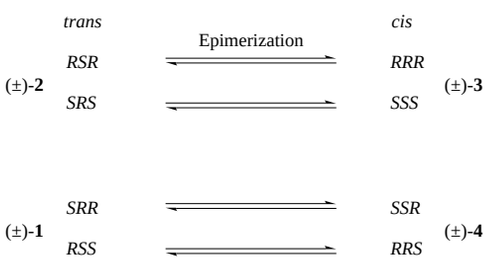
4

Structure of the 2,3-dihydrobenzofuran derivatives. The enantiomers produced when starting with pure *R* enaminone are shown.

Table 2 Conformational parameters of the 20 calculated conformers

Configuration $C^{\gamma}C^{\beta}C^{*}$	Configuration of N	Conformer no.	E/kcal mol^{-1}	Population (%[a])	Torsion angles/° $H^{C}–C^{\beta}–N–H^{E}$	$H^{D}–C–N–H^{E}$	$H^{D}–C–N–C^{\beta}$	J filter[b]	NOE rms-factor[c]
trans isomers									
RSR	*S*	1	44.94	94.7	−178.7	60.6	−60.3	+++	0.325
		2	47.13	2.2	−61.5	179.9	58.5	−−+	(0.943)
		3	47.48	1.2	177.2	−58.9	178.2	++−	(0.904)
		4	47.73	0.3	−177.6	−178.5	58.4	+−+	(0.302)
	R	5	47.63	0.4	−174.9	−60.4	62.7	+++	0.920
		6	47.54	0.8	61.9	60.21	−178.9	−+−	(0.842)
		7	47.69	0.4	59.5	176.1	−62.9	−−+	(0.933)
SRR	*S*	8	46.33	15.0	178.5	62.8	−59.9	+++	0.468
		9	46.59	9.6	−61.9	62.4	−57.5	−++	(0.395)
	R	10	45.40	74.0	63.2	179.9	−32.5	−−+	(0.434)
		11	47.70	1.4	179.9	64.2	−175.5	++−	(0.457)
cis isomers									
RRR	*S*	12	45.95	31.0	−61.05	62.7	−57.2	−++	(0.482)
		13	47.03	4.8	176.00	62.0	−60.58	+++	0.717
	R	14	45.73	45.2	179.80	178.8	−58.2	+−+	(0.523)
		15	46.26	18.2	62.6	−179.4	−57.5	−−+	(0.693)
		16	48.05	0.8	−179.7	−60.7	59.7	+++	1.055
SSR	*S*	17	44.68	71.8	179.7	60.5	−61.6	+++	0.714
	R	18	45.23	27.9	62.3	178.2	−60.2	−−+	(0.663)
		19	48.04	0.2	65.7	60.8	−178.7	−+−	(0.770)
		20	48.40	0.1	−177.1	−63.8	57.9	+++	0.805

[a] Calculated Boltzmann distribution for each diastereomer using the PCMODEL energies. [b] + Sign indicates acceptable angle, − sign indicates unacceptable angle; the three entries refer to the three previous angles respectively. The experimental ^{3}J values are given in Tables 1 and 5. [c] Calculated using average NOEs, as given in Table 3. Those conformations that fail the J filter are shown in parentheses.



Scheme 1 Configurational relationships between diastereomers. *cis*/*trans* Labels refer to the configuration about the $C^{\gamma}–C^{\beta}$ bond. The stereochemical assignments shown are those resulting from this work.

vicinal proton H^{C} listed in Table 1, showing clearly that **1** and **2** are *trans*, while **3** and **4** are *cis* at the $C^{\beta}C^{\gamma}$ carbons, or in other words that carbons C^{β} and C^{γ} have opposite chirality in **1** and **2**, but the same chirality in **3** and **4**. Thus, as shown in Scheme 1, the chiralities of all four diastereomers are readily deduced once the chirality of one of them is determined.

Calculation of low energy conformations

The program PCMODEL[16] was used to calculate low energy conformations of the dihydrobenzofuran with all possible relative stereochemistries. Only 20 are listed in Table 2, these being the lowest energy conformations for each stereochemical assignment. The nitrogen atom has tetrahedral geometry, and can be present either in the *R* or *S* configuration, though interconversion is expected to be rapid in solution.[17] We have therefore included both possible nitrogen configurations in each conformational family.

Steady-state NOEs

To determine appropriate parameters for NOE studies, ^{1}H and ^{13}C T_1 relaxation times were determined for **2** in degassed [$^{2}H_5$]-pyridine solution at 300 K. The longest ^{1}H T_1 time (for H^{B}) was 2.23 s, implying that an irradiation time of 15 s is long enough to produce steady state (>5T_1). ^{13}C NT_1 values were approximately constant throughout the molecule, showing that it tumbles isotropically, and indicated that the overall correlation time of **2** is about 0.3×10^{-10} s [*cf.* eqn. (2.19) of ref. 1].

Table 3 Experimental NOEs for isomer **2**[a]

	Observed						
Irradiated	H^{A}	H^{B}	H^{C}	H^{D}	NH^{E}	Me^{F}	Me^{G}
H^{A}	—	2.0	1.1	0.3	0.8	1.6	
H^{B}	0.8	—	0.3		0.4		
H^{C}			—	5.5	0.6	0.0	
H^{D}	−0.4		10.3	—	0.6	0.0	1.6
NH^{E}	0.0		2.9	6.2	—	1.3	1.5
Me^{F}	21.6	9.9	2.7	0.5	2.2	—	
Me^{G}	−0.4		0.1	12.6	1.5		—

[a] Empty entries have NOEs of less than 0.4% in all cases. The values given are the mean values for six experiments. Individual values varied in most cases by less than 1.0%; the standard deviation averaged over all NOEs is 0.31%. There is no difference in standard deviation between large and small NOE values.

Six independent sets of steady-state NOEs were recorded for **2**. Analysis of the NOEs showed that the assumption of constant variance of the error of each NOE is reasonable, and the errors are independent and are close to being normally distributed, as required by the statistical model used. Over the complete set, the mean standard deviation between independent measurements is 0.31%. Mean values of the NOEs are reported in Table 3.

Preliminary tests for best single structure

Before carrying out the test for the best stereochemical ensemble, we performed simpler preliminary tests on individual single structures. The root mean square (rms) NOE error between the calculated conformer and the experimental NOE results was calculated as shown in eqn. (1) for each of the 20 conformers, and is listed in Table 2. Only a few conformers (1, 4, 8, 9, 10, 11, 12) have sufficiently small rms error for them to be considered to be in adequate agreement with the experimental results. We also calculated the J values expected for the angles $H^{C}–C^{\beta}–N–H^{E}$, $H^{D}–C–N–H^{E}$ and $H^{D}–C–N–C^{\beta}$ for each conformer, and compared these to the experimental values. One can then use these comparisons as a 'filter' for acceptable con-

Table 4 Conformers giving the best fit for each possible stereochemistry

Stereochemistry	Conformers making up stereochemistry	Conformers involved in best fit	Proportions in best fit	Weighted residual sum of squares
RSR	1,2,3,4,5,6,7	1,4	0.47 0.53	194
SRR	8,9,10,11	8,9,11	0.23 0.59 0.18	399
RRR	12,13,14,15,16	12,14,16	0.74 0.19 0.07	1120
SSR	17,18,19,20	17,18,20	0.03 0.68 0.29	2103

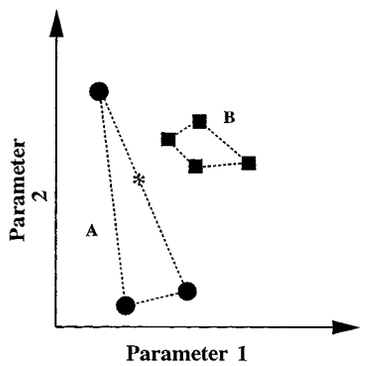


Fig. 1 Representation in two-dimensional space of the multi-dimensional parameter space defining a conformation. The Figure shows the parameters for model conformers in two relative stereochemistries: A (●) and B (■). The experimental result is shown by an asterisk. In this hypothetical representation, the experimental result is produced by an average of two of the A conformers. However, the closest single conformer to the experimental result is in the B ensemble. Therefore in this case, a simple choice of the stereochemistry that gives a conformer closest to the experimental result would yield the wrong stereochemical assignment.

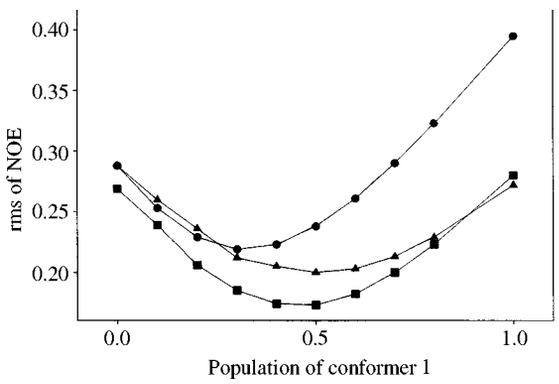


Fig. 2 Goodness of fit of the calculated NOEs to the experimental NOEs (as assessed by the rms value) as a function of the relative populations of conformers 1 and 4. Curves are shown for three different trial values of the external relaxation rate ρ*: (■) 0.1 s^{-1} (the optimum value); (●) 0.04 s^{-1} and (▲) 0.15 s^{-1} (which correspond to effective distances of relaxation sinks of 2.3, 2.7 and 2.2 Å respectively). A weighting of two on *J* errors was used.

formers, as shown in Table 2. Thus for example, the experimental coupling constant for the angle H^{C}–C^{β}–N–H^{E} is 13.3 Hz, which is consistent with an angle of around 180°, but not with angles of ±60°, for which *J* values of around 3 Hz are expected. Any conformer in which this torsion angle is not in the 180° staggered well is rejected by the *J* filter (symbol '−' in column 9 of Table 2). The combination of the requirement for a low rms error and passing the *J* filter leaves only conformers 1 and (less likely) 8 as acceptable, if the solution conformation is to be represented by only a single conformer.

Determination of 'best' conformational ensemble

As described in detail in the Experimental section, the above test for the best single conformer may produce an incorrect result for the overall stereochemistry if the solution conformation consists of an equilibrating mixture of two or more conformers which individually are quite different from the conformational average (Fig. 1). Therefore it is necessary to consider all possible combinations of the conformers within each stereochemical family (*RSR*, *SRR*, *RRR*, *SSR* at the three chiral centres $C^{\gamma}C^{\beta}C$* respectively). Thus, the problem is to find the set of weights *p* such that for the conformations *c* (*e.g.* the set of seven conformations with carbon chiralities *RSR*: see Table 2) the weighted set of conformations Σpc gives the best possible agreement between calculated and experimental parameters (NOEs and *J* couplings), subject to the constraints $\Sigma p = 1$, $p \geqslant 0$. The same must also be done for conformations $c_8 \cdots c_{11}$, $c_{12} \cdots c_{16}$, $c_{17} \cdots c_{20}$. The result of applying the iterative search described in the Experimental section is a single best-fitting mixture for each stereochemistry, shown in Table 4. By far the best fitting mixture comes from approximately equal parts of conformers 1 and 4, in the stereochemistry *RSR*. [A formal test of the hypothesis that the mixture of conformers 8, 9 and 11 is correct, against the hypothesis that the mixtures of conformers 1 and 4 is correct, overwhelmingly rejects the former with a *p*-value of less than 10^{-8} (see Experimental section)]. A standard *F*-test shows that the mixture fits significantly better than either 1 or 4 alone. When the relative weighting of NOE and *J* data was altered, the best fitting conformers were not affected, but the relative proportions were. Thus, giving extra weight to *J* couplings increased the proportion of conformer 1 in the best fit mixture, while giving extra weight to NOEs increased the proportion of conformer 4. A similar result was obtained when the value of the external relaxation rate ρ* was altered (Fig. 2): the overall goodness of the fit decreased, and the proportions of conformers 1 and 4 altered, but the conformers contributing to the best fit were not altered.

This analysis demonstrates with very high probability that the stereochemistry of diastereomer **2** is *RSR* at carbons $C^{\gamma}C^{\beta}C$* respectively. From the conformational analysis above, this implies that **1** is *SRR*, **3** is *RRR* and **4** is *SSR* (Scheme 1).

Minimisation

The analysis above makes it clear that, not only is the best stereochemistry *RSR*, but also the only conformers within this stereochemistry that need to be considered are 1 and 4, and that a mixture of the two is better than either alone. This is the limit of the information available using only NOEs and *J* values. However, the principle of this study is that each source of information should be used, weighted appropriately to our confidence in it. Thus, we note that the molecular mechanics calculations (Table 2) predict that conformer 1 will be of much lower energy than all other conformers within the stereochemistry *RSR*, and comprises roughly 95% of all conformers within this stereochemistry. The fact that conformer 1 is also one of the two required conformers increases our confidence in the molecular mechanics calculation, and suggests that there may be a single conformer within the cluster defined by conformer 1 that alone fits the experimental results at least as well as the combination of 1 and 4. We therefore carried out a refinement of conformer 1, minimising the difference between observed and calculated NOEs and *J* couplings. The result is conformer 1A, in which the four major torsional angles within the molecule have been rotated by 11.2° rms from conformer A (Table 5). On comparing this conformation to conformers 1 and 4, it fits better than either alone but not as well as the best

Table 5 Fit to experimental data of conformations 1 and 1A

Conformer	$H^C–C^\beta–N–H^E$/°	$^3J^a$/Hz	$H^E–N–C–H^D$/°	$^3J^a$/Hz	$H^D–C–N–C^\beta$/°	$^3J^b$/Hz	rms factor
1	−178.7	14.0	60.6	6.0	−60.3	2.0	0.325
1A	−183.3	14.5	72.7	3.6	−48.3	3.2	0.332
Experimental		13.3		1.9,[c] 2.2[d]		4.6[c]	

[a] Coupling constant calculated using eqn. (4) of ref. 3(*a*). [b] Coupling constant calculated using equation given in ref. 19. [c] [$^2H_6$]DMSO. [d] [$^2H_5$]Pyridine.

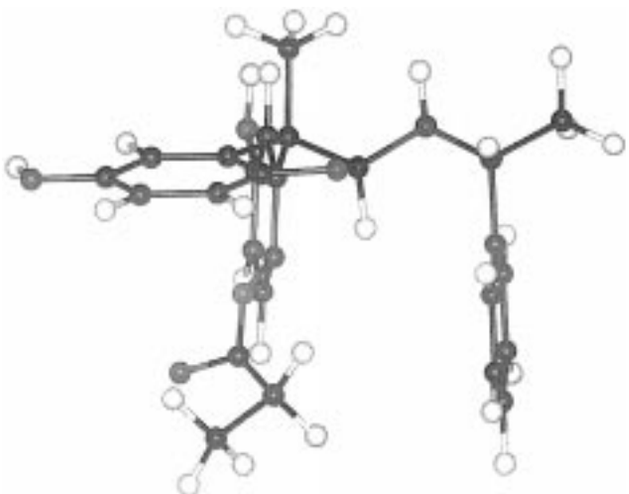


Fig. 3 The structure of the minimised best structure 1A

combination of 1 and 4. Conformer 4 differs from conformer 1 almost entirely in the torsion angle $H^E–N–C^*–H^D$, which is roughly 120° different. On minimisation of conformer 4, a conformer well outside the cluster defined by conformer 4 was always formed. Thus, there is a single conformer close to conformer 1 that alone can reproduce the experimental results reasonably well, while there is no single conformer near conformer 4 that can do so. Since the molecular mechanics calculations predict that conformer 1 is much lower in energy than conformer 4, we believe that the experimental results, together with the molecular mechanics calculations, imply that the most likely conformation in solution is dominated by the single conformational family shown in Fig. 3 centred around conformer 1A. However, other conformations are probably populated significantly; these may include conformations similar to conformer 4.

Discussion

The calculations presented here have allowed us to conclude that the conformations in solution are dominated by a conformer similar to conformer 1A, although no single structure can adequately fit all the data. The dihydrobenzofuran on which these studies have been made is a useful test case, because there is a crystal structure, obtained from a CH_2Cl_2–*n*-hexane mixture.[18] The most important conclusion to be obtained from comparing the crystal structure with conformer 1A is that the stereochemistry of both structures is *RSR*. Moreover, the stereochemistry at the tetrahedral nitrogen is also *S* in both structures. Thus, the method described here was able to reproduce the correct stereochemistry. On a more detailed comparison, the crystal structure is very similar to conformer 1A: the rms difference in the four major dihedral angles defined above is about 18°. Thus, there is good agreement between crystal and NMR structures.

The most important result presented here is that we are able to put confidence limits on the correctness of the determined chirality. This has been determined making no assumptions about the relative populations of the interconverting conformations within each chirality, although we have relied on the molecular mechanics to provide the structures of the conformations involved. Clearly, a poor set of initial conformers will result in a poorer result, which should manifest itself as a worse fit to the experimental data. From the analysis of conformers 1 and 4 presented above, it appears that the determination of the chirality is reasonably independent of the exact conformers used.

The relative importance of NOEs and coupling constants is determined by the variance associated with each measurement. For the NOEs, the variance comes directly from the experimental measurements, but for *J* couplings, the variance was derived from an estimate of the goodness of fit between measured values and the appropriate Karplus relationship, given in refs. 3 and 19 for $^3J_{HH}$ and $^3J_{CH}$ respectively. Such an estimate has some element of arbitrariness. Intuitively, one feels that *J* couplings should be reasonably well met, which implies that the importance of individual coupling constants is generally greater than that of individual NOEs. It is therefore necessary to give more weight to errors in *J* values than to errors in NOE values, particularly because there are many more NOE values measured than *J* values, and therefore the statistical procedure used here will automatically be dominated by the NOE fitting, unless steps are taken to increase the importance attached to *J* values. The refinement of conformer 1 to produce conformer 1A results mainly in a rotation of 12° around the C*–N bond, which has only a small effect on the NOEs, but reduces the error in the *J* value from 3.8 to 1.4 Hz.

It is worth pointing out that the statistical analysis carried out here gives the same result as the much simpler rms NOE error and *J* filter calculation listed in Table 2. This is a reassuring result, since it implies that the intuitively 'obvious' solution is correct, at least in this case. However, the two big advantages of the new approach presented here are that it can handle conformational ensembles in a much more satisfactory way, and that it puts certainty estimates on the results.

The statistical analysis used here deals only with random errors, and takes no account of systematic errors. These may arise from many sources, of which one is certainly incorrect values for NOEs resulting from incorrect values for external relaxation and the rotational correlation time. Preliminary studies suggest that this at least causes no major error. Sources of systematic error (particularly for *J* couplings) will generally lead to some calculated values being too low and some too high, in effect increasing the random error. Thus it may be that the estimates of the true error used here are too small. Experience with structure calculations of proteins suggests that the best way to combat systematic error or bias in the constraints is to increase the number of constraints rather than to tighten them: typical 'high resolution' NMR protein structures have roughly 3–4 constraints per rotatable bond.[20] It is clear that the addition of extra constraints to a structure calculation should serve only to further limit the conformational space, rather than introduce new allowed conformations. This will happen only if the initial constraints make proper allowance for systematic as well as random error, by increasing the estimate of the error associated with each constraint. Further studies are in progress.

This study has not addressed the question of the possible width of any conformational minimum, not least because such detail is beyond the limits of precision of current techniques. In effect, by using structures clustered in 15° groups, we have

assumed that structures within 15° of each other are indistinguishable. Conformational averaging within a minimum will tend to lead to a better fit of any calculated conformation to the experimental data, and requires a very detailed and accurate force field to allow any useful calculations. We therefore do not feel that it would be helpful to consider local averaging further.

The molecule studied here is well suited to our method, being limited in its conformational flexibility. In particular, calculations predict that it adopts essentially only one conformation in solution, which makes the testing of the calculation particularly straightforward. We are currently studying more difficult examples, where more than one conformation must be considered.

Experimental

Synthesis

A mixture of diastereomers was obtained from the reaction of *p*-quinone with the enaminone C_2H_5–CO–C(CH_3)=CH–NH–C*H(CH_3)C_6H_5 (which was used both as the pure enantiomer, with *R* configuration at C*, or as the racemate) in toluene, as described earlier.[15,18] Crystalline diastereomer **2** was separated out and the residual oil was purified by several recrystallisations from toluene.

Molecular modelling

The program PCMODEL[16] was used to calculate low energy conformations, using the force field MM3. Conformations were analysed using the program MMXCOMP[21] to produce a set of conformations within 3 kcal mol^{-1} of the minimum energy conformer. Placing a cut-off at 3 kcal mol^{-1} has been suggested to cover an acceptable range of conformations.[5] This program orders the calculated conformations in order of energy, and then clusters together conformations within a preset range of dihedral angles (in this case, within 15° of each other). It does this by setting the lowest energy conformer as a representative of the first cluster. The second lowest energy structure is added to the cluster if none of its dihedral angles differs by more than 15° from the first structure; if it does differ in at least one angle, it is taken as the representative of the second cluster. Each new structure is added in the same way.

NMR Experiments

Steady-state NOEs were performed at room temperature on a sample of (±)-**2** in [$^2H_5$]pyridine (17 mg ml^{-1}, 40 mM) on a Bruker AM-500 spectrometer. The sample was degassed to minimise external relaxation. A routine Bruker program was used for multiplet irradiation, using repetitive cycling round each line of a multiplet, and a total of 32 entries in the frequency list for each multiplet, with 128 scans per spectrum. Thus for example for a doublet, the frequency list contained ($[f_1 f_2]_{16}$), where f_1 and f_2 are the frequencies of the two lines of the doublet. The longest ^{1}H T_1 time (for H^B) was 2.23 s, implying that an irradiation time of 15 s is long enough to produce steady-state NOEs. The experimental conditions used were therefore 15 s total irradiation (made up of 15/32 = 0.47 s at each frequency in the list), 2 s acquisition, 5000 Hz spectral width, digital resolution 1 Hz. NOE intensities were calibrated by using a reference signal that was unaffected by the irradiation, and the same phase parameters and integration limits were used for reference and irradiated spectra. Experimental NOEs were obtained from six independent runs on the same sample. The mean standard deviation of the experimental NOE values was 0.31%, and the standard deviations were approximately normally distributed. There was no significant difference between the standard deviations for large and small NOEs. *J* Couplings were measured from 1D spectra. The J_{CH} coupling was measured from an undecoupled ^{13}C spectrum.

NMR Calculations

The theoretical NOEs were calculated using the programs NOE[21] and BUILDUP,[22] which are available on request from the authors. BUILDUP was used for scaling the theoretical NOEs to the experimental data. A best fit value for the external relaxation rate ρ^* of 0.1 s^{-1} was used, and the correlation time τ_c was set to 0.3×10^{-10} s on the basis of ^{13}C T_1 data. The program NOE is based on steady-state equations for multi-spin NOEs in the presence of external relaxation,[1] while BUILDUP uses numerical integration of the Solomon equations, and incorporates the Tropp equations[23] for methyl groups. As a quick and simple measure of the goodness of fit of calculated to observed NOEs, we used the rms factor, defined as eqn. (1) in a manner similar to its use in crystallography.[24]

$$\text{rms} = \{\Sigma(\text{NOE}_{\text{calc}} - \text{NOE}_{\text{obs}})^2/\Sigma(\text{NOE}_{\text{obs}})^2\}^{1/2} \quad (1)$$

J Couplings were calculated for ^{1}H–^{1}H couplings using the equations of Osawa *et al.*,[3] and for ^{1}H–^{13}C couplings using those of Wasylishen and Schaefer.[19] $^3J_{HH}$ is given by eqn. (2)

$$J = -1.3356 \cos\theta + 4.9649 \cos 2\theta + 1.0374 \cos\theta\, \Sigma(\Delta\chi_i \cos\varphi_i) - 0.2061\{(\omega_1 + \omega_2)/2 - 110\} + 5.8231 \quad (2)$$

where θ is the H–H dihedral angle, φ_i is the dihedral angle from the proton to an electronegative atom or group, ω_i is the bond angle of proton *i*, and $\Delta\chi_i$ is the difference in Mullay's group electronegativity between proton and atom, using $\chi_H = 2.08$, $\chi_C = 2.47$ (except $\chi_{Me} = 2.32$), $\chi_{OC} = -4.03$ and $\chi_{lone\ pair} = -8.13$. $^3J_{HC}$ is given by eqn. (3).

$$J = 3.56 \cos 2\theta - \cos\theta + 4.26 \quad (3)$$

Statistical methods

This study contains three separate statistical tests: (1) a test of which set(s) of stereochemistries best fits the NMR data, (2) a test of which conformations within the successful stereochemistry contribute significantly to the conformational ensemble as measured by NMR, and (3) a test of whether the 'minimised' best conformation resulting from test (2) is actually better than the starting conformation. Tests (1) and (2) can be done by essentially the same linear least squares (one-way analysis of variance) method, while test (3) is made using a standard *F*-test.

There are two aspects of this study that require more sophisticated treatment than normal, namely the requirements for multiple comparisons (*i.e.* simultaneous comparisons between a large number of possible solutions) and for comparisons of ensembles rather than single structures.

Multiple comparisons. These are not easy to handle, because, even though one may reject, at the 1% level say, the hypothesis that an event found in a single comparison arose from random variation, if a large number of such comparisons is made, the likelihood of such an event arising by chance rises roughly in proportion to the number of comparisons.[25] We use the result due to Scheffé quoted by Spjøtvoll and Stoline,[26] that intervals of the form given by eqn. (4) are simultaneous confidence inter-

$$\Sigma\, \lambda_i \hat{\mu}_i \pm \sqrt{s^2 k F_{\alpha,k,\nu} \Sigma\, \lambda_i^2}/n_i \quad (4)$$

vals for linear combinations of the μ_i values, where μ_i are the experimental values of conformational parameters (*e.g.* NOEs), λ_i are the differences in the value of the calculated parameters for the pair of conformers tested, s^2 is an estimate of the true variance σ^2, for example the mean square error from the analysis of variance, *k* is the number of pairs (*i.e.* the number of types of NOE measurement), n_i is the number of repetitions of measurements on pair *i*, ν is the degrees of freedom of the estimate s^2 of σ^2, with $\nu = \Sigma\, n_i - k$ in the analysis of variance case, and $F_{\alpha,k,\nu}$ is the upper α point for the *F* distribution with *k*

and ν degrees of freedom. That is, given a confidence coefficient $1 - \alpha$, in proportion $1 - \alpha$ of cases all these intervals will contain the true values of the appropriate linear combinations.

For the J measurements, essentially the same method can be used. The only differences are that the error term occurs in the calibration of the predicted values (rather than in the observation itself) and the error variance is known from past data (rather than estimated from repeated observations). NOE and J values can be combined together into the same fitting procedure by scaling the measurements so that they have equal variance. Since the H–C–N–C measurements are known to have a different variance from the other J measurements (*i.e.* C–H couplings have a smaller maximum range than H–H couplings), they were rescaled to satisfy the constant variance assumption. Note that this means that an error in an individual J value is treated the same as an error in an individual NOE value: in many cases, it may be more appropriate to handle these two errors differently, for example by placing more importance on an error in a J value. In such a situation, the method is readily adapted by weighting J values more heavily than NOEs.

The number of degrees of freedom ν when considering only the NOE values is simply equal to the number of NOE values. The value of ν when considering only the J values is determined differently, because the estimated standard deviation for the J values comes not from the data but from estimates of the errors in the Karplus curve. The standard deviation of values on the curve is judged to be 0.5 ± 0.25 Hz. This must be related to the usual form of error estimate, which consists of a point estimate plus the degrees of freedom of a χ^2 distribution representing the uncertainty in s^2. The point estimate of s^2 is just $(0.5)^2 = 0.25$. A χ^2 distribution for s^2 corresponds to a χ distribution for s,[27] and the degrees of freedom of the distribution can be determined by the coefficient of variation (defined to be the standard deviation divided by the mean). The coefficient of variation of a χ distribution with ν degrees of freedom is given by eqn. (5), where Γ is

$$\frac{\sqrt{2}\Gamma\{[(\nu + 1)/2]/\Gamma(\nu/2)\}}{\sqrt{\nu - 2\{\Gamma[(\nu + 1)/2]/\Gamma(\nu/2)\}^2}} \quad (5)$$

the standard mathematical function. This gives an estimated value of $\nu \approx 8$. The value of ν to choose for the final calculation of the Scheffé confidence intervals when using both NOE and J values is slightly problematical, since we are using an estimate of the variance to rescale the data. We have used the rather conservative choice of taking ν to be the degrees of freedom from the J values as well as the more realistic case of using the total degrees of freedom, with similar results.

Comparison of ensembles. The molecular modelling results in a number of low energy conformations for each of the four possible relative stereochemistries at the three chiral centres. We wish to compare the experimental data to the calculated conformations and test whether one stereochemistry is significantly better than the others. This is not a trivial exercise, for the following reason.

The simplest way to test for the best stereochemical assignment would be to test each conformation in turn and pick the stereochemistry represented by the best fit.[6d] However, it is conceivable that the molecule in solution exists as two or more conformations in fast exchange (on the chemical shift timescale), and that these conformations individually fit the experimental data badly, but in combination produce the experimental result (Fig. 1). In such a case, the simple test described above would produce the wrong answer. It is therefore necessary to test how well any combination of conformations from each chirality fits the data. In practice, this means that we must determine the best possible combination of conformers for each chirality (*i.e.* the combination that produces a set of calculated data closest to the experimental values), and then compare each combination to determine whether one fits significantly better than all others. This is a more conservative and reliable procedure than simply relying on the populations calculated from molecular mechanics.[10] This combination of conformers is essentially a linear combination, since J values are averaged in a linear manner, and the averaging of NOE values is surprisingly close to being linear, even when conformational exchange is fast compared to T_1, as it is here, when NOEs should actually be averaged as r^{-6}.[28] We may therefore think of this problem geometrically in conformation space as finding the best-fitting point in the convex hull of the points representing individual conformers. The convex hull is in reality a shape in a large number of dimensions, but is represented in two dimensions in Fig. 4. This fitting problem may be handled using the methods of linear regression, in a simple stepwise procedure.

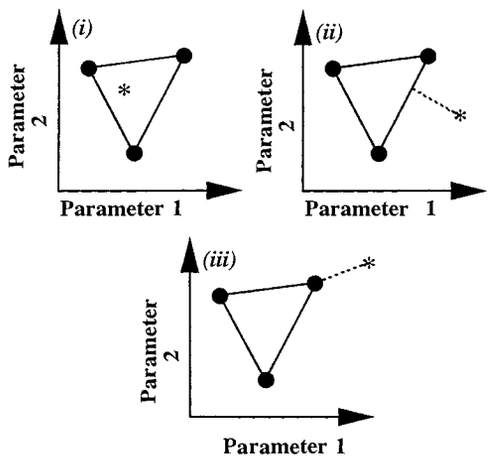


Fig. 4 Stepwise procedure for calculating the best set of contributing conformers to a stereochemical assignment (see text). The experimental result is marked by an asterisk, and the set of calculated parameters for the contributing conformers is shown as circles. Here we represent the multidimensional parameter space by three conformers in two-dimensional space. (*i*) If $P(S)$ consists only of positive populations p, the experimental result can be fully described by a linear combination of conformers, *i.e.* it is inside the convex hull defined by the conformers. If one or more of the p is negative, the experimental result must lie outside the convex hull. We therefore remove one element and calculate the fit again. If this results in both remaining p being positive then we have found the best fit, which is the linear combination of the two remaining conformers that is closest to the experimental result (*ii*). If one of the p is still negative, then the experimental point must be closest to one apex of the convex hull (*iii*).

Let S be the set of all conformers corresponding to a given stereochemistry. Write $P(S)$ for the best-fit set of estimates obtained on fitting the model with the set of conformers S. If $P(S)$ contains zero or positive fractional populations of each conformer, we immediately have the appropriate model, represented by a point in the interior of the convex hull [Fig. 4(*i*)]. If not, our point must be outside the hull, since at least one of the populations in the best fit $P(S)$ is negative. Thus, the model is not yet identified, and the solution must lie somewhere on the boundary of the convex hull. This necessarily implies that at least one of the conformer populations must be zero. We therefore consider each subset S' consisting of S with one element deleted, and for each S' such that $P(S')$ does not satisfy the constraints (*i.e.* all populations non-negative), further consider subsets S'' consisting of S' with one element deleted, and so on, until the model is satisfied, possibly going down to sets with just one element, which automatically satisfy the constraints. The mixtures thus defined represent mixtures which are locally best (*i.e.* better than any other mixture of the same subset of conformers). One of these, identified by the smallest residual sum of squares, will be the best model possible.

Statistically, the above procedure amounts to fitting the regression model, eqn. (6), where Y_{ij} is the jth repetition of

$$Y_{ij} = \sum_{l \in s} \rho_l \beta_i(l) + \varepsilon_{ij} \quad (6)$$

measurement type i, $\beta_i(l)$ is the prediction for measurement i for conformer l, and S is a set of conformers, for example the set of all conformers in a given stereochemistry, subject to the constraints [eqn. (7)] to ensure that a valid mixture is defined. To

$$\rho_l \geqslant 0, \sum_{l \in s} \rho_l = 1 \tag{7}$$

actually fit this model under these constraints, using standard methods, we first rewrite it as eqn. (8) where A is an arbitrarily

$$Y_{ij} = a_i + \sum_{l \in S,\, l \neq l_A} \rho_l [\beta_i(l) - a_i] + \varepsilon_{ij} \tag{8}$$

chosen conformer, indexed by $l_{\mathbf{A}} \in S$, a_i is the ith measurement for conformer A [so a_i is equivalent to $\beta_i(l_A)$], and we now have the constraints as eqn. (9). Fitting this model without con-

$$\rho_l \geqslant 0, \quad \sum_{l \in S,\, l \neq l_A} \rho_l \leqslant 1 \tag{9}$$

straints is now a straightforward problem in linear regression: fitting with constraints is done by the stepwise procedure outlined above.

Formal comparison of the results of the best fit ensembles for each stereochemistry amounts to testing for differences between two regression models that are not nested (*i.e.* one is not a restricted version of the other). We use the approach of Szroeter,[29] giving bounds for the *p*-value for testing one regression model against another; the results will be somewhat conservative (less likely to reject) in this case because of the constraints on the parameters in the regression model.

Minimisation

The final minimisation of conformer 1 to give conformer 1A was carried out by iteratively rotating each of the rotatable bonds to give the best fit between experimental and calculated data. The fitting procedure was based on the rms NOE and *J* errors, since these had been shown to behave very similarly to the linear regression method. The iteration was carried out in several different ways to check for consistency. Several different weightings of the rms NOE *vs.* rms *J* error were tried, with similar results in each case. The result reported here used a relative weighting of two on the *J* error. The significance of the improvement of 1A over 1 was assessed using a standard *F*-test.

Acknowledgements

This work is part of a research program sponsored by KBN grant no. 3T09A 071 09. We acknowledge the support of the British Council and the Royal Society for exchange visits (grants WAR/992/055 and 635054.P769 respectively).

References

1 D. Neuhaus and M. P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH, New York, 1989.
2 M. Karplus, *J. Chem. Phys.*, 1959, **30**, 11.
3 (*a*) K. Imai and E. Osawa, *Magn. Reson. Chem.*, 1990, **28**, 668; (*b*) E. Osawa, T. Ouchi, N. Saito, M. Yamato, O. S. Lee and M.-K. Seo, *Magn. Reson. Chem.*, 1992, **30**, 1104.
4 G. V. Nikiforovich, O. Prakash, C. A. Gehring and V. J. Hruby, *J. Am. Chem. Soc.*, 1993, **115**, 3399.
5 C. R. Landis, L. L. Luck and J. M. Wright, *J. Magn. Reson., Ser. B*, 1995, **109**, 44.
6 (*a*) C. R. Landis and V. S. Allured, *J. Am. Chem. Soc.*, 1991, **113**, 9493; (*b*) J. Wang, R. S. Hodges and B. D. Sykes, *J. Am. Chem. Soc.*, 1995, **117**, 8627; (*c*) D. O. Cicero, G. Barbato and R. Bazzo, *J. Am. Chem. Soc.*, 1995, **117**, 1027; (*d*) M. Reggelin, M. Köck, K. Conde-Frieboes and D. F. Mierke, *Angew. Chem., Int. Ed. Engl.*, 1994, **33**, 753.
7 L. Kozerski, R. Kawęcki, P. Krajewski, P. Gluziński, K. Pupek, P. E. Hansen and M. P. Williamson, *J. Org. Chem.*, 1995, **60**, 3533.
8 (*a*) L. Poppe, *J. Am. Chem. Soc.*, 1993, **115**, 8421; (*b*) R. E. Schirmer, J. P. Davis, J. H. Noggle and P. A. Hart, *J. Am. Chem. Soc.*, 1972, **94**, 2561.
9 (*a*) Ž. Džakula, A. S. Edison, W. M. Westler and J. L. Markley, *J. Am. Chem. Soc.*, 1992, **114**, 6200; (*b*) M. Reggelin, H. Hoffmann, M. Köck and D. F. Mierke, *J. Am. Chem. Soc.*, 1992, **114**, 3272.
10 J. T. Martin, P. O. Norrby and B. Åkermark, *J. Org. Chem.*, 1993, **58**, 1400.
11 W. Inman and P. Crews, *J. Am. Chem. Soc.*, 1989, **111**, 2822.
12 G. V. Nikiforovich, B. B. Vesterman and J. Betins, *Biophys. Chem.*, 1988, **31**, 101.
13 (*a*) M. J. Blackledge, R. Brüschweiler, C. Griesinger, J. M. Schmidt, P. Xu and R. R. Ernst, *Biochemistry*, 1993, **32**, 10960; (*b*) Y. Kim and J. H. Prestegard, *Biochemistry*, 1989, **28**, 8792; (*c*) H. Kessler, C. Griesinger, J. Lautz, A. Müller, W. F. van Gunsteren and H. J. C. Berendsen, *J. Am. Chem. Soc.*, 1988, **110**, 3393.
14 (*a*) R. U. Lemieux, S. Koto and D. Voisin, in *The Anomeric Effect: Origin and Consequences*; eds. W. A. Szarek and D. Horton, ACS Symp. Series No. 87, Washington DC, 1979; (*b*) H. Booth, J. M. Dixon, K. A. Khedhair and S. A. Readshaw, *Tetrahedron*, 1990, **46**, 1625.
15 L. Kozerski and Z. Urbańczyk-Lipkowska, *Bull. Pol. Acad. Sci., Chem.*, 1984, **32**, 159.
16 *PCMODEL, Molecular modelling software*, Serena Software; P. O. Box 3076, Bloomington, IN 47402-3076.
17 J. B. Lambert, *Top. Stereochem.*, 1971, **6**, 19.
18 L. Kozerski, P. Krajewski, Z. Urbańczyk-Lipkowska, P. Gluziński and A. Krówczyński, *J. Mol. Struct.*, 1994, **326**, 203.
19 R. Wasylishen and T. Schaefer, *Can. J. Chem.*, 1973, **51**, 961.
20 K. Wüthrich, *J. Biol. Chem.*, 1990, **265**, 22059.
21 L. Kozerski, R. Kawęcki and P. E. Hansen, *Magn. Reson. Chem.*, 1994, **32**, 517.
22 M. P. Williamson, *Magn. Reson. Chem.*, 1987, **25**, 356.
23 J. Tropp, *J. Chem. Phys.*, 1980, **72**, 6035.
24 D. Gonzalez, J. A. C. Rullmann, A. M. J. J. Bonvin, R. Boelens and R. Kaptein, *J. Magn. Reson.*, 1991, **91**, 659.
25 T. H. Wonnacott and R. J. Wonnacott, *Introductory Statistics*, J. Wiley and Sons, New York, 1972.
26 (*a*) H. Scheffé, *The Analysis of Variance*, John Wiley and Sons, New York, 1959; (*b*) E. Spjøtvoll and M. R. Stoline, *J. Am. Stat. Assoc.*, 1973, **68**, 975.
27 N. L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions*, John Wiley and Sons, New York, 2nd edn., 1994, vol. 1.
28 Ref. 1, p. 151 and Fig. 5.5 (p. 169).
29 J. Szroeter, *Stats. Prob. Letts.*, 1996, **29**, 9.

Paper 7/00149E
Received 7*th January* 1997
Accepted 30*th May* 1997