Available online at www.sciencedirect.com
ScienceDirect
Journal of Magnetic Resonance 188 (2007) 56–67
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Modeling of NMR processing, toward efficient unattended processing of NMR experiments

Dominique Tramesel [a,b], Vincent Catherinot [b], Marc-André Delsuc [a,*]

[a] Centre de Biochimie Structurale, 29 rue de Navacelles, CNRS UMR5048, INSERM U554, Université Montpellier 1 & 2, F34090 Montpellier, France
[b] NMRtec S.A.S., Bld. Sébastien Brandt, Bioparc - Bat. B, F67400 Illkirch-Graffenstaden, France

Received 10 December 2006; revised 27 April 2007
Available online 16 June 2007

Abstract

Many alternative processing techniques have recently been proposed in the literature. Most of these techniques rely on specific acquisition protocols as well as on specific data processing techniques, the need for an efficient versatile and expandable NMR processing tool would be a particularly timely addition to the modern NMR spectroscopy laboratory. The work presented here consists in a modeling of the various possible NMR data processing approaches. This modeling presents a common working frame for most of the modern acquisition/processing protocols. Two different data modeling approaches are presented, strong modeling and weak modeling, depending whether the system under study or the measurement is modeled. The emphasis is placed on the weak modeling approach. This modeling is implemented in a computer program developed in python and called NPK standing (standing for NMR Processing Kernel), organized in four logical layers (i) mathematical kernel; (ii) elementary actions; (iii) processing phases; (iv) processing strategies. This organisation, along with default values for most processing parameters allows the use of the program in an unattended manner, producing close to optimal spectra. Examples are shown for 1D and 2D processing, and liquid and solid NMR spectroscopy. NPK is available from the site: http://abcis.cbs.cnrs.fr/NPK

Keywords: Unattended NMR data processing; Data modeling; MaxEnt processing; Partial sampling; Generic shearing

1. Introduction

In 1966, the introduction of the Fourier transform spectroscopy by Ernst and Anderson [1,2], was one of the first steps in introducing computer processing in the NMR domain. Since this pioneering work, the introduction of multidimensional experiments and more recently of optimized acquisition protocols has confirmed the central role of data processing in high resolution NMR.

With the increase of the sensitivity of the modern spectrometers, many new processing techniques intended to minimize the acquisitions times, have been proposed in the recent literature. Most of these recently introduced techniques rely on specific acquisition protocols.

The following non exhaustive list of alternative processing protocols may be cited: (i) frequency sign determination (Hypercomplex acquisition [3,4]; echo/anti-echo modu-lation [5]); (ii) reduced dimensionality (Radial Sampling [6,7], G-Matrix FT [8]; Recursive Multidimensional Decomposition [9]); (iii) optimized acquisition scheme (sparse [10–13], reduced [10–12], or non-regular sampling [14,15]); Hadamard spectroscopy [16,17]; (iv) single scan acquisition [18,19]; (see Atreya and Szyperski [20] for a recent review on processing.)

This abundance of approaches places the potential user in front of a difficult choice indeed, and under the threat of depending on a particular implementation of this processing protocol, eventually incompatible with its own particular set-up. There is thus an acute need for an efficient versatile and expandable NMR processing tool in the modern NMR spectroscopy lab.

* Corresponding author. Fax: +33 467 41 79 13.
E-mail address: ma.delsuc@cbs.cnrs.fr (M.-A. Delsuc).

1090-7807/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2007.05.023

Most current NMR processing software [21–24] is based on a set of procedures which implement the various required processing. These procedures are usually implemented in a two level approach, with a low-level library implementing very efficiently the fundamental operations (mathematical, data management, etc.) and a higher level set of programs, usually written in some kind of script language, implementing the details of the NMR data-processing.

In this approach, the user has to implement a set of actions; these actions have to be optimized, and are described by a set of parameters. The order of the actions is also very important, and may depend on parameters values of other actions (e.g. the apodisation and zerofilling actions are usually different, depending whether a Linear Prediction FID reconstruction has been performed or not).

We call this approach the *cook-book* approach. The main limitation of this approach lies in the fact that each *recipe* requires specific *ingredients*. Each implementation of a new protocol is dependent on a new set of procedures which have to be developed.

Additionally, the new processing techniques may share procedures with more conventional ones, but might require different set-up (for instance MaxEnt requires that no or very mild apodisation is used on the FID), and the user may not be fully aware of this. There is an increased risk of wrong parameter values or of wrong procedure order.

Finally, each new protocol requires additional information either from the spectrometer or from the user (files, parameters, etc.), in addition to the actual NMR raw-data. This information may not be readily available in the required format and a specific format converter must be developed.

With the wealth of NMR processing protocols available today, this approach has reached a limit, and we wish to present here an alternative approach which, we believe, permits to handle most of the situations.

The work presented here consists in a *modeling* of the various possible NMR processing strategies. This modeling allows to present a common working frame for most of the modern acquisition/processing protocols, and thus to avoid the overwhelming number of independent and unrelated methods.

Rather than concentrating on the processing details the modeling presented here separates the processes into a series of fundamental phases, which are common to all processing techniques. Each phase is then further separated in fundamental actions. Along with this hierarchical approach, an implementation of each processing action is proposed.

This approach departs from the classical one in several aspects. Firstly, the strict separation of the parameters from methods allows a more formal description of the processing, independent of the details of the actual data-set being processed.

Secondly, all required parameters are given default values, eventually depending on the kind of experiment being processed, and incompatible actions and parameters are checked. Default actions and default parameters are defined for a known experiment type, and for each processing phase. As a consequence the user is not required to write the sequence of procedures which will be applied on the data-sets, but rather to give to the program the key elements of the processing to be applied. She/he can then safely use new methods not yet fully mastered. The processing of a given experiment can even be carried out in a completely automatic manner.

A higher level of processing abstraction, based on a detailed modeling of the actions taking place is proposed. At this abstraction level, the user does not have to bother on the details of the implementation such as action order, as this is modeled in the program itself.

Finally, from a program developer point of view, this modeling provides a strong and extensible framework, which helps the implementation of the newly developed method into the general processing line.

2. Theory

2.1. Two approaches to modeling

Processing of an NMR data-set consists in taking the experimental numeric data from the spectrometer, then in applying to this data a series of processing phases, and finally in delivering a result in a form which is manageable by the spectroscopist.

This final result can take on very different aspects: for instance, a graphic representation of the frequency spectrum of the spin resonances, or a listing of the concentration of the various compounds present.

Two different kinds of processing techniques can be recognized depending on what is being modeled: the system under study or the way the measurement is performed. This two approaches are called here, respectively, strong modeling and weak modeling, they are sometimes referred as system modeling and measurement modeling. They are detailed below.

2.1.1. Strong modeling

In strong modeling, a mathematical model of the system under study is considered, based on apriori assumptions on the system itself. This mathematical model is characterized by a certain number of parameters which are extracted from the experimental data by the processing process.

The number of parameters is usually much smaller than the number of experimentally measured values. For instance, if the studied system is a standard liquid sample studied by high resolution pulsed NMR, the model of the NMR phenomenon is a set of resonances with exponentially decaying time response. The parameters of this mathematical model are: the actual number of lines present in the spectrum, as well as the frequency, life-time, amplitude and phase of each resonance. The result of the processing thus consists in presenting these values to the user.

In this approach, the adequacy of the extracted mathematical model to the experimental data is generally computed through a statistical estimator such as a χ^2. This approach assumes that (i) an estimate of the experimental data is computationally available from the parameterized model, and that (ii) the number of parameters of the model is smaller than the number of independently acquired data points.

Processing based on a strong model, is sometime referred as a *parametric adjustment*.

The general mathematical description can be given as (i) an analytical (or algorithmic) form of the system model, (ii) a statistical test on the solution, in the data space, (iii) a minimizer engine permitting to optimize the statistical test.

Processing techniques such as Linear Prediction, Bayesian analysis or GFT analysis are examples of strong modeling.

2.1.2. Weak modeling

On the other hand, in weak modeling, no mathematical model of the system is assumed nor extracted from the data, but an image of the physical characteristics is built and presented graphically, and only very weak assumptions on the system under study itself are made.

For instance, for the same standard high resolution NMR study as above, the result of the processing presented to the user is an image of the resonance frequency spectrum. The mathematical model of the system in this case is minimum, and is mostly associated with the acquisition process (causality, finite spectral-width, finite power, etc.) with eventually general assumptions such as the positivity of the spectrum.

This approach assumes that (i) an estimate of the experiment can be computed from the image by numerically applying what is called the transfer function, and that (ii) the inverse of the transfer function can be estimated, either directly of iteratively.

Processing based on an acquisition model, is sometime referred as a *transform*.

The general mathematical description can be given as (i) an analytical (or algorithmic) form of the transfer function, (ii) a statistical test on the solution in data space (iii) a statistical test on the solution in spectrum space, (iv) a minimizer engine permitting to optimize the statistical test. Point ii is generally implemented with a χ^2 statistic. Point iii is needed here to determine a unique solution to the otherwise under-determined problem, it is usually called a regularizer. Examples of regularizers are: minimum energy; minimum curvature; maximum entropy, Tikhonov regularizer.

Processing techniques such Fourier transform, projection reconstruction or MaxEnt processing are examples of weak modeling.

2.1.3. Comparison of the two approaches

The transform approach produces an image of the system under study. This image is produced with few or no hypotheses on the underlying system, in consequence it cannot present any direct information on the sample. For instance, the concentration of a given sample, or the *J*-value of a scalar coupling cannot be read directly in the spectrum, but some additional processing based on a mathematical model of the system itself (lineshape and linewidth) will be required. Rather, it presents, in a graphic manner more readable for the user, all the information already present in the raw experimental data. As such, it can be understood as a mere change in the point of view with essentially no loss or addition of information to the data. This change in point of view usually introduces more degrees of freedom (more data points) as a result of a transform processing. These additional points no not contain additional information, but rather present some kind of data interpolation, with strong point to point correlation.

On the other hand results from an adjustment approach provide a clear answer to the problems, usually with a confidence level estimate.

These two aspects (strong modeling vs weak modeling) are implemented in NPK, however with a strong emphasis on transform methods and acquisition modeling. The following transforms are implemented: fast Fourier transform and its inverse, Fourier transform by MaxEnt processing, Hilbert transform and its inverse, Laplace transform, inverse Laplace transform by MaxEnt processing and shearing and tilt transforms based on Hilbert transform.

2.2. Processing data modeling

A complete computer implementation of an NMR processing program requires that one fully describes the data that is to be handled and detail each action that will be taken on this data. Two independent but related computer models should thus be used: a data model and an action model. They share in common the fact that they should be easily extended, since new acquisition and processing protocols are frequently proposed in the literature.

In this work, the data model is based on a clear-cut separation between the parameters which are invariant upon transformation and parameters which vary from one processing method to another. For instance the temperature, the field of the NMR spectrometer or spin frequency are considered as invariant parameters. On the other hand, the binary information containing the result of the representation, as well as parameters such as the number of data points, the phase or the intensity of a peak are considered as variable.

The data model used here is derived from the comprehensive data model proposed by the CCPN work group [25–27]. However a much simplified version is used, as only a small part of this model is currently needed. On the other hand, it had to be extended on the experimental side, for instance adding models for partial sampling acquisition protocols.

The processing data model is based on the separation of the whole process into independent processing phases.

Owing to the fact that the spectral analysis step (the Fourier transform) is generally central to the whole data processing approach, three phases are used: pre-processing, spectral analysis, and post-processing.

Each phase is further detailed into elementary actions such as apodisation, phase rotation, Fourier transform, baseline correction, etc.

The different actions handled are algorithmically implemented in a general manner. Along with the algorithm itself, all the required parameters are listed, and rules are given which describe contraints on values (for instance that the size of the FFT action must be a power of 2) or eventual relationships between parameters. For instance the causal correction should not applied on the spectrum if it was already corrected at the FID level. Finally, a boolean parameter is also associated to each action, which tells whether this action is to be taken in the current phase.

3. Organisation

When implementing a transform operation on an experimental data-set, for example the Fourier transform of a 1D FID, one has to handle different aspects of this operation. At the lowest level, the Fourier transform algorithm must be efficiently implemented in the program, with eventually various options (complex/real; FFT algorithm/sweeping algorithm [14]). These algorithms must be efficiently implemented, presumably in low level languages such as Fortran or C, and are usually available in mathematical libraries.

However this aspect is insufficient to fulfill the requirements of the NMR spectroscopist. The Fourier transform step should take into accounts details such as the sign of the frequency axis, the correction of the initial time-lag, or the extraction of the region of interest. While trivial to implement, these actions must be carefully described. Since this kind of control requires a detailed description of the actions to be taken, but does not usually present an important computer burden, a high level computer language is a better choice for these actions.

For this reason, the modeling of the NMR processing, presented by the NPK program, is built in with a four layers architecture: mathematical kernel, elementary actions, processing phases and complete processing strategies.

3.1. Four layers

The general organisation is described in Fig. 1. The lowest layer consists in a mathematical library, implementing all the mathematical methods required for the processing. This includes Fourier transforms, apodisation functions, Maximum Entropy processing, statistical analysis, as well as many standard methods.

Above this mathematical kernel, a library of the elementary actions required for NMR processing is implemented. A high level language has been chosen to implement these actions in order to allow a rapid development and an easy integration in a the laboratory environment. We have chosen to develop this part in python, because of its ease of use and of the large number of scientific codes already available. Each action comes with an algorithmic implementation, and a detailed description of the required parameters. They are implemented by making intensive use of the mathematical operations available at the lower layer. All these elementary actions are expressed in terms natural to the NMR spectroscopist. For instance, the methods take care of the implementation of the Fourier transform of a data-set depending on the acquisition mode (phase modulation, Hypercomplex [3], TPPI [4], PEP protocol [5], etc.).

The different elementary actions are then grouped in processing phases. These phases cover all the actions that will be taken on the various kind of NMR data. They are independent from each other and are applied successively. The phases depend on the type of data to process (1D, series of 1D, 2D, 3D), but also on the kind of processing to apply (FT spectral-analysis, MaxEnt spectral analysis, Inverse Laplace). As an example, the different phases defined for the processing of 1D data-set are presented in Table 1, along with the elementary actions on which they are built. All possible actions are listed for each phase, but most are optional, so a flag is stored along with each action indicating if this action is to be applied or not. Additionally, it can be seen that certain actions are incompatible (for instance modulus, phase and autophase). These constraints between actions are actually defined within the phase definition and enforced in the implementation.

The list of all currently defined phases is shown in Table 2. The phases are built such that each action is independent of the preceding and following phases. These phases can then be easily chained to produce a useful processing. The way the phases are chained depends only on the kind of processing to be executed.

The phases are then chained to realize complete data-set processing strategies. As a general approach, the analysis of a spectral axis is performed in three phases: pre-processing, spectral analysis, and post-processing. In the current implementation, based on the different phases detailed in Table 2, nine strategies are proposed: 1D Spectral analysis by Fourier or by MaxEnt; 1D Peak analysis; 2D Spectral analysis (FT or MaxEnt); Series of 1D; 2D DOSY analysis; 2D Peak Analysis; 3D Spectral Analysis. The details are presented in Table 3. The separation into phases permits to easily build these strategies, and to add new ones if required. Additional strategies are currently being developed: 3D DOSY; 3D Peak Analysis; 2D Series, Peak tracking over experiment series.

As was stated earlier, each action requires a set of parameters which fully describes the processing to be applied. Examples of parameters are : size for the Fourier transform; phase corrections. The set of all the parameters actually describes the complete processing strategy to be applied, and is completely independent of the details of the algorithmic implementation.

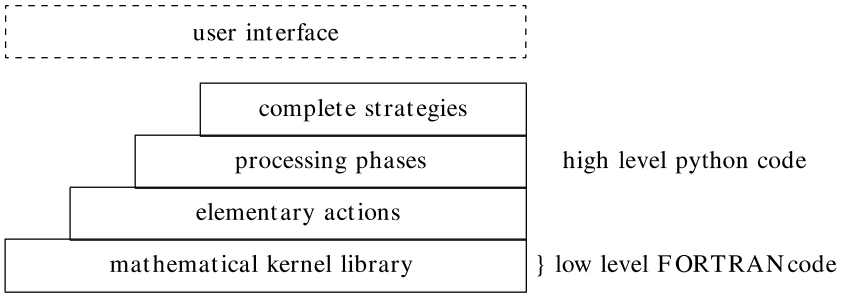


Fig. 1. Scheme presenting the overall organisation of NPK.

Table 1
List of 1D actions by phases

Phases for 1D processing	Various actions implemented
1D Pre processing	Evaluates noise and offset levels on FID
	Corrects for constant offset
	Corrects non-causality on DSP processed data-sets
	Removes solvent signal
	Drops first points
	Adds empty points on the beginning of the FID
	Reconstructs missing points in the beginning of the FID by LP analysis
1D FT analysis	Truncates the FID by removing the last points
	Extends FID with a Linear Prediction algorithm
	Applies apodisation
	Performs the Fourier transform
	Performs causal correction on the spectrum on DSP processed data-sets
	Reverses the spectral axis after FT
1D Post processing	Computes modulus
	Applies phase correction
	Computes automatic phase correction
	Applies inverse Hilbert transform
	Calibrates the ppm scale
	Extracts spectral zone
	Applies a baseline correction
	Applies a smoothing or median filter
	Computes the *n*th derivative
	Evaluates noise level on spectrum
1D Inverse FT	Applies apodisation
	Performs inverse Fourier transform
	Reverses spectral axis after *i*FT
1D MaxEnt analysis	Computes a temporary Fourier transform
	Truncates the FID by removing the last points
	Applies a preconvolution before analysis
	Sets-up for processing data partially sampled in the time domain
	Applies deconvolution during analysis
	Applies MaxEnt analysis
1D Peak picking	Evaluates noise on spectrum
	Applies a pre-smoothing of the spectrum, (*modification is not stored permanently*)
	Restricts the peak picking to a certain spectral zone
	Peak picks, by detecting local extrema
	Sorts peak list to aggregate peaks close from each other
1D Integration	Computes integral positions from peaks

3.2. Default values, and default processing

There are numerous actions already available, and many more will be implemented in the next version of NPK. If all parameters are to be given for each action, the set of parameters defining a given processing cannot be used when the processing program is extended or modified. So it is of tremendous importance that each parameter has a default value which is used if a value is not explicitly provided. The default values for each parameter have been carefully chosen to ensure to the extent possible that a meaningful spectrum is produced from the default

Table 2
List of defined phases

Dimension	Kind	Phase[a]
1D	Pre processing	Pre processing
	Spectral analysis	FT analysis
		MaxEnt spectral analysis
	Post processing	Post processing
		Post MaxEnt processing
		Inverse FT
		Peak picking
		Integration
2D	Pre processing	t_1t_2 Pre processing
		t_1 Pre processing
		t_2 Pre processing
		t_1 Pre ILT processing
	Spectral analysis	F_1 FT analysis
		F_2 FT analysis
		F_1F_2 MaxEnt spectral analysis
	Laplace analysis	Pre MaxEnt ILT processing
		MaxEnt ILT analysis
		Post MaxEnt ILT processing
	Post processing	F_1 Post processing
		F_2 Post processing
		F_1F_2 Post processing
		Post MaxEnt processing
		F_1 Inverse FT
		F_2 Inverse FT
		F_1F_2 Peak picking
		F_1F_2 Integration
3D	Pre processing	$t_1t_2t_3$ Pre processing
	Axis processing	F_1 1D processing
		F_2 1D processing
		F_3 1D processing
	Plane processing	F_1F_2 2D processing
		F_2F_3 2D processing
	Post processing	$F_1F_2F_3$ Post processing

[a] By convention t_x is related to the x axis before processing, and F_x to the x axis after processing.

processing. With this set-up, in many cases, it is possible to generate a useful spectrum without providing any parameter. For instance, the default size to be used for the Fourier transform step is defined as being twice the smallest power of two value larger than the actual FID size (unless a LP extension was computed), thus realizing a nearly optimum zero-filing operation in most cases.

3.3. The NPK program

The computer modeling presented above has been implemented in a program call NPK (standing for NMR Processing Kernel).

The four levels of modeling presented above are reflected in the code organisation. The mathematical library is based on a corpus of Fortran codes, and also include links to readily available libraries such as LAPACK [28]. A large part of the code present in this library reuse the development which have been done in the previous *Gifa* project [23].

The remainder of the code is exclusively developed in the python language. Actions and phases are algorithmically described, and the complete set of default values are auto-documented in the code and automatically extracted when required. Finally, the different strategies are implemented at the user interface level, permitting to apply a complete processing with a single command.

The whole program is then developed within a Java framework. This approach permits to easily link the binary library as well as to implement a complete python 2.1 shell by using the jython library [29].

The java wrapper program also provides access to a large number of possibilities, such as the choice of alternative user interfaces. The default interface consists in the python interpreter, however thanks to the versatility of the Java environment, several alternative could easily be built: WEB interface, graphic interface.

NPK is available from the following site: http://abcis.cbs.cnrs.fr/NPK

4. Examples of use

4.1. Unattended processing

As stated earlier, a complete separation of the processing algorithms from the processing parameters, and meaningful default values for processing parameters allows automatic processing of NMR data. As an example, for a given 1D FID, an optimized apodisation is applied, the correct Fourier transform is chosen and an automatic phase correction is computed. The spectrum is then subjected to a baseline correction. Finally, peak picking is accomplished and integration zones are determined from the peak list. All these actions being applied with default values.

Fig. 2 presents several spectra, obtained from the automatic unattended processing available in NPK. This unattended processing is possible because of the strong modelisation available, and thanks to the meaningful default values.

However, it is worth mentioning that unattended automatic processing presents some dangers. The obtained spectra are always the results of some trade-off and of some choices which are implemented in the strategies and which might not be correct for the present problem. For this reason, all the actions taken are fully documented in an audit-trail file, and all the default actions can be either modified or switched-off.

In any case, the default processing can then be used as a starting point for an improved processing quality.

4.2. Generic shearing

The hierarchical approach to processing makes very simple to develop high level functionalities. The generic shearing tool is presented here as an example. Shearing requires a frequency shift along the columns of a 2D spectrum. The value of the shift depends on the position of the column in the spectrum. As the slope of this

Table 3
List of defined statergies

Pre processing	Analysis	Post processing
1D FT analysis		
1D Pre processing	1D FT analysis	1D Post processing
1D MaxEnt spectral analysis		
1D Pre processing	1D MaxEnt analysis	1D Post MaxEnt processing
1D Peak analysis		
Peak picking	Integration	
1D series FT analysis		
t_1t_2 Pre processing		
	F_2 1D analysis (any 1D strategy)	
		F_1F_2 Post processing
2D FT analysis		
t_1t_2 Pre processing		
t_2 Pre processing	F_2 FT analysis	F_2 Post processing
t_1 Pre processing	F_1 FT analysis	F_1 Post processing
		F_1F_2 Post processing
2D MaxEnt spectral analysis		
t_1t_2 Pre processing[a]	F_1 F_2 MaxEnt analysis	F_1F_2 Post MaxEnt processing
2D DOSY analysis		
t_1t_2 Pre processing		
t_2 Pre Processing	F_2 FT analysis	F_2 Post processing
t_1 Pre MaxEnt ILT processing [a]	MaxEnt ILT analysis	Post MaxEnt ILT analysis
2D Peak analysis		
F_1F_2 Peak picking	F_1F_2 Integration	
3D FT analysis		
$t_1t_2t_3$ Pre processing		
	F_2F_3 2D analysis (any 2D strategy)	
	F_1 1D analysis (any 1D or partial 2D strategy)	
		$F_1F_2F_3$ Post processing

[a] This phase is actually implemented in several independent phases.

dependence is not an integer, it is a requisite to be able to do shifts by a fractional number of spectral points. This is done in NPK by switching to the time domain with a causality preserving transform, and by multiplying the data by a given frequency (realized with the `phase()` command). This allows a continuous shifting with no loss of information as long as the FT/inverse FT operations conserve the causality. Fig. 3 presents this shearing implementation. Tilt is realized in NPK in the same manner, by realizing the frequency shift operation along the F_2 axis instead of the F_1 one.

4.3. MaxEnt processing

MaxEnt processing is implemented in NPK, and available as a tool in several processing strategies. It is rapidly presented here in the frame of the weak modeling as presented in the Theory section, more details can be found elsewhere [30,13]. For a given experimental data-set d_i and a transfer function T_f the function Q described below is defined as the weighted sum of the entropy S and of the χ^2.

$$Q = S - \lambda\chi^2$$

$$\chi^2 = \sum_{i=1}^{M}\left(\frac{d_i - T_f(f_j)}{\sigma_i}\right)^2$$

$$S = -\sum_{j=1}^{N} p_j \log p_j$$

where σ_i is the uncertainty on the ith experimental point and p_j is the normalized image: $p_j = \frac{f_j}{\sum_k^N f_k}$.

For a given Lagrange multiplier λ, the spectral image $\{f_j\}$ which maximizes Q is chosen iteratively. Starting from $\lambda = 0$ and a flat image, λ is progressively raised until $\chi^2 \approx M$. The advantage of this approach is that the λ value and the default spectral level are algorithmically determined, and the only input from the user is the data uncertainty which is generally taken equal to the noise level. In the case of partially sampled data-sets, the χ^2 summation only runs on the actually acquired data points.

The above definition of the entropy implies image positivity. However positive/negative spectral reconstruction is possible by using the multi-channel approach [31] (see

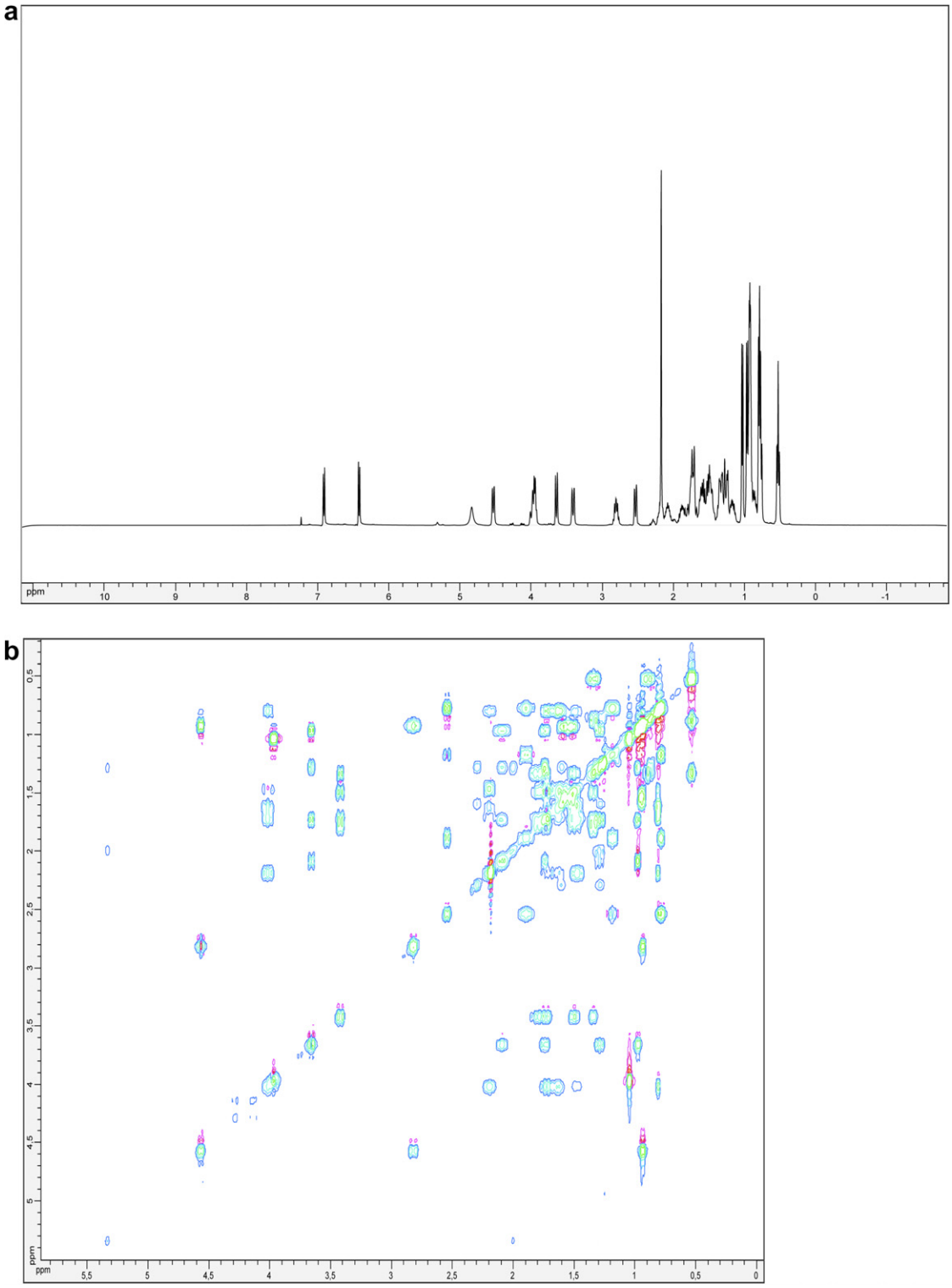
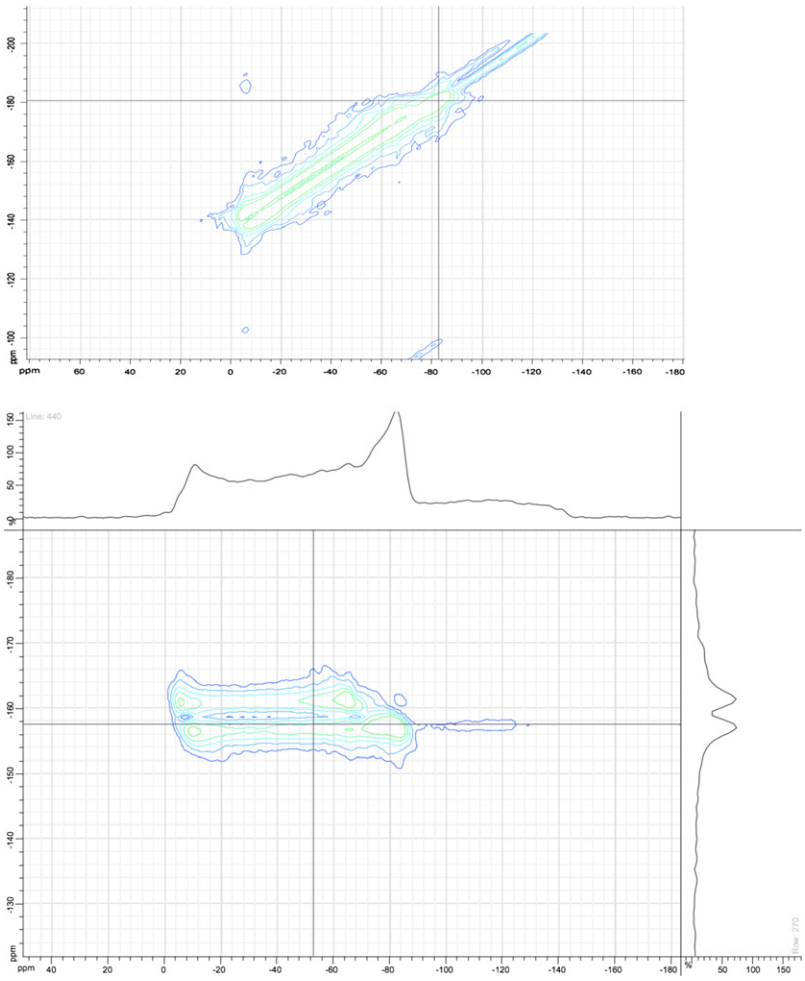


Fig. 2. Results from unattended processing of NMR data. (a) 1D data from a Lasalocid sample in $CDCl_3$, acquisition was performed on a Bruker Avance 400 spectrometer, FID is 2048 complex points, the spectral width is 5208.3 Hz. A 1.0 Hz apodisation was applied on the FID; a Fourier transform with zerofilling to 4096 complex points and an automatic phase correction was applied by minimizing the negative wings of the spectrum; an automatic offset baseline correction was then applied to the data-set (FT1D command). (b) 2D TOCSY data from the same sample in the same conditions. The dataset is 164×1024 hypercomplex points acquisition in States-TPPI protocol. A 20% shifted sine-bell apodisation was applied on the FID; a States-TPPI Fourier transform with zero filling to 512×2048 hypercomplex points and an automatic phase correction was applied by minimizing the negative wings of the spectrum; an automatic moving-average baseline correction was applied along both axes of the data-set (FT2D command). Note the slight phase error along the F_1 axis.

a



b

```
def shear( slope, pivot ):
    """
 shearing 2D
 slope removed to the spectrum by the operation.
   1 corresponds to going from the main diagonal to the horizontal
 pivot : 0..1
   position of the invariant column. 0: lower side - 1: upper side
 shear(1,0.5)       :  goes from SECSY to COSY
 shear(-1,0.5)      : goes from INADEQUATE to COSY
 shear(7.0/9,0.5)   : shears a Spin 3/2 MQMAS
 see also : tilt.g
    """
    if ( get_dim() != 2 ):
        raise "To be applied on 2D only"
    si = get_si1_2d()
    sisi = 2*power2(si-2)
    chsize( sisi, get_si2_2d())
    if ( get_itype_2d() == 0 or get_itype_2d() == 1 ):
        thetype = "real"
        iftbis("f1")
    else:
        thetype = "complex"
        ift("f1")
    si2 = get_si2_2d()
    sw2 = get_specw_2_2d()
    sw1 = get_specw_1_2d()
    for i in range(1, si2+1 ):
        col(i)
        dim(1)
        f2 = (i-pivot*si2)*sw2/si2
        decal = f2 * si / sw1
        coef = -180.0 * decal * slope
        phase( 0.5*coef, coef )
        dim(2)
        put("col", i)
    if ( thetype == "real" ):
        ftbis("f1")
    else:
        ft("f1")
    chsize( si, get_si2_2d())
```

Fig. 3. Presentation of the generic shearing tool. (a) the effect of the shearing on a ^{27}Al 3Q-MAS spectrum of $GeAl_{12}$. (b) the python code implementing this shearing operation. The shearing is implemented by an inverse Fourier transform conserving the causality (`iftbis()`) followed by a complex multiplication (`phase()`) equivalent to a continuously defined shift along the spectral axis.

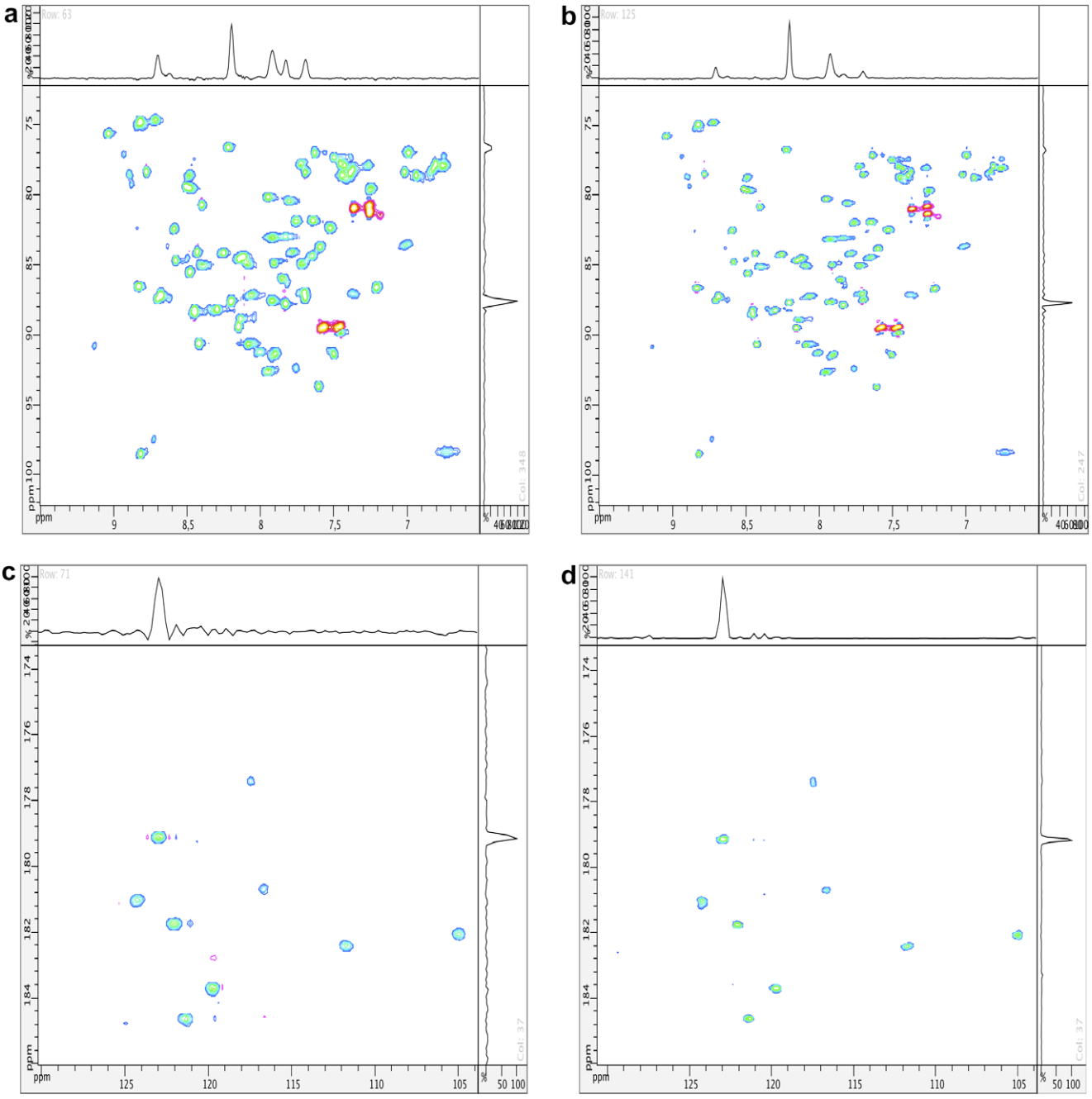


Fig. 4. MaxEnt processing. (a) 2D ^{15}N–^{1}H HSQC experiment of the wheat ns-LTP liganded with Lyso-myristoyl-phosphatidyl-glycerol [40], acquired at 500 MHz on a Bruker Avance spectrometer equipped with a cryoprobe. FID is 64×1024 hypercomplex and acquisition time took 50 min. Regular FT processing (FT2D command), on 128×2048 real points. The row and column displayed are extracted from the 2D at the location of the largest peak: ($\delta^{15}N = 87.65$ ppm–$\delta^{1}H = 8.20$ ppm). (b) The same data as in (a) processed by MaxEnt (MaxEnt2D command), (0.0,2.0) Hz preconvolution and a (2.0,5.0) Hz deconvolution was used. Image is reconstructed on 256×512 real points (only this spectral zone was processed, resolution is equivalent to 256×2048). Positive/negative multichannel method was used, and both images were added at the end of the processing. Processing took 9 s on a 1.66 GHz ppc processor, for 11 iterations and a final $\chi^2/M = 0.68$. (c) plane extracted at $\delta^{1}H = 9.30$ ppm from the HNCO experiment on the RPF domain of RV1009 from *M. tuberculosis* [41,42] acquired on the same spectrometer as in (a) FID is $64(^{13}C) \times 45(^{15}N) \times 512$ hypercomplex and acquisition time took 13 h. The row and column displayed are extracted from the plane at the location of the largest peak: ($\delta^{13}C = 179.14$ ppm–$\delta^{15}N = 122.88$ ppm). (d) The same plane as in (c) partially sampled at 20% with a random sampling following a 10 Hz exponential law on both axes [13] corresponding to a 2.6 h acquisition time. The plane was processed by MaxEnt, (10.0, 10.0) Hz preconvolution and a (10.0, 10.0) Hz deconvolution was used. Image is reconstructed on 128×256 real points. Processing took 113 s for 1000 iterations and a final $\chi^2/M = 1.02$.

Fig. 4). Transfer functions based on Fourier as well as Laplace transforms are implemented for 1D and 2D, and several deconvolution functions are provided. In consequence, 2D and 3D spectra, eventually partially sampled, as well as DOSY experiments can conveniently be processed by the MaxEnt algorithm.

The convergence algorithm is based on the Gifa algorithm [30,23,13] and shares the same properties, however the control logic has been fully rewritten.

5. Conclusion

Processing has been central in FT NMR since the early days of this spectroscopy. The constant increase in processing power afforded by the scientific computers, and the recent development in spectroscopic techniques have multiplied the use of computer processing in NMR spectroscopy rendering the need for efficient processing more and more acute. We have presented here a new NMR processing program which attempts to cover most of the processing situation met in the present NMR laboratory. The program does not present all the possible techniques presented so far in the literature, but allows a rapid implementation of new methods thanks to a modular organisation, and an easy to use programming interface.

This program is based on an intensive modeling of the processing. Within this modeling, regular operations as well as more sophisticated ones are implemented on a library of elementary actions.

The list of the actions which are actually implemented is biased against a weak modeling of the acquisition process, rather than the strong modeling of the system under study. As a consequence, transform operations such as Fourier, Hilbert, or Laplace transforms are favored in the underlying mathematical kernel on which NPK is built.

Thanks to this modeling, new operations not yet implemented, or even not yet described in the literature, can easily be added with a minimum burden.

The conjunction of comprehensive processing modeling with the definition of default values for every action permits very simply defined automatic processing. These default processing approaches, while clearly sub-optimum, are however sufficiently correct to permit a rapid inspection of the data. They can also be easily used as a starting point of a more detailed one.

Such a set-up can be of potential interest for developing automatic NMR processing pipe-lines which are required presently in fields as varied as protein structure [32,33], automatic analysis [34,35], or metabonomics [36].

Finally it should be pointed out that the purpose of this work is not to provide the user with a graphic tool permitting to rapidly interact graphically with the data. Included with the program is a rudimentary 1D graphic utility which permits to verify the quality on the data-set but which cannot, by any mean, be used comfortably on a regular basis. However the date-sets created by NPK are in a quite standard format and can be read without modifications by graphic softwares such as NMRview [37], AURELIA [38], Gifa [23] or NMRnotebook [39].

Acknowledgments

This work was supported by the NMRtec S.A.S. company (Illkirch-Graffenstaden, France), by university Montpellier I and by CNRS. Dr. Francis Taulelle is acknowledged for the data-set presented in Fig. 3. Dr. Philippe Barte and Pr. Christian Roumestand are acknowledged for the data-set presented in Fig. 4b. The authors thank C. Royer for her assistance with the english.

References

[1] R. Ernst, Sensitivity enhancement in magnetic resonance, Adv. Magn. Reson. 2 (1966) 1.
[2] R. Ernst, W. Anderson, Application of Fourier transform spectroscopy to magnetic resonance, Rev. Sci. Instrum. 37 (1966) 93.
[3] D.J. States, R.A. Haberkorn, D.J. Ruben, A two dimensional nuclear overhauser experiment with pure absorption phase in four quadrants, J. Magn. Res. 48 (1982) 286–292.
[4] D. Marion, K. Wüthrich, Application of phase sensitive 2D correlated spectroscopy (COSY) for measurements of ^{1}H–^{1}H spin–spin coupling constants in proteins, Biochem. Biophys. Res. Commun. 113 (3) (1983) 967–974, jAIJA.
[5] L. Kay, P. Keifer, T. Saarinen, Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, J. Am. Chem. Soc. 114 (26) (1992) 10663–10665.
[6] J. Simorre, B. Brutscher, M. Caffrey, D. Marion, Assignment of NMR spectra of proteins using triple-resonance two-dimensional experiments, J. Biomol. NMR 4 (1994) 325–333.
[7] E.K. Ji Won Yoon, Simon Godsill, R. Freeman, Deterministic and statistical methods for reconstructing multidimensional NMR spectra, Magn. Reson. Chem. 44 (2006) 197–209.
[8] H.S. Atreya, A. Eletsky, T. Szyperski, Resonance assignment of proteins with high shift degeneracy based on 5d spectral information encoded in g2ft NMR experiments, J. Am. Chem. Soc. 127 (2005) 4554–4555.
[9] V. Jaravine, I. Ibraghimov, V. Orekhov, Removal of a time barrier for high resolution multidimensional NMR spectroscopy, Nat. Meth. 3 (8) (2006) 606–609.
[10] G. Lippens, P.R. Bodart, F. Taulelle, J.-P. Amoureux, Restricted linear least square treatment processing of heteronuclear spectra of biomolecules using the anafor strategy, J. Magn. Reson. 161 (2003) 174–182.
[11] V. Orekhov, I. Ibraghimov, M. Billeter, Optimizing resolution in multidimensional NMR by three-way decomposition, J. Biomol. NMR 27 (2) (2003) 165–173.
[12] J. Chen, D. Nietlispach, A. Shaka, V. Mandelshtam, Ultra-high resolution 3D NMR spectra from limited-size data sets, J. Magn. Reson. 169 (2) (2004) 215–224.
[13] M. Delsuc, D. Tramesel, Application of maximum entropy processing to NMR multidimensional datasets, partial sampling case, C.R. Acad. Sci. 9 (3-4) (2006) 364–373.
[14] D. Marion, Fast acquisition of NMR spectra using Fourier transform of non-equispaced data, J. Biomol. NMR 32 (2005) 141–150.
[15] B.E. Coggins, P. Zhou, Sampling of the NMR time domain along concentric rings, J. Magn. Reson. 184 (2007) 207–221.
[16] E. Kupče, T. Nishida, R. Freeman, Hadamard NMR spectroscopy, Prog. NMR Spectrosc. 42, 95–122.
[17] J. Ashida, E. Kupče, J. Amoureux, Hadamard NMR spectroscopy in solids, J. Magn. Reson. 178 (2006) 129–135.
[18] L. Frydman, T. Scherf, A. Lupulescu, The acquisition of multidimensional NMR spectra within a single scan, Proc. Natl. Acad. Sci. USA 99 (2002) 15858–15862.
[19] P. Pelupessy, Adiabatic single scan two-dimensional NMR spectroscopy, J. Am. Chem. Soc. 125 (40) (2003) 12345–12350.

[20] H.S. Atreya, T. Szyperski, Rapid NMR data collection, Methods Enzymol. 394 (2005) 78–108.
[21] W. Boucher, Azara, (1993). Available from: <http://www.bio.cam.ac.uk/azara>.
[22] F. Delaglio, S. Grzesiek, G. Vuister, G. Zhu, J. Pfeifer, A. Bax, Nmrpipe: a multidimensional spectral processing system based on unix pipes, J. Biomol. NMR 6 (3) (1995) 277–293, article F Delaglio, Niddk, Chem Phys Lab, Bldg 5 B2-31, 5 Ctr Dr Msc 0505, Bethesda, MD 20892 US.
[23] J. Pons, T. Malliavin, M. Delsuc, Gifa v 4: A complete package for NMR data set processing, J. Biomol. NMR 8 (4) (1996) 445–452.
[24] J. Hoch, A. Stern, Rowland NMR toolkit, in: NMR Data Processing, Wiley-Liss, New-York, 1996.
[25] R. Fogh, J. Ionides, E. Ulrich, W. Boucher, W. Vranken, J. Linge, M. Habeck, W. Rieping, T. Bhat, J. Westbrook, K. Henrick, G. Gilliland, H. Berman, J. Thornton, M. Nilges, J. Markley, E. Laue, The CCPN project: an interim report on a data model for the NMR community, Nat. Struct. Biol. 9 (6) (2002) 416–418.
[26] R. Fogh, W. Boucher, W. Vranken, A. Pajon, T. Stevens, T. Bhat, J. Westbrook, J. Ionides, E. Laue, A framework for scientific data modeling and automated software development, Bioinformatics 21 (8) (2005) 1678–1684.
[27] W. Vranken, W. Boucher, T. Stevens, R. Fogh, A. Pajon, M. Llinas, E. Ulrich, J. Markley, J. Ionides, E. Laue, The CCPN data model for NMR spectroscopy: development of a software pipeline, Proteins 59 (4) (2005) 687–696.
[28] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, D. Sorensen, LAPACK Users' Guide, third ed., Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999.
[29] J. Hugunin, B. Warsaw, F. Bock, The jython project, (2005). Available from: <http://www.jython.org>.
[30] M. Delsuc, A new maximum entropy processing algorithm, with applications to nuclear magnetic resonance experiments, in: J. Skilling (Ed.), Maximum Entropy and Bayesian Methods, 1989th ed., Kluwer Academic, Dordrecht, 1989, pp. 285–290.
[31] V. Stoven, J. Annereau, M. Delsuc, J. Lallemand, A new n-channel maximum entropy method in NMR for automatic reconstruction of ''decoupled spectra'' and *J*-coupling determination, J. Chem. Inform. Comput. Sci. 37 (2) (1997) 265–272, article Stoven V, Ecole Polytech, Dcso, Lab Rmn, F-91190 Palaiseau, France.
[32] M. Baran, H. Moseley, J. Aramini, M. Bayro, D. Monleon, J. Locke, G. Montelione, Spins: a laboratory information management system for organizing and archiving intermediate and final results from NMR protein structure determinations, Proteins 62 (4) (2006) 843–851.
[33] B. López-Méndez, P. Güntert, Automated protein structure determination from NMR spectra, J. Am. Chem. Soc. 128 (40) (2006) 13112–13122.
[34] L. Griffiths, Towards the automatic analysis of ^{1}H NMR spectra NMR spectra, Magn. Reson. Chem. 38 (2000) 444–451.
[35] L. Griffiths, Towards the automatic analysis of ^{1}H NMR spectra: Part 2. accurate integrals and stoichiometry, Magn. Reson. Chem. 39 (2001) 194–202.
[36] K. Ott, N. Aranibar, Nuclear magnetic resonance metabonomics: methods for drug discovery and development, Meth. Mol. Biol. 358 (2006) 247–272.
[37] B. Johnson, R. Blevins, Nmrview: A computer program for the visualization and analysis of NMR data, J. Biomol. NMR 4 (603) (1994) 614.
[38] K.-P. Neidig, M. Geyer, A. Görler, C. Antz, R.S.W. Beneicke, H.R. Kalbitzer, Aurelia, a program for computer-aided analysis of multidimensional NMR spectra, J. Biomol. NMR 6 (3) (1995) 255–270.
[39] NMRtec, Nmrnotebook software, (2007). Available from: <http://www.NMRnotebook.com>.
[40] J.-L. Pons, F.d. Lamotte, M.-F. Gautier, M.-A. Delsuc, Refined solution structure of a liganded type 2 wheat non-specific Lipid Transfer Protein, J. Biol. Chem. 16 (2003) 14249–14256.
[41] M. Cohen-Gonsaud, P. Barthe, F. Pommier, R. Harris, P. Driscoll, N.H. Keep, C. Roumestand, ^{1}H, ^{15}N, and ^{13}C chemical shift assignments of the resuscitation promoting factor domain of rv1009 from *Mycobacterium tuberculosis*, J. Biomol. NMR 30 (2004) 373–374.
[42] M. Cohen-Gonsaud, P. Barthe, C. Bagnéris, B. Henderson, J. Ward, C. Roumestand, N.H. Keep, The structure of a resuscitation-promoting factor domain from *Mycobacterium tuberculosis* shows homology to lysozymes, Nat. Struct. Biol. 12 (3) (2005) 270–273.