ASCOM Tutorial

Ewen LESCOP, Bernhard BRUTSCHER

March 1, 2007

Laboratoire de RMN
Institut de Biologie Struturale
41, rue Jules Horowitz
38027 GRENOBLE Cedex 1
FRANCE

1 Introduction

The ASCOM protocole has been described in the article: LESCOP, E., Schanda, P., Rasia, R. & Brutscher, B. (2007) Automated Spectral Compression for Fast Multidimensional NMR and Increased Time Resolution in Real-Time NMR Spectroscopy", *J. Am. Chem. Soc.*

ASCOM stands for Automated Spectral COMpression and is currently a set of three PERL scripts (ASCOM1.pl, ASCOM2.pl and ASCOM3.pl) for the automated optimal settings of spectral parameters for different kinds of experiments:

ASCOM1: Spectral width optimization for the indirect dimension (X) of a 2D $^1H - X$ type experiment, such as the ^{15}N dimension of a $^1H - ^{15}N$ HSQC. The X nucleus can be either ^1H, ^{15}N and ^{13}C. This application can deal with real as well as complex acquisition schemes.

ASCOM2-2proj: Projection angle and tilted spectral width optimizations for the indirect dimension (NCO) of a 2D ^1H-NCO type experiment where the NCO dimension corresponds to the projection of the N and CO chemical shifts. Two 2D spectra are obtained that contain the information about the chemical shifts of $cos(\alpha) * \omega_N \pm sin(\alpha) * \omega_{CO}$. For this reason, and in contrast with the two other experiments, the peak positions and spectral widths are defined in Hz. In ASCOM2-2proj, the angle α is defined so that for $\alpha = 0$, we end up with the $^1H - ^{15}N$ plane, and for $\alpha = 90°$, we have the $^1H - ^{13}CO$ plane. Only complex acquisition is currently supported.

ASCOM2-2sw: Combined ^{15}N and ^{13}CO spectral widths optimization for the indirect dimensions of a HNCO-type experiment, such as the 3D HNCO, or of a 4D HNCOCACB. Only complex acquisition is currently supported.

ASCOM3: Combined spectral width optimization of the indirect dimensions (X and 1H) of a 3D $^1H - X$ HSQC edited experiment, such as the indirect ^1H and ^{15}N dimensions of a ^{15}N HSQC-NOESY. The X nucleus can be either ^1H, ^{15}N and ^{13}C. This application can deal with only complex acquisition schemes.

The optimization is based on the knowledge of the chemical shifts of the cross-peaks in the related spectra, and makes uses of the aliasing properties of the Fourier Transform for real and complex data. ASCOM1.pl, together with the file param_ASCOM1.par are designed to be used for ASCOM1, whereas ASCOM2.pl together with the file param_ASCOM2.par should be run for ASCOM2-2proj and ASCOM2-2sw. ASCOM3.pl, together with the file param_ASCOM3.par are designed to be used for ASCOM3. The scripts are called through the command line:

./ASCOM1.pl param_ASCOM1.par

./ASCOM2.pl param_ASCOM2.par

./ASCOM3.pl param_ASCOM3.par

Make sure that the first line of the perl script contain the correct path to perl executable (#! /usr/bin/perl). Some general information will be written in the terminal and should be explicit.

1.1 Modes

The scripts can work under two modes:

Static mode

For a given sets of spectral parameters (carrier frequency, spectral width(s)), a spectrum ("spectr.dat") is generated and can be inspected with another software. The spectr.dat file is a matrix defined by the number of points in any dimension and filled with values 0,$\pm$1, $\pm$2,... The new chemical shifts, i.e. taking aliasing into account are written in a file ("new_attribution.dat"). This mode is useful to check the spectrum that will be acquired later on the spectrometer.

Gridsearch mode

One or two loops will be run on spectral parameters such as spectral width(s) , carrier frequencies (for TPPI) and projection angle (ASCOM2-2proj). The ouput file ("score.dat") contains the results for the generated spectra, including the set of spectral parameters and the number of important peaks (see below) that are resolved. The score.dat can be inspected in another software, such as Xmgrace for further inspection.

Two different approaches to analyze the spectra hold for the two modes :

1.2 Grid search mode

For the gridsearch mode (ASCOM1 and ASCOM2), a first check will be carried on the chemical shift list. Indeed, cross-peaks with different chemical shifts along the direct ^1H dimension will never overlap. A "contact" matrix containing this information is then first generated. It will simply contain for each cross-peak the list of other cross-peak that may overlap during the gridsearch ("contact" cross-peaks). Then, for any given set of spectral parameters, the new peak positions will be generated according to the aliasing function. Then, for each cross-peak, the pseudo-distances with all "contact" cross-peaks are calculated with the following equation:

$$d = (\frac{\omega_{H,ref} - \omega_{H,contact}}{RH})^2 + (\frac{\omega_{X,ref} - \omega_{X,contact}}{RX})^2 \text{ (ASCOM1 and ASCOM3)}$$

$$d = (\frac{\omega_{H,ref} - \omega_{H,contact}}{RH})^2 + (\frac{\omega_{NC,ref} - \omega_{NC,contact}}{RNC})^2 \text{ (ASCOM2-proj)}$$

$$d = (\frac{\omega_{H,ref} - \omega_{H,contact}}{RH})^2 + (\frac{\omega_{N,ref} - \omega_{N,contact}}{RN})^2 + (\frac{\omega_{C,ref} - \omega_{C,contact}}{RC})^2 \text{ (ASCOM2-2sw)}$$

Here, indices ref refer to the reference cross-peaks compared with the contact cross-peaks. The R_H, R_X,R_N,R_C are the linewidths defining the minimum distance allowed in each dimension, and are given as parameters by the user. Cross-peaks are overlapping if $d < 1$. The user can define three sets of R parameters that will give information about overlapping on three different shells at the same time. Thus the output will present the number of overlapping cross-peaks for the three shells.

1.3 Generation of the spectrum (Static Mode)

In the static mode, the spectra are represented by 2D or 3D matrices containing the cross-peaks. Each cross-peak is defined by an ellipse (ellipsoid) that is digitized. The level of digitization, that is the number of points in each dimension, is set by the user and should be a compromise between the speed of calculation and the accuracy of the predicted spectra. For matter of comparison, the results of the number of overlapping cross-peaks are given for both direct distance calculation, and for overlapping estimated from the predicted spectra.

2 Input files

2.1 Chemical shifts

The chemical shift information is entered as a separate file (for example "ass.tab") with the following format:

peak0 8.279 116.665 180.766 0

peak1 7.534 112.475 180.557 0

peak2 6.867 112.502 180.560 1

peak3 7.488 111.543 180.381 0

The columns have the following meanings:

peak name -^{1}H (ppm) -^{15}N (ppm) -^{13}CO (ppm) - flag

The third column should be omitted when working with ASCOM1 and ASCOM3. The flag is related to the important peaks that the user wants to keep resolved in the final spectra. Thus, a value of 1 will be assigned to important peaks and a value of 0 to the others, that will be allowed to overlap. Of note, the script ASCOM2 can work only for H,N and CO nuclei (i.e. HNCO correlations) whereas ASCOM1/ASCOM3 affords H, N or C nuclei in the indirect dimension (i.e. for^{1}H-^{1}H, ^{1}H-^{15}N, and ^{1}H-^{13}C) . For ASCOM1/ASCOM3, the γ values for the nucleus X will be automatically changed while setting the parameter $X to "H", "N" or "C".

2.2 Parameter files

Two files containg the parameters for the two applications are provided. They are quite similar but different meanings can be assigned to the parameters. Common parameters are found:

$RH_n, $RN_n $RC_n: Linewidths (in Hz) for the crosspeaks in the given dimensions for three shells of different sizes. These values can be interpreted either as the natural size of the cross-peaks (in Static Mode), either as the cutoff distance that will be used as a criterium to define if two peaks are overlapped or not (Gridsearch Mode).

$field: Magnetic field of the spectrometer, given in MHz. Ex: 500, 600, 800...

$type_fold: flag to mention if real or complex acquisition is used (only in ASCOM1). 1 for TPPI (real), -1 for complex or Echo-Antiecho

$sign_fold: flag to set the sign of peaks folded an odd number of times (1: positive, -1:negative)

When working with the Static mode, the number of points in the indirect dimension(s) is set by the parameters with "_fin" extensions ($Npts_fin, $Cpts_fin for 2D and 3D applications, and $NCpts_fin for the tilted dimension in ASCOM2-2proj). The spectral width, carrier frequencies and projection angle are set by $dof_fin, $sw_fin, $dofN_fin, $swC_fin, $dofC_fin, $sw_fin, $swtilt_fin, $alpha_deg_fin.

When working with the Gridsearch mode, the gridsearch steps are generally defined by three parameters: $minSWX, $maxSWX and $incr_SWX for the spectral widths, and $min_alpha_deg, $max_alpha_deg, and $incr_alpha_deg for ASCOM2-2proj. For real acquisitions (in ASCOM1, $type_fold=1), the gridsearch on the carrier frequency will be automatically done between $dof2_def-$sw_trial/2 and $dof2_def+$sw_trial/2 by $incr_dof steps, where $sw_trial represents the current ^{15}N spectral width. For ASCOM2-2proj. the gridsearchs on α and SW_{tilt} can be run separately by setting either $grid_search_alpha=1 or $grid_search_SW=1. If both flags are set to 1, the 2D gridsearch on α and SW_{tilt}will be run.

3 Example of a ASCOM1 run

Let's assume that you want to optimize the spectral width of the ^{15}N dimension of an HSQC. This may be for rapid HSQC experiments or for the ^{15}N dimension of 3D ^{15}N edited-spectra, such as 3D-NOESY-HSQC experiments. You may use the followings steps.

1. Gerenate the file containing the chemical shift information (four columns: peak name, ^{1}H and ^{15}N chemical shifts (in ppm), and the flag). Make sure that you don't have any folded peaks in your list.

2. Modify the names for input and output files if need at the end of the file param_ASCOM1.par. Set the general parameters (field, type of aliasing, sign of folded peaks).

3. Set RH_n and RN_n (in Hz; n=1,2,3) according to your need. Usually, these values can range from 15*15Hz (3D spectra) to 50*50Hz (SOFAST-HMQC). This is easily estimated from the reference (i.e. full spectral width) spectrum. Be careful, the biggest RH will be used to define the number of contact residues. As a consequence, the larger it is, the slower the computation is.

4. Set the parameters for the gridsearch on ^{15}N spectral width ($minSWX, $maxSW and $incr_SW), and for the carrier frequency in case of real acquisition ($incr_dof).

5. Start the Gridsearch Mode by setting $gridsearch=1 and ./ASCOM1 param_ASCOM1.par. Have a look to the terminal: the results of the analysis of the spectra are given with the number of important peaks that are resolved. This has to be compare with the total number of important peaks.

6. Analyze the output "score.dat" that mostly contains the same information as displayed in the terminal (carrier frequency | spectral width | number of resolved important peaks). In particular, you will plot the number of resolved peaks versus the spectral width.

7. When you have chosen the optimized spectral width (and carrier frequency) , you can predict the new spectrum by using the Static Mode. As output, you will get the spectrum as a matrix, as well as the new chemical shifts ("new_assignment.dat"). This file contains the apparent chemical shift values together with a flag indicating the overlapping residues. This may be used to easily map the assignment of the acquired spectrum.

8. Go to the spectrometer and acquire the spectrum with the optimized parameters. Since aliasing-induced overlapping is carrier-frequency independent for complex acquisition, you don't have to worry about the carrier frequency. However, in real acquisition, the carrier frequency will have to be adjusted carefully on the spectrometer.

9. Read the spectrum in your preferred visualisation software. NMRView is fine since it automatically plots the peak picking at the correct position (for complex acquisition).

4 Example of a ASCOM2-2proj run

Let's assume that you want to optimize the projection angle and the tilted spectral width of the 3D-HNCO experiments along the^{15}N and ^{13}CO dimensions. This may be useful for (4-3D)-HNCOCACB experiment, which end up with two 3D spectra containing ^{1}H and $^{13}C_{\alpha\beta}$ dimensions and the two $^{15}N/^{13}CO$ subspectra. You may use the followings steps.

1. Gerenate the file containing the chemical shift information (five columns: peak name, ^{1}H, ^{15}N and ^{13}CO chemical shifts (in ppm), and the flag). The dimensions should be in this order. Make sure that you don't have any folded peaks in your list.

2. Modify the names for input and output files if need at the end of the file param_ASCOM2.par. Set the general parameters (field, sign of folded peaks) and $ASCOM=2.

3. Set RH_n, RC_n and RN_n (in Hz; n=1,2,3) according to your need. Usually, these values can be quite short. 30 (H)*15 (N)*15 (C) Hz are usually sufficient to separate the maximum of the cross-peaks in the (4-3D) experiment. The linewidth of a tilted cross-peaks will be calculated as $\$RN * cos^2(\alpha) + \$RC * sin^2(\alpha)$. Be careful, the biggest RH will be used to define the number of contact residues. As a consequence, the larger it is, the slower the computation is.

4. Set the parameters for the gridsearch on the tilted angle ($min_alpha_deg, $max_alpha_deg, and $incr_alpha_deg) and $gridsearch_alpha=1 and $gridsearch_SW=0. Set also the tilted spectral width at maximum. As indication, the minimum spectral width to prevent folding for any α is written in the terminal, together with the associated projection angle.

5. Start the Gridsearch Mode on α and run ./ASCOM2 param_ASCOM2.par. Have a look to the terminal: the results of the analysis of the spectra are given with the number of important peaks that are resolved for each shell. This has to be compare with the total number of important peaks.

6. Analyze the output "score.dat" that mostly contains the same information as displayed in the terminal. The format is: ^{15}N and ^{13}CO carrier frequencies, tilted spectral width, α, and for each shell: number of peaks resolved in both subspectra, number of peaks resolved in exactly one subspectrum, number of peaks never resolved, number of peaks resolved in at least one subspectrum. Plot the number of resolved peaks (the one you prefer) versus the projection angle and select the optimized angle.

7. Set $alpha_deg_fin to the selected value. Set $gridsearch_alpha=0 and $gridsearch_SW=1.

8. Start the Gridsearch Mode for the tilted spectral width and run ./ASCOM2 param_ASCOM2.par. Have a look to the terminal: the results of the analysis of the spectra are given with the number of important peaks that are resolved. for decreasing values of SW_{tilt}.

9. Run the Static Mode ($gridsearch=0) to have a look to the message in the terminal. Two spectra will be generated with cross-peaks located at $cos(\alpha) * \omega_N + sin(\alpha) * \omega_{CO}$ ("spectr_p.dat") and at $cos(\alpha) * \omega_N - sin(\alpha) * \omega_{CO}$ ("spectr_m.dat"). As output, you will also get the new chemical shifts ("new_assignment.dat") in the subspectra. This may be used to easily map the assignment of the acquired spectrum.

10. Select the optimized spectral width to run the experiment on the spectrometer.

As an alternative, one may also run the full gridsearch on both α and SW by setting $gridsearch_alpha=1 and $gridsearch_SW=1.

5 Example of a ASCOM2-2sw run

Let's assume that you want to optimize the spectral widths of the ^{15}N and ^{13}CO dimensions of a 3D -HNCO before going to the 4D-HNCOCACB experiment, as illustrated in the article. You may use the followings steps.

1. Gerenate the file containing the chemical shift information (five columns: peak name, ^1H, ^{15}N and ^{13}CO chemical shifts (in ppm), and the flag). The dimensions should be in this order. Make sure that you don't have any folded peaks in your list.

2. Modify the names for input and output files if need at the end of the file param_ASCOM2.par. Set the general parameters (field, sign of folded peaks) and $ASCOM=3.

3. Set RH_n, RC_n and RN_n (in Hz, n=1,2,3) according to your need. Usually, these values can be quite short. 30 (H)*15 (N)*15 (C) Hz are usually sufficient to separate the maximum of the cross-peaks for a 4D spectrum. This is easily estimated from the reference (i.e. full spectral width) spectrum. Be careful, the biggest RH will be used to define the number of contact residues. As a consequence, the larger it is, the slower the computation is.

4. Set the parameters for the gridsearch on ^{15}N and ^{13}C spectral widths ($minSWX, $maxSWX and $incr_SW).

5. Start the Gridsearch Mode by setting $gridsearch=1 and run ./ASCOM2 param_ASCOM2.par. Have a look to the terminal: the results of the analysis of the spectra are given with the number of important peaks that are resolved for each shell. This has to be compare with the total number of important peaks.

6. Analyze the output "score.dat" that mostly contains the same information as displayed in the terminal (carrier frequencies | spectral widths | number of resolved important peaks). One way is to extract the combination of spectral widths in both dimensions leading to the maximum of resolved cross-peaks and to plot the results on a 2D graph as in the article.

7. Go to the spectrometer and acquire the 4D spectrum with the optimized parameters. Since aliasing-induced overlapping is carrier-frequency independent for complex acquisition, you don't have to worry about the carrier frequency.

6 Example of a ASCOM3 run

Let's assume that you want to optimize the spectral widths of the ^1H and^{15}N dimensions of the first HSQC of the 4D 15N-HSQC NOESY HSQC spectrum. You may use the followings steps.

1. Gerenate the file containing the chemical shift information (four columns: peak name, ^1H and ^{15}N chemical shifts (in ppm), and the flag). Make sure that you don't have any folded peaks in your list.

2. Modify the names for input and output files if need at the end of the file param_ASCOM3.par. Set the general parameters (field, sign of folded peaks).

3. Set RH_n and RN_n (in Hz; n=1,2,3) according to your need. Usually, these values can range from 15*15Hz to 100*100Hz. This can be estimated from the reference (i.e. full spectral width) spectrum.

4. Set the parameters for the gridsearch on ^{15}N / ^1H spectral widths ($minSW, $maxSW, $incr_SW/ $minSWH, $maxSWH and $incr_SWH).

5. Start the Gridsearch Mode by setting $gridsearch=1 and ./ASCOM3 param_ASCOM3.par. Have a look to the terminal: the results of the analysis of the spectra are given with the number of important peaks that are resolved. This has to be compare with the total number of important peaks.

6. Analyze the output "score.dat" that mostly contains the same information as displayed in the terminal (carrier frequency | spectral width | number of resolved important peaks for each shell). In particular, you will plot the number of resolved peaks versus the spectral width.

7. When you have chosen the optimized spectral width (and carrier frequency) , you can predict the new spectrum by using the Static Mode. As output, you will get the spectrum as a matrix, as well as the new chemical shifts ("new_assignment.dat"). This file contains the apparent chemical shift values together with a flag indicating the overlapping residues. This may be used to easily map the assignment of the acquired spectrum.

8. Go to the spectrometer and acquire the spectrum with the optimized parameters. Since aliasing-induced overlapping is carrier-frequency independent for complex acquisition, you don't have to worry about the carrier frequency.

9. Read the spectrum in your preferred visualisation software. NMRView is fine since it automatically plots the peak picking at the correct position (for complex acquisition).