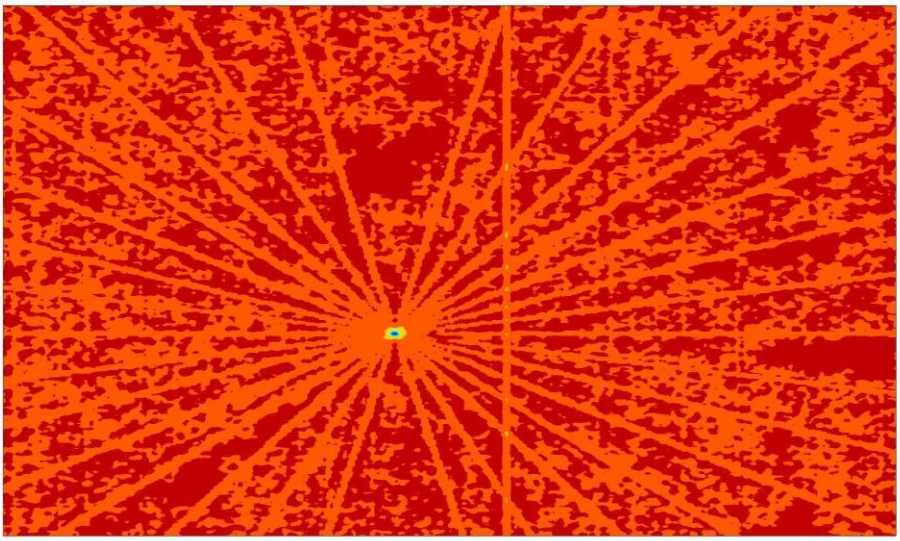
Projection Reconstruction in NMR and Intrinsically Unfolded Proteins

Yared Paalvast



Summary

Projection Reconstruction in NMR is a technique that allows us to record multidimensional NMR-experiments in less time than when recorded in the classical way. This is particularly advantageous for the study of intrinsically unfolded proteins whose spectra suffer from severe peak-overlap. To further reduce the time needed to obtain a NMR-spectrum, it is possible to use a small amount of a paramagnetic substance in the sample like MnS so that the delay between pulse-sequences can be made shorter without having to worry about saturating the sample. Because this also results in an increase in radiation per time unit this does ask for a pulse sequence with as little radiation as possible, in order to prevent overheating of the sample. In practice this means substituting decoupling-sequences with strategically placed 180-degrees pulses. It would be exciting if these techniques could be combined to do a backbone-assignment of the protein tau, the largest intrinsically unfolded protein on which this is done to date. To test whether we could make this work we first did a (Ha)$Ca_{i-1}CO_{i-1}$ N_iH_i –experiment on a calbindin sample, which gave satisfying results. However, another experiment we needed for the back-bone assignment and two other pulse-sequences tried since then, failed to give sufficient signal. This is something that has to be looked into for the future.

Introduction

Nuclear Magnetic Resonance (NMR) is one of the most important techniques available today to elucidate the structure (and some other characteristics) of proteins. For example in the period 2008-2009 7848 protein structures have been submitted to the protein databank of which 7.5 % were determined using NMR and 91.5 % protein structures have been determined using x-ray crystallography[1]. A great advantage of NMR over x-ray crystallography though is that the protein doesn't have to be crystallized and isn't destroyed by the measurement.

Using a high magnetic field and the spin property of nuclei allows us to examine what nuclei are close to one another and thus what the molecule on the whole must look like. The high magnetic field puts the system in disequilibrium, creating a bias for the spins to align with the magnetic field, thus creating a magnetic field of their own which we can measure by hitting it with radiofrequency(RF)-pulses to put the magnetization in the xy-plane and subsequently record the signal this creates by circular RF-coils in the NMR-apparatus. After doing a Fourier transform on the recorded signal, we get to see a spectrum of a multitude of nuclei precessing at a different frequency under the influence of the applied magnetic field and the direct environment of the nucleus. If we do this for only the protons in a protein, the spectrum obtained shows a complicated pattern of peaks with a lot of overlap, making it impossible to assign the different peaks to a single nucleus.

So if we want to have a chance to assign all the peaks to a specific nucleus of the protein, we would have to make sure that for a given nucleus there is only one specific frequency, or chemical shift on which it resonates. This is not possible. What is possible however is to label proteins with ^{15}N and ^{13}C and to manipulate the magnetization in such a way that you get resonances that are defined by the frequencies of the closely linked ^{1}H, ^{15}N and ^{13}C – nuclei, making it multidimensional and decreasing the probability of overlap. This is especially important for investigating intrinsically unfolded proteins, proteins with a less defined structure. Because of the floppy nature of these proteins, the surroundings of the examined nuclei tend to average out, thus leading to more crowded spectra.

Recording multidimensional NMR-spectra comes at a price. For each dimension added the recording time increases exponentially. While 3D-experiments take hours, 4D-experiments take weeks and 5D-experiments would take several years to complete.

This is where projection reconstruction shows its value.

Projection Reconstruction

Before delving into the particulars of projection reconstruction however, it might be best to recap what happens during a 'normal' multi-dimensional NMR-experiment.

Consider a 3D-experiment with two indirect domains and (of course) one direct domain. This means that the signal we record (direct domain) is 'modulated' independently by the two indirect domains. Using Euler's formula the signal being detected by the RF-coil can be described by:

$$S(t1, t2, t3) = \sum_{j=1}^{N} A_j \, e^{i\,\omega 1 j\, t1} \; e^{i'\,\omega 2 j\, t2} \, e^{i''\,\omega 3 j\, t3}$$

Where ω_i is the frequency in the i-th time domain, t_i is the evolution time in the i-th domain and A_j is the amplitude of the j-th resonance. Since the third domain is recorded directly, we get cosine and sine phase by simply using the quadrature detection in the RF-coil. To get the cosine as well as the sine modulated signals of the two indirect domains, we have to record a variance of pulse-sequences to get a phase-shift in the desired domain. Getting both sine and cosine modulation of each time domain is important to establish the sign of a resonance.

To get from the time domain to the frequency domain we have to perform a Fourier transform. This is done first for the third time domain (direct time domain).

$$S(t1, t2, \omega 3) = \sum_{j=1}^{N} M_j(t1, t2)\; A_j(resonance\ of\ spin\ j\ in\ 3\omega)$$

The M is used to emphasize that the signal is amplitude modulated by $M(t_1, t_2)$.

Now imagine a (t_1 x t_2) square with each (t_1,t_2) point in that square hiding a ω_3-spectrum in a third dimension. Subsequently, we take each row of the (t_1 x t_2) square (in each row t2 is constant) and perform a Fourier transform on that row (or t_1 x ω_3-plane) so we obtain a (ω_1 x t_2) square. If we then do the same for the columns, and recall the hidden ω_3 – dimension, we get the full (ω_1 x ω_2 x ω_3) - spectrum.

If we do these Fourier transforms, in turns so to say, for every time domain, we need a full set of increments for every other increment in the other domains. On top of that every time we add an indirect domain recording time will double because we need separate pulse sequences to record the cosine and sine modulated signal. If we now, for the fun of it, assume that the recording of a 1D-spectrum takes 1 second, and every indirect time domain added brings a score of 60 increments, we come up with the following list of recording times for n-D experiments.

1D = 1s

2D = 2 x 60 x 1 s = 120 s = 2 min.

3D = 2^2 x 60 x 60 x1 s = 1440 s = 240 min. = 4 hours

4D = 2^3 x 60 x 60 x 60 x 1 s = 1728000 s = 480 hours = 20 days = 2.9 week

5D = 2^4 x 60 x 60 x 60 x 60 x 1 s =207360000 s = 2400 days = 343 weeks = 6.6 years

Or:

$$nD = 2^{n-1} \times 60^n \times 1\,s$$

While 3D recordings are quite feasible this way, 4D recordings will need a lot of effort on the experimenter's behalf to convince his colleagues that it is imperative a 4D recording is done and 5D experiments would require an amount of patience (and money) people just won't have. (6D experiments would outlive the experimenter.)

Now that we are again reminded why we need a trick to obtain multidimensional spectra in a reasonable amount of time, let us examine what projection reconstruction is actually about.[2,3,4] As we have just seen, the biggest reason for multidimensional experiments to take so long is that we have to sample all increments of the other domains, for each increment in an indirect domain independently. The trick is now to sample the increments in the different indirect domains not independently but dependently. So instead of stepwise increasing the evolution time in one domain while keeping the evolution time in the other domains constant, we do a stepwise increase in the evolution time in all the different indirect time domains at the same time. If you then take different size constants for the steps in each indirect time domain, you get different projections. So if we have:

$$S(t1, t2, \omega 3) = \sum_{j=1}^{N} A_j \, e^{i\,\omega 1 j\, t1} \; e^{i'\,\omega 2 j\, t2} \, (resonance\ of\ spin\ j\ in\ \omega 3)$$

We can rewrite this as:

$$S\left(\frac{p1\,n}{sw1}, \frac{p2\,n}{sw2}, \omega 3\right) = \sum_{j=1}^{N} Aj \, e^{i\,\omega 1 j \frac{p1\,n}{sw1}} \, e^{i\,\omega 2 j \frac{p2\,n}{sw2}} \, (resonance\ of\ spin\ j\ in\ \omega 3)$$

Where sw is the spectral width, n is the increment counter and p is the projection value. For instance, if we take (p1,p2) = (0.5 , 0.5) and for convenience we define the reduced frequency φ as ω/sw, we get:

$$S(0.5\,\varphi1\,n, 0.5\,\varphi2\,n, \omega3) = \sum_{j=1}^{N} A_j\, e^{i\,0.5\,\varphi1 j\,n}\, e^{i\,0.5\,\varphi2 j\,n} (resonance\ of\ spin\ j\ in\ \omega3)$$

Fourier transformation will yield for every 3D resonance j: (0.5 φ1+ 0.5 φ2, ω3)
In this way we effectively project 3 dimensions onto 2 dimensions. Making multiple projections will give you different linear combinations of the reduced frequencies of the indirect domains and make it possible to reconstruct the original reduced frequencies and thus the spectrum. In fact, when we record the (0.5, 0.5) projection, we can combine the datasets so created in different ways, creating a (0.5, 0.5) and a (0.5, -0.5) projection.

To find out how exactly those datasets should be combined to get those different projections, we have to make the sines and cosines in the last formula, a little more explicit using the double angle formulas.

Since,

$$e^{i\,(a+b)} = \cos(a+b) + i\ \sin(a+b)$$

And,

$$\cos(a+b) = \cos a \cos b - \sin a \sin b$$

$$\sin(a+b) = \sin a \cos b + \cos a \sin b$$

We get:

$$\sum_{j=1}^{n} Aj\, e^{i\,0.5\,\varphi1 j\,n}\, e^{i\,0.5\,\varphi2 j\,n}$$

$$=$$

$$\sum_{j=1}^{n} Aj\ (\cos(0.5\,\varphi1 j\,n)\cos(0.5\,\varphi2 j\,n) - \sin(0.5\,\varphi1 j\,n)\sin(0.5\,\varphi2 j\,n) + i\sin(0.5\,\varphi1 j\,n)\cos(0.5\,\varphi2 j\,n) + \cos(0.5\,\varphi1 j\,n)\sin(0.5\,\varphi2 j\,n))$$

And if we write cos (0.5 n φ1) cos (0.5 n φ2) like cc we get:

$$S\,(p1\;\varphi 1\;n, p2\;\varphi 2\;n) = \sum_{j=1}^{n} Aj\;((cc - ss) + i(sc + cs))$$

And because:

$$\cos(-a) = \cos(a)$$

$$\sin(-a) = -\sin(a)$$

We have to multiply each term by -1 for each occurrence of sin (- n p_2 φ_2) to get:

$$S\,(p1\;\varphi 1\;n, -p2\;\varphi 2\;n) = \sum_{j=1}^{n} Aj\;((cc + ss) + i(sc - cs))$$

Also, the coefficient before a term (where term is cc, ss etc.) is easily found by substituting c by 1 and s by i because for every complex signal you get the real part of it as the cosine and the imaginary part as the sine. So if we take the product of these functions the coefficient can still be found by multiplying by 1 for each cosine in a term and by i for each sine in a term. This is illustrated by:

$$e^{ia}e^{ib} = (\cos a + i \sin a)(\cos b + i \sin b) = (cc - ss) + i(cs + sc)$$

If we now take for instance 3D we would get the terms:

$$ccc + ccs + \mathrm{csc} + css + scc + scs + ssc + sss$$

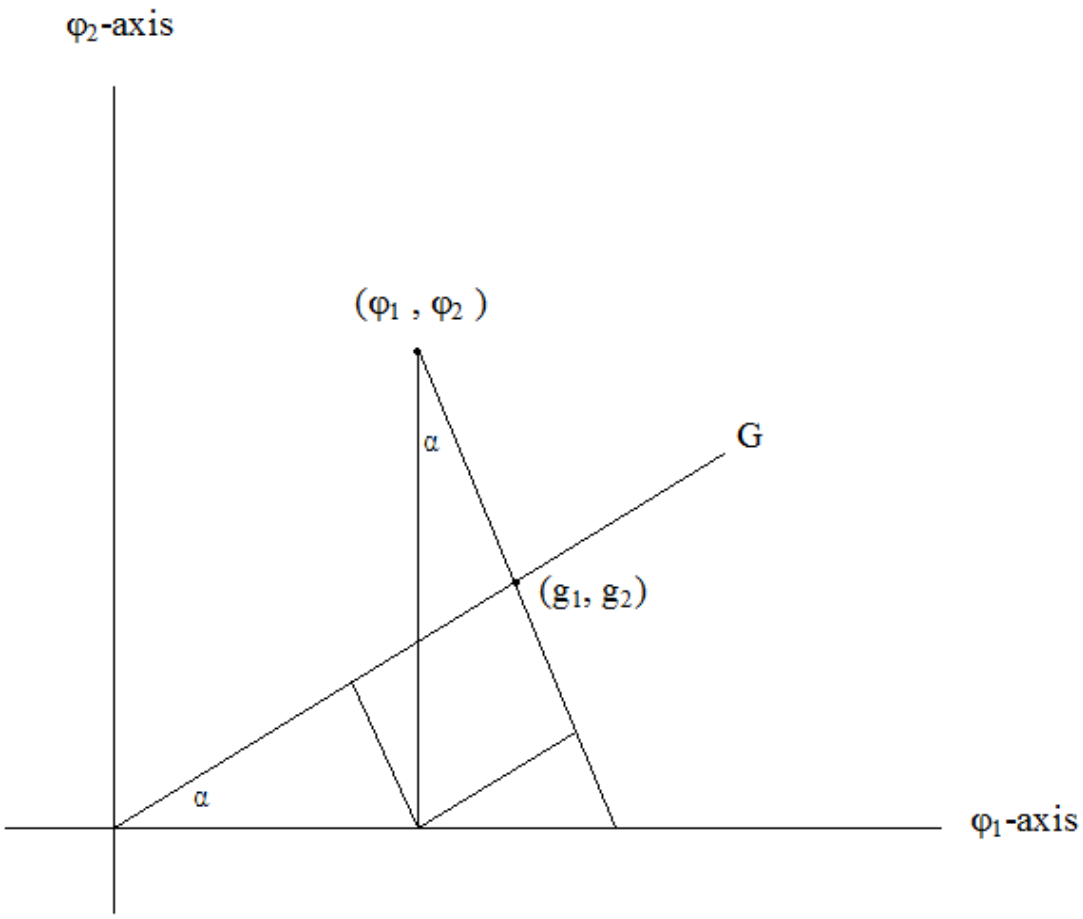
Then we find for the coefficients of S(-p1, p2, -p3):

$$ccc + css + scc - ssc + i(-ccs + csc - scc + sss)$$

So this is how you combine the datasets to obtain the data for the different projections, which will transform in the desired projection after a Fourier transform. But how do you use those projections to retrieve the original spectrum?

An algorithm that can be used for this purpose starts with going along each point in the reduced frequency space (φ1, φ2, …, φn) to find where it projects on all the projections and sums the intensity of all the points on the projections to that point in reduced frequency space. This is called simple back-projection and is illustrated in figure 1, along with some mathematical descriptions of the points and lines involved.

Figure 1



Where G can be described by : $\varphi_2 = a\,\varphi_1$

$$\varphi_2 = \varphi_1 \cos\alpha$$

If we have a random point (φ_1 , φ_2) as shown in figure 1, it projects on G as:

$$\left(\frac{\varphi_1}{\sqrt{1+a^2}} + \frac{a\varphi_2}{\sqrt{1+a^2}}\right)$$

$$(\varphi_1 \cos\alpha + \varphi_2 \sin\alpha)$$

Or simply: $(\sqrt{g_1^2 + g_2^2})$

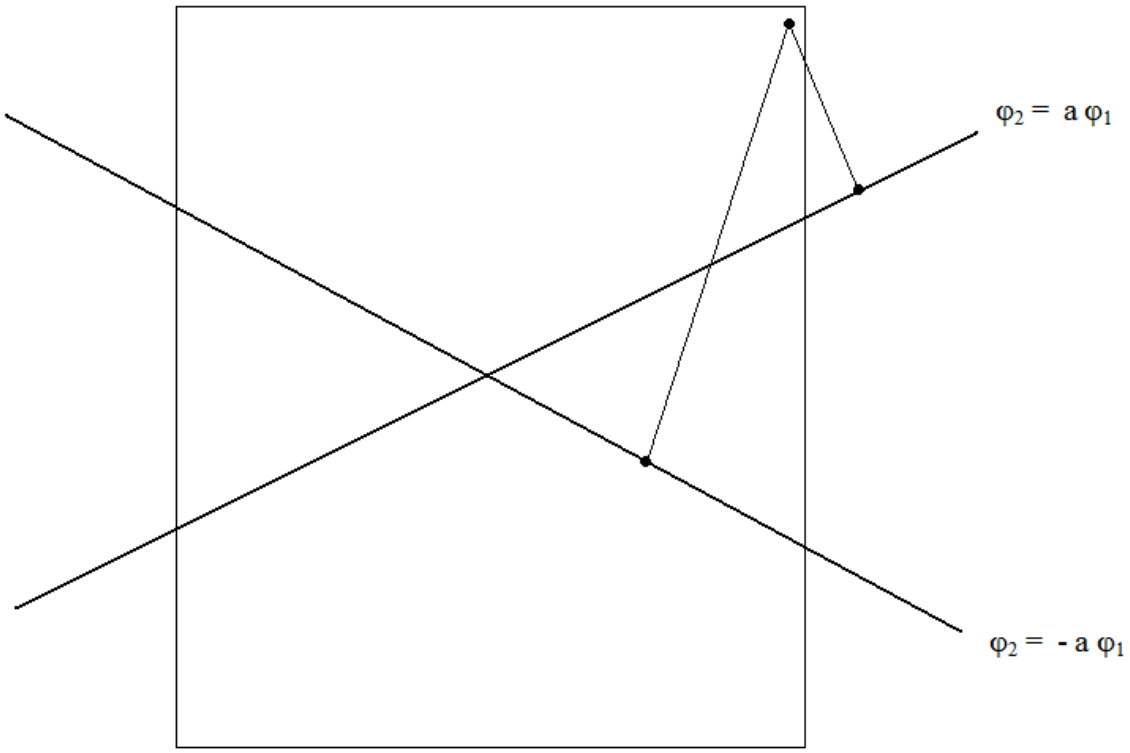
Where : $(g_1, g_2) = (g_1, ag_1) = \left(\frac{\varphi_1 + a\,\varphi_2}{1+a^2}, a\,\left(\frac{\varphi_1 + a\,\varphi_2}{1+a^2}\right)\right)$

This would all be for the projection:

$$(p_1, p_2) = \left(\frac{1}{1+a}, \frac{a}{1+a}\right)$$

If we now define the reduced frequency space as the square (φ_1 x φ_2) we get the situation as in figure 2:

Figure 2



It is evident that especially if you reach the corners of the reduced frequency space, the point will project outside the reduced frequency space on one projection and inside the reduced frequency space on the projection that is mirrored in the x-axis. So it looks as though we need a different reduced frequency space for the projections.

Now if you see to it that for the projection (p1, p2):

$$p_1 + p_2 = 1$$

And in general:

$$\sum p_i = 1$$

And you define a new reduced frequency space for the projections that includes all possible projections for all corners, you make sure that at the corners of the reduced frequency space (0.5 , 0.5) you are still able to project the resonance on a projection without folding, because the dot product **p.φ** will never exceed 0.5.[5]

Because the exact place where (0.5 , 0.5) will project on will be dependent on the angle of projection (or slope), this effectively creates a clover-like form of reduced frequency space for the projections of points in the reduced frequency space (of the spectrum that is to be reconstructed) where points can project on without folding or aliasing.

The point (g1, g2) in (or out) the reduced frequency space was given by:

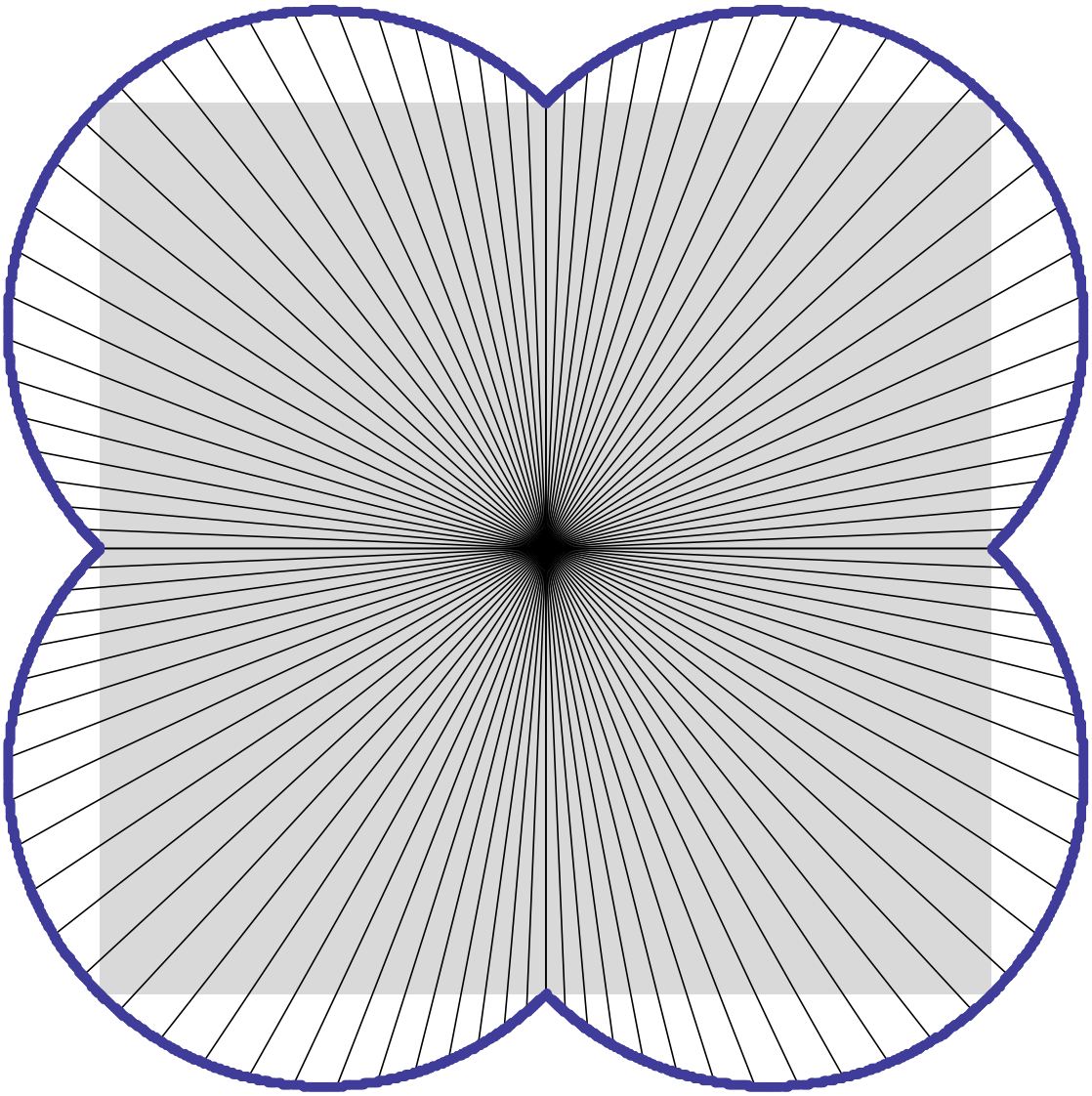
$$(g_1 , g_2) = (g_1, a\ g_1) = \left(\frac{\frac{\varphi_1}{a} + \varphi_2}{a + \frac{1}{a}} , \frac{\varphi_1 + a\ \varphi_2}{a + \frac{1}{a}} \right)$$

If (φ1, φ2) = (0.5 , 0.5), this reduces to:

$$(g'_1 , g'_2) = (g'_1, a\ g'_1) = \left(\frac{\frac{0.5}{a} + 0.5}{a + \frac{1}{a}} , \frac{0.5 + a\ 0.5}{a + \frac{1}{a}} \right)$$

To see the shape of the clover you can form it by making a function of (g1,g2) for all the corners of the reduced frequency spectrum for different projections (figure 3).

Figure 3



Of course, you can still have peaks that are outside the reduced frequency space of projections if the peaks are outside the reduced frequency space of the original spectrum, peaks that are also folded in the orthogonal projections. In this case they will also fold or alias in the projections of this peak. But if you just use the folded peaks from the projections for the back-projection they do not form (folded) peaks in the reduced frequency space.

If we would consider a point (f_1 , f_2) in reduced frequency space that now lies outside the reduced frequency space. It would fold in on the projection to:

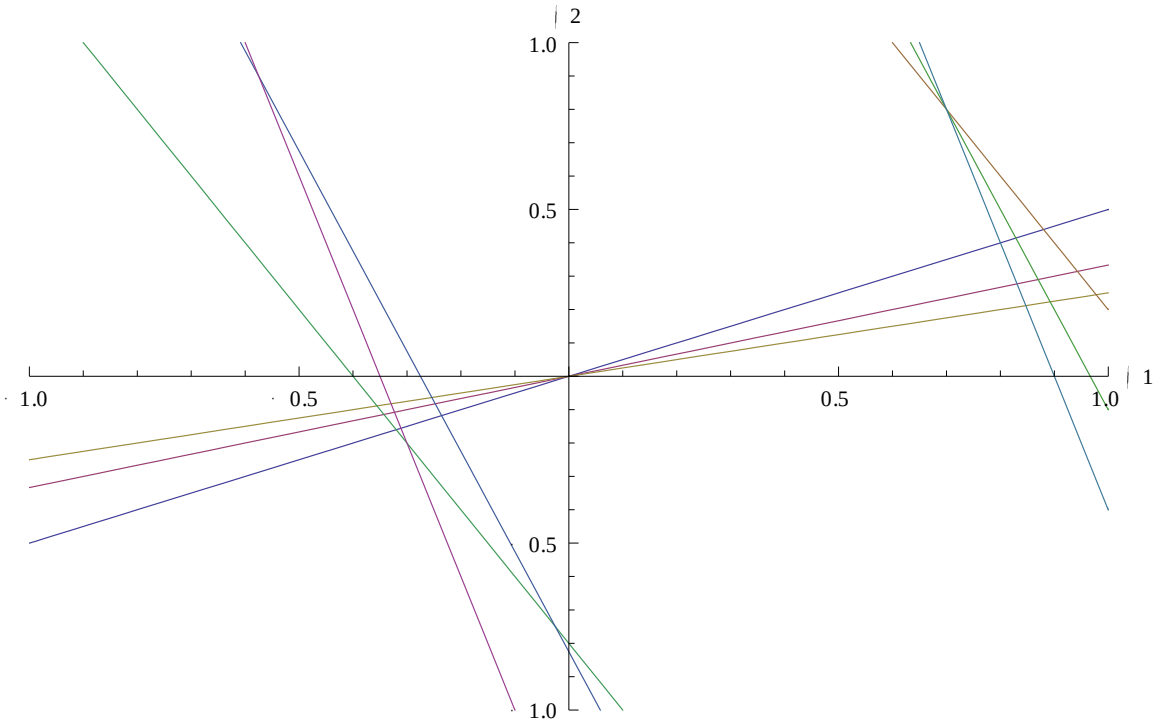
$$(g_1 - 2g_1', g_2 - 2g_2') = \left(\frac{(f_1 - 1) + a(f_2 - 1)}{1 + a^2}, \frac{a(f_1 - 1) + a^2(f_2 - 1)}{1 + a^2}\right)$$

If we do a back-projection from this folded peak, you would get a line orthogonal to the projection that can be described as:

$$\varphi_2(\varphi_1, a) = \left(a + \frac{1}{a}\right)\frac{(f_1 - 1) + a(f_2 - 1)}{1 + a^2} - \frac{\varphi_1}{a}$$

If we now have these back-projections for several slopes a, we don't get a new point where all those lines intersect (figure 4).

Figure 4



To account for this we make a bigger square around the reduced frequency area where the peaks that would have been folded in the projections will be on the projections as though the reduced frequency area, the clover, would have been bigger. Then, if those points of the projections are used for the back-projection they do form peaks. And when you start cleaning the raw back-projection spectrum up you account for this. So if you remove intensity from the artifacts around a peak outside the reduced frequency space (the clover) but inside the bigger square, you also remove intensity from the associated

artifacts from those folded in peaks inside the reduced frequency space. This ensures that no artifacts remain in the cleaned up spectrum.

Simple back-projection can also be seen as taking all the projections, and for each point on that projection extending ridges of intensity orthogonal to that line of projection. To do this, you first have to know where that point is in reduced frequency space. So if you have point S on the projection line:

$$s = \sqrt{g_1^2 + a^2 g_2^2}$$

Which leads to:

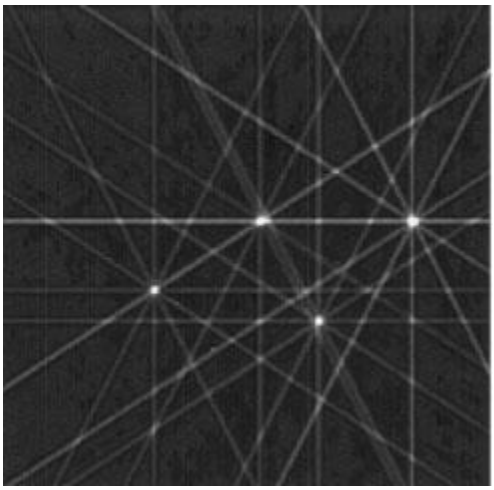
$$g_1 = \frac{s}{\sqrt{1 + a^2}}$$

Then, the line orthogonal to (g1, g2) = (g1, a g1) will simply be:

$$\varphi_2 = \frac{g_1 - \varphi_1}{a} + a g_1$$

Because each single projection has no information whatsoever on where the signal comes from on the axis orthogonal to that projection, the method of projection reconstruction has the inheritable effect of producing troughs of ‘signal’ orthogonal to the projections (figure 5 and front image). Only at places where all projections cross real peaks are found, the other ‘peaks’ are mere artifacts. So an algorithm is needed to reliably locate the real peaks, once the simple back-projections is made.

Figure 5



A way to do this is to let the algorithm start with the highest peak and then subtract 1% (or a similar small percentage) of that peaks intensity and of the associated artifacts and stores the point of maximum intensity in a new artifact-free spectrum. It continues to do so till another peak has the highest intensity, and another, and so on till all the intensities above a certain threshold value have disappeared from the original spectrum. The 'new' spectrum is then rid of all the artifacts and the 'old' spectrum contains noise only.

Intrinsically unfolded proteins

Now that we have investigated the technique of projection-reconstruction it is time we spend some time familiarizing with the proteins whose 'structure' we want to analyze. Namely intrinsically unfolded (unstructured) proteins and specifically the tau protein.

Intrinsically unfolded proteins often have a role as signal peptide, in cell-cycle regulation, gene-expression or function as a chaperone protein. Because they are so flexible in structure, it is easier for them to bind to a set of different targets, making them much more suitable for the roles stated above. The increased flexibility also enables an increased speed of interaction. These ways of 'flexible-interaction' are sometimes referred to as 'fly-casting' or 'protein-fishing'.[6]

Tau proteins physiological role seems to be stabilization of microtubule structures in neuronal cells. They are better known however, for their presence in the neurofibrillary tangles in the brain cells of Alzheimer patients. For our purposes, the main property of interest of tau is its intrinsically unfolded nature and its size (441 residues), making it challenging to obtain an unambiguous 'structure' because of severe peak overlapping in spectra and simply that the protein has very little structure by definition.[7]

A first attempt was made by Smet e.a. (2004) who figured they could use the degeneracy of the chemical shift to their advantage. Since every residue has about the same chemical shift for ^{13}C-alpha and ^{13}C-beta across the protein, but differ enough to be able to reconstruct what residue we are dealing with, you can make 'product planes' of certain residues. For instance, you can make a product plane of lysine residues by taking all the intensities in the plane of 56.2 ppm for C-alpha and 33.1 ppm for C-beta. This simplifies the spectrum considerably, highlighting only the lysine residues in the spectrum. Only difficulty is that there are always exceptions and peaks appear that do not belong to lysine but to a serine shifted upfield by a preceding proline. Also, the chemical shift differences between different residues aren't very big, as there seems to be some overlap between the methionine and lysine product plane. In the end, they resorted to doing their experiments on small peptides instead of the whole protein, reasoning that the spectrum would look the same because of the intrinsically unfolded properties. Indeed, many of the peptide spectra had good overlap with the whole protein spectrum, allowing a reasonable assignment of the different peaks, though not yet complete.[8]

A more recent and more succesfull attempt was made by Mukrasch e.a. (2009) who were able to assign 98% of the non-proline backbone resonances using HSQC-spectra, 3D (Ha)CaNNH and HNN experiments on a 21.14 T machine. To confirm their assignment they performed the same experiments on three overlapping fragments of the full tau protein. Subsequently they subtracted the random coil values from the chemical shifts and used these secondary chemical shifts for an analysis of any residual secondary structure. Doing this, they found a propensity for beta-sheet structure for some parts of the R1, R2, R3 and R4 region. Also some propensity for alpha-helixes were found in region between I2 and P2 and near the C-terminus. In addition a series of sign-inversed residual couplings were measured for the C-terminus, confirming the propensity for alpha-helix formation. The high residual coupling for a part of the R3 region, which is usually observed for a more extended structure, also confirms the propensity for beta-sheet formation in that part of the protein. In the N-terminus, a lot of single negative RDC's were found for residues, suggesting that in this part (residue 1-200), the protein has a lot of turns. [9]

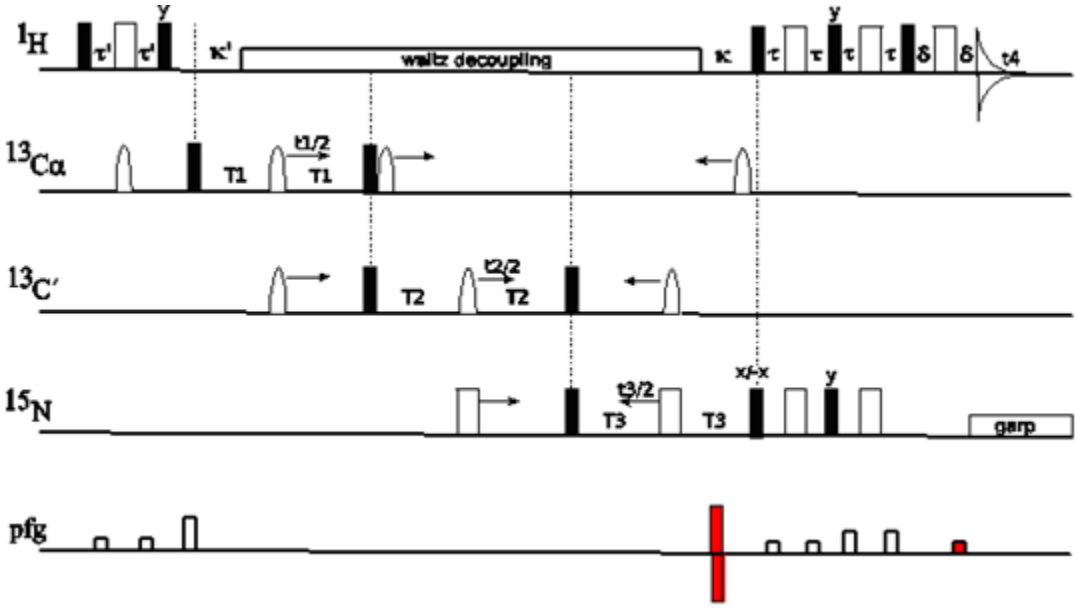
Pulse sequences

To investigate the structure of tau protein we first have to do a backbone assignment. What follows next is a description of the pulse-sequences that should give us the resonances that allow us to do a 'sequential-walk' along the backbone of the protein.

The first described experiment (figure 6) involves the interresidual (Ha)$Ca_{i-1}CO_{i-1}$ N_iH_i experiment. The sequence starts with the INEPT-transfer from Ha to Ca, giving HazCax magnetization. Subsequently, the antiphase magnetization is allowed to become in-phase at which point a waltz decoupling is applied and at the same time, the Ca transverse magnetization couples with CO for the constant time element T1 and the chemical shift evolves for t1, giving CaxCOz. This antiphase in Cax becomes antiphase in COy after a simultaneous 90 degree pulse to Ca and CO, thus creating CazCOy magnetization. During the second constant time element T2, the CaCO coupling brings the CazCOy in phase (COx) and the CON coupling brings this coherence in antiphase, giving COyNz with a chemical shift evolution of t2. Then, the magnetization is transferred to the N with a simultaneous 90 degree pulse to CO and N, giving CozNy. In the third constant time element T3, this antiphase is put back in phase and modulated by a chemical shift evolution of t3. Also, the waltz decoupling ends in this period so Nx can be influenced by the NH coupling, giving the NyHz coherence modulated by cos(ΩN t3). A simultaneous 90 degree pulse on H and N then puts the magnetization back on H, which can then become in phase after an echo sequence, giving Hx modulated by cos(ΩN t3). The rest of the pulse sequence is sensitivity-enhancement. The Hx is parked along the z-axis using a 90 degree pulse, which stays Hz during the echo sequence. Then the H magnetization is put back in the xy-plane, giving Hy back for acquisition. The 'other' sequence in the enhancement makes the cos(ΩN t3) Hy signal change sign. The fate of the sin(ΩN t3)

modulated HyNx signal after T3 is a change of sign because of the 180 degrees pulse on Hx during the first echo. Then the two 90 degree pulses over the y-axis produce HyNz, which evolves into Hx during the next echo, producing the sin(ΩN t3) Hx. So exp. 1 gives cos(ΩN t3) Hy + sin(ΩN t3) Hx and exp. 2 gives - cos(ΩN t3) Hy + sin(ΩN t3) Hx. So exp. (1+2) gives 2 sin (ΩN t3) Hx and exp.(1-2) gives 2 cos(ΩN t3) Hy. Of course, all these signals are also modulated by cos(ΩCa t1) cos (ΩCO t2). To get the sin(ΩCa t1) modulated signals, we have to give the 90 degree pulse for the Ca after the T1 constant period along the y-axis instead of the x-axis. Similarly, if we want the sin(ΩCO t2) modulated signal we have to give the 90 degree pulse for the CO after the T2 constant time period along the y-axis instead of the x-axis. In this way we can get the sine and cosine modulated signal for every indirect time domain.

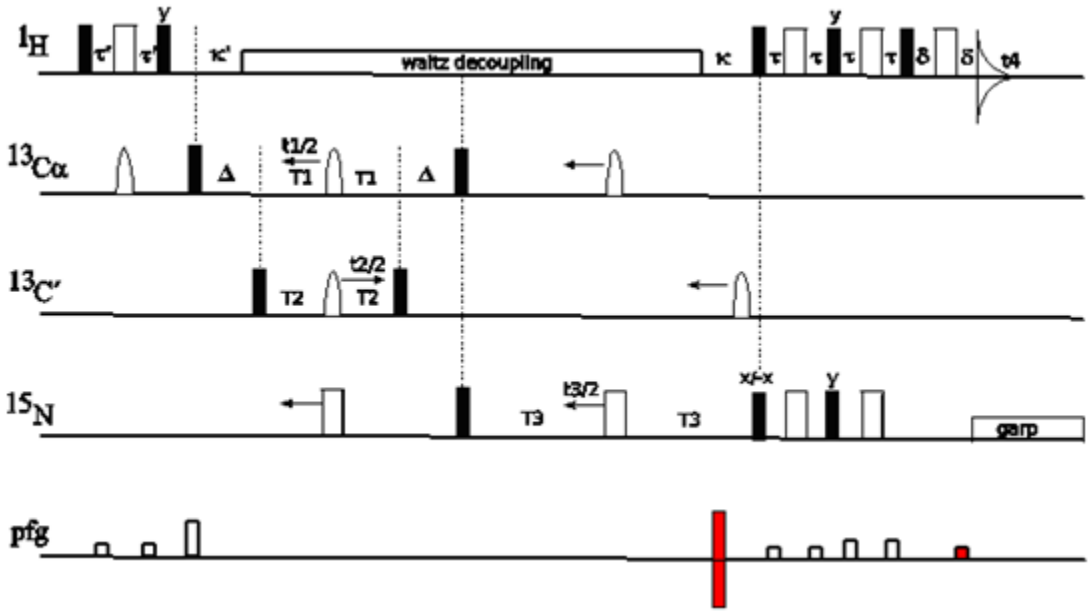
Figure 6



The second experiment described (figure 7) involves the intraresidual $HaCa_iCO_{i-1}N_iH_i$ experiment. Just as the experiment described above, it starts with an INEPT-transfer giving HazCax magnetization, which is also allowed to become in phase after which waltz decoupling is started. From then on, things start to get complicated. This is because two constant time domains, T1 and T2 evolve at the same time in the sequence. During period Delta (=1/2 JCaCO), the transverse Ca magnetization is brought in antiphase with CO (CayCOz). A 90 degree pulse to CO then creates a multiple quantum coherence (CayCOy) that is modulated by the CO-evolution inside the T2 domain for t2. At the end of T2 the multiple quantum-coherence is brought back in antiphase magnetization by a

CO 90 degrees pulse to CayCOz. Then for another period Delta the antiphase magnetization evolves back into Cax. However, during the total of the period of the Delta's and the T2 domain there is also a coupling between Ca and N, if we make this period 1/2JNCa, the coherence becomes CayNz. By sliding the 180 degree pulse on Ca to the left together with the 180 degree pulse on the N, the coherence also gets modulated in t1 by the Ca evolution.

Figure 7



There are some more remarks to be made of this part of the pulse sequence.
Since the full CayNz coherence is only acquired if we set t1=t2=0, what would happen if we let Ca and CO evolution take place? To find a solution we should also know for what time each coupling is active. The CaN coupling runs for the entire period of 2 x Delta + 2 T1. The CaCO coupling runs for 2x Delta + 2 T1 minus period b (T1 - T2 + t1/2 + t2/2), which amounts to : 2 x Delta + (T1 +T2) – (t1/2+t2/2) . The CON coupling runs for a – b + c which amounts to 2 T2 – t1 – t2. Because all these couplings commute and if we follow the CayCOy coherence to the point where it should be CayCOz we find the following. It turns out that we only have to take into account the CaN-coupling. This is because the CayCOy coherence is insensitive to CaCO-coupling, so it only couples during the Delta period to bring antiphase back in phase, giving the following observable signal:

$$\sin(\pi\, JCaN\, t)\ \cos\ (\Omega CO\ t_2)\ \cos\ (\Omega Ca\ t_1)\ \ CayNz2$$

Where t = 2 Delta + 2T1, so the total time the JCaN coupling is active. In the above reasoning we have assumed there is only one N that we have to take into account. In reality things are more complicated since we have to deal with two N couplings, one for the N on the same residue as the Ca (N_i = N1) and one for the N on the residue after that Ca (N_{i+1} = N2). Because these couplings commute we get the following observable coherences.

$$\sin(\pi\, JCaN1\; t)\cos(\pi\, JCaN2\; t)\;\cos\;(\Omega CO\; t_2)\;\cos\;(\Omega Ca\; t_1)\;\; CayN1z2$$

$$\sin(\pi\, JCaN2\; t)\cos(\pi\, JCaN1\; t)\;\cos\;(\Omega CO\; t_2)\;\cos\;(\Omega Ca\; t_1)\;\; CayN2z2$$

What we can surmise from the above expressions, is that we get optimal signal for N1 if we find the maximum of the functions:

$$\sin(\pi\, JCaN1\; t)\cos(\pi\, JCaN2\; t)$$

$$\sin(\pi\, JCaN2\; t)\cos(\pi\, JCaN1\; t)$$

If we look at the figure of the plot of these functions, with JCaN2 = 7 Hz and JCaN1 = 12 Hz, we see a maximum around 30 ms (figure 8). And if we take into account relaxation an even shorter period, 25 ms, should be optimal to get a nice signal for N1 (figure 9), but we have to consider the amount of time available for the T1-period if we choose a value. Since 25 ms = 2 Delta + 2 T1 and Delta (1/2 JCaCO) is 9.1 ms, this leaves only 7 ms for 2 T1. So to have enough time for t1 and t2 evolution a higher value should be taken, but this would result in signal loss. To (at least partly) get around this, we can make the constant-time period ‘semi’-constant instead. Meaning that for the situation when the evolution time is 7ms or shorter, we use a constant-time period of the optimal value of 25 ms. And if we have the situation where the evolution time is larger than 7ms, we simply increase the constant time bit by bit to give room for the evolution time, and at the same time getting as little signal loss as possible because of relaxation.

Looking at the figures, it is obvious that it’s much harder to get a good signal for the N2. Though for the sequential walk we are more interested in the N1 coupling, it could still be convenient to have both correlations.

Figure 8 (blue is N1 signal, red is N2 signal, relaxation not taken into account)

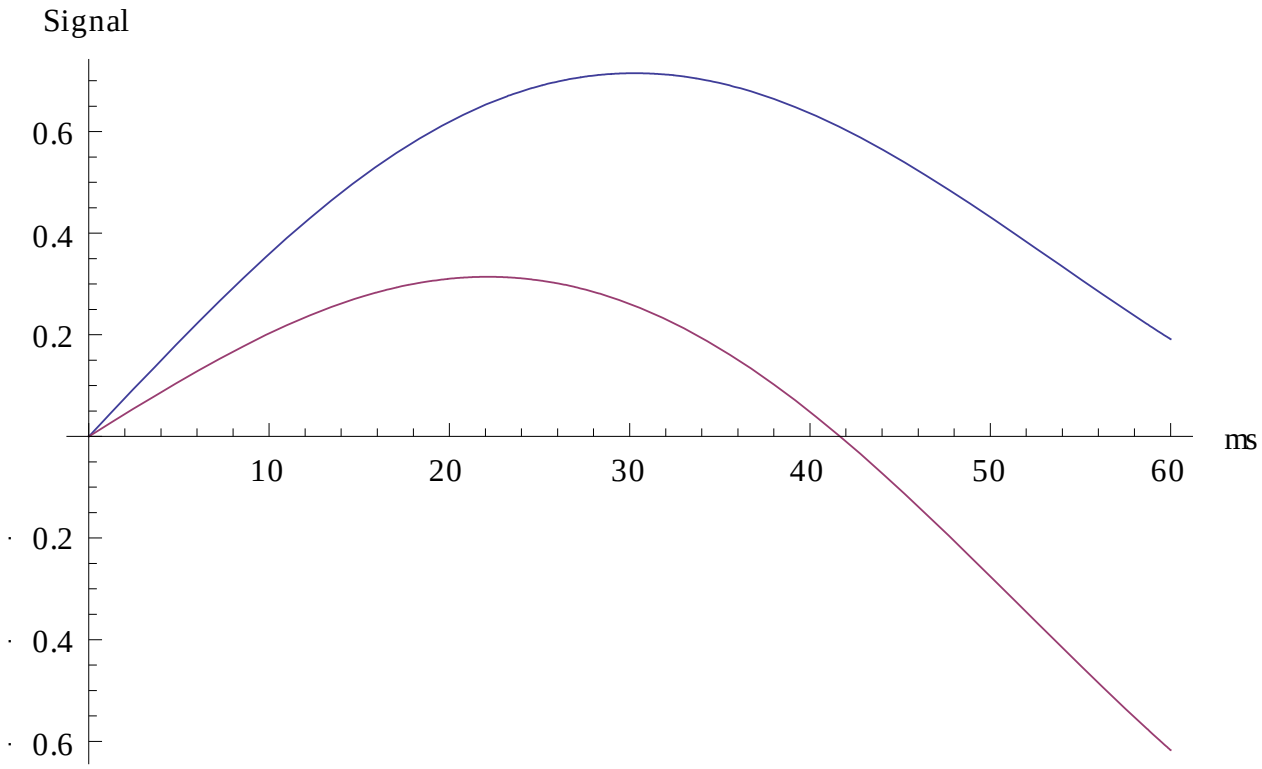
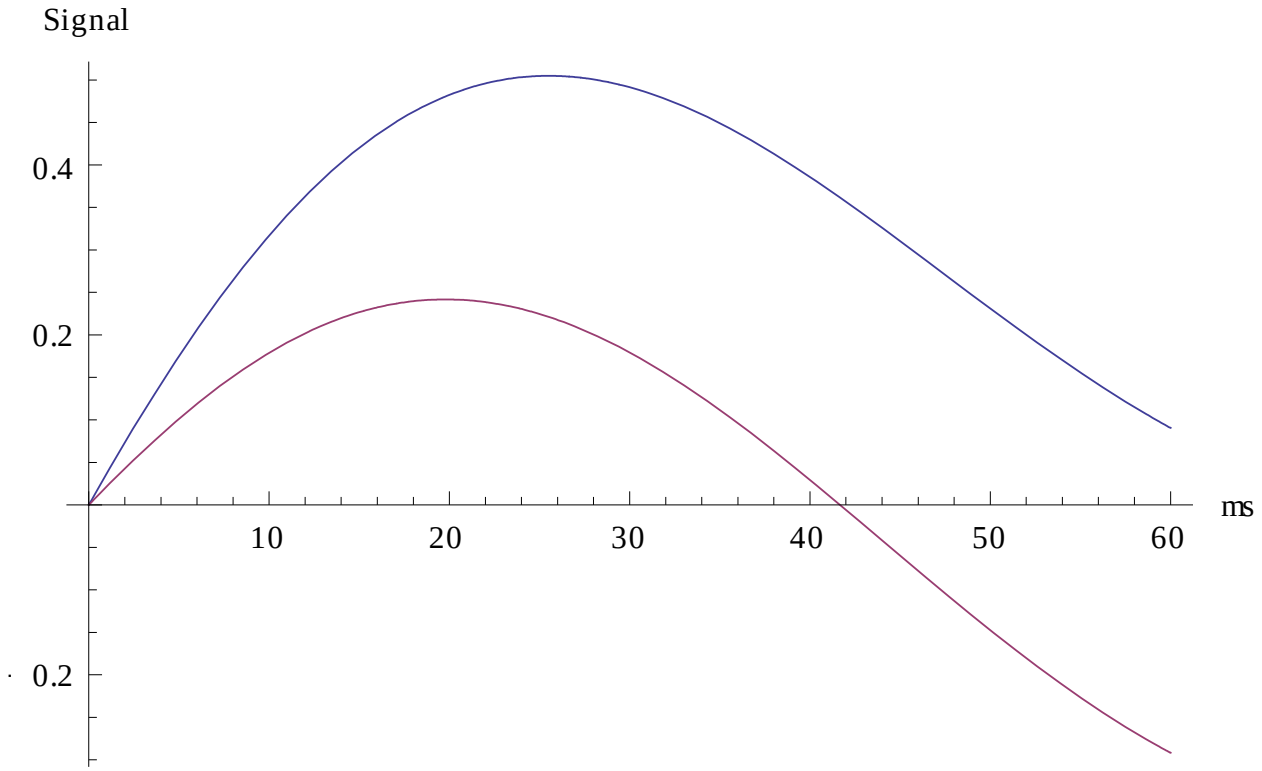


Figure 9(blue is N1 signal, red is N2 signal ,now with relaxation using a $t_{1/2}$ of 80ms)



Once the CayNz coherence is created, the sequence becomes rather straightforward. Two simultaneous 90 degree pulses make it CazNy antiphase which then evolves for t3 in the T3 constant time domain. The T3 time so chosen that it becomes in phase to Nx and at the same time, when the waltz decoupling stops, evolves into NyHz antiphase. The subsequent sensitivity enhancement pulse sequence is the same as for the first experiment.

In practice however, the signal/noise ratio of this pulse sequence is disappointing. Because we can't do a backbone assignment with just the first experiment, other pulse sequences needed to be explored for its use in this particular situation.

The third experiment (figure 10) or the second experiment we tried for the second experiment is a $HNCa_iCO_{i-1}$ sequence.[10-11] First the magnetization is transferred to the nitrogen by an inept-transfer, creating –HzNy at point a in the sequence. Then this coherence is coupled for 2 tau b by JNCO and JNCa. The antiphase thus created is first modulated by chemical shift evolution in period t1 creating:

$$\cos(\Omega\, CO\ t1)\ \ CO_y N_z 2$$

$$-\sin(\Omega\, CO\ t1)\ \ CO_x N_z 2$$

These antiphases are further modulated in period t2 by the chemical shift evolution of Ca, by first applying a 90 degrees pulse over the x-axis to bring the Ca spins in the xy-plane. It should be noted that we do not bring the CO magnetization back to the z-axis, the 90 degrees pulse to CO is applied over the x-axis:

$$-\cos(\Omega\, Ca\ t2)\cos(\Omega\, CO\ t1)\ \ Ca_y CO_y N_z 4$$

$$\sin(\Omega\, Ca\ t2)\cos(\Omega\, CO\ t1)\ \ Ca_x CO_y N_z 4$$

Of course, the JNCO coupling also has an effect during this time creating:

$$\sin(\pi\, JCON\ t2)\cos(\Omega\, Ca\ t2)\cos(\Omega\, CO\ t1)\ \ Ca_y CO_x 2$$

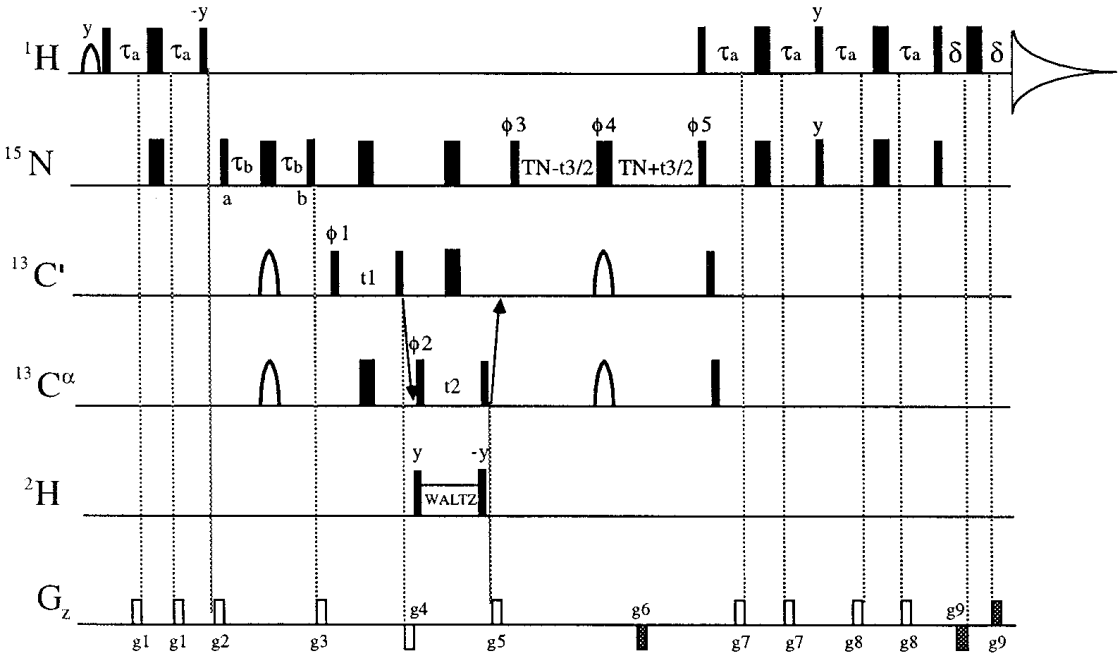
$$\sin(\pi\, JCON\ t2)\sin(\Omega\, Ca\ t2)\cos(\Omega\, CO\ t1)\ \ Ca_x CO_x 2$$

And similar expressions for the sine modulated signal in t1 with CayCOy and CaxCoy magnetization.

Then in period t3, all these antiphase and multiple quantum coherences are coupled for 2 tau b but then in opposite direction. By looking at the pulse sequence we can see this is achieved by the 90 degrees pulse on N over the x-axis just after b which effectively forms a 180 degrees pulse with the 90 degrees pulse on N over the x-axis (φ_3) at the start of T3. So it's brought back in phase with the nitrogen, and at the same time the coupling with protons brings it in antiphase with the protons. The rest of the sequence is the same as for

the other pulse sequences (i.e. the sensitivity enhancement). This sequence though also gave very poor signal.

Figure 10



Some spectra

Some spectra are attached, they give a nice idea about the quality of spectra using projection reconstruction and also illustrate the difference in spectra between folded and unfolded proteins.

The first two spectra are from the (Ha)$Ca_{i-1}CO_{i-1}$ N_iH_i –experiment performed on calbindin. We did not record the Ca-evolution though, so it was a 3D experiment. That being said, with projection-reconstruction we could do it overnight instead of taking an entire weekend and as you can see with plenty of signal.

The other two spectra are from the same experiment, but then slightly adjusted so we could obtain the chemical shift of the Ha, performed on α-synuclein, an intrinsically unfolded protein. The ‘unfoldedness’ is clearly reflected in the small range of chemical shifts and the overlap of peaks in the Ha-domain.

Enhanced relaxation and 'cool' pulse-sequences

Though the method of projection reconstruction saves us quite some time, we can further increase the speed of the process by shortening the delay between pulse sequences. This delay is necessary to prevent the 'saturation' of the sample, that is, the magnetization of the sample. When a delay between pulse sequences would be omitted, the sample has no time to get into a new equilibrium with the magnetic field of the NMR-apparatus, thus no net z-magnetization is formed and no pulse sequence will ever create a signal. It is possible though to shorten the time the sample needs to get in equilibrium with the external magnetic field. This is done by adding small amounts of a paramagnetic substance to the sample, for instance a salt containing Mn^{2+} or Gd^{3+}.[12]

A side-effect of decreasing the delay between pulse-sequences is that on the whole, the sample gets more radiation per time unit. This can result in overheating of the sample. To prevent this from happening, we have to make the pulse-sequences themselves a bit 'cooler' by minimizing the amount of irradiance. In practice, this means substituting the decoupling sequences like 'waltz' by a series of well placed 180-pulses, achieving the same goal (decoupling) with less radiation.

If we consider the first experiment, the interresidual (Ha)$Ca_{i-1}CO_{i-1}$ N_iH_i as shown in figure 6, the first constant time period where waltz decoupling is applied is during T1. To achieve the same coupling without using waltz, we have to insert a 180 pulse on H in such a way that the total amount of coupling during T1 is exactly $1/(2\ JHCa) = 2\ \tau_1$, because we want the HzCay and HzCax antiphases to evolve into transverse Ca magnetization. At the same time the coupling of Ca with CO should also be $1/(2\ JCaCO) = 2\ \tau_2$, because we want this transverse Ca magnetization to become antiphase with CO.

This is achieved by changing the pulse sequence from:

p90[Cax]//delay[τ_2 + t1/2]//p180[Cax + Cox]//delay[τ_2 – t1/2]

To:

p90[Cax]//delay[τ_1 + t1/2]//p180[Hx]//delay[τ_2- τ_1]//p180[Cax + Cox]//delay[τ_2-t1/2]

We can see this is correct because (π JHCa t) is now equal to (τ_1 + t1/2) – (τ_2- τ_1)+(τ_2-t1/2) = 2 τ_1. And also (π JCaCO t) is equal to (τ_1 + t1/2) + (τ_2 – τ_1) + (τ_2 –t1/2) = 2 τ_2.

A similar trick can be done with the third constant time period T3, where we need a total amount of coupling for JNH of 1/(2 JNH) = 2 tau4 and a total amount of coupling of JCON of 1/(2 JCON) while the coupling between Ca with N and CO remains zero.

Starting with:

p90[Cox + Nx]//delay[τ_3-t3/2]//p180[Cox + Nx]//delay[τ_3]//p180[Cax]//delay[t3/2]

We can make this cool by changing it to:

p90[Cox + Nx]//delay[τ_3-t3/2]// p180[Cox + Nx]//

delay[τ_3- τ_4]//p180[Hx]//delay[τ_4]//p180[Cax]//delay[t3/2]

We can see that this is correct because:

(π JNH t) = (τ_3-t3/2)-(τ_3-τ_4)+(τ_4)+(t3/2) = 2 τ_4

(π JCoN t) = (τ_3-t3/2)+(τ_3-τ_4)+(τ_4)+(t3/2) = 2 τ_3

(π JCaN t) = (τ_3-t3/2)-(τ_3-τ_4)-(τ_4)+(t3/2) = 0

(π JCoN t) = (τ_3-t3/2)-(τ_3-τ_4)-(τ_4)+(t3/2) = 0

Future Plans

Projection reconstruction is a technique that allows us to record multidimensional NMR-spectra in less time than when recorded in the classical way. Being able to do multidimensional NMR-experiments in less time is especially important if we want to investigate unfolded proteins. Because of the extensive peak overlap expected in unfolded proteins the need for 4D or even 5D-experiments arises, which would be either highly impractical or simply unobtainable if done in the classical way. By recording a 3D experiment (we used the (Ha)$Ca_{i-1}CO_{i-1}$ N_iH_i – sequence, experiment 1, but then without evolution in the Ca-domain) with projection reconstruction for a sample of calbindin, we have shown that using this technique saves time and does not involve quality loss. To show that this technique can also be used to record the spectra for an unfolded protein and do a backbone-assignment, we want to investigate tau-protein, the biggest unfolded protein on which this has been done to date. One of the necessary experiments for being able to do so failed to give sufficient signal however. So this is something we have to look into for the future.

Reference List:

[1] PDB databank, annual report 2010

[2] Coggins, Venters and Zhou: Generalized Reconstruction of n-D NMR Spectra from Multiple Projections: Application to the 5-D HACACONH Spectrum of Protein G B1 Domain. *J Am Chem Soc* 126:1000-1001 (2004)

[3] Kupce & Freeman: Reconstruction of the three-dimensional NMR spectrum of a protein from a set of plane projections. *Journal of Biomolecular NMR* 27:383-387 (2003)

[4] Freeman, Kupce: Distant echoes of the Accordion: Reduced Dimensionality, GFT-NMR, and Projection-Reconstruction of Multidimensional Spectra. *Concepts Magn Reson Part A* 23A:63-75 (2004)

[5] Scheek, vd Meulen, Mulder et al.: Projection Reconstruction applied to a 4d HaCACONH NMR experiment on the intrinsically unfolded protein alfa-synuclein (unpublished)

[6] Tompa: The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters* 579:3346-3354 (2005)

[7] Skrabana, Sevcik and Novak: Intrinsically Disordered Proteins in the Neurodegenerative Processes: Formation of Tau Protein Paired Helical. *Cellular and Molecular Neurobiology* 26:1085-1097 (2006)
Filaments and Their Analysis

[8] Smet, Leroy, Lippens et al.: Accepting its Random Coil Nature allows a Partial NMR Assignment of the Neuronal Tau Protein. *ChemBioChem* 5:1639-1646 (2004)

[9] Mukrasch, Bibow, Zweckstetter et al.: Structural polymorphism of 441-residue tau at single residue resolution. *PLOS Biology* 7(2):399-414 (2009)

[10] Konrat, Yang and Kay: A 4D TROSY-based pulse scheme for correlating $^{1}HN_{i}$,$^{15}N_{i}$,$^{13}Ca_{i}$, $^{13}C0_{i-1}$ chemical shifts in high molecular weight, ^{15}N,^{13}C, ^{2}H labeled proteins. *Journal of Biomolecular NMR* 15:309-313 (1999)

[11] Yang & Kay: 4D Trosy Triple-resonance Four-dimensional NMR Spectroscopy of a 46 ns Tumbling Protein. *J Am Chem Soc* 121:2571-2575 (1999)

[12] Chilton, Jackels, Hinson and Ekstrand: Use of a Paramagnetic Substance, a Colloidal Manganese Sulfide, as an NMR Contrast Material in Rats. *J Nucl Med* 25:604-607 (1984)