JOURNAL OF MAGNETIC RESONANCE **132,** 1–12 (1998)
ARTICLE NO. MN971347

Toward a Generalized Algorithm for the Automated Analysis of Complex Anisotropic NMR Spectra

F. Castiglione, M. Carravetta, G. Celebre, and M. Longeri

Dipartimento di Chimica, Universita' della Calabria, 87036 Rende (CS), Italy

Received August 7, 1997

An existing algorithm, founded on the works of Stephenson and Binsch, for the automatic analysis of isotropic or simple anisotropic NMR spectra has been improved to treat very complex NMR spectra of molecules dissolved in nematic solvents. The main options added to the original algorithm are a wider choice of smoothing functions; the use of the principal component regression method; and the possibility of selecting molecular coordinates, order parameters, and spectral parameters as variables of the problem. By means of these new options, it has been possible to analyze automatically NMR spectra (even depending on 27 spectral parameters) of 16 molecules in an anisotropic environment. Details of each case are discussed.

INTRODUCTION

NMR spectroscopy of molecules dissolved in liquid-crystalline phases (*1*) provides a powerful method for determining detailed structural and conformational information. Dipolar coupling constants depend directly on internuclear distances and, for flexible molecules, on the population of each conformer. The major drawback against a widespread diffusion of this technique is the complexity of ^{1}H single-quantum (SQ) NMR spectra: the number of independent spectral parameters (chemical shifts and direct couplings D_{ij}, the indirect ones J_{ij} usually being taken from isotropic spectra) increases with the size of the molecular spin system or the lowering of its symmetry. Therefore, it is a formidable task to obtain spectral parameters in all but the simplest cases. In practice, anisotropic spectra can be analyzed only if a good set of starting parameters are known at the beginning of the iterative procedure, based on the Castellano Bothner-By (CB-B) algorithm (*2*). The different experimental approaches to obtaining such a good set have been reported elsewhere (*3*).

As an alternative approach, powerful algorithms for a fully automated analysis of isotropic NMR spectra (i.e., DAVINS (*4*), PERCH (*5*)) have been developed and are becoming increasingly used in commercial packages (PERCH, WINNMR (*5, 6*)), with very good results.

There is a basic difference between the traditional CB-B approach and automatic analysis programs since, in the former, many hours of an experienced operator's time are required, whereas what is needed in the latter is (ideally) just CPU time.

There is not yet an automatic procedure for anisotropic spectra with performance comparable to that for isotropic spectra. The difference in behavior can be easily understood by recalling that anisotropic spectra differ from isotropic spectra in that they lack the typical cluster structure of the latter. As far as we know, the only programs of this kind are DANSOM (*7*) and DAISY (*8*), both implementations of DAVINS, where the spin Hamiltonian has been modified in order to deal with dipolar couplings (as a further improvement to speed up the calculations, in the latter the Hamiltonian matrix is further symmetry-factorized according to the spin system point group); however, their efficiency is very low.

Any automatic program for spectral analysis requires a scalar quantity φ, which estimates the degree of agreement between computed and experimental spectra. For example, in DAVINS this function is the error hypersurface, defined as $\varphi = (\mathbf{s} - \hat{\mathbf{s}})^{\mathrm{T}}(\mathbf{s} - \hat{\mathbf{s}})$, where the experimental spectrum $\mathbf{s}$ and the calculated spectrum $\hat{\mathbf{s}}$ are represented as N-dimensional vectors whose elements are spectral intensities measured at N discrete and equally spaced frequencies. The usual root mean square (RMS) could be related to φ by the simple relation, $\mathrm{RMS} = \sqrt{\varphi/m}$, m being the number of independent spectral parameters.

There is no general algorithm for finding the global minimum of an arbitrary function. Even if the field of global optimization is very young (see (*9a–9c*) for global optimization background), efficient methods have been devised for classes of functions which satisfy particular conditions (i.e., partially separable functions (*9d*)): unfortunately φ does not belong to this class. It follows that very efficient algorithms are needed to find the global minimum (that is, the deepest one), starting with any initial set of parameters, since the global minimum must be located in the presence

1090-7807/98 $25.00

of a myriad of local minima in an $(m + 1)$-dimensional space, each minimum in the error hypersurface arising from each experimental–computed line overlap.

DAVINS tries to overcome this difficulty using a SPIRAL-like algorithm (*10*) (well suited to local optimization), performing the search for the global minimum through a stepwise local optimization of an ordered sequence of l functions $\varphi'_1, \varphi'_2, \cdots, \varphi'_l = \varphi$ such that there is only one minimum on φ'_1 and, for each $k > 1$, the global minimum of φ'_k lies inside the convergence radius of φ'_{k+1} global minimum. Each hypersurface can be defined as $\varphi'_k = (\mathbf{s} - \hat{\mathbf{s}})^T \mathbf{W}_k (\mathbf{s} - \hat{\mathbf{s}})$, where $\mathbf{W}_l = \mathbf{I}$. The $\mathbf{W}_k$ matrices are such as to have a decreasing smoothing effect on the error hypersurface. Because of the complexity of φ, and since the correct sequence of smoothing matrices cannot be determined a priori, there are very few chances for the program to end up right in the global minimum, and the only way to increase its success probability is to start with a good set of spectral parameters.

Multiple-quantum (MQ) spectroscopy (*11*) is a powerful tool for obtaining a good set of spectral parameters in a relatively simple way (*12*); these parameters must then be refined on the corresponding single-quantum spectrum. This can be done using either a traditional or an automatic approach. One method that was recently proposed for automatically analyzing anisotropic spectra (*13*) involves in a first step the automated analysis of MQ spectra to obtain a good set of parameters, which are then refined by using DAISY to fit the corresponding single-quantum spectrum.

Since high-order MQ spectra are likely to give good values of linear combinations rather than single values of parameters, the starting set to be used as input for DAISY might not be good enough to allow the minimization of φ. The aim of this paper is to improve the performance of DAISY in order to extend the proposed procedure (*13*) to much more complex spin systems. To attain this goal the original algorithm has been modified (*14*) by:

(i) allowing for a wider range of the kinds of smoothing matrices $\mathbf{W}_k$;

(ii) introducing the principal component regression (PCR) method (*15*); and

(iii) using a mixed set of molecular coordinates, order parameters, and spectral parameters as DAISY variables (a similar option has been independently inserted into a commercial package, WIN-DAISY (*16*)).

The efficiency of the improved code has been tested on increasingly complex spectra (see Table 1) starting without guessed values of spectral parameters; in particular, spectra of simple rigid molecules of high symmetry have been used to test the modified code performance. Then we approached spectra of rigid molecules of lower symmetry with a greater number of spectral parameters and spectra of molecules with internal motions.

TABLE 1
Spectra Analyzed with the Modified Version of DAISY

Molecules	Spin system	No. of parameters	Point group
(**1**) Bromobenzene[a]	AA′BB′C	9	C_{2v}
(**2**) Ethynylbenzene	AA′BB′CD	13	C_{2v}
(**3**) 1,4-Naphthoquinone	AA′BB′CC′	12	C_{2v}
(**4**) Naphthalene[b]	AA′A″A‴BB′B″B‴	12	D_{2h}
(**5**) Biphenylene	AA′A″A‴BB′B″B‴	12	D_{2h}
(**6**) 9,10-Dibromoanthracene	AA′A″A‴BB′B″B‴	12	D_{2h}
(**7**) Benzofuran	ABCDEF	21	C_s
(**8**) 1-Bromonaphthalene[c]	ABCDEFG	28	C_s
(**9**) 1-Chloronaphthalene[c]	ABCDEFG	28	C_s
(**10**) 2-Bromonaphthalene[c]	ABCDEFG	28	C_s
(**11**) Benzaldehyde	AA′BB′CD	13	
(**12**) 1,3-Benzodioxole	AA′BB′C_2	10	C_{2v}
(**13**) Benzylbromide[b]	AA′BB′CD_2	14	
(**14**) Benzylchloride[b]	AA′BB′CD_2	14	
(**15**) Benzyliodide[b]	AA′BB′CD_2	14	
(**16**) THF	AA′A″A‴BB′B″B‴	12	

Note. Unless stated otherwise, the spectra were recorded using ZLI1132 as solvent.

[a] ZLI1132 (**a**) and I52 (**b**) as solvents.

[b] ZLI1132 (**a**) and ZLI1132 ÷ EBBA 55% wt (**b**).

[c] ZLI1132 (**a**) and I35 (**b**).

The cases treated should provide an indication of the spectral complexity that the procedure can handle with a high expectation of success.

Obviously, for spectra with a high number of independent parameters, the probability of straightforward analysis decreases, but we defined a multistep procedure, requiring sequential runs of the code, able to give a very high overall chance of success.

The entire procedure is quite CPU time consuming and, for the most demanding conditions, it is not fully automated, needing user intervention; in any case, with respect to the CB-B approach, the role played by the operator, as we will explain in the following sections, is completely different.

EXPERIMENTAL

Some of the spectra reported in Table 1 (**1–3** (*13*), **4** (*17*), **13–15** (*18*)) have been taken from previous works. All the other NMR spectra have been acquired at room temperature on a 7.4-T Bruker AC300 spectrometer. A one-night temperature equilibrium period was adopted, when possible, before recording the spectra. Solutes (Aldrich) and liquid-crystalline solvents (Merck-Darmstadt), with the exception of EBBA (*18*), are commercially available and were used without further purification.

The computational tests were performed on an Indy Silicon Graphics workstation by an interactive graphical shell developed in our laboratory (*19*) whose last version, written in C language, uses the OSF/Motif (*20*) Graphical Library.

CUSTOMIZATION OF DAISY FOR ANISOTROPIC SPECTRA

As previously stated, anisotropic spectra differ from isotropic spectra in two primary ways:

(1) many of the indirect couplings can safely be assumed to be zero since they are smaller than the experimental linewidth; and

(2) the D_{ij} most affecting the spectrum are larger than the chemical shift differences (δ), so that we can write $|D_{ij}| > |\delta_i| > |J_{ij}|$.

It follows that:

(1) for isotropic spectra both the parameter space and the number of resolvable experimental lines are much smaller, giving thereafter fewer and more sharply located minima; and

(2) the typical cluster structure of isotropic spectra is lost and so the shape of the correlation matrix becomes critical since lines ''belonging'' to different isochronous or nearly isochronous nuclei are deeply interloped.

In our opinion the program efficiency could be extended by the possibility of choosing among a wider range of smoothing functions and using a limited set of independent parameters in the first explorative runs, until the main features of the experimental spectrum are well reproduced. The global minimum will then be searched for by iterating over the entire set of parameters, using as starting values those just found. This multistep analysis still requires operator intervention, but only for choosing among the different results by comparison of calculated and experimental spectra (the boring operation of carrying out the assignment of experimental–calculated lines, peculiar to the CB-B method of analysis, is definitively bypassed). Both the introduction of PCR and the iteration over order parameters, in place of some dipolar couplings, are aimed at reducing the number of independent parameters.

(a) Variation of the Form of the Smoothing Function

The smoothing matrices, as previously defined, should allow the achievement of the global minimum by the optimization algorithm starting from an arbitrary set of parameters. The expressions of these matrices can only be inferred: they can be considered functions of the distance between the correlated points of the spectrum array, $W_{ij} = f(|i - j|)$, where f is a monotonically decreasing, nonnegative function such that $f(x) \leq 1, \forall x$. There are many functions satisfying these conditions. The authors of DAVINS tested that the exponential function

$$f(x) = \exp\{\alpha x\} \quad [1]$$

and the Lorentzian function

$$f(x) = \frac{1}{1 + \alpha x^2}, \quad [2]$$

where α is positive for a Lorentzian function and negative for an exponential function, are the best ones for the analysis of isotropic NMR spectra. Actually, there is no theoretical reason for assuming that these are also the most satisfactory functions for the analysis of anisotropic spectra. In order to use smoothing functions with a more flexible form, we replaced the original functions with a generalized exponential

$$f(x) = \exp\{\alpha x^b\} \quad [3]$$

and with a generalized Lorentzian

$$f(x) = \frac{1}{1 + \alpha x^b}, \quad [4]$$

where b is a positive parameter that can be fixed. We found that, for anisotropic spectra, the original functions work quite well for simple spectra, but as the complexity of the spectrum increases, the smoothing efficiency of the exponential rapidly decreases: different choices of the smoothing function can be decisive for analyzing complex anisotropic spectra.

(b) Combination of the Existing Optimization Code with the Principal Component Regression Method

During the first iterative steps, when the error hypersurface is heavily smoothed, spectral parameters are strongly correlated and variations of all but a few linear combinations of parameters will affect significantly the resulting φ'. For example, that is what happens for 2–6 (*ortho–ortho*) and 3–5 (*meta–meta*) direct couplings in a monosubstituted benzene: the spectrum depends mainly on their sum rather than on their individual values. Only when the spectrum is well reproduced and almost all the experimental lines are assigned does the correlation degree decrease and the spectrum become dependent on their difference too.

Principal component analysis (*15*) can be used to distinguish linear combinations whose influence on the spectrum is significant from others which can be safely kept constant: in this way it is possible to reduce the effective number of independent variables.

Let $\mathbf{p} \in R^m$ be the parameter array; the spectrum variation, **ds**, due to a variation of parameters, **dp**, is given by

$$\mathbf{C}\mathbf{dp} = \mathbf{ds}, \quad [5]$$

where **C** is an $(N \times m)$ matrix (N being the number of points of the spectrum) whose elements are defined according to

$$C_{ij} = \partial s_i / \partial p_j. \quad [6]$$

For a least-squares problem **dp** can be calculated as

$$\mathbf{dp} = (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{ds}, \quad [7]$$

where $\mathbf{C}^{\mathrm{T}}$ is the transposed matrix, $\mathbf{H} = (\mathbf{C}^{\mathrm{T}}\mathbf{C})$ (*15*) corresponds to the variance–covariance matrix (Hessian matrix), and $\mathbf{C}^{\mathrm{T}}\mathbf{ds}$ is the gradient array **G**, that is, $\mathbf{dp} = \mathbf{H}^{-1}\mathbf{G}$.

Diagonal elements of **H** give an estimate of the dependence of the spectrum on each parameter; a generic out-diagonal element H_{ij} is a measure of the correlation degree between the ith and jth parameters. The fundamental principle of the PCR method is that **p** can be replaced by a new set of parameters, **q**, each one being a linear combination of the old set, such that the new Hessian matrix $\mathbf{H}'$, given by $\mathbf{H}' = \mathbf{R}^{\mathrm{T}}\mathbf{H}\mathbf{R}$, is diagonal, that is, $H'_{ij} = h_i\delta_{ij}$, h_i being **H** eigenvalues and $\mathbf{q} = \mathbf{R}^{\mathrm{T}}\mathbf{p}$.

Let **p** be our starting set of parameters, containing some unimportant or strongly correlated parameters: some eigenvalues must be close to zero and so the corresponding linear combinations q_i will undergo unrealistic variations since the spectrum is almost insensitive to their changes. Therefore, it is convenient to reduce the number of variables, keeping constant all the linear combinations corresponding to eigenvalues lower than a threshold value, determined according to an appropriate criterion. This threshold value is calculated at each iteration as a small percentage ξ (usually 1 or 0.1%) of the trace of **H**, which measures the total variance of the spectrum with respect to parameters. The correlation degree between parameters depends both on our choice of the limiting percentage and on the form of the smoothing function: the smoother the hypersurface, the greater the number of linear combinations that will be fixed. During each run, the smoothing level slowly decreases and the code itself can gradually reduce the number of correlated parameters.

Anyway, this simple trace-based criterion for discriminating between important and unimportant linear combinations is inefficient since usually too many combinations are fixed during the first iterations of the code, leading to small variations of the parameters. We developed a method (that we will call the *modified* PCR method) which takes into account not only the eigenvalues of **H** but also the information contained in the eigenvector matrix in order to consider the existing correlations without losing too much information.

Let the combinations q_i be ordered according to their eigenvalues so that $\forall i < j,\ h_i < h_j$. The generic element of q_i, q_i^k, measures the importance of the p_k parameter in the ith linear combination. In this modified PCR algorithm we consider the n linear combinations such that $h_i > \xi \mathrm{Tr}(\mathbf{H})$ and, starting from the $(m - n - 1)$th position, we include one by one other combinations such that their eigenvalues are greater than an arbitrarily chosen, trace-independent threshold value of 10^{-7} ($h_i > 10^{-7}$), until all the parameters are considered with a nonnegligible weight in at least one combination ($\forall k, \exists j : |q_j^k| > 0.1, j \in [i, m]$) (in our experience, eigenvalues less than 10^{-7} can be safely neglected without losing information).

Several tests point out a better behavior of this algorithm (for at least two cases, left unsolved using the standard PCR method, the global minimum was obtained by the modified algorithm); these and other possible strategies for taking correlations into account must be tested further.

(c) Introduction of Order Parameters and Coordinates as DAISY Variables

For rigid molecules, if geometry were exactly known, the set of independent parameters could be noticeably reduced since all dipolar couplings can be expressed as

$$\begin{aligned} D_{ij} = -\frac{\gamma_i\gamma_j h}{8\pi^2\langle r_{ij}^3\rangle} \{ & S_{aa}\langle 3\cos^2\vartheta_{ija} - 1\rangle \\ & + (S_{bb} - S_{cc})\langle \cos^2\vartheta_{ijb} - \cos^2\vartheta_{ijc}\rangle \\ & + 4S_{ab}\langle \cos\vartheta_{ija}\cos\vartheta_{ijb}\rangle \\ & + 4S_{ac}\langle \cos\vartheta_{ija}\cos\vartheta_{ijc}\rangle \\ & + 4S_{bc}\langle \cos\vartheta_{ijb}\cos\vartheta_{ijc}\rangle \} \end{aligned} \quad [8]$$

in terms of internuclear distances r_{ij} and of a set of ''order parameters'' $S_{\alpha\beta}$ (*1*), defined as

$$S_{\alpha\beta} = \tfrac{1}{2}\langle 3\cos\vartheta_{\alpha Z}\cos\vartheta_{\beta Z} - \delta_{\alpha\beta}\rangle, \quad [9]$$

which describe the average orientation of a Cartesian reference frame fixed on the molecule with respect to a laboratory frame (usually the Z axis of the laboratory frame is parallel to the static magnetic field $\mathbf{H}_0$).

For flexible molecules the dependence of D_{ij} couplings on geometry is more complex since the conformational equilibrium must be taken into account. Direct couplings between nuclei belonging to the same rigid subunit can be calculated in terms of averaged orientational parameters of that rigid fragment, reducing again the total number of independent parameters.

The original code could iterate only on chemical shifts (CS) and direct and indirect coupling costants (plus baseline and linewidth parameters). Indirect couplings are usually kept fixed to their isotropic values in the analysis of anisotropic spectra. The program has been modified in order to iterate also on order parameters and/or Cartesian coordinates. Given an assumed, fixed geometry, the global minimum that can be found varying order parameters is not exactly the correct one but it is likely to be a good starting point for further refining in one or more steps, according to a strategy that depends on the complexity of the spectrum and on the reliability of the geometrical assumptions.

Symmetry can further reduce the number of independent coordinates so that sometimes it can be more convenient to iterate directly on order parameters, coordinates, and chemical shifts, the total number of independent parameters still being lower than the full set of spectral parameters (CS and D_{ij}). Geometrical constraints, easy to set, can help to avoid local minima due to wrong combinations of spectral parameters.

As a summary, the program can iterate on CS, D_{ij}, S, and coordinates (X) at the same time, just specifying the type of variable parameters in the input file.

RESULTS AND DISCUSSION

We can summarize the strategy adopted for analyzing different classes of molecules by the flow diagram shown in Fig. 1.

In each step of the analysis, the general approach consists in preparing a list of runs, using different smoothing functions: as outlined before, locating the global minimum (GM) is *a problem with no general solution,* and we do not know a priori which are the best conditions for obtaining the convergence of the code. By visual inspection, the operator discriminates among the different runs to choose the ''best looking'' spectrum and the corresponding parameter set to be subjected to further refining.

When using the PCR approach (PCRA) the entire set of spectral parameters is generally varied, but the introduction of ''selection'' criteria quoted above is aimed at reducing the number of variables to more easily locate the GM.

On the contrary, using the geometrical approach (GA), we reduce as much as possible the number of iteration variables (order parameters S and CS, usually) in the first runs; the best solution is then refined, gradually removing constraints so that the number of independent parameters increases in the following steps (the full set of spectral parameters is changed in the last step).

Preliminary Tests

Initial tests were performed on extremely simple spin systems, that is, bromobenzene (**1**), ethynylbenzene (**2**), and 1,4-naphthoquinone (**3**). As starting conditions we assumed:

(i) a perfect hexagon (C–C = 1.4 Å) for the phenyl ring;

(ii) standard lengths for acetylenic bonds;

(iii) the same value for aromatic CS;

(iv) isotropic ΔCS between aromatic and acetylenic protons; and

(v) null order parameters S_{zz} and $S_{xx} - S_{yy}$.

Bromobenzene. It was possible to analyze spectra of **1** (Table 1) using both geometrical and PCR strategies:

(a) GA: iterating on CS, S, and X the GM can be reached in one shot ($b = 1$ and $\alpha = -10^{-4}$);

(b) two-step PCRA: it is possible to obtain at least one good set of parameters ($\alpha = 10^{-8}$, $b = 2$) when starting with equal CS and null D_{ij}, varying all the parameters with a threshold factor ξ equal to 0.01 (1% of the Hessian trace). The best solution found is then refined in a second step (it may be interesting to note that we also performed some tests using a *cosine square* function, but it proved unsuitable even for this simple system, so we decided to reject it).

Ethynylbenzene. It was possible to reach the GM by:

(a) GA: in a single step, iterating on CS, S, and X ($\alpha = -10^{-6}$, $b = 2$);

(b) two-step PCRA: in the first step $\alpha = -10^{-13}$, $b = 3$, $\xi = 0.01$, keeping $D_{ortho-ortho} = D_{meta-meta}$.

1,4-Naphthoquinone. In this case it was not possible to locate the GM immediately, as we did for **1,** iterating over CS, S, and X. The global solution could be reached by either the GA or the PCRA. We found at least two different analysis strategies for obtaining the GM of **3** according to the GA. The analysis settings are reported in Table 2.

The values of spectral parameters for spectra **1–3** are reported elsewhere (*13*).

In all of our tests on **1,** both Gaussian and original DAVINS functions led to good results, even if the original functions frequently had a better behavior. On the other hand, as the number of independent parameters and interacting spins increases, the number of local minima on the error hypersurface increases sharply. So the choice of a proper smoothing function is essential: the runs on spectra **2** and **3** pointed out that a simple exponential function is inappropriate for the analysis, even for these six-spin systems, while both Gaussian and Lorentzian functions gave good results.

In the following we will give a description of the procedure adopted for analyzing increasingly complex spectra of rigid molecules and a few simple flexible molecules.

Six-Spin Systems

Both 1,3-benzodioxole (**12**) and benzofuran (**7**) are six-proton molecules but the number of independent spectral

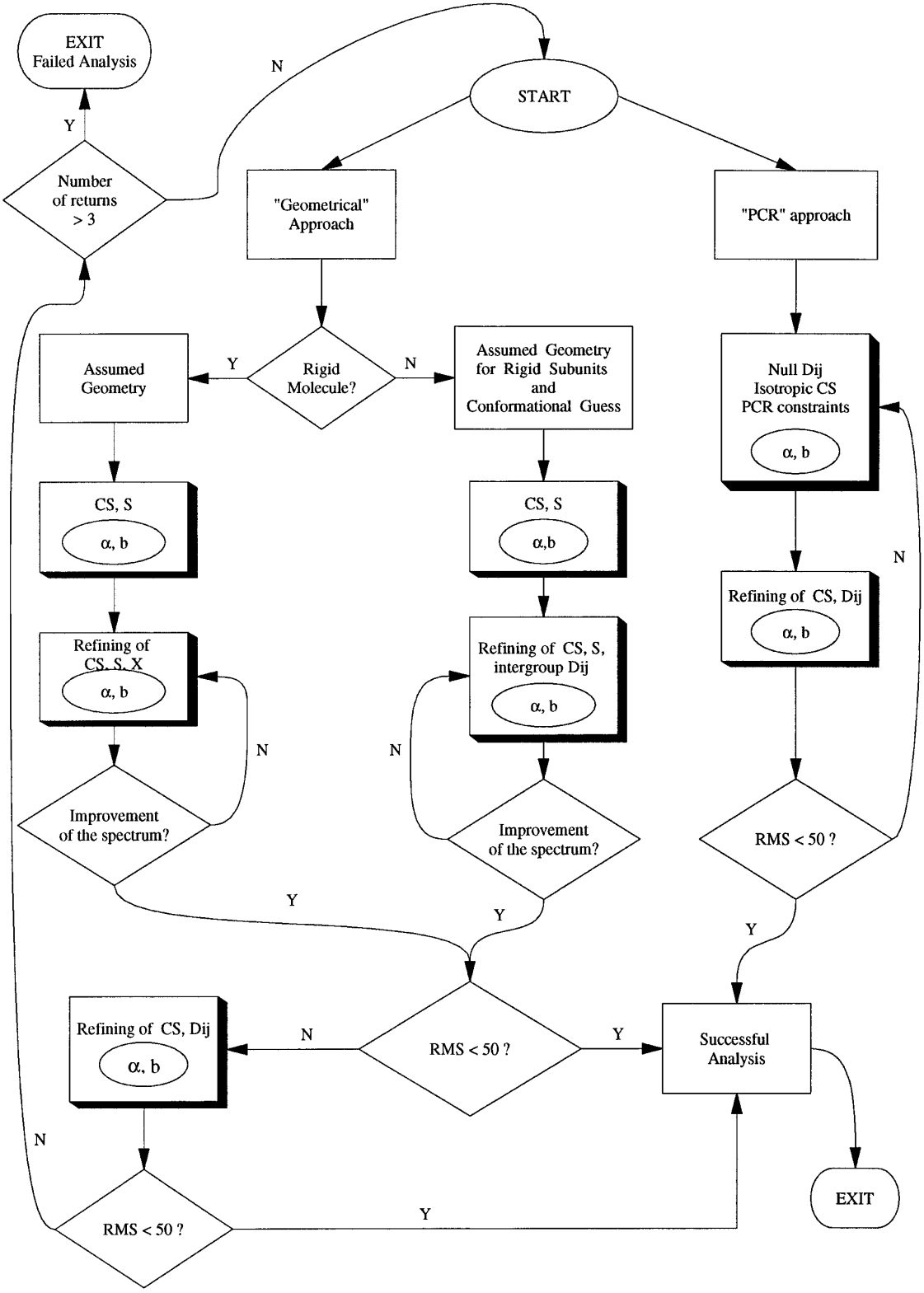


FIG. 1. The flow diagram summarizes the analysis strategy of anisotropic spectra using the modified DAISY code. The shaded boxes contain the variables of each iteration step. The ellipse within the shaded boxes stands for making a series of runs using different smoothing conditions (i.e., varying the α and b parameters). The RMS < 50 (''experience''-based threshold) test is a global minimum location criterion: lower RMS values can be obtained, if needed, with further refining of the linewidth and baseline.

TABLE 2
Analysis Settings for Spectrum 3 of Table 1

Strategy	First step variables	Successful run conditions	Refining step variables
GA	CS, S	$\alpha = -10^{-12}$, $b = 2$	CS, D_{ij} or CS, S, X
GA	S, X, and all CS together	$\alpha = -10^{-9}$, $b = 2$	CS, D_{ij}
PCRA	CS, D_{ij}	$\alpha = -10^{-12}$, $b = 2$, $\xi = 0.01$	CS, D_{ij}

parameters is significantly higher for the latter, as a consequence of its lower symmetry.

1,3-Benzodioxole. The structure of the carbon skeleton of **12** was previously calculated (*21*), and on this basis, we obtained approximated proton coordinates using standard angles and lengths for C–H bonds. In this case our initial assumptions are as follows:

(i) both order parameters positive (we put arbitrarily $S_{zz} = S_{xx} - S_{yy} = 0.2$), as can be guessed from geometrical consideration; and

(ii) isotropic ΔCS between aromatic and aliphatic protons.

The GM was reached by:

(a) GA: iterations over CS and S led to at least one good set of parameters ($\alpha = -10^{-6}$, $b = 3$). We were not able to refine this set by usual methods. It might be due to bad reproduction of dipolar couplings by a simple geometry optimization, so we iterated independently over CS, S, X, and D_{56} (geminal coupling): D_{56} is actually the largest coupling, very sensitive to geometrical assumptions and strongly affected by vibrational corrections. In this way it was possible to get a good enough solution ($\alpha = -10^{-12}$, $b = 2$) to allow the convergence in one more run, by variation of CS and D_{ij};

(b) PCRA: it was not possible to analyze this spectrum using the trace-based PCR method; the modified PCR algorithm proved more efficient and allowed the location of the GM in two steps, using Gaussian functions.

Final parameters are reported in Table 3.

Benzofuran. Proton coordinates of **7** were determined using interprotonic distances derived from a previous analysis of the spectrum of benzofuran in anisotropic media (*22*). The GM was reached in a three-step analysis, starting with iterating on null order parameters ($S_{zz} = S_{xx} - S_{yy} = S_{xz} = 0$) and equal CS ($\alpha = 10^{-12}$, $b = 2$), refining CS, S, and X ($\alpha = -10^{-9}$, $b = 2$), and then iterating on the entire set of spectral variables. The final parameters are reported in Table 4.

Eight-Spin Systems

The spectra of four molecules belonging to the AA′A″A‴ BB′B″B‴ spin system have been analyzed: naphthalene, biphenylene, 9,10-dibromoanthracene, and tetrahydrofuran.

We assumed that naphthalene and anthracene were formed by fused benzene rings, and biphenylene was formed from two coplanar rings 1.53 Å apart.

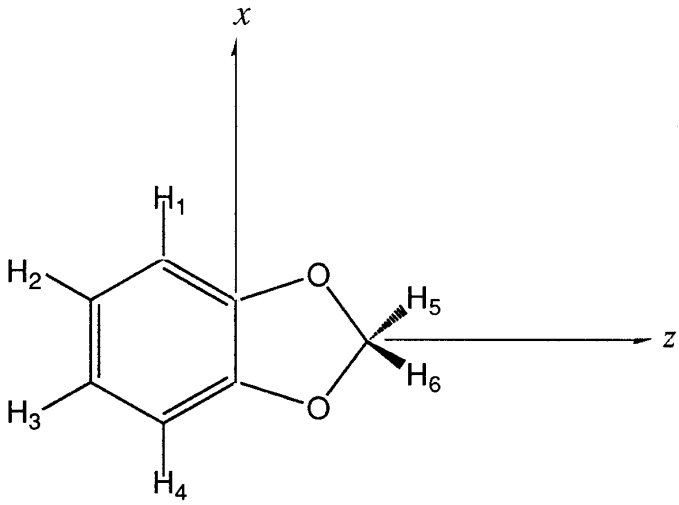
As usual, we put equal CS and null S. Both GA and PCRA proved successful, following the strategy reported in Fig. 1.

Naphthalene. The spectrum of **4b** was easily analyzed in two steps using the GA ($b = 2$ functions) without X optimization. Using the same geometry, the two-step strategy *does not work* for **4a** and it needs an intermediate refining step, optimizing X, CS, and S.

The spectrum of **4a** was also analyzed by PCRA ($\xi = 0.01$, $\alpha = -10^{-7}$, $b = 2$).

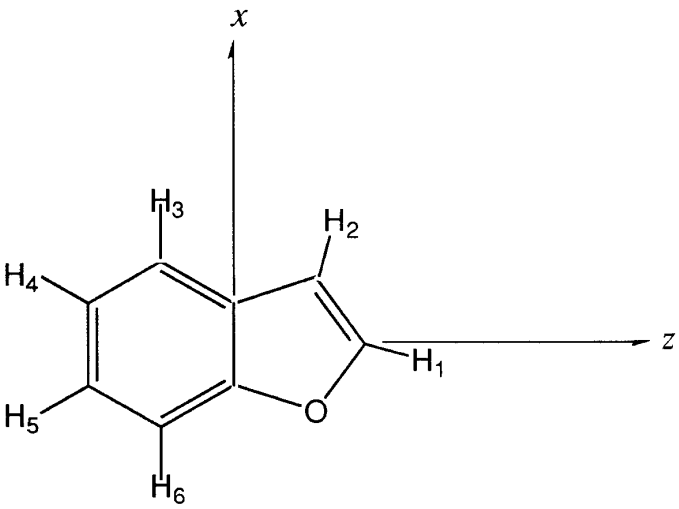
It is worthwhile to notice that when PCRA was used

TABLE 3
Dipolar Couplings, D_{ij}, and Chemical Shifts, δ_i, Both in Hertz, Obtained by the Analysis of the Spectrum of 1,3-Benzodioxole (12)



D_{12}	−1516.10
D_{13}	−101.13
D_{14}	5.23
D_{15}	−127.45
D_{23}	−1.50
D_{26}	−198.83
D_{56}	4401.79
$\delta_1 - \delta_3$	193.80
$\delta_2 - \delta_3$	149.11

TABLE 4
Dipolar Couplings, D_{ij}, and Chemical Shifts, δ_i, Both in Hertz, Obtained by the Analysis of the Spectrum of Benzofuran (7)



D_{12}	−205.03
D_{13}	−129.50
D_{14}	−92.19
D_{15}	−119.31
D_{16}	−264.83
D_{23}	−1040.83
D_{24}	−212.35
D_{25}	−124.97
D_{26}	−95.72
D_{34}	−1835.42
D_{35}	−171.04
D_{36}	−13.66
D_{45}	−106.40
D_{46}	−61.13
D_{56}	−1235.03
$\delta_1-\delta_2$	175.37
$\delta_3-\delta_2$	62.38
$\delta_4-\delta_2$	43.70
$\delta_5-\delta_2$	82.14
$\delta_6-\delta_2$	46.01

starting from the best **4a** solution, found after the first series of runs, the analysis failed. All this shows that the different efficiency of the PCR method dramatically depends on the starting parameter set; moreover, the chance of locating the GM does not necessarily increase with the nearness between the GM and the starting point. As said before, PCR analysis lets the code vary only the most important linear combinations, relaxing later the others; but, when combined with DAISY, the PCR method can be inefficient if the disagreement between computed and experimental spectra is mainly due to parameters the spectrum is less sensitive to: in such a situation DAISY must be used without the PCR option.

Biphenylene and 9,10-dibromoanthracene. These spectra were analyzed both using the PCRA and in three sequential runs according to the GA ($b = 2$ and $b = 3$ functions).

Tetrahydrofuran (THF). A virtual planar geometry was assumed for the skeleton of THF with averaged proton positions to give a C_{2v} symmetry to the molecular spin system (two independent order parameters).

This spectrum was analyzed:

(a) by the PCRA; and
(b) in three steps, using the GA: proton coordinates were changed, keeping a C_{2v} symmetry during the geometry optimization in the second step of the analysis ($b = 2$ and $b = 3$ functions).

Final parameters of **4, 5, 6,** and **16** are reported in Table 5 and Table 6.

Monosubstituted Naphthalenes

The geometry of naphthalene was adopted to describe its halo derivatives. We started with all *CS* set equal, $S_{zz} = S_{xx} - S_{yy} = 0.2$ and $S_{xz} = 0.01$ (arbitrary values). In order to analyze these spectra we iterated at the beginning on as few parameters as possible, removing the constraints gradually. In most cases the code could locate the GM after three to four sequential runs. The first step always consisted in varying the three independent order parameters and one CS (keeping all CS equal). On the contrary, starting from the ''best looking'' spectrum, it was not possible to follow a common *modus operandi*, the most effective analysis strategy depending on the particular case (see below) and on the comparison of computed and experimental spectra after the first series of runs. Smoothing functions worked quite well with $b = 2$ and $b = 3$.

1-Bromonaphthalene. The spectrum of **8a** was analyzed by three more series of runs:

(i) varying independently S, one CS, and a constrained subset of coordinates;
(ii) varying all CS; and
(iii) iterating over CS and D_{ij}.

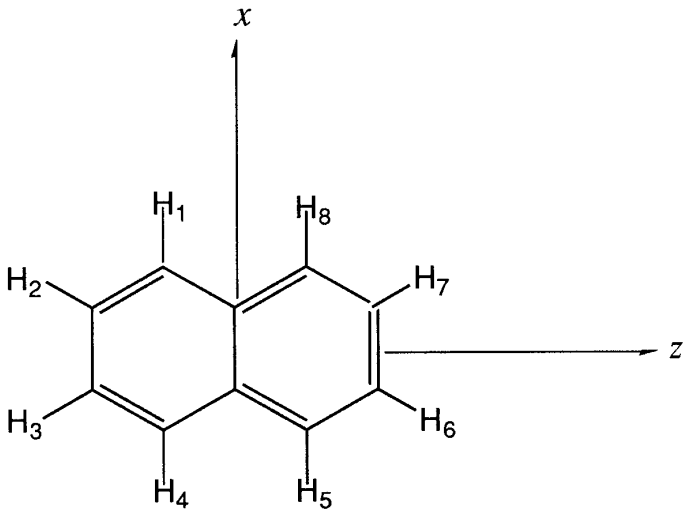
The spectrum of **8b** was easily analyzed in two steps, iterating on the final CS and S of **8a** (using the improved geometry) and then refining. It was also analyzed in three steps without an initial guess, refining all CS and S as a second step and then iterating over all the spectral parameters.

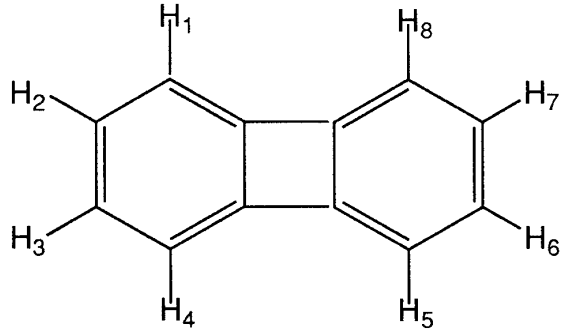
1-Chloronaphthalene. The analysis of **9a** was straightforward, by an **8b**-like two-step approach, starting from the final parameters for **8a.**

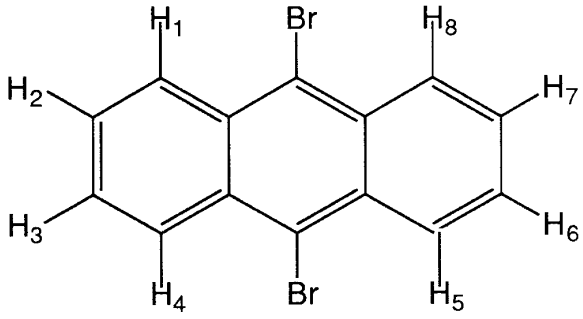
The GM could also be reached without a clever guess about the values of the parameters. After the first series of runs, the GM was reached by:

TABLE 5

Dipolar Couplings, D_{ij}, and Chemical Shifts, δ_i, Both in Hertz, Obtained by the Analysis of the Spectrum of Naphthalene (4), Biphenylene (5), and 9,10-Dibromoanthracene (6)







	4a	4b	5	6
D_{12}	−1588.32	−1957.18	−2069.79	−2299.82
D_{13}	−80.51	−117.34	−122.66	−137.41
D_{14}	18.73	5.63	13.14	2.29
D_{15}	−27.63	−43.4	−61.90	−66.81
D_{16}	−91.95	−115.65	−92.95	−70.85
D_{17}	−283.51	−344.61	−191.45	−120.46
D_{18}	−2181.06	−2642.41	−805.47	−371.39
D_{23}	159.30	61.8	140.08	52.21
D_{26}	−78.17	−95.7	−69.68	−50.26
D_{27}	−108.70	−131.84	−86.52	−60.06
$\delta_2-\delta_1$	34.19	58.73	114.01	360.35

(i) varying independently one CS, S, and a subset of coordinates;

(ii) varying all CS, S, and X; and

(iii) final refining.

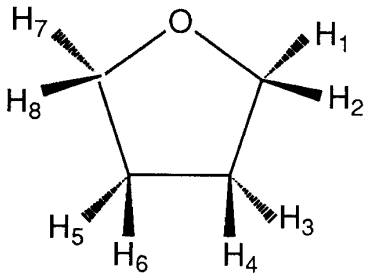
The spectrum of **9b** was analyzed in two steps starting from the final parameters of **9a.** The ''unguessed'' approach was performed by:

(i) varying one CS and S;

(ii) varying all the CS and S;

(iii) varying CS, S, and X; and

TABLE 6
Dipolar Couplings, D_{ij}, and Chemical Shifts, δ_i, Both in Hertz, Obtained by the Analysis of the Spectrum of THF (16)



D_{12}	1443.82
D_{13}	119.59
D_{14}	180.74
D_{15}	−187.48
D_{16}	−63.06
D_{17}	−385.95
D_{18}	−151.73
D_{34}	1548.68
D_{35}	−736.28
D_{36}	−206.27
$\delta_1-\delta_3$	472.13

(iv) refining.

2-Bromonaphthalene. For **10a,** the best solution found after the first series of runs was further improved by:

(i) varying all CS and *S*;
(ii) varying CS, *S*, and *X*;
(iii) a second refining, varying the same parameters, in order to obtain a better location of the GM; and
(iv) final refining.

The final parameters of **10a** are not a good starting point for the analysis of **10b.** The spectrum of **10b** is much wider and characterized by a different distribution of signals, so that a simple scaling of *S*, in order to take into account differences in spectral width between the spectra, does not even allow one to roughly reproduce the experimental spectrum. After the first series of runs, it was possible to obtain an estimate of order parameters, but there was no way to improve this solution.

As a general consideration, 1-halonaphthalenes proved much easier to analyze, maybe because the position of the halogen atom was such as to cause less spreading of CS and a minor distortion of the molecule with respect to naphthalene. The final parameters of **8, 9,** and **10** are reported in Table 7.

Flexible Molecules

We considered two kinds of flexible molecules: benzaldehyde (**11**) and benzyl halides (**13–15**). In both cases, as a starting hypothesis, we assumed the geometries corresponding to the conformational minimum (aldehydic and aromatic fragments are coplanar, C–*X* bond lies in a plane perpendicular to the ring), using standard lengths and angles, null order parameters, and a CS difference between aromatic and nonaromatic protons taken from standard isotropic values. The existence of a deep minimum in the rotation potential function makes it possible to deal with this kind of molecule using the same approach that worked for rigid molecules during the first runs, in order to obtain a good set of starting parameters.

Benzaldehyde. This spectrum has been analyzed by:
(a) GA:

(i) iterating independently on *S*, δ_6, and δ_1 (see Table 8) and constraining all aromatic CS to be equal ($\alpha = 10^{-14}$, $b = 3$);
(ii) iterating on all CS, the two ring order parameters (related to aromatic dipolar couplings), and aromatic–aldehydic (mobile) couplings; and
(iii) refining all the spectral parameters.

(b) PCRA allowed the location of the GM in two steps, as for **3.**

The final parameters are reported in Table 8.

Benzyl halides. The NMR spectrum of **13a** has been analyzed by:

(a) GA: iterating on CS and *S* to calculate direct couplings with the exception of D_{67} (geminal coupling), fixed to a value that can be roughly estimated from the experimental spectrum, and then refining; and
(b) PCRA: as for **12, 13a** could be analyzed only with the modified PCR algorithm.

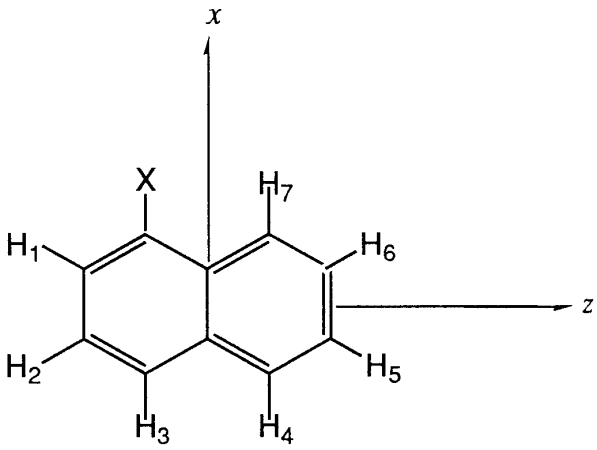
The spectra of **14a** and **15a** could be easily analyzed in two steps, starting from the final parameters of **13a,** first varying CS, ''mobile'' dipolar couplings, and *S* (in place of the aromatic direct couplings), and then refining CS and D_{ij}. The spectra of **13b–15b** have been analyzed with a similar strategy, starting from CS, *S*, and mobile D_{ij} of the corresponding benzyl halides in ZLI1132 but scaling *S* and D_{ij}. The final parameters for spectra **13–15** are reported elsewhere (*18*).

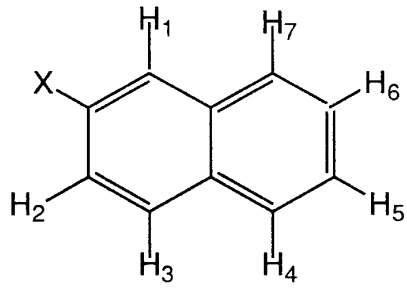
CONCLUSION

Lorentzian-like and exponential-like functions with greater *b* values (i.e., 2–3) proved more powerful for analyzing anisotropic spectra, leading to an improvement of the code performance. When applied to spectra with a low number of independent parameters, the PCR-based analysis strat-

TABLE 7

Dipolar Couplings, D_{ij}, and Chemical Shifts, δ_i, Both in Hertz, Obtained by the Analysis of the Spectrum of 1-Bromonaphthalene (8), 1-Chloronaphthalene (9), 2-Bromonaphthalene (10)



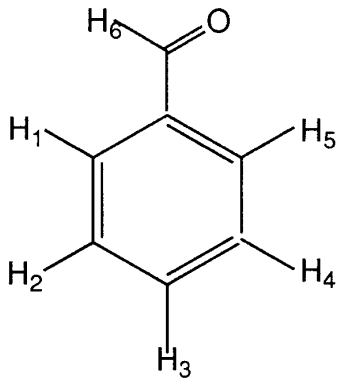


	8a	8b	9a	9b	10a
D_{12}	−955.43	−764.45	−738.94	−639.88	99.64
D_{13}	−303.69	−292.75	−268.29	−254.90	37.51
D_{14}	−137.04	−140.06	−133.28	−126.97	−119.08
D_{15}	−78.09	−79.98	−81.25	−75.12	−216.26
D_{16}	−78.00	−79.32	−86.57	−77.07	−489.75
D_{17}	−172.90	−166.47	−203.13	−168.66	−3107.40
D_{23}	−1873.49	−1935.18	−1879.06	−1776.90	−3311.31
D_{24}	−264.02	−272.84	−280.81	−258.46	−489.07
D_{25}	−79.47	−79.47	−87.84	−77.29	−155.41
D_{26}	−47.65	−43.82	−57.21	−45.58	−79.20
D_{27}	−60.21	−47.66	−67.68	−49.35	−41.89
D_{34}	−1640.82	−1666.23	−1822.24	−1608.63	−3111.66
D_{35}	−178.20	−172.19	−203.35	−169.59	−317.55
D_{36}	−60.42	−48.09	−68.59	−51.26	−41.74
D_{37}	−58.13	−36.44	−52.05	−34.72	53.65
D_{45}	−958.01	−839.03	−1122.82	−871.87	−1172.26
D_{46}	−126.83	−80.69	−118.73	−79.22	98.78
D_{47}	−119.18	−93.86	−93.99	−80.15	38.08
D_{56}	−929.04	−740.56	−725.42	−625.70	324.63
D_{57}	−301.92	−289.49	−265.88	−254.36	−301.88
D_{67}	−1920.03	−1980.29	−1917.73	−1812.31	−3310.21
$\delta_1-\delta_7$	135.66	171.45	57.51	38.09	−40.36
$\delta_2-\delta_7$	29.81	20.99	−4.39	−64.00	84.32
$\delta_3-\delta_7$	142.65	120.58	57.67	1.02	−0.61
$\delta_4-\delta_7$	160.10	141.37	88.54	37.89	69.88
$\delta_5-\delta_7$	161.45	161.76	103.42	53.28	249.20
$\delta_6-\delta_7$	85.13	84.31	41.98	−6.22	88.39

egy works well too. It is possible to increase the usefulness of PCR by choosing a proper criterion (finer than the percentage of trace) to discriminate between important and unimportant linear combinations. Anyway, as the spin system complexity increases, PCR by itself cannot always locate the global minimum and a multistep approach must be used. The procedure proposed proved to be really powerful for analyzing spectra of rigid solutes (geometrical approach), but its efficiency steeply decreases when applied to rigid systems with a bad geometrical guess or a very high number of independent spectral parameters (as in the case of monosubstituted naphthalenes); moreover, when dealing with flexible molecules, a good conformational guess is necessary.

TABLE 8

Dipolar Couplings, D_{ij}, and Chemical Shifts, δ_i, Both in Hertz, Obtained by the Analysis of the Spectrum of Benzaldehyde (11)



D_{12}	−1994.56
D_{13}	−303.71
D_{14}	−85.63
D_{15}	−43.82
D_{16}	−660.56
D_{23}	−43.68
D_{24}	−1083.53
D_{26}	−204.63
D_{36}	−144.59
$\delta_1-\delta_2$	146.34
$\delta_3-\delta_2$	159.72
$\delta_6-\delta_2$	855.55

REFERENCES

1. (*a*) P. Diehl, *in* ''Nuclear Magnetic Resonance of Liquid Cristals'' (J. W. Emsley, Ed.), Reidel, Dordrecht (1985); (*b*) J. W. Emsley and J. C. Lindon, ''NMR Spectroscopy Using Liquid Crystal Solvents,'' Pergamon Press, Oxford (1975); (*c*) C. A. Veracini and M. Longeri, *in* ''Nuclear Magnetic Resonance of Liquid Cristals'' (J. W. Emsley, Ed.), Reidel, Dordrecht (1985).
2. S. Castellano and A. A. Bothner-By, *J. Chem. Phys.* **41,** 3863 (1964).
3. M. Longeri and G. Celebre, *in* ''Encyclopedia of NMR'' (D. M. Grant and R. K. Harris, Eds.), p. 2774, Wiley, New York (1996).
4. D. S. Stephenson and G. Binsch, *J. Magn. Reson.* **37,** 395 (1980).
5. (*a*) R. Laatikainen, *J. Magn. Reson.* **92,** 1 (1991); (*b*) R. Laatikainen, M. Niemitz, J. Sundelin, T. Hassinen, and J. Vepsäläinen, *J. Magn. Reson. A* **120,** 1 (1996); (*c*) PERCH is a copyright by University of Kuopio, Finland.
6. WIN-NMR, a trademark by Bruker Spectrospin, uses a modified version of DAVINS.
7. (*a*) D. S. Stephenson and G. Binsch, *Org. Magn. Reson.* **14,** 226 (1980); (*b*) D. S. Stephenson and G. Binsch, *Mol. Phys.* **43,** 697 (1981).
8. G. Hagèle, M. Engelhardt, and W. Boenigk, ''Simulation und automatisierte Analyse von Kernresonanzspektren,'' VCH, Weinheim (1987).
9. (*a*) A. H. G. Rinnooy Kan and G. T. Timmer, *in* ''Optimization'' (G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, Eds.), p. 631, North-Holland, Amsterdam (1989); (*b*) R. Horst and H. Tuy, ''Global Optimization,'' Springer-Verlag, Berlin (1990); (*c*) R. Horst and P. M. Pardalos (Eds.), ''Handbook of Global Optimization,'' Kluwer Academic, Dordrecht (1994); (*d*) R. H. Byrd, E. Eskow and R. B. Schnabel, Technical Report CU-CS-630-92, Department of Computer Science, University of Colorado, Boulder (1992).
10. (*a*) D. W. Marquardt, *J. Soc. Ind. Appl. Math.* **11,** 431 (1963); (*b*) A. Jones, *Comput. J.* **13,** 301 (1970); (*c*) D. S. Stephenson and G. Binsch, *J. Magn. Reson.* **32,** 145 (1978).
11. (*a*) L. Field, *in* ''Encyclopedia of NMR'' (D. M. Grant and R. K. Harris, Eds.), p. 3172, Wiley, New York (1996); (*b*) D. P. Weitekamp, *Adv. Magn. Reson.* **11,** 111 (1983).
12. (*a*) J. M. Polson and E. E. Burnell, *J. Magn. Reson. A* **106,** 223 (1994); (*b*) J. C. T. Rendell and E. E. Burnell, *J. Magn. Reson. A* **112,** 1 (1995).
13. G. Celebre, F. Castiglione, M. Longeri, and J. W. Emsley, *J. Magn. Reson. A* **121,** 139 (1995).
14. M. Carravetta, Tesi di Laurea, Università della Calabria, Rende (1996).
15. R. Laatikainen and K. Tuppurainen, *Computers Chem.* **14,** 109 (1990).
16. WIN-DAISY is a trademark by Bruker Spectrospin.
17. G. Celebre, G. De Luca, and A. Ferrarini, *Mol. Phys.* **92,** 1039 (1997).
18. G. Celebre, G. De Luca, M. Longeri, and A. Ferrarini, *Mol. Phys.* **83,** 309 (1994).
19. G. Celebre, G. De Luca, M. Longeri, and E. Sicilia, *J. Chem. Inf. Comput. Sci.* **34,** 539 (1994).
20. OSF/Motif is an Open Software Foundation Inc. trademark.
21. B. Vidal, J. Y. Conan, G. Lamaty, and J. Vardin, *Aust. J. Chem.* **41,** 1107 (1988).
22. S. Arumugam, A. C. Kunvar, and C. L. Khetrapal, *Org. Magn. Reson.* **18,** 157 (1982).