Available online at www.sciencedirect.com
ScienceDirect
Chemical Physics Letters 441 (2007) 342–346
CHEMICAL PHYSICS LETTERS
www.elsevier.com/locate/cplett

A simple two-step automatic assignment procedure for complicated NMR spectra of solutes in liquid crystals using genetic algorithms

W. Leo Meerts [a,b], C.A. de Lange [b,*], A.C.J. Weber [c], E.E. Burnell [c]

[a] *Molecular and Biophysics Group, Institute for Molecules and Materials, Radboud University Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands*
[b] *Laser Centre, Vrije Universiteit, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands*
[c] *Chemistry Department, University of British Columbia, 2036 Main Mall, Vancouver, BC, Canada V6T 1Z1*

Received 28 February 2007; in final form 6 May 2007
Available online 10 May 2007

Abstract

The complexity of ^{1}H NMR spectra of solutes in partially ordered solvents such as liquid crystals increases rapidly with the number of spins. Spectra of simple solutes with sufficient symmetry and containing not too many spins (typically $\leqslant 6$) are readily analysed. The analysis of larger spin systems is more difficult, and often impossible. In this Letter, we present the application of a general automated genetic algorithm to solving highly complex spin systems with minimal operator intervention. The robustness of the method is demonstrated for the nine-spin system *p*-bromo-biphenyl, a solute interconverting between two symmetry-related conformations.
© 2007 Elsevier B.V. All rights reserved.

1. Introduction

NMR spectra of molecules partially oriented in an anisotropic liquid-crystalline environment provide detailed information about interactions that cannot be observed in the isotropic phase. An important anisotropic interaction that usually dominates the NMR spectra is the dipole–dipole coupling between every pair of nuclei with spin. From these couplings accurate relative solute geometries and information about solute conformational change can be obtained. Moreover, such studies have contributed greatly to a better understanding of intermolecular interactions between liquid-crystal solvent and solute. In addition, the anisotropic observables are relevant in the context of theory, because they allow a comparison with the results of computer simulations [1,2].

Since the direct dipole–dipole couplings usually dominate the anisotropic NMR spectra, the number of independent intramolecular pair interactions is important. This number is small for molecules with a limited number of spins ($\leqslant 6$) and high symmetry. For solutes with more spins and less symmetry, the number of different dipole–dipole interactions increases rapidly, thus making the analysis of complex NMR spectra notoriously difficult. Many such spectra have therefore remained untractable for decades.

Various means of simplifying the complex spectra arising from such solutes have been developed. We mention selective deuteration in conjunction with heteronuclear decoupling, multiple quantum NMR (MQ–NMR) techniques, computer experiments such as molecular dynamics or Monte Carlo type simulations to estimate the degree of orientational order, and the use of 'magic mixtures' in combination with simple phenomenological models to estimate order parameters. All these methods require considerable

* Corresponding author.
E-mail addresses: Leo.Meerts@science.ru.nl (W.L. Meerts), cdelange@few.vu.nl (C.A. de Lange), elliott.burnell@ubc.ca (E.E. Burnell).
URLs: http://www.leomeerts.nl (W.L. Meerts), http://www.chem.ubc.ca/personnel/faculty/burnell (E.E. Burnell).

0009-2614/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.cplett.2007.05.027

expertise and sophistication, and are very time-consuming [2,3].

Attempts to develop automated computer routines to efficiently search parameter space in order to find unique values for the spectral parameters that reproduce an observed NMR spectrum were made at an early stage, initially for isotropic liquids. Such an alternative approach to the more conventional methods of spectral analysis was first suggested by Diehl et al. [4,5]. In a similar spirit Stephenson and Binsch developed a basic algorithm which relies on a matrix method derived from a general formulation of the least-squares problem [6]. This approach was implemented in the computer programme DAVINS (Direct Analysis of Very Intricate NMR Spectra) [7], and was reasonably successful. Subsequently, the more challenging task of applying the method to the NMR spectra of solutes in anisotropic liquids was undertaken. The programme DANSOM (Direct Analysis of NMR Spectra of Oriented Molecules) was first used in the analysis of the spectra of a number of allyl halides [8] and cyclopentene [9]. However, the limitations of the method in these applications became apparent, because background corrections proved troublesome and for the more complicated cases, the intervention of skilled operators during the fitting procedure was usually required. Later, the Cosenza group in Italy achieved some notable successes, despite the need for significant operator intervention to avoid the problem of trapping in local minima [10,11].

In a separate development the use of a genetic algorithm (GA) for the analysis of NMR spectra in integrated form of solutes in liquid crystals was reported in a short Letter that gave little detail [12]. It appears that trapping in local minima during the convergence process was problematic. In a follow-up paper [13] it was made clear that GA-fitting methods were employed to obtain a first approximation to the experimental NMR spectrum. The best GA fit obtained was then used as a starting point for the assignment of individual lines and subsequent least-squares adjustment and refinement. Apart from these two publications, we have found no further applications of this method in the literature.

In this Letter, we apply a sophisticated and very robust genetic algorithm to fit highly complex anisotropic NMR spectra that avoids problems commonly encountered in the automated analysis of NMR spectra of solutes in liquid-crystalline solvents. In particular, the GA-method is inherently capable of circumventing situations where the convergence procedure gets trapped in local minima, without ever reaching the desired global minimum. Our routine is extremely efficient, requires minimal operator interference, and constitutes essentially an intelligent parameter search method. In the absence of reasonable initial guesses for order parameters or dipole–dipole couplings and chemical shielding parameters, convergence is still reached and excellent fits are obtained. Sometimes, additional information about order or spectral parameters is available which can then be used to narrow the search range and accelerate the convergence process. Owing to the enormous advances in computer technology and parallel processing, the analysis of complex anisotropic NMR spectra now becomes eminently feasible.

2. Theoretical background

The Hamiltonian (in Hz) of a molecule dissolved in a uniaxial nematic liquid crystal is given by:

$$H = \sum_{\mu} \nu_{\mu} I_{Z;\mu} + \sum_{\mu<\nu} J_{\mu\nu} \mathbf{I}_{\mu} \cdot \mathbf{I}_{\nu} + \sum_{\mu<\nu} D_{\mu\nu} (3 I_{Z;\mu} I_{Z;\nu} - \mathbf{I}_{\mu} \cdot \mathbf{I}_{\nu}) \tag{1}$$

where ν_{μ} is the resonance frequency of nucleus μ and $J_{\mu\nu}$ and $D_{\mu\nu}$ are the indirect spin–spin and the direct dipole–dipole couplings between nuclei μ and ν. The chemical shielding term ν_{μ} can have both isotropic and anisotropic parts, whereas the $D_{\mu\nu}$ terms are completely anisotropic and hence do not contribute to the familiar spectra of isotropic liquids. We omit effects of the anisotropy in $J_{\mu\nu}$ which are to modify slightly the observed values of $D_{\mu\nu}$ and which are negligible for proton–proton couplings. Since we only concern ourselves with nuclear spins 1/2 the quadrupolar term has been omitted.

In order to describe the orientational order of the solute, the Saupe order matrix S_{kl} is employed, with S_{kl} defined as averages over reorientational motion of the second-order Legendre polynomials:

$$S_{kl} = \frac{1}{2} < 3 \cos\theta_{Z,k} \cos\theta_{Z,l} - \delta_{kl} > \tag{2}$$

where $\theta_{Z,k}$ is the angle between the molecule-fixed k-axis and the space-fixed Z-axis which lies along the liquid-crystal director. The S-tensor is traceless and symmetric, hence at most five independent parameters exist. Symmetry reduces this number. The S-tensor components can only take the following values:

$$\begin{aligned} -1/2 \leqslant S_{xx}, S_{yy}, S_{zz} \leqslant 1 \\ -3/4 \leqslant S_{xy}, S_{xz}, S_{yz} \leqslant 3/4 \end{aligned} \tag{3}$$

For the common situation in nematic liquid crystals for which the director lies along the magnetic field direction we can express the direct dipole–dipole interaction as:

$$D_{\mu\nu} = -\frac{h\gamma_{\mu}\gamma_{\nu}}{4\pi^2} \left\langle r_{\mu\nu}^{-3} \right\rangle \sum_{k,l} S_{kl} \cos\theta_k \cos\theta_l \tag{4}$$

where θ_k is the angle between the k molecule-fixed Cartesian axis and the vector connecting nuclei μ and ν at internuclear distance $r_{\mu\nu}$. The angular brackets denote averaging over all internal motions [2,14].

The Hamiltonian of Eq. (1) is an excellent predictor of observed NMR spectra in anisotropic liquids. Such spectra show great complexity in the case of solutes with a low degree of symmetry and many spins 1/2. In the case of the liquid crystal itself there are so many protons and hence so many overlapping transitions that fine structure is no longer observed.

In the case of an experimental spectrum measured for spin 1/2 nuclei (often ^{1}H) in a solute in a nematic phase the objective of spectral analysis is to find the spectral parameters that can faithfully reproduce the spectrum when the Hamiltonian of Eq. (1) is employed. The indirect couplings are usually taken from the corresponding isotropic spectra of the solute or from the literature, and kept fixed. The purpose of the analysis is then to find the dipole–dipole couplings and chemical shielding parameters that reproduce the experimental spectrum. This approach, although leading to a perfect fit in principle, is hampered by the fact that the number of adjustable parameters can be very large, and that little previous knowledge about them is available. The associated search space may then be too extensive to obtain solutions in a reasonable time.

From Eq. (4) it is clear that the dipolar couplings depend on geometrical information about the solute and on a maximum of five independent order parameters. If we assume a relatively crude geometrical structure for the molecule, the order parameters are now unknowns, but parameter space has been greatly reduced. However, this method comes at a price, because employing a trial geometry leads to a Hamiltonian that is not capable of reproducing the experimental spectrum faithfully. Hopefully, a fitting procedure in which only the order parameters and the chemical shielding parameters are varied leads to a correspondence between calculated and experimental spectra close enough to recognize a sufficient number of common features. If this is the case, good estimated values for the dipolar couplings can be derived that serve as trial parameters in a fitting procedure based on Eq. (1) in which all the dipole–dipole couplings and chemical shielding parameters are varied.

3. The GA method

A fit using genetic algorithms (GA) mimics the concepts of natural reproduction and selection processes. For a detailed description of the GA the reader is referred to the original literature on evolutionary or genetic algorithms [15–17]. A detailed description of the GA used in this investigation can be found in [18–20].

The molecular parameters are encoded in binary or real type, each parameter to be optimized representing a gene. The vector of all genes, which contains all molecular parameters, is called a chromosome. In an initial step the values of all parameters are set to random values between lower and upper limits which are chosen by the user. The quality of the solutions is then evaluated by a fitness function. A proper choice of this fitness function is of vital importance for the success of the GA convergence. In Refs. [18,19] the fitness function F_{fg} has been defined as:

$$F_{fg} = \frac{(\boldsymbol{f},\boldsymbol{g})}{\|\boldsymbol{f}\|\|\boldsymbol{g}\|} \tag{5}$$

Here $\boldsymbol{f}$ and $\boldsymbol{g}$ are the vector representations of the experimental and calculated spectrum, respectively. The inner product $(\boldsymbol{f}, \boldsymbol{g})$ is defined with the metric $\boldsymbol{W}$ which has the matrix elements $W_{ij} = w(|j-i|) = w(r)$ as

$$(\boldsymbol{f},\boldsymbol{g}) = \boldsymbol{f}^T\boldsymbol{W}\boldsymbol{g} \tag{6}$$

and the norm of $\boldsymbol{f}$ as $\|\boldsymbol{f}\| = \sqrt{(\boldsymbol{f},\boldsymbol{f})}$; similar for $\boldsymbol{g}$. For $w(r)$ we used a triangle function [18] with a width of the base of Δw:

$$w(r) = \begin{cases} 1 - |r|/\left(\frac{1}{2}\Delta w\right) & \text{for } |r| < \frac{1}{2}\Delta w \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The above defined fitness function is able to smooth in a controlled way the error landscape, and therefore allows the GA to locate the global minimum. The width of the function $w(r)$ critically determines the ability of the GA to converge to the global minimum and also the speed of convergence. The smoothing of the error landscape allows to sense regions far from the minimum. The GA convergence is obtained in a well-defined procedure. At first, the function $w(r)$ should be chosen relatively broad; $\Delta w \approx 15$–20 times the line width of an individual NMR transition in the spectrum. In this way, a first set of parameters is obtained, which still has to be refined. This is done by decreasing Δw and narrowing the limits of the parameter space to be searched in the fit. Decreasing Δw improves the accuracy of the molecular parameters obtained from the fit, while narrowing the parameter space leads to an improved sampling in the region of the minimum. This of course is critical in the procedure, but can in most cases be done automatically. In a final calculation Δw is set to zero. Usually full GA convergence to the best set of parameters is achieved by narrowing Δw to zero in one or two steps.

One optimization cycle, including evaluation of the fitness of all solutions, is called a generation. Pairs of chromosomes are selected for reproduction and their information is combined via a crossover process. Since crossover combines information from the parent generations, it basically explores the error landscape. The value of a small number of bits is changed randomly by a mutation operator. Mutation can be viewed as exploration of the fitness surface. The best solutions within a generation are excluded from mutation. This elitism prevents already good solutions from being degraded. Mutation prevents the calculation from being trapped in local minima, as is often the case with more conventional fitting routines.

The GA described above together with the library PGAPack version 1.0 [21] is very well suited for massive parallel computation. The larger the starting population for a given problem, the faster is the convergence in terms of generations. On the other hand, a large starting population tends to slow the algorithm down due to computational demands. A large number of fast processors will circumvent this problem and will lead to fast and straightforward assignments of the spectra. The computational speed scales inversely proportional to the number of processors. Therefore, this kind of algorithm is perfectly applicable for parallel processing. Modern computer cluster systems

make these calculations very feasible, even for complex spectra.

The automatic GA-fitting of anisotropic NMR spectra represents one step up in complexity compared to the fitting of high-resolution laser induced fluorescence (LIF) spectra [18–20] that was performed with the same GA procedure before. The reason is that the LIF spectra often show some kind of regularity. This is in general completely absent in anisotropic NMR spectra because of the complex nature of the spin Hamiltonian.

4. Results

As an illustrative example we focus on the nine-spin system *p*-bromo-biphenyl (see Fig. 1) dissolved in three different nematic phases, *viz*. Merck ZLI 1132 which is a eutectic mixture of 1,4-(*trans*-4′-*n*-alkylcyclohexyl)-cyanobenzene (alkyl = propyl, pentyl, heptyl) and 1,4-(*trans*-4′-*n*-pentylcyclohexyl)-cyanobiphenyl, N-(4-ethoxybenzylidene)-4′-*n*-butylaniline (EBBA), and a 'magic mixture' with composition 55 wt% Merck ZLI 1132 / EBBA. The orientational order of the molecule can be described by the *S*-parameters S_{xx}, S_{zz} and S_{yz}. Since for the related biphenyl there are only two order parameters, S_{yz} is taken to be close to zero. Inter-ring couplings are averaged by rotation about the central CC bond, and thus the molecule has 16 independent dipolar couplings and five chemical shielding parameters. NMR spectra were obtained at room temperature (298 K) by standard techniques on a Bruker Avance 400 MHz Inverse spectrometer.

In our GA-fit, a two-step procedure was implemented. In the first step, a relatively crude geometrical structure of the solute is assumed and two independent order parameters, chemical shielding parameters and the dihedral angle between the two rings are varied in a GA-fitting procedure until the best correspondence with the experimental spectrum is obtained. Owing to the approximations made in this step, complete agreement cannot be obtained. However, in practice the choice of geometry does not have to be perfect in order to obtain sufficient common features between calculated and experimental spectra. Approximate dipolar couplings are calculated in this step. In a second GA-calculation procedure, these approximate couplings are used as starting values for a fit in which all dipole–dipole couplings and chemical shielding parameters are varied until convergence is reached. Exceptionally good agreement is obtained in step two because no constraints are placed on the dipolar coupling and chemical shielding parameters. For this problem a modern cluster of PC's was used. The CPU time required amounted to ~64 min per GA-run, which with 16 processors came to 4 min wall clock time per run. A typical result is presented in Fig. 2.

The robustness of the GA-fitting method has been demonstrated for solutes that belong to various categories of increasing complexity [22]: rigid solutes with a single conformation (e.g., the eight-spin systems azulene and biphenylene), solutes interconverting between a number of symmetry-related conformations (e.g., *p*-bromo-biphenyl), and solutes undergoing conformational change between non-symmetry-related conformations (e.g., the ten-spin system butane).

At this point, it is interesting to consider whether there are fundamental limits to the applicability of GA-fitting.

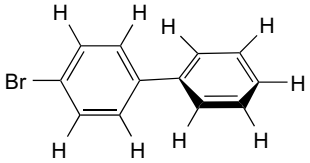

Fig. 1. The molecule *p*-bromo-biphenyl has the *x*-axis along the central CC-bond, with the *y*- and *z*-axes dissecting the dihedral angles.

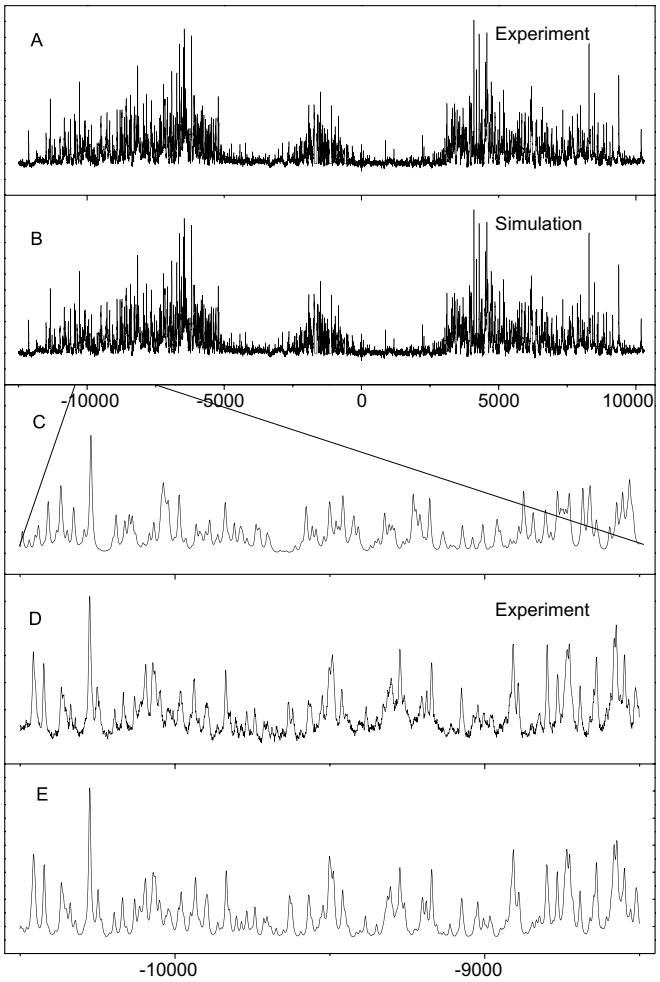

Fig. 2. ^{1}H NMR spectrum of *p*-bromo-biphenyl dissolved in 'magic mixture'. The experimental spectrum (A) is compared with the final fit obtained by varying 16 dipolar couplings and five chemical shieldings (B); in the next frames the spectrum is enlarged to show the fit obtained by varying 2 *S*-parameters, five chemical shieldings and the dihedral angle (C), the experimental spectrum (D), and the fit obtained by varying 16 dipolar couplings and five chemical shieldings (E). All horizontal scales are in Hz.

It is well known that liquid-crystal molecules themselves, that usually contain 20 or more spins 1/2, show anisotropic NMR spectra in which separate transitions can no longer be observed and where extensive overlap leads to broad, structureless spectral features. In such a case, the combination of a very large number of independent dipolar couplings and an NMR spectrum without distinct features would be prohibitive for the successful application of GA-fitting. This is not to say that spectral overlap per se presents an insurmountable problem. However, the success of the GA-fitting procedure hinges on the number of independent parameters in combination with the presence of well-defined structure in the spectrum. At present, we are attempting to fit the well-resolved but extremely complex spectrum of pentane (12 spins 1/2) in a liquid crystal. The success or failure of this attempt will provide some further answers toward the applicability limits of GA-fitting methods in anisotropic NMR.

5. Conclusions

With the application of the current GA-fitting algorithm to the analysis of highly complex NMR spectra of solutes in liquid-crystalline environments, the study of such systems has obtained a new lease of life. Owing to the extensive use that is made of parallel processing and advanced computer technology, and with the development of new intelligent search methods, the present two-step GA-method is very robust, requires minimal operator intervention, and can be considered to be automated in the true sense of the word. The results of our GA-fitting technique show that both positions and intensities of the calculated NMR spectral transitions perfectly match the experimental transitions. Moreover, it should be noted that no assignments of individual transitions and subsequent least-squares refinements are needed. The assignments of all individual transitions follow directly from our final GA fit. This is in marked contrast to the GA method used previously by Takeuchi et al. [12,13].

Until now, the extreme difficulties associated with the spectral analysis of anisotropic NMR spectra of many-spin systems provided a bottleneck that hampered progress in this field. With the removal of this bottleneck, a significant advance in the study of such complex NMR spectra and the physics and chemistry underlying them is to be expected.

Acknowledgements

E.E.B. acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada. W.L.M. thanks the National Computer Facilities of the Netherlands Organization of Scientific Research for a Grant on the Dutch supercomputing facility SARA.

References

[1] E.E. Burnell, C.A. de Lange, Prediction from molecular shape of solute orientational order in liquid crystals, Chem. Rev. 98 (1998) 2359.
[2] E.E. Burnell, C.A. de Lange (Eds.), NMR of Ordered Liquids, Kluwer Academic, Dordrecht, The Netherlands, 2003.
[3] E.E. Burnell, C.A. de Lange, Solid State NMR 28 (2005) 73.
[4] P. Diehl, S. Sýkora, J. Vogt, J. Magn. Reson. 19 (1975) 67.
[5] P. Diehl, J. Vogt, Org. Magn. Reson. 8 (1976) 638.
[6] D. Stephenson, G. Binsch, J. Magn. Reson. 37 (1980) 395.
[7] D. Stephenson, G. Binsch, J. Magn. Reson. 37 (1980) 409.
[8] D. Stephenson, G. Binsch, Org. Magn. Reson. 14 (1980) 226.
[9] D. Stephenson, G. Binsch, Mol. Phys. 43 (1980) 697.
[10] F. Castiglione, M. Carravetta, G. Celebre, M. Longeri, J. Magn. Reson. 132 (1998) 1.
[11] F. Castiglione, G. Celebre, G. De Luca, M. Longeri, J. Magn. Res. 142 (2000) 216.
[12] H. Takeuchi, K. Inoue, Y. Ando, S. Konaka, Chem. Lett. 11 (2000) 1300.
[13] K. Inoue, H. Takeuchi, S. Konaka, J. Phys. Chem. A 105 (2001) 6711.
[14] R.Y. Dong, Nuclear Magnetic Resonance of Liquid Crystals, Second ed., Springer-Verlag, New York, 1994.
[15] J.H. Holland, Adaption in Natural and Artificial Systems, MI: The University of Michigan Press, Ann-Arbor, 1975.
[16] D.E. Goldberg, Genetic Algorithms in Search, Optimisation and Machine Learning, Addison-Wesley, Reading, MA, 1989.
[17] I. Rechenberg, Evolutionsstrategie – Optimierung technischer Systeme nach Prinzipien der biologischen Evolution, Frommann-Holzboog, Stuttgart, 1973.
[18] J.A. Hageman, R. Wehrens, R. de Gelder, W.L. Meerts, L.M.C. Buydens, J. Chem. Phys. 113 (2000) 7955.
[19] W.L. Meerts, M. Schmitt, G. Groenenboom, Can. J. Chem. 82 (2004) 804.
[20] W.L. Meerts, M. Schmitt, Int. Rev. Phys. Chem. 25 (2006) 253.
[21] D. Levine, PGAPack V1.0, PgaPack can be obtained via anonymous ftp from: ftp://ftp.mcs.anl.gov/pub/pgapack/pgapack.tar.z (1996).
[22] A.C.J. Weber, E.E. Burnell, W.L. Meerts, C.A. de Lange, unpublished.