Analysis of Complex High-Resolution NMR Spectra by Sophisticated Evolutionary Strategies

W. Leo Meerts[1,2], Cornelis A. de Lange[2], Adrian C.J. Weber[3] & E. Elliott Burnell[4]

[1] *Radboud University, Nijmegen, The Netherlands*
[2] *Vrije Universiteit, Amsterdam, The Netherlands*
[3] *Brandon University, Brandon, Manitoba, Canada*
[4] *University of British Columbia, Vancouver, British Columbia, Canada*

The NMR spectra of orientationally ordered molecules increases rapidly in complexity with the number of coupled nuclear spins. For example, the spectrum of *n*-pentane as solute in a nematic liquid-crystal solvent (12 coupled proton spins) consists of roughly 20,000 transitions, many of which overlap. The analysis of such complicated spectra by line-assignment techniques is fraught with difficulty. However, application of the ideas of genetic evolution via evolutionary strategies makes possible the analysis of spectra, which are quite probably not solvable by ordinary methods. In particular, the covariance matrix adaptation evolution strategy uses the idea of mutation and the results of previous trials to give solutions to quite complicated spectra. The techniques described herein have already led to solution of NMR spectra that were not achieved with older techniques, and show promise for cracking even harder problems in the future. An important aspect with NMR spectra of solutes in liquid-crystal solvents is the presence of a broad underlying baseline from the solvent protons. An automatic baseline removal method is described.

Keywords: genetic algorithm, evolutionary strategy, evolutionary algorithm, NMR, spectral analysis, baseline removal, liquid crystal, ordered liquids, anisotropic

Introduction

NMR of molecules in the liquid phase has developed over the years into a unique tool for the characterization of chemical species. Despite its limited sensitivity, the technique has proved invaluable and is used extensively. The reasons for its success are threefold. First, the Hamiltonian describing the NMR spectrum in the isotropic liquid phase is simple, well-known, and given by[1,2]

$$\hat{H} = -\frac{B_Z}{2\pi} \sum_i \gamma_i (1 - \sigma_i^{\text{iso}}) \hat{I}_{i,Z} + \sum_{i<j} J_{ij}^{\text{iso}} \hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j \qquad (1)$$

with B_Z the external magnetic field in the Z direction, γ_i the magnetogyric ratio of nucleus i, σ_i^{iso} the isotropic part of the chemical shielding of nucleus i, and J_{ij}^{iso} the isotropic part of the indirect spin−spin coupling between nuclei i and j. The operators $\hat{\mathbf{I}}$ signify nuclear angular momentum operators. Second, this simple Hamiltonian is an excellent predictor for NMR spectra, because the reservoir of nuclear spins is very weakly coupled to spatial coordinates, making perturbations to the Hamiltonian negligible. Third, although indirect couplings in principle exist between any pair of nuclei, the magnitudes of these couplings decrease rapidly with distance. Usually, indirect couplings over four or more bonds are much smaller than typical NMR line widths that result from relaxation processes in the liquid.

When solutes are dissolved in liquid-crystal solvents that have cylindrically symmetric phases that are apolar, the molecular tumbling of the solute molecules is no longer isotropic. In addition to terms that occur in the isotropic Hamiltonian, a number of new anisotropic terms arise, and, in the limit of high magnetic field, the Hamiltonian becomes[3−5]:

$$\hat{H} = -\frac{B_Z}{2\pi} \sum_i \gamma_i (1 - \sigma_i^{\text{iso}} - \sigma_i^{\text{aniso}}) \hat{I}_{i,Z}$$
$$+ \sum_{i<j} J_{ij}^{\text{iso}} \hat{\mathbf{I}}_i \cdot \hat{\mathbf{I}}_j$$
$$+ \sum_{i<j} (2D_{ij}^{\text{aniso}} + J_{ij}^{\text{aniso}}) \left[\hat{I}_{i,Z} \hat{I}_{j,Z} - \frac{1}{4} \left(\hat{I}_{i,+} \hat{I}_{j,-} + \hat{I}_{i,-} \hat{I}_{j,+} \right) \right]$$
$$+ \sum_i \frac{q_i^{\text{aniso}}}{4I_i (2I_i - 1)} (3\hat{I}_{i,Z}^2 - \hat{\mathbf{I}}_i^2) \qquad (2)$$

Here, σ_i^{aniso} is the anisotropic contribution to the chemical shielding, J_{ij}^{aniso} the anisotropic contribution to the indirect spin coupling between nuclei i and j, and D_{ij}^{aniso} the anisotropic direct dipolar spin coupling between nuclei i and j. The final term with q_i^{aniso} signifies the quadrupolar coupling contributions for nuclei i with spin $I \geq 1$. In this article, we shall focus on ^{1}H magnetic nuclei. The indirect coupling between two ^{1}H nuclei is to a very good approximation isotropic, and

the anisotropic contribution is therefore commonly neglected. Also, the quadrupolar coupling vanishes for ^{1}H nuclei.

A particularly challenging situation arises with molecules that undergo conformational change. Usually, the timescale for conformational change is much faster than the NMR timescale. Hence, NMR spectra are observed that are an average over all conformations. In particular, the observed dipolar couplings can be expressed as[6−8]

$$D_{ij} = \sum_n p^n \sum_{k,l} d_{kl,ij}^n S_{kl}^n \qquad (3)$$

where we drop the superscript aniso, where p^n is the probability of conformation n, and $d_{kl,ij}^n$ is given by

$$d_{kl,ij}^n = -\frac{h\gamma_i\gamma_j}{4\pi^2}(\cos\theta_{ij,k}^n \cos\theta_{ij,l}^n/r_{n,ij}^3) \qquad (4)$$

with $\cos\theta_{ij,k}^n$ the cosine of the angle between the ij direction in the molecule in conformation n, and k a conformer-fixed axis.

$$S_{kl}^n = \left\langle \frac{3}{2}\cos\theta_{k,Z}^n \cos\theta_{l,Z}^n - \frac{1}{2}\delta_{kl} \right\rangle \qquad (5)$$

are the Saupe order parameters for each conformation n, where $\cos\theta_{k,Z}^n$ is the cosine of the angle between the molecular k direction in conformation n and the space-fixed Z direction of the external magnetic field B_Z. There is a maximum of five independent order parameters for every conformation, and molecular symmetry can reduce this number.

So far, we have assumed that the direction of the mesogen director is parallel to that of the external magnetic field, as is the case in most nematics. When there is an angle Ω between the director and the magnetic field, there is a factor $P_2(\cos\Omega)$ that scales the dipolar couplings in equation 3.[3]

A crucial issue in NMR liquid-crystal spectroscopy is how to describe the average degree of the orientational order of a molecule that undergoes interchange among several symmetry-unrelated conformations. Initially, attempts were made to

assume an average molecule, which then possesses a single set of at most five orientational order parameters. Unfortunately, this idea has been shown to be fallacious, because every conformation requires its own set of at most five orientational order parameters.[6−8] For a molecule with a significant number of conformations, the number of *a priori* unknown order parameters therefore can be quite large. This situation presents a serious bottleneck for the analysis of complex spectra.

In contrast to isotropic NMR spectra, spectra of solutes with a large number of spins dissolved in liquid crystals are usually extremely complicated as shown in Figure 1 for the series of solutes from methane to *n*-pentane.[9−14] The reason is that the anisotropic direct dipolar couplings that exist between any pair of nuclei tend to be much larger than the NMR line widths, thus causing spectra with many transitions that are observed separately. The analysis of such often extremely complex spectra is the topic of this article. Although the theory developed so far equally holds for mesogenic molecules themselves, in this article we shall focus on the analysis of the often extremely complex spectra of solutes in cylindrically symmetric apolar anisotropic phases.

The analysis of NMR spectra of solutes in both isotropic and nematic phases has a long history. For simple solutes, the ^{1}H NMR spectrum can be easily obtained by calculating the matrix elements of the corresponding Hamiltonian, followed by matrix diagonalization to calculate the nuclear spin energy levels. By computing the transition probabilities between all energy levels, a theoretical spectrum is obtained. Comparing the theoretical spectrum with the experimental one, transitions can be assigned and values for chemical shieldings as well as indirect and direct coupling constants can be extracted by hand.[1,2] For somewhat larger solutes, molecular symmetry can be employed to block out and simplify the Hamiltonian matrix and to calculate the ^{1}H NMR spectrum as before.[1,2] For larger solutes without symmetry, this analytical approach fails. One then has to rely on computer methods to calculate
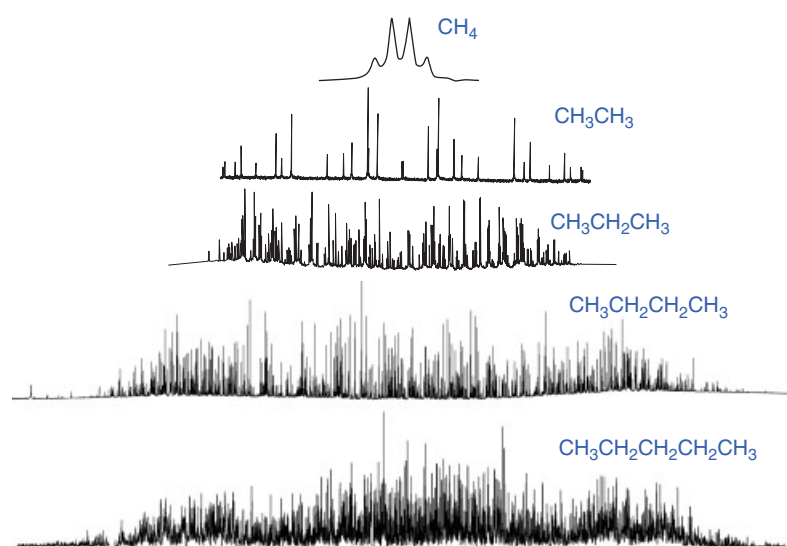

Figure 1. ^{1}H NMR spectra of orientationally ordered methane, ethane, propane, *n*-butane, and *n*-pentane in nematic solvents

a trial spectrum from a set of guessed spectral parameters. This trial spectrum is then compared to the experimental one, and an attempt is made to assign transitions. Next, with least-squares methods the spectral parameters are varied to obtain the best correspondence between the calculated and measured spectra. When the initial assignment is correct, this procedure converges. If not, new starting parameters must be tried and the procedure repeated.[15] For solutes with up to approximately eight spins, the spectra increase in complexity with the number of spins, and this method becomes extremely time consuming. Many months of calculating and assigning trial spectra were usually invested, and spectral analysis at that level was a true art, which required great experience and expertise. For solutes with eight or more spins, the number of transitions increases tremendously (about 20,000 for n-pentane with 12 ^{1}H spins[13]), many lines lie close together, and the assignment problem looks insurmountable. Despite the large efforts in many NMR laboratories put into measuring and analyzing very complex spectra of solutes with many spins, it seemed that the range of applicability of NMR liquid-crystal spectroscopy had reached its limit.

In addition to attempts to analyze complex single-quantum spectra, various experimental means of simplifying such spectra have been developed. We mention selective deuteration in combination with heteronuclear decoupling, multiple-quantum (MQ)-NMR, and the use of 'magic mixtures' in combination with simple phenomenological models to estimate order parameters. All these methods have in common that they require considerable expertise and sophistication, and are very time consuming.[7,8,16] A detailed discussion of these methods is outside the scope of this article.

Since the days that computers were employed to aid in understanding NMR spectra, the concept of automated spectral analysis has been explored. Ideally, an automated computer algorithm should search the parameter space efficiently in order to find unique values for the spectral parameters that reproduce an observed NMR spectrum without or with minimal operator intervention. Such an automated approach was first suggested by Diehl *et al.*[17,18] In a similar spirit, Stephenson and Binsch[19] developed a basic algorithm for isotropic NMR spectra that relies on a matrix method derived from a general formulation of the least-squares problem. This approach was implemented in the algorithm DAVINS (direct analysis of very intricate nuclear magnetic resonance spectra)[20] and was reasonably successful. Subsequently, the more challenging task of applying automated analysis to the NMR spectra of solutes in nematic solvents was undertaken. The program DANSOM (direct analysis of nuclear magnetic resonance spectra of oriented molecules) was first used in the analysis of spectra of a number of alkyl halides[21] and cyclopentene.[22] In a review article published in 2007, the work carried out with DAVINS and DANSOM was summarized.[23] However, the method met with serious difficulties, because background corrections proved troublesome and, for more complicated cases, operator intervention was often required. Later, the Cosenza group in Italy achieved some notable successes, despite the need for significant operator intervention to avoid the problem of getting

trapped in local minima.[24,25] In an article published in 2007, the strategies employed in spectral analysis were reviewed.[26]

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 paper that gave little detail.[27] Apparently, trapping in local minima presented problems. In a follow-up paper,[28] it was made clear that GA fitting methods were only employed to obtain a first approximation to the NMR spectrum. The parameters associated with the best GA fit calculated were then used as a starting point for the assignment of individual lines, followed by the more conventional least-squares refinement. Apart from these two papers (and those from our own work, discussed later), we have found no further applications of the GA method in the literature.

In recent years, we have exploited the enormous advances in computer technology and parallel processing, as well as recent developments in the field of GAs and evolution strategies (ESs), to develop robust new approaches toward solving very complex spectra of larger solutes dissolved in nematic solvents.[12−14,29,30] Problems previously encountered with less sophisticated methods such as getting trapped in local minima without ever reaching the desired global minimum, problems with accounting for spectral background, and the need for frequent operator intervention can now be efficiently circumvented. Our routine essentially represents an efficient and intelligent search of an extensive parameter space.

To limit the search range and to speed up the convergence process, ideally one should start with reasonable guesses of shielding parameters and of direct dipolar couplings. Alternatively, guesses of order parameters in combination with reasonable estimates of molecular geometries (and conformer probabilities for flexible molecules) can be used to generate the guesses of direct dipolar couplings. Indirect couplings are usually known well enough from NMR in the isotropic phase. Guesses for order parameters for each conformer can often be obtained from simple phenomenological models.[7,8] Direct dipolar couplings can be estimated from a relatively crude knowledge of the geometries of these conformations and by averaging over all conformations.[6,12,13] Even in the absence of reasonable initial guesses for the parameters, convergence is still often reached and excellent final fits are obtained. It must be noted that the spectral fitting is facilitated by the acquisition of high-quality NMR spectra with narrow lines throughout; in order to achieve this, attention to the removal of concentration and temperature gradients (which lead to inhomogeneous broadening of lines in the wings of the spectra) is important.

In this article, we shall discuss modern ESs and show how the application of these novel methods to liquid-crystal NMR is eminently successful and has given a completely new impetus to an old field of research.

Theory of Evolutionary Algorithms

In the past decades, computational tools have been developed that are able to handle complex multiparameter optimizations intelligently and within an acceptable investment in computer time. The use of evolutionary algorithms (EAs) for solving
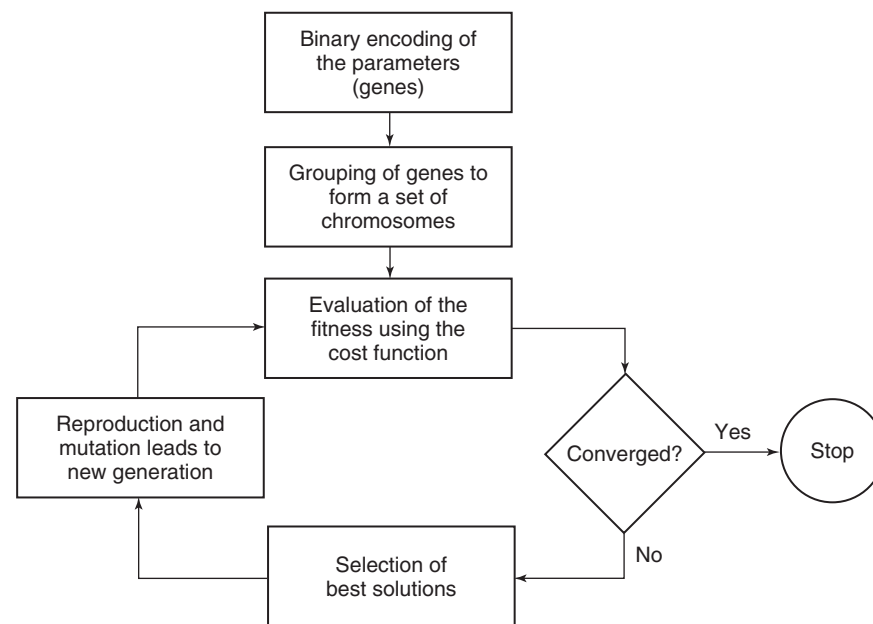
Figure 2. Schematics of the GA process

such highly nonlinear and complex processes in science and engineering has become widespread. Although they are conceptually simple, their ability in avoiding local optima and instead finding the global optimum is remarkable and makes them suitable to handle complex optimization problems.

EAs represent a set of general-purpose probabilistic search methods based on natural evolution. These algorithms mimic the concepts of natural reproduction and selection processes. The basic idea in EAs is to create an artificial environment that encodes the search problem into biology-like terms.

From a spectroscopic point of view, the EA approach reproduces this behavior to fit an experimental spectrum with a model based upon transitions between spin energy levels calculated as eigenvalues of the Hamiltonian such as $\hat{H}$ given by equation (2). Each of the molecular parameters in $\hat{H}$ can be thought of as a gene, encoded in a binary or real type. The vector of all genes, which contains all molecular parameters, is called a *chromosome*.

Three main different EAs have been developed over the years: the GA,[31] the ES,[32] and the evolutionary programming.[33]

For the automatic assignment and analysis of high-resolution NMR spectra, we have initially used the GA.[34] However, lately we have been mainly using an ES in our procedures.

The Genetic Algorithm

A detailed description of the GA used in the automatic assignment and fitting of the spectra described in this article can be found in the literature.[34]

The molecular parameters are encoded in binary or real type, each parameter to be optimized representing a gene. As mentioned previously, 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 then is evaluated by a fitness function. A flowchart of the procedure is shown in Figure 2.

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 fitness 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 Evolution Strategy

The ES algorithm starts with one or more parents. A parent is a trial solution that corresponds to a set of parameters like in the GA. From this parent, an offspring of multiple children is generated. The quality or performance of these children is checked and, depending on the strategy, the next parent is generated. There are several different strategies for the generation of the offspring as well as the next parent. For a more detailed description, see the literature.[32]

The offspring is created from the parents in a mutative step-size control. A drawback of the standard ES is that the mutations of the decision and the strategy parameters are subject to independent random processes. If, for example, an individual with a large step size undergoes only a very small change in the decision parameters and this small change turns out to yield a high fitness, the large step size will be passed on to the next generation. As a result, the fitness in the next mutations may worsen. This problem is resolved in derandomized (DR) algorithms, which make the random mutations in decision
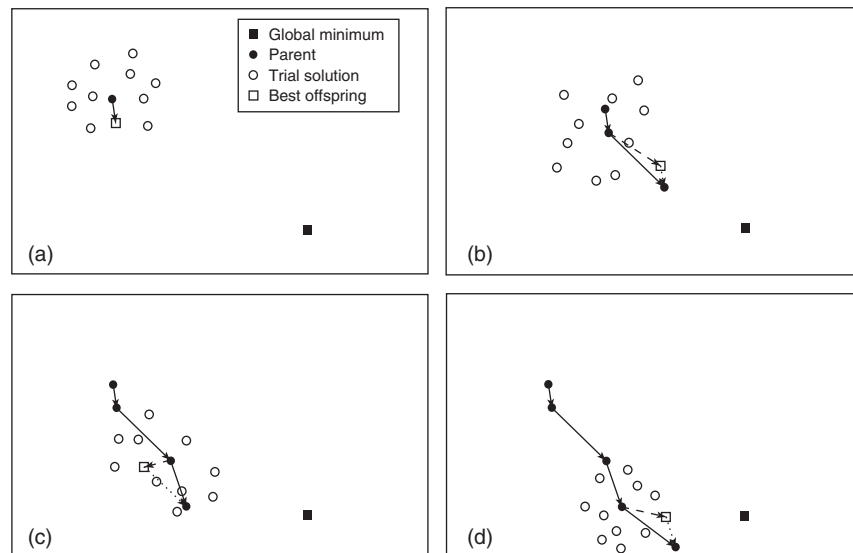
Figure 3. The first four generations of an evolution strategy: (a) An initial population is generated, and the best offspring is used as the next parent. (b) The offspring is spread over a larger area in the second generation because of the relatively large step made in the previous generation. The vector from the parent to the best offspring (dashed line) is combined with the (shortened) mutation vector of the last generation (dotted line) to generate the new parent (solid line). (c) Owing to the correlation between the past two mutations, the search range has been extended again in the general direction of both mutations while it has been limited in the perpendicular direction. The best offspring is now a local minimum. The memory effect of the evolutionary algorithm, which incorporates past mutation vectors into the calculation of the next parent, helps to overcome the local minimum, and the next parent is still closer to the global minimum. (d) The barrier between the local and global minima has been overcome, and the optimization is progressing toward the global minimum. (Reprinted with permission from W.L. Meerts, C.A. de Lange, A.C.J. Weber and E.E. Burnell, J. Chem. Phys., 130, 044504, (2009). Copyright 2009, American Institute of Physics)

and strategy parameters dependent on each other. This idea was implemented initially as DR1 and soon improved by the concept of accumulated information,[35] which will be called DR2. The history of the optimization is recorded, and the evolution of the mutation ellipsoid is partially governed by past successful mutations.

A further improvement was achieved by Hansen and Ostenmeier[36] with the covariance matrix adaptation evolution strategy (CMA-ES). It turns out to be a particularly reliable and highly competitive EA for local optimization and, surprisingly at first sight, also for global optimization.[37] The CMA-ES does not leave the choice of strategy parameters open to the user −only the population size can be set. Finding good strategy parameters is considered to be part of the algorithm design.

Figure 3 depicts the first four generations of an ES and demonstrates the effect of the chosen strategy. In general, the ESs converge faster and are more robust than the GA.

The Fitness Function for the Analysis of Spectra

A proper choice of the fitness function is of vital importance for the success of the EA convergence. In the literature,[38] a fitness function F_{fg} was defined as

$$F_{fg} = \frac{(f, g)}{\|f\| \|g\|} \tag{6}$$

Here, f and g are the vector representations of the experimental and calculated spectrum, respectively. The inner product (f, g) is defined with the metric W, which has the matrix elements

$W_{ij} = w(|j - i|) = w(r)$ as

$$(f, g) = f^T W g, \tag{7}$$

and the norm of f as $\|f\| = \sqrt{(f, f)}$; similar for g. For $w(r)$, a triangle function was used[39] with a width of the base of Δw given by

$$w(r) = \begin{cases} 1 - |r| / \left(\frac{1}{2}\Delta w\right) & \text{for} |r| < \frac{1}{2}\Delta w \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

The above-defined fitness function is able to smooth in a controlled way the fitness landscape and therefore allows the EA to locate the global minimum. The width of the function $w(r)$ critically determines the ability of the EA to converge to the global minimum and also the speed of convergence. The smoothing of the fitness landscape allows sensing regions far from the minimum. The EA 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 widths of an individual 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 EA convergence to the best set of parameters is achieved by narrowing Δw to zero in one or two steps.

Examples of Application of Evolutionary Algorithms to Extract NMR Parameters from Spectra of Orientationally Ordered Species

The successful application of ESs to the analysis of rather complicated spectra of orientationally ordered solutes in liquid-crystal environments has opened the door to problems ranging from those that were very time consuming to those that were seemingly intractable. We now look at some examples.

The first test of the method was on the nine-spin system *p*-bromobiphenyl[29] (Figure 4). Earlier versions of automatic spectral analysis programs that were applied to similarly complicated spectra seemed to require operator intervention in order to achieve a successful fitting to the experimental spectra. The important advance with this first test is that convergence between the experimental and calculated spectra is achieved with no operator intervention. In the case of *p*-bromobiphenyl, an initial GA was run by assuming a molecular structure and
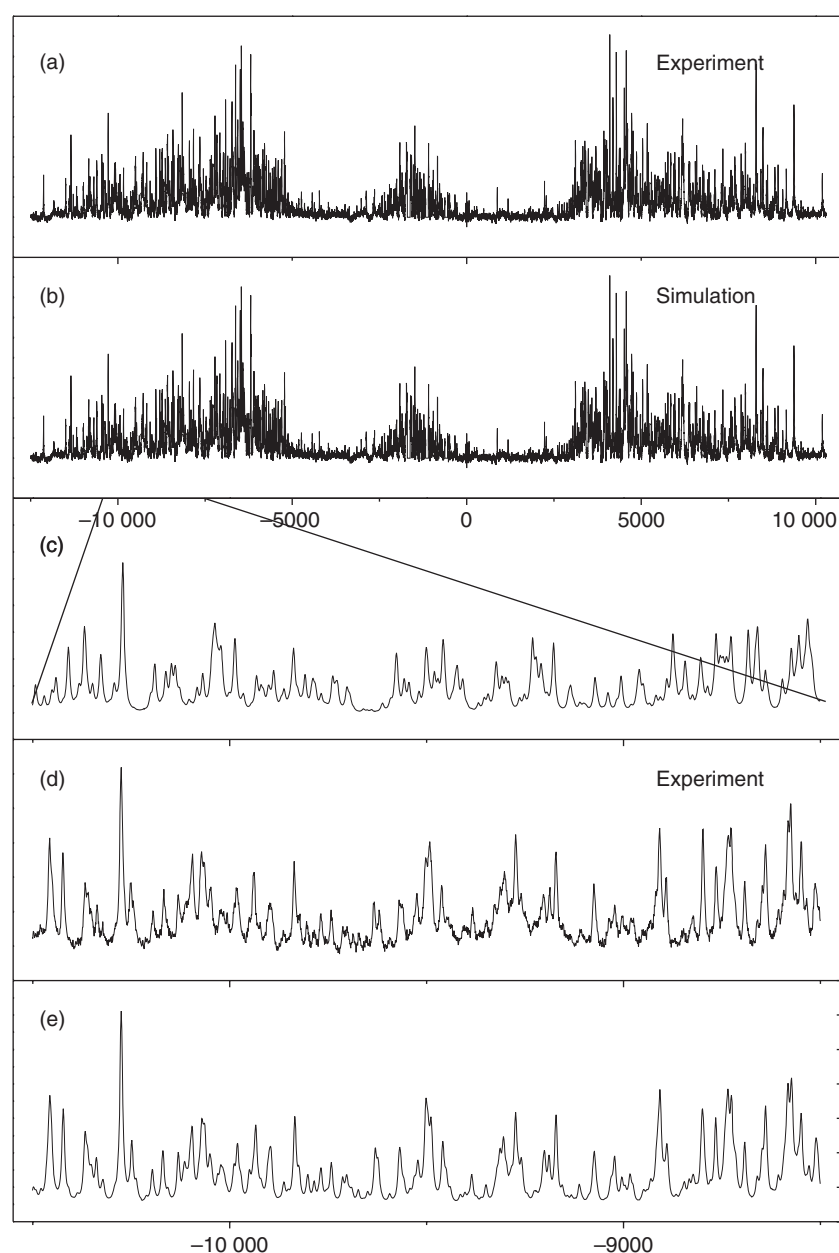

Figure 4. ^{1}H NMR spectrum of *p*-bromo-biphenyl dissolved in 'magic mixture.' The experimental spectrum (a) is compared with the final fit obtained by varying 15 dipolar couplings and 5 chemical shifts (b). In the next frames, the spectrum is enlarged to show the fit obtained by varying 2 *S*-parameters, 5 chemical shifts, and the dihedral angle (c), the experimental spectrum (d), and the fit obtained by varying 15 dipolar couplings and 5 chemical shifts (e). All horizontal scales are in hertz. (Reprinted from Chem. Phys. Lett., 441, W.L. Meerts, C.A. de Lange, A.C.J. Weber and E.E. Burnell, A simple two-step automatic assignment procedure for complicated NMR spectra of solutes in liquid crystals using genetic algorithms, 342–346, Copyright (2007), with permission from Elsevier.)

then fitting the spectrum to the order parameters and chemical shifts. The result (displayed in Figure 4c) demonstrates the good agreement between the experimental and fitted spectrum. A much superior fit (Figure 4e) is obtained when the fit is to the actual spectral parameters, *i.e.*, dipolar couplings instead of order parameters. However, the initial step (with only eight adjustable parameters) provided excellent starting values for the dipolar couplings, which expedited the final fitting to these couplings. This first fitting was with GA, and in many later cases using CMA-ES, the initial step of fitting to order parameters (rather than dipolar couplings) was no longer needed. In the case of flexible solutes (such as *n*-butane and *n*-pentane), the number of independent dipolar couplings equals or is less than the number of independent order parameters plus conformer probabilities, and the fitting was carried out to the dipolar couplings themselves.

The initial GA fitting was also applied to two other solutes, namely azulene and biphenylene, each of which is an eight-spin problem.[40] While these molecules had been studied earlier, the GA demonstrated the ease of the spectral analysis of these solutes in three different liquid-crystal solvents. The dipolar couplings obtained were later used to investigate nonrigid effects on the dipolar couplings and the potential errors introduced when these interactions (between vibrational and reorientational motions) were neglected. A prescription for estimating the error was presented, and it was demonstrated that, when the error was accounted for in a realistic manner, the structural parameters obtained for both solutes were the same in all three liquid-crystal solvents. Such agreement among solvents is not always the case when the nonrigid effects are ignored. Such discrepancies among solvents have led in the past to claims of distortion of the solute by the liquid-crystal environment.[41] Our experiments show that there is no need to invoke such molecular distortions.

Another example where the ability of the CMA-ES to simplify spectral analysis is exploited is in the investigation of multiple solutes in the same sample tube.[42,43] Such experiments are desirable when investigating the anisotropic intermolecular forces present in orientationally ordered liquids, as they ensure that all solutes experience precisely identical conditions. When solutes are studied in separate sample tubes, the effect of differing concentrations (and other experimental conditions) can lead to differences in liquid-crystal order parameter among samples, making it difficult to compare accurately results for the different solutes. Hence, the CMA-ES was easily adapted for the analysis of the spectrum that results from multiple solutes, and an example is shown in Figure 5.

The use of CMA-ES with multiple solutes in the same sample has facilitated the investigation of liquid crystals that experience both a higher temperature nematic phase and a lower temperature smectic A phase (and in one instance an even lower temperature reentrant nematic phase[44]); these experiments have allowed the determination of the solute smectic order parameters (τ, a measure of the preference for a solute to place itself in some particular region of the smectic layers, or to be evenly distributed if $\tau = 0$) and the coupling κ between nematic
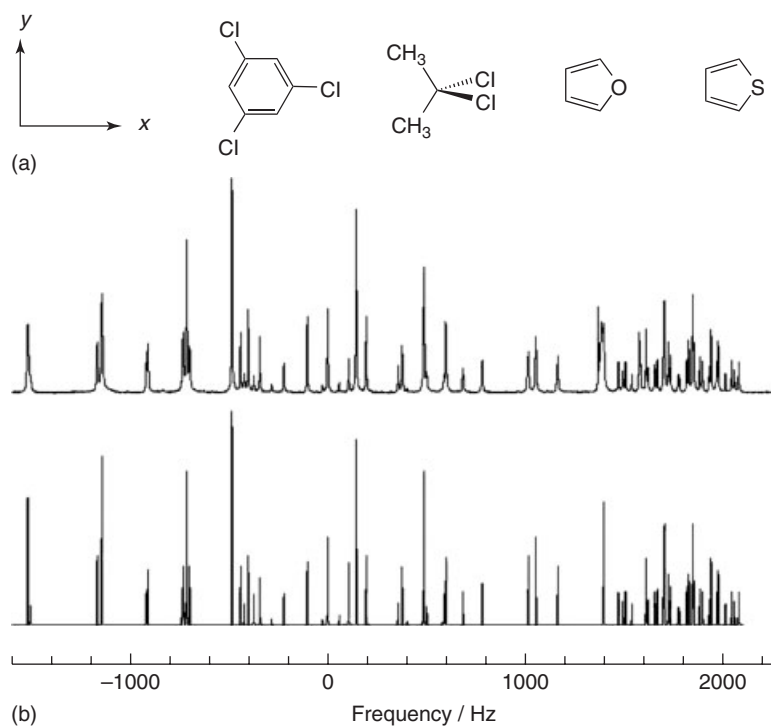

Figure 5. (a) The experimental 400 MHz NMR spectrum and (b) the spectrum found using the CMA-ES. The peaks of the solutes are interspersed with one another. (Reprinted from Chem. Phys. Lett., 476, A.C.J. Weber, X. Yang, R.Y. Dong, W.L. Meerts and E.E. Burnell, Solute order parameters in liquid crystals from NMR spectra solved with evolutionary algorithms: Application of double Maier−Saupe Kobayashi−McMillan theory, 116−119, Copyright (2009), with permission from Elsevier)

orientational ordering and smectic layering potentials (the nematic potential can vary with position in the smectic layer).

Perhaps the greatest success of the CMA-ES is the tremendous advance made possible in the analysis of solutes that can exist in several symmetry-unrelated conformers. To date, studies of *n*-butane (10 spins with 3 conformers, 2 of which are symmetry-unrelated),[12] and *n*-pentane (12 spins with 9 conformers, 4 of which are symmetry-unrelated)[14] have been carried out, while extension to *n*-hexane (14 spins with 27 conformers, 10 of which are symmetry-unrelated) may also be possible.

n-Butane was originally analyzed in one liquid-crystal solvent in a two-step process.[45] First, MQ-NMR spectra were accumulated and analyzed to provide rough values of the spectral parameters. The high-order multiple-quantum spectra contain far fewer lines than the single-quantum high-resolution spectra that are required to obtain very accurate values of the spectral parameters –for technical reasons the lines associated with the multiple-quantum transitions are broad, and the intensities depend strongly on the values of the spectral parameters used for the spectral calculation. For *n*-butane (as well as several other multispin solutes, including biphenylene[46] and biphenyl[47]), the high-order MQ spectra yielded to analysis and provided excellent starting points for analysis of the high-resolution single-quantum (normal) spectra. This process was tedious and time consuming, and has, for example, never been applied to *n*-pentane or *n*-hexane.

Spectra of *n*-butane were readily analyzed by the original GA as long as the parameter ranges set in the program included the actual values of the dipolar couplings and chemical shieldings; otherwise the GA failed.[12] When the initial parameters were chosen and ranges set by examining dipolar couplings that were calculated using model potentials to estimate values of order parameters (three for the gauche and three for the trans conformer), the spectrum was fitted 'over lunch.' Although it is in principle possible to use very large parameter search ranges, the time taken to achieve convergence increases dramatically when large ranges are chosen, and it is certainly wise to 'help' the analysis procedure by paying attention to reasonable starting values along with their search ranges before embarking on a trial. In the case of *n*-pentane, where the 4 symmetry-unrelated conformers require a total of 13 order parameters as well as relative probabilities of all 4 conformers, estimating values for all these parameters presents a problem. Indeed, in this case there are only 11 independent dipolar couplings, so a search using the dipolar couplings involves fewer parameters than one connected with order parameters and conformer probabilities. Initial values of dipolar couplings and their ranges were chosen based on a model potential that rationalizes the order parameters of a host of solutes in a special liquid-crystal mixture for which spectra of di-deuterium indicate the absence of an average electric field gradient. Results for solutes in this special 'magic mixture' are consistent with model potentials for the main interaction leading to solute orientational order being that due to short-range size and shape interactions. Although the liquid crystal used for this first study was not a magic mixture, it has been argued that hydrocarbons such as *n*-pentane are 'magic solutes' and that electrostatic interactions (such as those between molecular quadrupoles and mean solvent electric field gradients, or between molecular polarizability anisotropies and average solvent electric fields squared) are of minor importance and the dominant ordering mechanism involves solute short-range size and shape interactions.[48] Conformer probabilities were calculated by assuming a value for the energy difference between trans and gauche conformers, E_{tg}. Dipolar coupling ranges were then set by examining the change in dipolar coupling with change in anisotropic intermolecular potential parameters and of changes in E_{tg}. Ranges were set by comparing the overall appearance of the calculated spectrum with the experimental one. The important point here is that, for these flexible solutes, some couplings have a large conformer dependence, and sometimes the experimental coupling is small while the component conformer contributions are large, but of opposite sign. Only when such couplings are given a large range does the GA converge. For *n*-pentane, only when CMA-ES was employed with carefully chosen starting parameters and parameter ranges was a fit obtained, and it was achieved very quickly[14] (Figure 6).

In this case, there was a small problem with baseline for part of the spectrum. After the application of a simple baseline-correction algorithm, the entire spectrum could be used in the fitting, and an excellent fit of the calculated spectrum to the experimental one was achieved giving very precise values of the dipolar couplings, see Figure 4 of the corresponding paper[14] and discussion therein.

Because of the ease of analysis with CMA-ES, it was possible to investigate *n*-butane (and *n*-pentane, unpublished) as a function of temperature in several different liquid-crystal solvents.[13,49] Changing the temperature changes the Boltzmann factor and hence conformer probabilities, thereby providing the key to the investigation of the effect of the anisotropic environment on these probabilities. Our recent analysis of this problem exploits the finding that, in a given liquid-crystal solvent, the ratios of dipolar couplings in ethane to those in propane are constant over a large temperature range; if we now assume that the same holds for ratios of *n*-butane conformer dipolar couplings to ethane dipolar couplings, we have a way of fixing the temperature dependence of these *n*-butane conformer couplings and are thus able to extract conformer probabilities. In particular, we obtain separate estimates for the isotropic intra- and intermolecular conformer potentials and their effect on conformer probabilities, as well as the contribution to conformer probabilities from the anisotropic intermolecular potential.

In recent work (unpublished), we have used CMA-ES to analyze the spectra of *n*-butane, propane, and ethane present in the same sample tube, thereby providing a more accurate estimate of the *n*-butane conformer couplings. The ability of CMA-ES to handle three very complicated spectra simultaneously demonstrates the power of this spectral analysis tool.

Even with the power of CMA-ES, the selection of reasonable starting parameters is still key to solving a complicated spectrum in a reasonable time. One approach for choosing the starting parameters is to use computer simulation of the liquid-crystal/solute system. To this end, molecular dynamics (MD) simulations of *n*-pentane in the nematic phase of 5CB
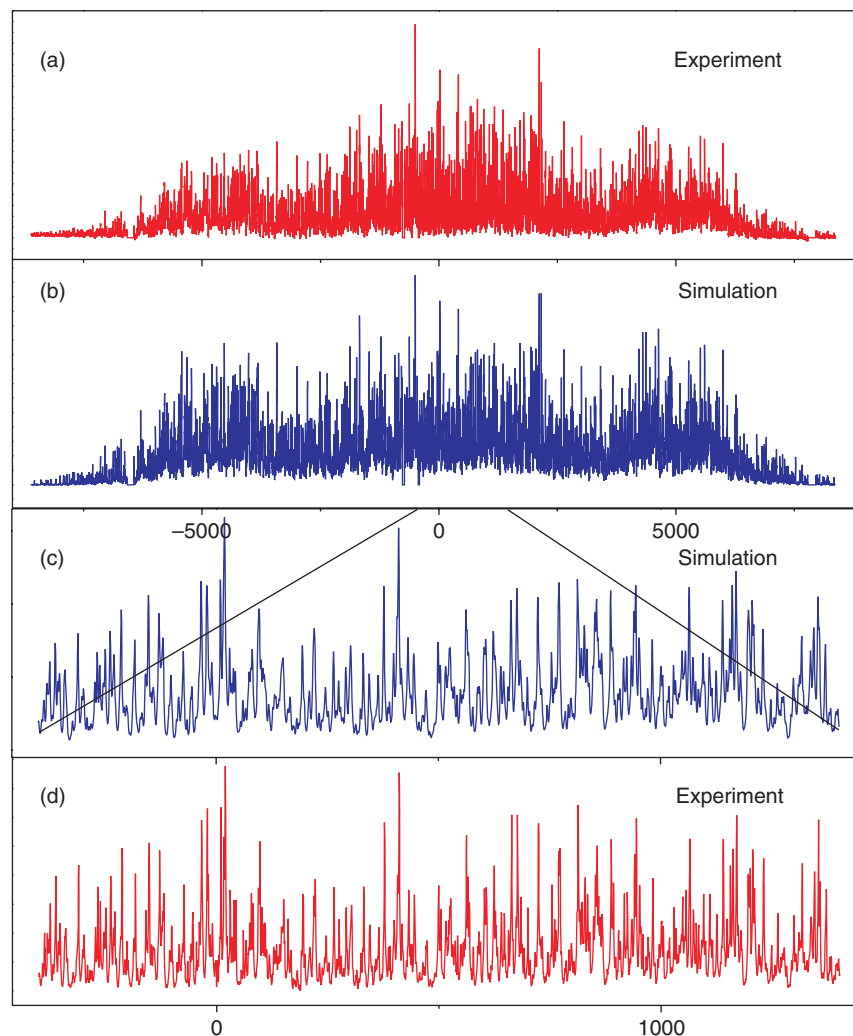
Figure 6. ^{1}H NMR spectrum of n-pentane dissolved in the nematic liquid-crystalline solvent Merck ZLI 1132. The experimental spectrum (a) is compared with the final fit obtained by varying 11 dipolar couplings, 9 indirect spin–spin couplings, and 3 chemical shieldings (b). In the next frames, the spectrum is enlarged to show the experimental spectrum (d) and the same fit as above (c). All horizontal scales are in hertz. The 11 indirect spin–spin couplings were varied only in the final fitting procedure; in earlier iterations they were given their isotropic values. (Reprinted with permission from W.L. Meerts, C.A. de Lange, A.C.J. Weber and E.E. Burnell, J. Chem. Phys., 130, 044504, (2009). Copyright 2009, American Institute of Physics)

(4-n-pentyl-4′-cyanobiphenyl) were carried out as a function of temperature.[50] The 2000 molecules in the simulated sample were kept at the same relative concentration as the NMR sample, which was also studied as a function of temperature spanning the nematic phase. The simulated and experimental nematic-to-isotropic phase transitions were essentially coincident. The temperature and chain-position dependence of the simulated n-pentane order parameters agreed with expectation. However, for a given temperature, the simulated and experimental spectra differed in the overall width (Figure 7). In order to proceed, the NMR result at 283.5 K and the MD result at 295 K were chosen for analysis, these being the spectra that showed the best agreement with respect to overall widths.

Comparison of the dipolar couplings from MD and NMR spectra gave a root-mean-square (RMS) difference of 45 Hz. This is an impressive starting point for CMA-ES analysis,

especially considering that the only 'fitting parameter' was the comparison between experimental and theoretical spectra. Next, a search range needs to be defined. For larger dipolar couplings, a range of ±15% is selected, while for smaller couplings the range is set to ±50 Hz. This value is chosen because these small couplings often result from averaging among different conformers that have relatively large couplings of opposite signs. Upon defining the problem in such a way, quick convergence is obtained and spectral parameters of the experimental spectrum involving over 20 000 transitions are obtained.

The success of the MD predictions for n-pentane raises the question whether this process can be used to solve problems of still greater complexity. To this end, the CMA-ES has been tested to see whether convergence can be obtained on an artificial n-hexane spectrum involving 188 000 transitions with search ranges similar to those used in the case of n-pentane.
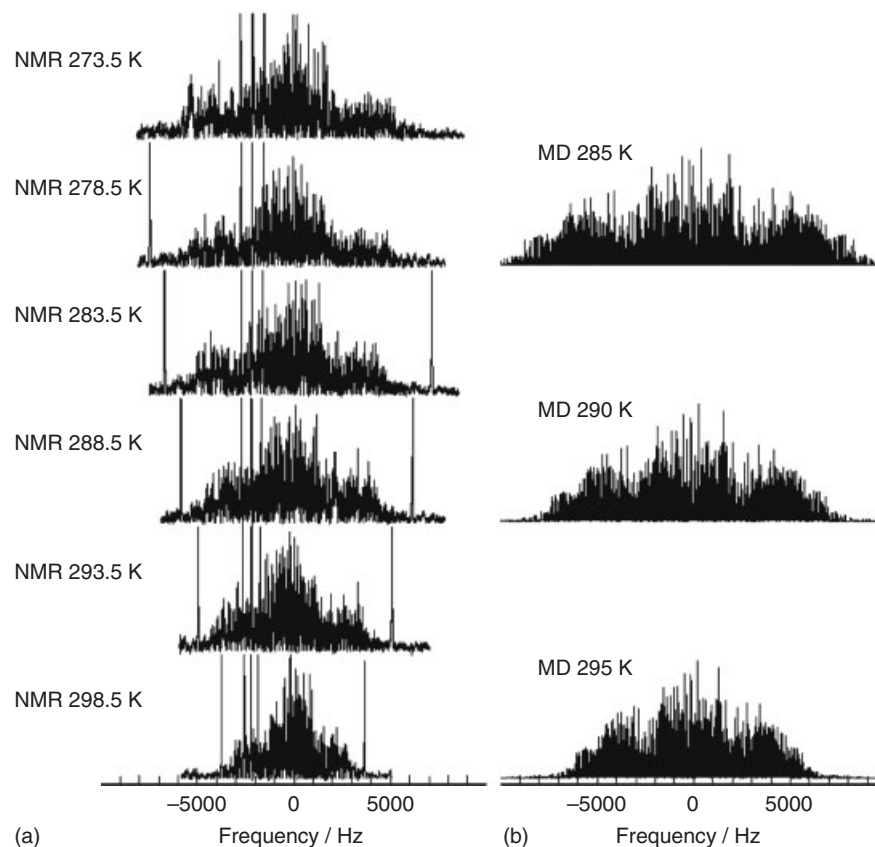
Figure 7. (a) Experimental and (b) calculated from MD simulations spectra of *n*-pentane in 5CB. The intense lines in the outer regions of the NMR spectra arise from an unknown impurity in the 5CB, and the other three intense lines are the triplet from 1,3,5-trichlorobenzene that was added as an orientational reference. (Reprinted with permission from A.C.J. Weber, A. Pizzirusso, L. Muccioli, C. Zannoni, W.L. Meerts, C.A. de Lange and E.E. Burnell, J. Chem. Phys., 136, 174506, (2012). Copyright 2012, American Institute of Physics)

Convergence was obtained, indicating the possibility of extending the CMA-ES to NMR spectra of solutes (such as *n*-hexane) that are considerably more complicated than those investigated so far.

Automatic Baseline Removal in NMR Spectra

Slow baseline fluctuations in NMR spectra of solutes in liquid-crystal solvents are a well-known problem. For sparse spectra such as in the upper three spectra of Figure 1, a manual removal of the background is quite feasible. Furthermore, the baseline problem mentioned in the previous section can be corrected with relative ease because the spectra still show a clear baseline. A non-flat baseline arising from (broad) NMR signals originating from the solvent and experimental imperfections becomes more pronounced for weak NMR spectra. Furthermore, if the number of lines further increases such that overlap of the transitions occurs, manual removal is no longer possible. The combination of the above-mentioned problems gave rise to the need to automatically remove the background originating from both the solvent and overlapping lines. This was achieved by removal of a smoothed spectrum from both the experimental and calculated spectra.

The smoothed experimental spectrum $f^S = (f_1^S, f_2^S, \ldots, f_N^S)^T$ is defined as

$$f_i^S = \frac{1}{n_s + 1} \sum_{j=i-n_s/2}^{i+n_s/2} f_j \qquad (9)$$

Here, n_s is the number of points over which the smoothing should be performed. The smoothing over the calculated spectrum g is carried out identically, resulting in $g^S = (g_1^S, g_2^S, \ldots, g_N^S)^T$:

$$g_i^S = \frac{1}{n_s + 1} \sum_{j=i-n_s/2}^{i+n_s/2} g_j \qquad (10)$$

The actual spectrum that is fitted is $f^R = f - f^S$ against $g^R = g - g^S$, i.e., in equation (6), f and g are replaced by f^R and g^R, respectively.

The idea behind this method is that removal of a smoothed spectrum removes the baseline originating from experimental sources not present in the calculated spectrum. At the same time, features in the spectrum that are on top of a background caused by overlapping transitions become more pronounced. Hence, in the global fit the local features contribute more effectively to the fitness. This increases the dynamics of the
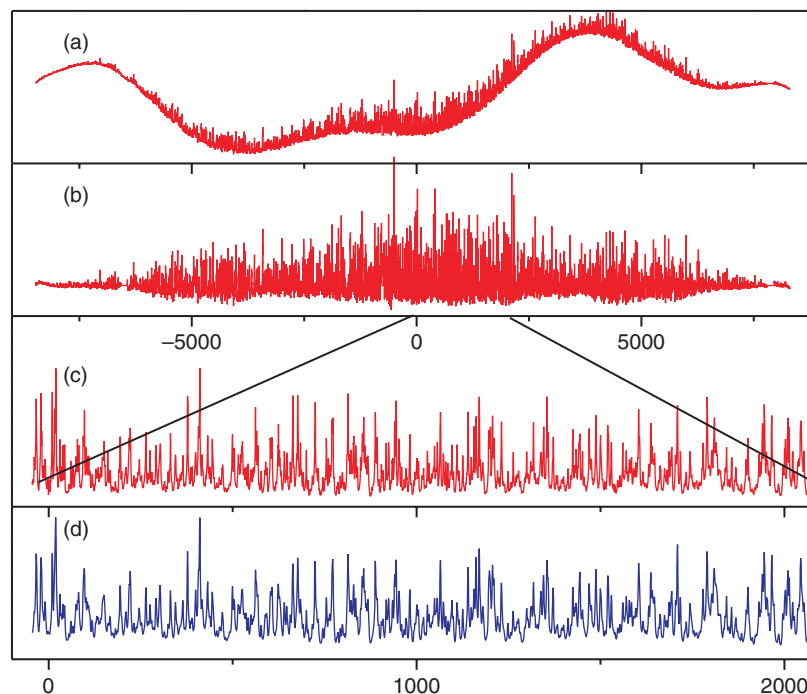
Figure 8. Demonstration of automatic baseline removal from an experimental spectrum. To the ^{1}H NMR spectrum of *n*-pentane from Figure 6 we have added an artificial baseline (a). Trace (b) shows spectrum (a) after automatic baseline removal by a smoothing over 250 Hz. The two lower traces show (c) an enlarged section of spectrum (b) compared to the fitted spectrum (d). All horizontal scales are in hertz

fitness function. The effect of automatic background removal is demonstrated in Figure 8.

Conclusion

NMR spectroscopy of solutes dissolved in liquid-crystal solvents yields highly accurate dipolar couplings, which are very sensitive to molecular structure and behavior. However, traditional line-assignment methods become intractable when considering molecules with more than eight spins, especially those with low symmetry and those that exist in several symmetry-unrelated conformers. Such spectra consist of thousands of (often overlapping) transitions. By coupling the ideas of reproduction and mutation, the GA method circumvents this problem by simply trying to match a population of potential solutions to the overall form of the experimental spectrum. For the more difficult problems, the ES (and CMA-ES), which is similar in spirit but focuses on mutation and cross-generational memory, is the key to success.

In this article, we have pointed out some examples where EAs have either simplified spectral analysis or have made it possible at all. We probed nematic and smectic liquid-crystal environments with relatively complicated solutes in the same sample tube by simply solving for multiple-spin systems simultaneously. We demonstrated the fallacy of the notion of solute structures being distorted by liquid-crystal environments. The ease of spectral analysis facilitated the investigation of alkanes and the effect of condensed phases on conformational statistics; it should be noted that the analysis required the measurement of *n*-butane spectra over an 80° temperature range, which were

analyzed in a few days (and would probably have taken months to solve with older methods).

One important aspect of the use of these strategies is the need to find reasonable starting values of the parameters. To this end, we demonstrated the power of MD simulations. Since the CMA-ES converged on a synthetic *n*-hexane spectrum with realistically applicable search ranges, the possibility of spectral analysis of this very complicated spectrum from a solute with 10 symmetry-unrelated conformers is real.

Biographical Sketches

W. Leo Meerts. b 1948. B.Sc., 1968 and M.Sc., 1970, University of Nijmegen, The Netherlands, PhD, 1975, Nijmegen, followed by a staff position at the physics department (University of Nijmegen). From 2006, a joint professor position at the Vrije Universiteit, Amsterdam. Over 200 papers on subjects varying from high-resolution radiofrequency (RF) to extreme-ultraviolet (XUV) spectroscopy. Since 1998, successfully applying evolutionary algorithms to the analysis of high-resolution UV laser spectra as well as solid-state and liquid-crystal NMR spectra.

Cornelis A. de Lange. b 1943. B.Sc., 1963 and M.Sc., 1966 (both cum laude), University of Amsterdam. PhD, 1969, University of Bristol. Employment: Shell Research Amsterdam, Vrije Universiteit Amsterdam, University of Amsterdam. Currently emeritus professor of physics and member of the Dutch Senate. Approximately 200 papers on the application of NMR techniques to the investigation of orientationally ordered liquids, and laser spectroscopy, in particular photoelectron spectroscopy.

Adrian C.J. Weber. b 1981. B.Sc., 2004, University of Manitoba, with research assistantships at the National Microbiology Laboratory and National Research Council spectroscopy institute, Winnipeg. PhD, 2010, University of British Columbia (UBC), followed by postdoc also

with Prof. E. Elliott Burnell at UBC. Joined Department of Chemistry, Brandon University, in 2011 as assistant professor. Eleven Papers on NMR applications to the study of liquid crystals and solutes dissolved therein.

E. Elliott Burnell. b 1943. B.Sc., 1965 and M.Sc., 1968, Memorial University of Newfoundland, PhD, 1970, University of Bristol. Following postdocs in physics (University of British Columbia, UBC) and physics (Basel), joined Department of Chemistry, UBC, in 1972 (currently professor). Approximately 150 papers on various aspects of the application of NMR techniques to the investigation of orientationally ordered liquids, especially the understanding of anisotropic intermolecular forces.

RELATED ARTICLES

Bothner-By, Aksel A.: Computer Analysis of High-Resolution NMR Spectra; Emsley, James W.: Having Fun with Liquid Crystals; Hägele, Gerhard: Analysis and Simulation of High-Resolution 1D NMR Spectra. Some Steps Closer to Automation; Khetrapal, C. L.: Development of NMR of Oriented Systems; Analysis of High-Resolution Solution State Spectra; Analysis of Spectra: Automatic Methods; Liquid Crystalline Samples: Spectral Analysis; Liquid Crystalline Samples: Structure of Nonrigid Molecules; Liquid Crystals: General Considerations; Computerized Analysis of High-resolution Solution-state Spectra

References

1. J. A. Pople, W. G. Schneider, and H. J. Bernstein, High-resolution Nuclear Magnetic Resonance, McGraw-Hill: New York, 1959.

2. P. L. Corio, Structure of High-resolution NMR Spectra, Academic Press: New York, 1966.

3. A. D.. Buckingham and K. A. McLauchlan, Progress in Nuclear Magnetic Resonance Spectroscopy, Pergamon Press: Oxford, 1967, Vol. 2, p 63.

4. P. Diehl and C. L. Khetrapal, NMR Basic Principles and Progress, Springer-Verlag: Berlin, 1969, Vol. 1, p 1.

5. J. W. Emsley and J. C. Lindon, NMR Spectroscopy Using Liquid Crystal Solvents, Pergamon Press: Oxford, 1975.

6. E. E. Burnell and C. A. de Lange, *Chem. Phys. Lett.*, 1980, **76**, 268.

7. E. E. Burnell and C. A. de Lange, *Chem. Rev. (Washington, DC)*, 1998, **98**, 2359, references therein.

8. E. E. Burnell and C. A. de Lange, eds, NMR of Ordered Liquids, Kluwer Academic Publishers: Dordrecht, The Netherlands, 2003, ISBN 1-4020-1343-4.

9. E. E. Burnell and C. A. de Lange, *J. Chem. Phys.*, 1982, **76**, 3474.

10. E. E. Burnell, C. A. de Lange, J. B. S. Barnhoorn, I. Aben, and P. F. Levelt, *J. Phys. Chem. A*, 2005, **109**, 11027.

11. E. E. Burnell, L. C. ter Beek, and Z. Sun, *J. Chem. Phys.*, 2008, **128**, 164901.

12. A. C. J. Weber, C. A. de Lange, W. L. Meerts, and E. E. Burnell, *Chem. Phys. Lett.*, 2010, **496**, 257.

13. E. E. Burnell, A. C. J. Weber, C. A. de Lange, W. L. Meerts, and R. Y. Dong, *J. Chem. Phys.*, 2011, **135**, 234506.

14. W. L. Meerts, C. A. de Lange, A. C. J. Weber, and E. E. Burnell, *J. Chem. Phys.*, 2009, **130**, 044504.

15. S. Castellano and A. A. Bothner-By, *J. Chem. Phys.*, 1964, **41**, 3863.

16. E. E. Burnell and C. A. de Lange, *Solid State NMR*, 2005, **28**, 73.

17. P. Diehl, S. Sýkora, and J. Vogt, *J. Magn. Reson.*, 1975, **19**, 67.

18. P. Diehl and J. Vogt, *Org. Magn. Reson.*, 1976, **8**, 638.

19. D. S. Stephenson and G. Binsch, *J. Magn. Reson.*, 1980, **37**, 395.

20. D. S. Stephenson and G. Binsch, *J. Magn. Reson.*, 1980, **37**, 409.

21. D. S. Stephenson and G. Binsch, *Org. Magn. Reson.*, 1980, **14**, 226.

22. D. S. Stephenson and G. Binsch, *Mol. Phys.*, 1981, **43**, 697.

23. D. S. Stephenson, Encyclopedia of Magnetic Resonance, John Wiley & Sons, Ltd, 2007. DOI: 10.1002/9780470034590.emrstm0013.

24. F. Castiglione, M. Caravetta, G. Celebre, and M. Longeri, *J. Magn. Reson.*, 1998, **132**, 1.

25. F. Castiglione, G. Celebre, G. De Luca, and M. Longeri, *J. Magn. Reson.*, 2000, **142**, 216.

26. M. Longeri and G. Celebre, Encyclopedia of Magnetic Resonance, John Wiley & Sons, Ltd, 2007. DOI: 10.1002/9780470034590.emrstm0268.

27. H. Takeuchi, K. Inoue, Y. Ando, and S. Konaka, *Chem. Lett.*, 2000, **29**, 1300.

28. K. Inoue, H. Takeuchi, and S. Konaka, *J. Phys. Chem. A*, 2001, **105**, 6711.

29. W. L. Meerts, C. A. de Lange, A. C. J. Weber, and E. E. Burnell, *Chem. Phys. Lett.*, 2007, **441**, 342.

30. E. E. Burnell, C. A. de Lange, and W. L. Meerts, Novel strategies for solving highly complex NMR spectra of solutes in liquid crystals, in Nuclear Magnetic Resonance Spectroscopy of Liquid Crystals, ed. R. Y. Dong, World Scientific: Singapore, 2010, Chapter 1, p 1.

31. J. H. Holland, Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence, MI: The University of Michigan Press: Ann-Arbor, MI, 1975.

32. I. Rechenberg, Evolutionsstrategie - Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution, Frommann-Holzboog: Stuttgart, 1973.

33. H.-P. Schwevel, Evolution and Optimum Seeking, John Wiley & Son, Inc.: New York, 1993.

34. W. L. Meerts and M. Schmitt, *Int. Rev. Phys. Chem.*, 2006, **25**, 353.

35. A. Ostermeier, A. Gawelczyk, and N. Hansen, Step-Size Adaptation Based on Non-Local Use of Selection Information, Lecture Notes in Computer Science: Parallel Problem Solving from Nature (PPSN III), Springer-Verlag, 1994, pp 189–198.

36. N. Hansen and A. Ostermeier, *Evolut. Comput.*, 2001, **9**, 159.

37. N. Hansen and S. Kern, Evaluating the CMA evolution strategy on multimodal test functions, in Parallel Problem Solving from Nature PPSN VIII, of *LNCS*. 8th International Conference, Birmingham, UK, September 18–22, 2004, Proceedings, eds X. Yao, E. Burke, J.A. Lozano, J. Smith, J.J. Merelo-Guervós, J.A. Bullinaria, J. Rowe, P. Tino, A. Kabán, H.-P. Schwefel, Springer-Verlag, Berlin/Heidelberg, 2004, Lecture Notes in Computer Science, Vol. 3242, p 282, Also available online. http://www.springer.com/computer/theoretical+computer+science/book/978-3-540-23092-2.

38. W. Leo Meerts, M. Schmitt, and G. Groenenboom, *Can. J. Chem.*, 2004, **82**, 804.

39. J. A. Hageman, R. Wehrens, R. de Gelder, W. Leo Meerts, and L. M. C. Buydens, *J. Chem. Phys.*, 2000, **113**, 7955.

40. C. A. de Lange, W. L. Meerts, A. C. J. Weber, and E. E. Burnell, *J. Phys. Chem. A*, 2010, **114**, 5878.

41. A. G. Shahkhatuni, A. A. Shahkhatuni, H. A. Panosyan, G. H. J. Park, R. W. Martin, and A. Pines, *J. Phys. Chem. A*, 2004, **108**, 6809.

42. A. C. J. Weber, X. Yang, R. Y. Dong, W. L. Meerts, and E. E. Burnell, *Chem. Phys. Lett.*, 2009, **476**, 116.

43. A. C. J. Weber, X. Yang, R. Y. Dong, and E. E. Burnell, *J. Chem. Phys.*, 2010, **132**, 034503.

44. E. E. Burnell, R. Y. Dong, A. C. J. Weber, X. Yang, and A. Yethiraj, *Can. J. Chem.*, 2011, **89**, 900.

45. J. M. Polson and E. E. Burnell, *J. Chem. Phys.*, 1995, **103**, 6891.

46. J. M. Polson and E. E. Burnell, *J. Magn. Reson., Ser. A*, 1994, **106**, 223.

47. T. Chandrakumar, J. M. Polson, and E. E. Burnell, *J. Magn. Reson., Ser. A*, 1996, **118**, 264.

48. A. F. Terzis, C.-D. Poon, E. T. Samulski, Z. Luz, R. Poupko, H. Zimmermann, K. Müller, H. Toriumi, and D. J. Photinos, *J. Am. Chem. Soc.*, 1996, **118**, 2226.

49. A. C. J. Weber and E. E. Burnell, *Chem. Phys. Lett.*, 2011, **506**, 196.

50. A. C. J. Weber, A. Pizzirusso, L. Muccioli, C. Zannoni, W. L. Meerts, C. A. de Lange, and E. E. Burnell, *J. Chem. Phys.*, 2012, **136**, 174506.