Research Article

Magnetic Resonance in Chemistry

Received: 16 September 2008 Revised: 30 October 2008 Accepted: 29 November 2008 Published online in Wiley Interscience: 3 February 2009

(www.interscience.com) DOI 10.1002/mrc.2394

Automatic phase correction of 2D NMR spectra by a whitening method

Giuseppe Balacco[a] and Carlos Cobas[b]*

A new method for the automatic phase correction of multidimensional NMR spectra is described. It is based on the *whitening* concept formulated as the 'maximization of the number of white pixels into a bitmap that corresponds to the spectrum'. This process of maximization can be factorized along the individual axes of the spectrum and this property makes the method robust and fast. It employs a statistic measure based on a large number of spectral data points and, for this reason, is very tolerant to low signal-to-noise ratio (SNR) and local artifacts. The algorithm can efficiently phase either homonuclear or heteronuclear experiments and, unlike other previous methods, it can also process automatically spectra containing positive or negative peaks so that it is not necessary to deal with individual or special cases

Keywords: NMR; phase correction; 2D; automatic processing; whitening

Introduction

Traditionally, 2D NMR spectra were presented in magnitude or power mode for reasons of operational simplicity. However, nowadays, the majority of multidimensional NMR spectra of liquid samples is routinely acquired in phase-sensitive mode because of its inherent advantages in spectral resolution and the ability to resolve resonances which would have otherwise overlapped in a magnitude/power spectrum.[1] Furthermore, many multidimensional experiments (NOESY, edited HSQC, etc) have peaks with both positive and negative intensity, and this information would be lost in the magnitude or power modes.

Phasing an *n*-dimensional spectrum involves forming a linear combination of the 2^n components in the hyper complex space to get absorptive lines in all orthogonal dimensions. Although it has been shown that a phase correction can be rendered unnecessary by a precise fine tuning of experimental acquisition parameters,[2] a post-FT phase correction will normally still be required due to potential instrumental instabilities or to spectra being acquired under different experimental conditions. Traditionally, the trigonometric coefficients of the linear combination (two parameters per each dimension) are determined empirically by a process called 'manual phase correction' with the help of interactive software.

In older applications, the correction is not applied directly on the matrix. What the user can do, instead, is to extract selected rows and columns from it, correct the phase for these traces, then ask the software to apply the same correction on the whole matrix. Modern applications allow, in addition to the method above, the direct manipulation, in real time, of a bi-dimensional plot. The experience with both methods tells that phase correction of a 2D spectrum is not more complicated than phase correction of a 1D spectrum, despite the higher number of parameters to be adjusted (4 instead of 2). In other words, it is never necessary to adjust all the parameters simultaneously, but it is possible instead to sequentially perform the correction along the individual axis.

While manual correction, as permitted by today's software, can be extremely accurate, the availability of an automatic method can simplify the job and reduce processing time. Automatic correction of 2D NMR spectra is of particular importance in platforms aimed to work in a fully unattended mode, such as in fields as varied as metabonomics,[3] structure analysis and verification of small molecules,[4] protein structure[5] and open access laboratories.

The literature is rich in articles[6–18] dealing with the automatic phase correction of 1D NMR spectra, but only a few papers were found that directly address the multidimensional problem. Cieslar *et al*.[19] proposed the first algorithm for automatic phase correction of 2D NMR spectra, which maximizes peak asymmetry and peak height of the diagonal peaks, and which therefore is limited to homonuclear experiments only. A different approach was taken by Hoffman *et al*.[20] based on the Dispersion vs Absorption (DISPA) method, which is a plot of the dispersion component of the signal against the absorption component. As in all DISPA based approaches, the algorithm is critically dependent on the accuracy of peak selection.

An enhanced procedure was introduced in the PROSA program,[21] which has been designed to correct higher dimensional spectra taking advantage of the fact that a typical 3D- or 4D spectrum is sparse and presents numerous sufficiently isolated resonances. A more recent method called Phase Angle Measurement from Peak Areas (PAMPAS)[22] calculates the phases of isolated peaks across individual traces in the *n*D spectrum by using Fourier analysis of a series of peak areas measured with systematically incremented phase shifts. The zero- and first-order corrections are then found by linear regression.

We have recognized that manual phase correction, when performed by a skilled operator, is a methodical process, evidently driven by a principle. We have investigated if this inner principle could be expressed as an algorithm, and found that the principle could be formulated in different ways. A simple one, referring to the bitmap representation of a 2D spectrum is: 'maximize the number of white points'.

* *Correspondence to: Carlos Cobas, Mestrelab Research, Xose Pasín, 15706, Santiago de Compostela, Spain. E-mail: carlos@mestrelab.com*

a *Nucleomatica, Via Francesco Carabellese, 10/C, 70056 Molfetta BA, Italy*

b *Mestrelab Research, Xose Pasín, 15706, Santiago de Compostela, Spain*

In this article we present a new algorithm for automatic 2D phase correction based on the above concept which, unlike other methods, works directly on the 2D matrix, taking advantage of all informations contained in the hypercomplex space and thus making it robust to local artifacts and low signal-to-noise ratio (SNR). Despite the apparent high-computational cost involved in this approach, the algorithm herein presented is very fast and simple, producing very good results for wide ranges of multidimensional NMR experiments.

Discussion

In the following text, for the sake of simplicity, we will consider the case of automatic processing of 2D NMR spectra, but the approach is extensible to higher dimensional spectra. 2D Fourier transform of phase-sensitive experiments yields a hypercomplex matrix consisting of four components RR, RI, IR, II (R = Real; I = Imaginary). In general only the RR quadrant is displayed. Phasing of this hypercomplex matrix is carried out through a rotation of the 2D matrix in the hypercomplex space by applying the following linear combination:

$$RR' = \cos\psi_1 \cos\psi_2 RR - \sin\psi_1 \cos\psi_2 IR - \cos\psi_1 \sin\psi_2 RI + \sin\psi_1 \sin\psi_2 II \quad (1)$$

$$IR' = \sin\psi_1 \cos\psi_2 RR + \cos\psi_1 \cos\psi_2 IR - \sin\psi_1 \sin\psi_2 RI - \cos\psi_1 \sin\psi_2 II \quad (2)$$

$$RI' = \cos\psi_1 \sin\psi_2 RR - \sin\psi_1 \sin\psi_2 IR + \cos\psi_1 \cos\psi_2 RI - \sin\psi_1 \cos\psi_2 II \quad (3)$$

$$II' = \sin\psi_1 \sin\psi_2 RR + \cos\psi_1 \sin\psi_2 IR + \sin\psi_1 \cos\psi_2 RI + \cos\psi_1 \cos\psi_2 II \quad (4)$$

where

$$\Psi_2 = \alpha_2 + (k/N)\beta_2 \quad (5)$$

$$\Psi_1 = \alpha_1 + (j/M)\beta_1 \quad (6)$$

$k = 0, \ldots, N-1$; $j = 0, \ldots, M-1$ and N and M are the number of points along F2 and F1, respectively.

The aim of automatic phase correction is to find the four parameters $\alpha_2, \beta_2, \alpha_1$ and β_1 in order to get an absorption spectrum. This could be achieved by an iterative process optimizing some objective function. In previous articles, different functions have been proposed for the automatic phase correction of 1D spectra, including maximizing the lowest point of the spectrum[7] or the entropy of the spectrum.[18] In this work, we introduce a new objective function, designed specifically for 2D spectra, which we call *whitening* and which is defined as the *maximization of the number of white pixels in a 2D image*. A 2D NMR spectrum can be considered as a computer image formed by *pixels*. Intuitively, it is easy to appreciate that, if we are going to represent both the dispersion and the absorption component of a spectrum with two bitmaps, the number of white pixels will be higher in the latter one. Generalizing, we can say that the number of white pixels reaches the maximum when the phase is corrected (Fig. 1).

Looking at the 2D spectrum as an image (Fig. 1), it can be observed that the number of colored pixels decreases continuously during the correction. This intuitive concept can be translated into a computer algorithm by defining an interval $[-t, t]$, with t being a positive quantity which segments the data points in the spectrum into two sets:

1. Points whose intensity falls inside the interval (they correspond to the white points of the image) and
2. Points whose intensity falls outside the interval (they correspond to the colored points).

The only requirement for t is that neither set ever becomes empty. Under this condition, a decrease of the population of set 2 signals an improvement of the phase correction. In other words, if P is the number of points whose absolute intensity is higher than t, minimization of P will yield a phase corrected spectrum. P is a function of the number of phase parameters by the number of dimensions. In the case of a 2D spectrum,

$$P = f([\alpha_2, \beta_2]^{f2}, [\alpha_1, \beta_1]^{f1}) \quad (7)$$

In principle, this would require simultaneous optimization of the four parameters, a process which is both time consuming and prone to converging to local minima. However, following the same approach used while correcting the phase manually, in which one dimension is corrected at a time, automatic phase correction can also be performed along both orthogonal axes independently. For example, when correcting the phase along the

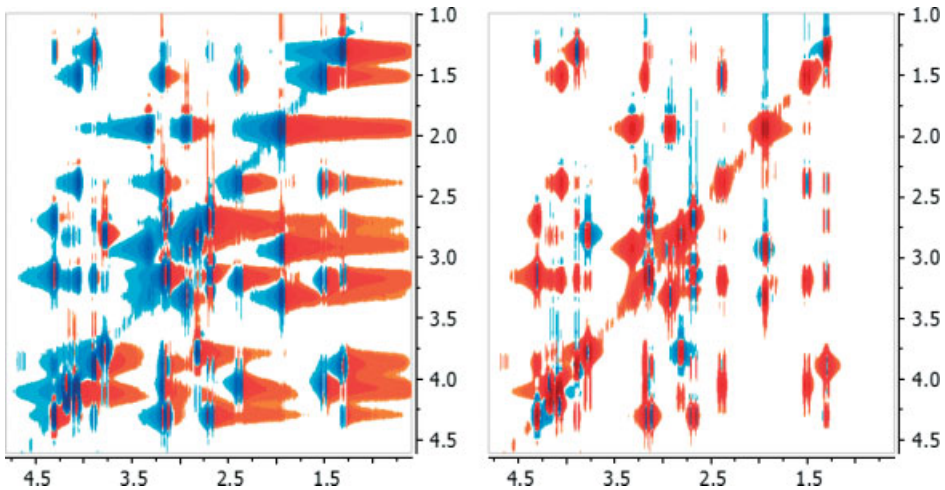

Figure 1. Unphased (left) and phased (right) spectra. It can be observed that the number of white pixels in the phased spectrum is larger.

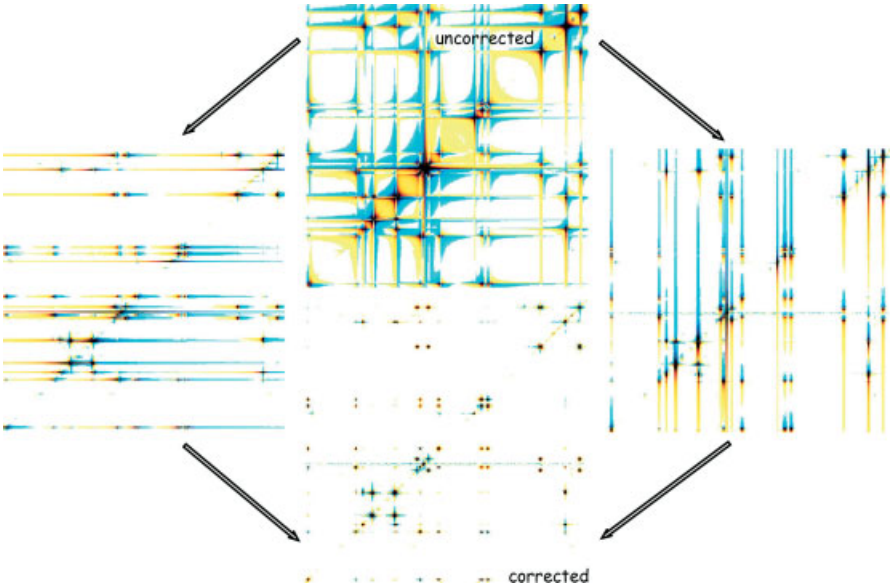

Figure 2. When correcting the phase along the *X*-axis, the number of white pixels increases regardless of the state of the phase along the *Y*-axis and vice versa.

X-axis, *P* always decreases, regardless of the fact that the phase along the remaining axes is already correct or not (Fig. 2). The independency of the corrections not only simplifies the algorithm, but it also reduces the computational time and increases the chances of success.

Therefore, optimization of the objective function *P* can be carried out independently along both dimensions. In this work, we have chosen the well-established simplex algorithm by Nelder and Mead[23] used to minimize a generic function of *n* variables, like *P*, which has been already successfully applied to the phase correction of 1D NMR spectra.[7]

The main merits of the whitening principles proposed in this work are:

- It can be applied to a wide class of spectra, because it assumes very little about them; for example, it does not discriminate between homonuclear and heteronuclear experiments, positive and negative peaks, etc.
- It employs a statistic measure based on thousands or, more likely, millions of data points, therefore it is little affected by noise and local artifacts;
- It can be applied to spectra of all dimensionality or to a single, selected dimension if required.
- Being based on a single and simple principle, there is an ample space for modifications and optimizations, as will be shown in the rest of this work.

Experimental

In order to calculate the value of *P* corresponding to a given phase correction it is, in theory, necessary to recalculate the real part of the spectrum. Considering that this real part is recalculated many times during the minimization process, a computationally effective trick involves storing, for each point, a couple of values: the absolute value *A* and the phase φ along the dimension that is being corrected. When the phase correction δ at this frequency has been determined, the new real component for the corresponding point can be calculated as:

$$R = A\cos(\varphi + \delta) \tag{8}$$

If $|R| \geq t$, this point contributes to the value of *P*, increasing it by one unit. The advantage of using this representation of the complex spectrum lies in the fact that, for most of the points, there is no necessity to calculate *R*. We see, in fact, that:

$$|R| \leq A \tag{9}$$

thus, whenever $A < t$, the calculation of *R* can be skipped because necessarily, $|R| \leq t$. There is some freedom in choosing the value *t*, a possible solution would be:

$$t = \bar{A} \text{ (the average value of all } A\text{s)} \tag{10}$$

In a typical homonuclear spectrum, the large and prominent diagonal keeps the average value significantly higher than the noise level, therefore more than the half of the points will fall below $t = \bar{A}$ and it would not be necessary to calculate them. This single observation can boost the speed of the algorithm by a factor greater than 2. Even if the weakest cross-peaks are ignored, correcting the region around the diagonal is enough to correct the whole spectrum. In other cases, however, the choice of *t* can be a little more critical. For example, a common HSQC spectrum of a small molecule contains a few-weak peaks diluted into an ocean of noise. Here the average value of the spectrum is almost coincident with the average value of noise. If a low value of *t* is chosen, we will be sampling too many regions of noise which can mask the effect of the phase changes on *P*; if the noise cannot be filtered out, *P* is no more a correct estimator of phase correction. If, instead, *t* is too high, *P* becomes a constant (zero), regardless of the phase.

In practice, we have found satisfactory results by setting $t = 2\,\bar{A}$. In this way, the noisy regions of the spectrum are automatically ignored and they are prevented from affecting the accuracy of our

method. There are also lesser calculations to perform. At the same time, in all our tests, *P* has always remained far from zero with this value of *t*.

The simplex algorithm, when used to optimize two parameters, as in this case, requires three starting guesses (they define the starting position of the simplex into the solution space). When the starting position is too far from the global minimum, there is the risk of falling into a local minimum of *P* that does not correspond to a correct phase. This fact explains a small percentage of failures, that can however be prevented. We have devised two different ways to generate a starting guess that is near enough to the proper minimum:

1. Applying a mono-dimensional automatic correction on the orthogonal projections.
2. Using the whitening principle itself for a preliminary scan of the solution space.

Method 1 involves the calculation of the internal projection as the sum of all columns (or rows) in the 2D matrix. In the case of homonuclear correlation spectra, the projection is analogous to the diagonal. The phase correction calculated by an automatic 1D algorithm can be used as a rough estimate of the phase correction of the whole 2D matrix. In this work we have used the algorithm originally devised by Siegel.[7] This method is satisfactory in the case of homonuclear correlation, but inapplicable in the case of an edited-HSQC (with both positive and negative peaks and much more noise).

The second method is based on an effective way to exploit the whitening principle itself, but without the simplex algorithm, to generate a rough estimate of the phase correction. When we calculate the absolute value of all the points in the matrix, we register the position of the highest one (the pivot point). Then we select a narrow band centered around this frequency value along the axis to be corrected. We calculate the function *P* neglecting the rest of the matrix, first without any phase correction, then again applying two zero-order phase corrections of $\pm 15^{\circ}$. If *P* decreases we keep correcting with identical amounts of zero-order correction. In other words, we perform a rapid scan along the axis corresponding to the zero-order correction. The step must be large enough to overcome random effects, like noise. We stop the scan as soon as *P* starts increasing again. When we have found the zero-order correction for the pivot point, we can find, in a similar way, the first-order phase correction for the whole matrix. The calculation of *P* is no longer restricted to a band, but performed in the normal way, paying attention to keeping the phase correction at the frequency of the pivot constant. The step used for scanning the first-order axis is higher ($\pm 30^{\circ}$). It is easy to see that, apart from the larger steps, this kind of phase correction mimics closely the manual process. Once we have determined our starting point with the described sequential scans, we build a little triangular simplex around it and are ready to apply the Nelder–Mead algorithm. A possible weakness of our preliminary scan is that it stops prematurely when the spectrum is extremely noisy and the starting position is very far away from the minimum. In this case there are too many steps to take and, just because each step entails the risk of interrupting the scan prematurely, the overall risk becomes sizeable.

In our experience, these two methods for generating the starting position of the simplex can optionally be combined, in the same order in which we have described them.

Results

The whitening algorithm has been programmed in C/C++ and integrated into iNMR[24] and Mnova.[25] All the spectra have been acquired in Bruker and Varian spectrometers to reflect the diversity commonly found in an NMR laboratory. In all cases, a cosine bell-weighting function was applied along the acquired dimension and a cosine squared bell-weighting function in the indirect dimension. No baseline correction has been applied to any of the spectra. All spectra in the figures below are displayed with the same contour levels.

The performance of the algorithm is demonstrated by the following examples, which were acquired from different NMR spectrometers and have different experimental conditions and artifacts. These spectra were chosen to clearly show the algorithm's ability to efficiently deal with spectra with positive and/or negative peaks and its robustness in conditions of low SNR, presence of baseline artifacts or large dispersive solvent signals.

For our first example, we applied the algorithm to a TOCSY spectrum of the human acidic fibroblast growth factor (FGF) protein[26] with DIPSI Spinlock and Watergate water suppression, acquired in a 750-MHz Inova Varian spectrometer. As Fig. 3 shows,

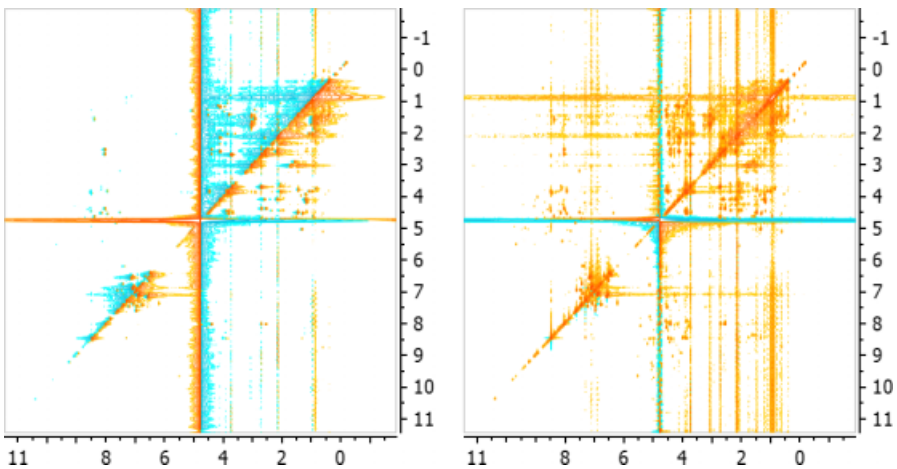

Figure 3. TOCSY spectrum of the human acidic fibroblast growth factor (FGF) protein before (left) and after application of the automatic phase correction algorithm (right).

the spectrum contains a large residual dispersive signal of water and baseline artifacts along both dimensions despite which our algorithm yields a perfectly phase corrected spectrum. In this case the phase of the water peak is incompatible with the rest of the spectrum, but the diagonal is the driving force. Another explanation is that the water peak is so strong that the tails never become white, therefore they are not 'seen' by the algorithm.

In the second example we illustrate the performance of the algorithm with noisy spectra. The spectrum showed in Fig. 4 corresponds to the HSQC experiment of Strychnine recorded in a 400-MHz Varian instrument. Application of the automatic phase correction algorithm shows an excellent behavior under these conditions of low SNR.

A further example corresponding to the TOCSY spectrum of Strychnine acquired in a 500-MHz Bruker DRX spectrometer is depicted in Fig. 5. In this case, the raw spectrum exhibits large phase distortions along both F2 and F1 dimensions. Our whitening algorithm finds a solution very close to the optimal one found by careful manual adjustment.

The *whitening* principle does not make any assumption on the line shape or sign of the peaks. This is illustrated in Fig. 6 with a multiplicity-edited HSQC echo-antiecho spectrum recorded in a 400-MHz DRX-Bruker spectrometer using the Bruker HSQCEDETGP pulse sequence. As it can be observed, what really matters is the fact that the number of white pixels into the bitmap is the highest possible.

A valuable feature of the *whitening* algorithm is its ability to properly phase data containing crowded antiphase resonances, such as in DQF–COSY spectra. Such data represent a serious stumbling block to less experienced NMR users, as these spectra can be difficult to phase manually. This is shown in Fig. 7 with the DQF–COSY of Taxol collected in a 300-MHz AMX-Bruker instrument. The spectrum at the left of the figure shows the original unphased spectrum while the spectrum at the right shows the results obtained after applying the automatic phase correction routine. It is clear that our *whitening* algorithm yields a correctly phased spectrum, nearly identical to the manually corrected counterpart (spectrum not showed). Similar results have been obtained with other DQF–COSY spectra analyzed by the authors.

Conclusions

The *whitening* algorithm outlined in this paper has been designed specifically to deal with characteristic phase distortions commonly found in 2D NMR spectra. Despite being based on an iterative simplex algorithm, it has been highly optimized in such a way that, in most cases, automatic phase correction takes less than 2–3 s in up-to-date personal computers. It is easy to implement and robust

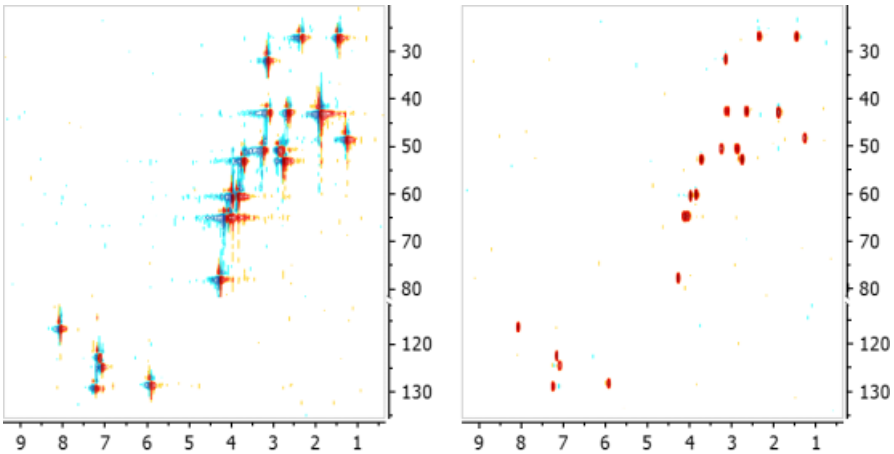

Figure 4. A noisy HSQC spectrum (left) and the result (right) after automatic phase correction by the whitening algorithm.

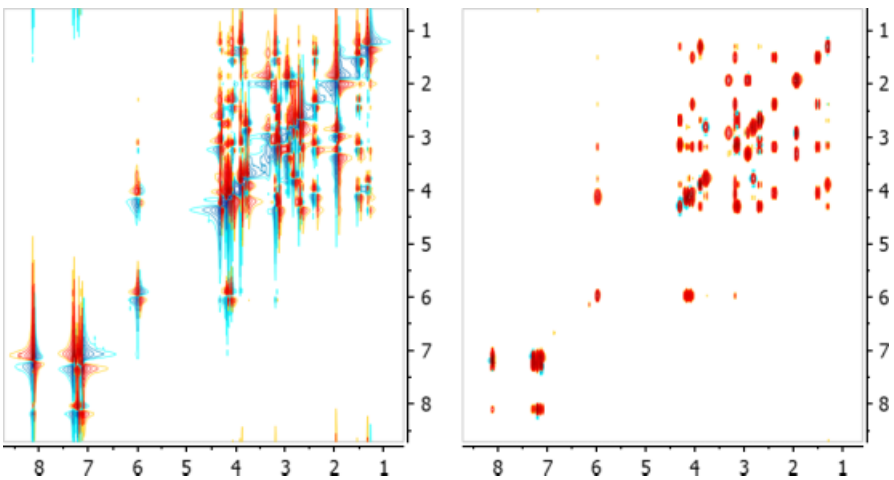

Figure 5. A TOCSY spectrum of strychnine before (left) and after automatic phase correction (right).

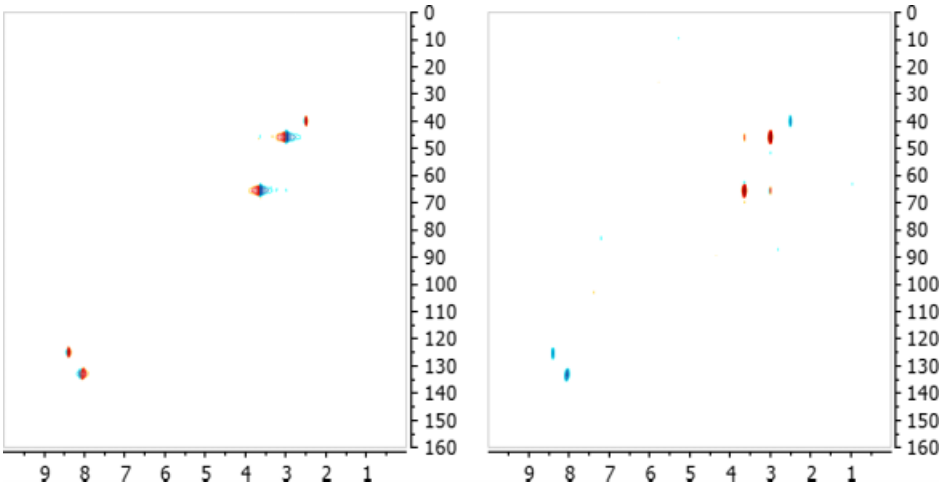

Figure 6. An edited-HSQC spectrum before (left) and after automatic phase correction (right).

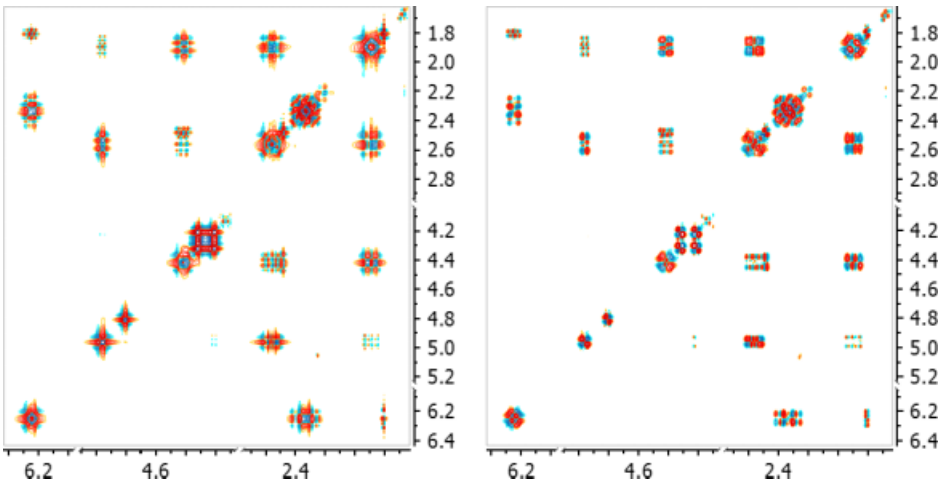

Figure 7. Unphased DQF-COSY of Taxol (left) and after phase correction with the whitening algorithm (right).

with respect to experimental noise and baseline imperfections, peak overlaps and line shapes. It can therefore process either homonuclear or heteronuclear experiments or spectra containing only positive peaks or a combination of positive and negative peaks.

An implementation of the algorithm has been incorporated into a new autophase routine within the iNMR[24] and Mnova[25] software packages for NMR processing and analysis.

Acknowledgements

C.C is grateful for the support and funding granted for this project by the Ministerio de Educacion y Ciencia (Spain - Programa Torres Quevedo) and the Xunta de Galicia (PGIDIT06SIN005E).

References

[1] M. H. Levitt, R. Freeman, *J. Magn. Reson.* **1979**, *34*, 675.
[2] Q. Teng, *Structural Biology*, Practical NMR Applications, **2005**.
[3] K. Otto, N. Aranibar, *Methods Mol. Biol.* **2006**, *358*, 247.
[4] L. Griffiths, H. H. Beeley, R. Horton, *Magn. Reson. Chem.* **2008**, *46*, 818, references therein.
[5] B. López-Méndez, P. Güntert, *J. Am. Chem. Soc.* **2006**, *128*, 13112.
[6] R. Ernst, *J. Magn. Reson.* **1969**, *1*, 7.
[7] M. M. Siegel, *Anal. Chim. Acta* **1981**, *133*, 103.
[8] C. H. Sotak, C. L. Dumoulin, M. D. Newsham, *J. Magn. Reson.* **1984**, *57*, 453.
[9] F. G. Herring, P. S. Phillips, *J. Magn. Reson.* **1984**, *59*, 489.
[10] L. F. Gladden, S. R. Elliott, *J. Magn. Reson.* **1986**, *68*, 383.
[11] C. N. Chen, L. S. Kan, *Comput. Methods Programs Biomed.* **1988**, *26*, 81.
[12] E. C. Craig, A. G. Marshall, *J. Magn. Reson.* **1988**, *76*, 458.
[13] D. E. Brown, T. W. Campbell, R. N. Moore, *J. Magn. Reson.* **1989**, *85*, 15.
[14] E. A. Wacher, E. Y. Sidky, T. C. Farra, *J. Magn. Reson.* **1989**, *82*, 352.
[15] A. Heuer, *J. Magn. Reson.* **1991**, *91*, 241.
[16] G. Balacco, *J. Magn. Reson.* **1994**, *110*, 19.
[17] P. Koehl, C. Ling, J. F. Lefèvre, *J. Chim. Phys.* **1995**, *92*, 1929.
[18] L. Chen, Z. Weng, L. Y. Goh, M. Garland, *J. Magn. Reson.* **2002**, *158*, 164.
[19] C. Cieslar, G. M. Clore, A. M. Gronenborn, *J. Magn. Reson.* **1988**, *79*, 154.
[20] R. E. Hoffman, F. Delaglio, G. C. Levy, *J. Magn. Reson.* **1992**, *98*, 231.
[21] P. Güntert, V. Dötsch, G. Wider, K. Wütrich, *J. Biomol. NMR* **1992**, *2*, 619.
[22] D. Dzakula, *J. Magn. Reson.* **2000**, *146*, 20.
[23] J. A. Nelder, R. Mead, *Comput. J.* **1965**, *7*, 308.
[24] iNMR, version 3, Nucleomatica, Molfetta, Italy, www.inmr.net, **2008**.
[25] MestReNova (Mnova), version 5.3, Mestrelab Research, SL, Santiago de Compostela, Spain, www.mestrelab.com, **2008**.
[26] A. Pineda-Lucena, M. A. Jiménez, J. L. Nieto, J. Santoro, M. Rico, G. Giménez-Gallego, *J. Mol. Biol.* **1994**, *242*, 81.

