

Protein Assignment

Dr. Albert A. Smith-Penzel

5 May, 2020

A protein consists of many amino acids (~100 or more), connected by peptide bonds. Each nucleus of a given type (^1H , ^{13}C , ^{15}N) will then yield a unique signal in a one-dimensional spectrum. But, this may yield hundreds of signals (ex. Figure 1). We have no way of obtaining site-specific structural information from a spectrum, then, without knowing where in the protein each signal comes from. Therefore, we must solve two problems:

1. Obtain *resolution*, that is, spread out the signals enough that they are sufficiently separated to distinguish them.
2. Obtain the *assignment* of the signals, in order to know where in the protein a signal comes from.

We achieve this by using multi-dimensional spectra. The dense 1D spectra that are obtained for proteins can be spread across two or more dimensions, yielding greater resolution. Second, we can determine what resonances are in close proximity to each other. This information, along with the chemical shifts of those resonances, can be used to assign the resonances in an NMR spectrum. We will discuss this approach today, starting with a review of the amino acids and peptide bonds which connect amino acids together in a protein.

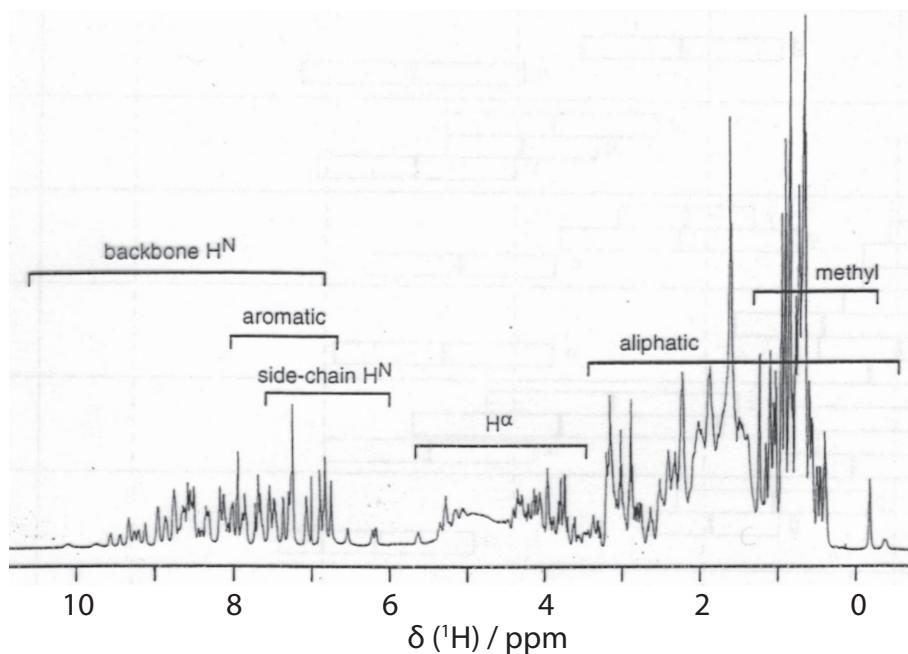


Figure 1: Ubiquitin 1D ^1H spectrum.

1 Amino acids and Peptide bonds

When studying proteins, we generally discuss the protein's primary, secondary, tertiary, and quaternary structure.

1. *Primary*: This is the sequence of amino acids that appears in a protein, connected by peptide bonds
2. *Secondary*: This is the structure of local segments of the protein, determined by the pattern of hydrogen bonds; the most common elements are α -helices and β -sheets, but one also commonly has unstructured regions (tending to be dynamic, without a well-defined structure).
3. *Tertiary*: This is the overall structure of a single protein molecule.
4. *Quaternary*: If a protein molecule is part of a larger complex, the quaternary structure defines how different components of such a complex fit together.

To obtain the protein resonance assignment, we will focus on the connectivity between amino acids, that is, the primary structure. We will discuss secondary, tertiary, and quaternary structures in more detail next week. Then, the first question is what are the amino acids making up proteins, and how are they connected to each other.

1.1 The Amino Acids

20 amino acids are coded for by DNA (although other mechanisms allow for additional amino acids to appear in proteins). These 20 amino acids (+selenocysteine) are shown in Figure 2. The sequence of a protein determines its structure to large extent (although in some cases, a protein may take on different structures, depending on conditions, ex. fibrils).

- We are interested in how we may use the chemical shifts of nuclei in different amino acids to identify which amino acid a given shift corresponds to.
- We will primarily use differences in ^{13}C chemical shift in order to identify amino acids, although ^1H shifts and to lesser extend ^{15}N shifts may also be used. ^1H and ^{13}C shift ranges as a function of amino acid and position in the amino acid are shown in Figure 3.
- A single chemical shift is not sufficient to identify an amino acid, given the significant overlap in chemical shift ranges (Figure 3). Amino acids can be identified based on several shifts (ex. $\text{C}\alpha$, $\text{C}\beta$, etc. shifts).
- We use J-couplings, that is, through-bond couplings, to connect multiple shifts together, discussed in detail in the next section.
- Identifying the type of amino acid isn't enough to assign resonances, since many amino acids appear multiple times in a given protein. We must also determine where observed resonances fall in the sequence of amino acids of a protein.

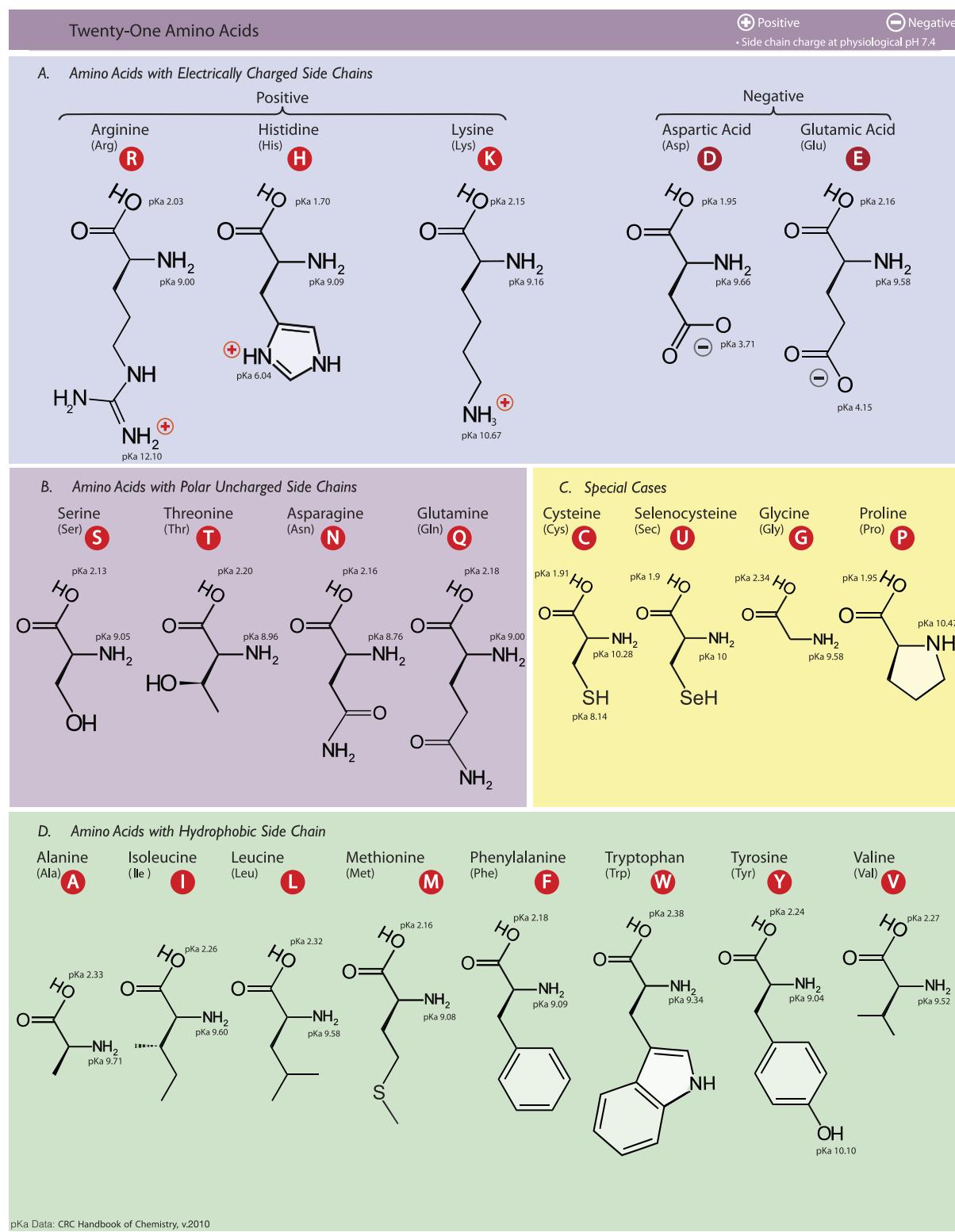


Figure 2: The 21 most common amino acids

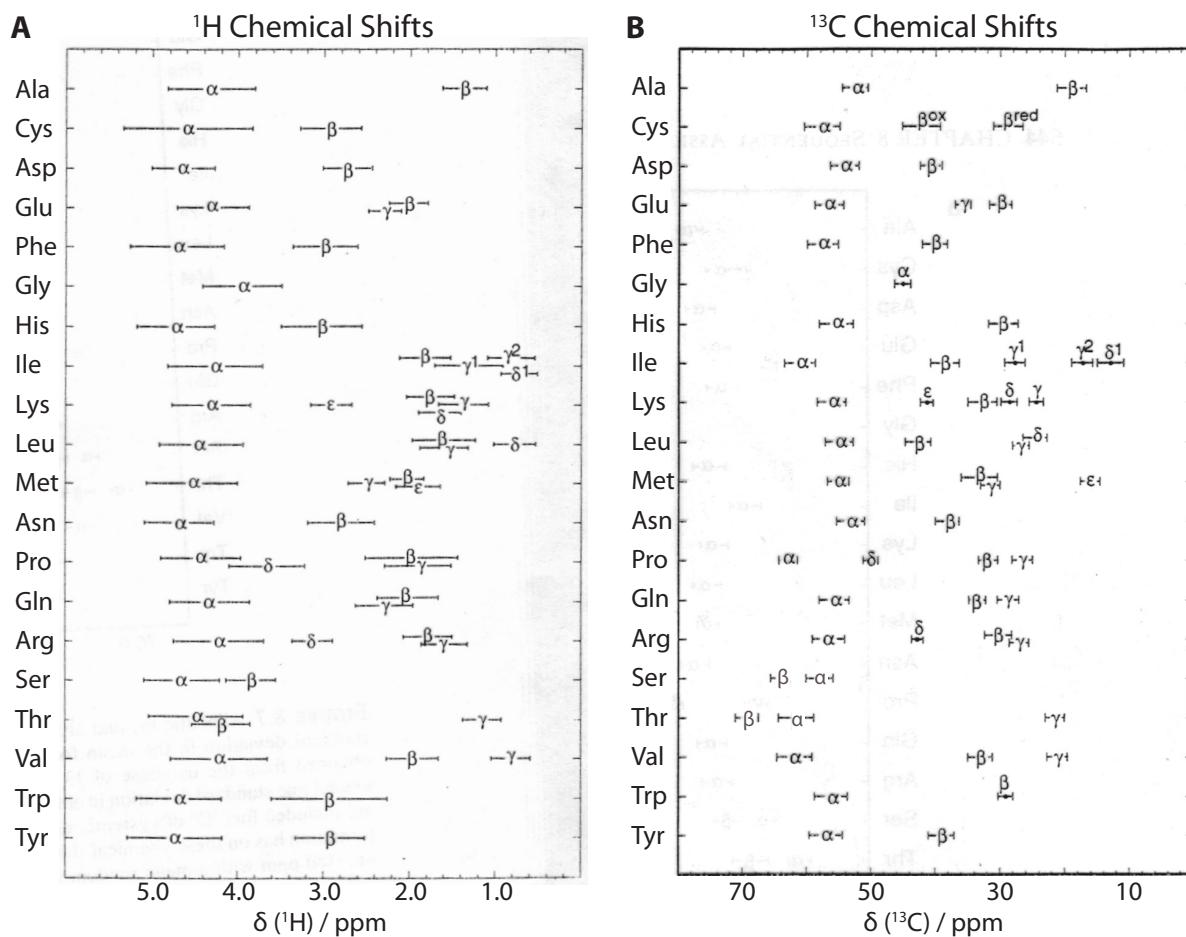


Figure 3: ^1H (A) and ^{13}C Chemical shift ranges for each amino acids at the various positions in the side chain. Bars indicate one standard deviation of the chemical shift.

1.2 Peptide bonds

That brings us to connectivity between amino acids. Amino acids are joined by a peptide bond, whose formation is illustrated in Figure 4A. The result is shown in Figure 4B. We refer to the chain of atoms (N-C α -CO-N-C α -CO-...) as the protein backbone.

- We can connect chemical shifts of nuclei via through-bond interactions (J -couplings). For example, we may determine for a given $^{13}\text{C}\alpha$ chemical shift, chemical shifts of the neighboring ^{15}N nuclei, and the chemical shifts of the bonded ^1H .
- We may also connect those ^{15}N and ^1H resonances to the next $^{13}\text{C}\alpha$, and work our way through the protein backbone. Given enough $^{13}\text{C}\alpha$ resonances in a series, we may determine, based on chemical shift ranges in Figure 3B, where the sequence of chemical shifts fits in the protein. One may also obtain additional chemical shifts (ex. $^{13}\text{C}\beta$) in case ^1H , ^{15}N , and $^{13}\text{C}\alpha$ shifts are insufficient to perform an assignment. We'll detail this in the following sections, first investigating how to connect chemical shifts via multi-dimensional spectroscopy.

A

Aminosäuren sind gewöhnlich über eine Amidbindung miteinander verknüpft, die man Peptidbindung nennt.

Peptidbindung: Die vier Atome in jedem grauen Kasten bilden eine steife, ebene Einheit. Um die C-N-Bindung gibt es keine freie Drehbarkeit.

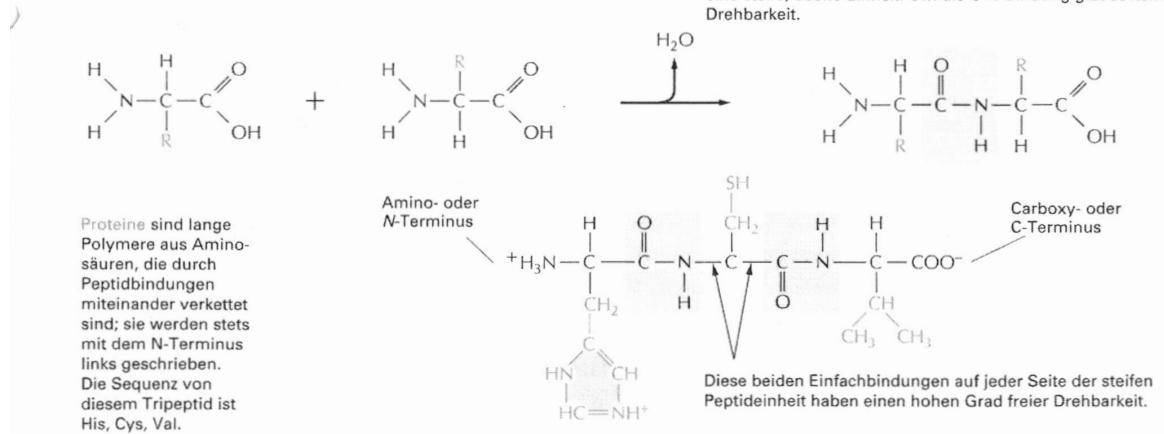
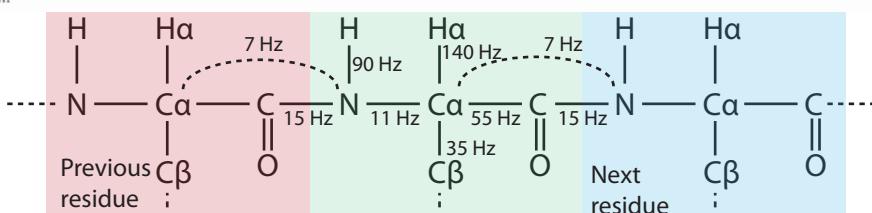
**B**

Figure 4: Peptide bonds. **A** Peptide bond formation. **B** Backbone structure after formation and J-couplings.

Table 1: J-couplings in Proteins

Spin Coupling	Typical J value	$1/2J$	$1/4J$
$^1J(C,H)$	140 Hz	3.57 ms	1.79 ms
$^1J(N,H)$	90 Hz	5.56 ms	2.78 ms
$^1J(C,CO)$	55 Hz	9.10 ms	4.55 ms
$^1J(C_\alpha, C_\beta)$	35 Hz	14.3 Hz	7.14 ms
$^1J(N,CO)$	15 Hz	33.3 ms	16.7 ms
$^1J(N,C_\alpha)$	11 Hz	45.5 ms	22.7 ms
$^2J(N,C_\alpha)$	7 Hz	71.4 Hz	35.7 ms

2 Correlation Spectroscopy

So, how do we correlate the chemical shifts of different resonances in NMR? We start reviewing the basic one-dimensional NMR experiment, which requires excitation of spins of a given type (sometimes called preparation) followed by observation of the evolution of the magnetization of that spin. This is illustrated in Figure 5, and in `wk2_videos.ppt`, slides 1-2.

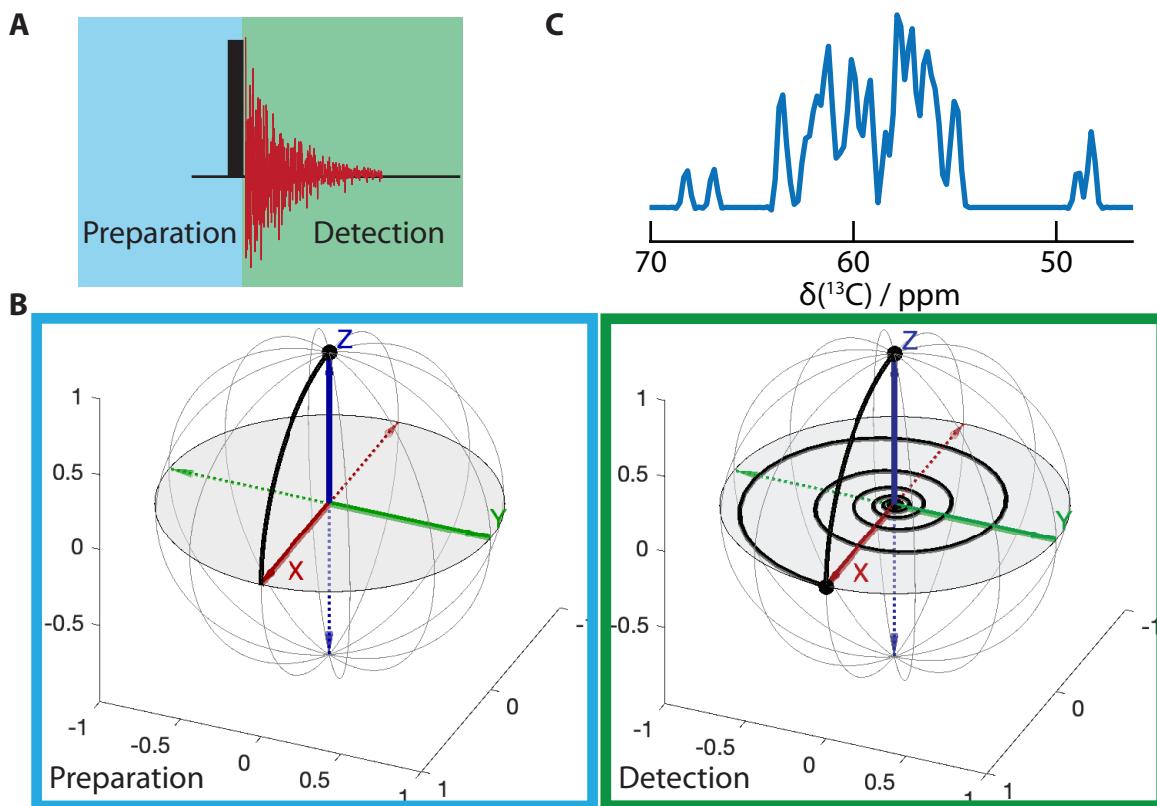


Figure 5: Basics of a one-D experiment. **A** The preparation period (here, a simple $\pi/2$ -pulse, that is, a B_1 field generated with the NMR coil, on-resonance with the spins) generates magnetization along x (in the rotating frame), followed by signal detection (detected with the NMR coil). **B** The trajectory of the magnetization during the pulse (under the applied B_1 field) and after the pulse (detected signal, evolving under the chemical shift in the xy-plane). **C** A processed signal, obtained by Fourier-transform of the detected magnetization. Also see `wk2_videos.ppt`, slides 1-2.

2.1 Basic Principles of Multi-dimensional spectroscopy

To correlate chemical shifts, that is, to determine what chemical shifts correspond to neighboring nuclei, we have to transfer magnetization between those nuclei. We also have to allow the magnetization to oscillate under each chemical shift. For a multi-dimensional experiment, then a pulse sequence begins with a preparation sequence, where magnetization is generated on some nucleus. This is followed by evolution of that magnetization, followed by transfer to another nucleus, evolution on that nucleus, etc. This is illustrated in Figure 6.

Note, however, we only actually detect signal during the evolution of the *last* spin. In principle, we could turn on the receiver during earlier evolutions in the sequence. This would give us a one-D spectra of each spin detected, but it would not tell us how the measured chemical shifts are correlated. Instead, acquisition of

a 2D (for example) sequence proceeds as follows, with illustration in Figure 7, and as a video in wk2_videos.ppt, slides 3-4.

- Execute pulse sequence with $t_1 = 0$, acquire direct dimension (t_2).
- Increment t_1 (Figure 7A), re-acquire direct dimension, repeat (Figure 7B).
- Fourier transform t_2 (Figure 7C). This yields the chemical shifts of the the second spin ($^{13}\text{C}\alpha$ in this example).
- The peaks obtained by Fourier transform of t_2 oscillate during incrementation of t_1 , with oscillation frequency determined by the chemical shift of the first spin (^{15}N , Figure 7C).
- Fourier transform t_1 , to yield the two-dimensional spectrum (Figure 7D).

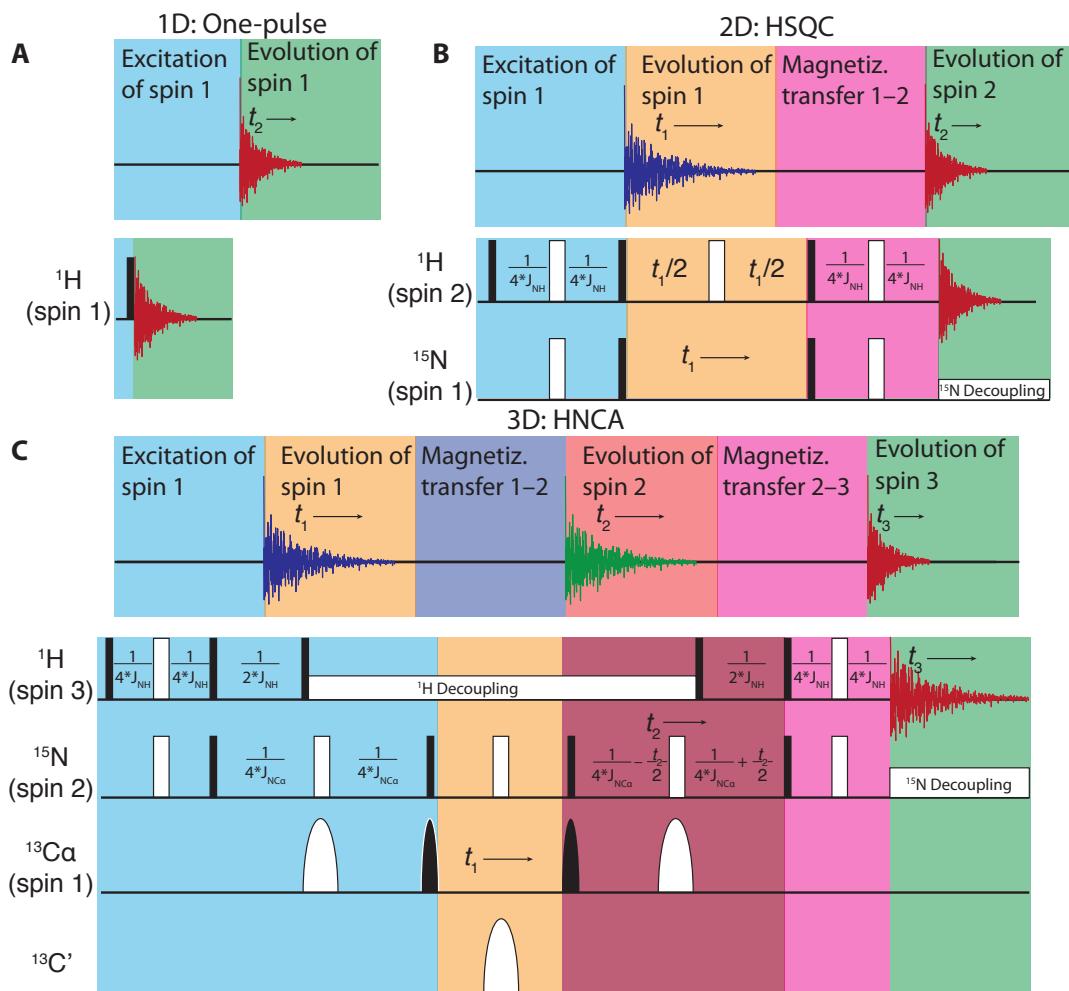


Figure 6: Dissecting correlation-spectroscopy experiments. Each subplot specifies the sections found in a sequence of that dimension, followed by an example sequence. **A** 1D experiment, with a one-pulse sequence. **B** 2D experiment, including excitation of spin 1, followed by evolution of spin 1 (t_1), transfer of magnetization to spin 2, and evolution of magnetization on spin 2. The example is a Heteronuclear Single Quantum Coherence (HSQC) experiment [1]. **C** 3D experiment, including two transfers. Note in the example, HNCA [2], the magnetization transfer between spins 1 and 2 and the evolution of spin 2 happen simultaneously (dark purple region). $^{13}\text{C}\alpha$ and ^{13}CO nuclei are excited with selective pulses, that is, they only act on the $\text{C}\alpha$ or the CO , indicated by non-rectangular pulses. Note, this is a simplified HNCA, where a few pulses and gradients are omitted.

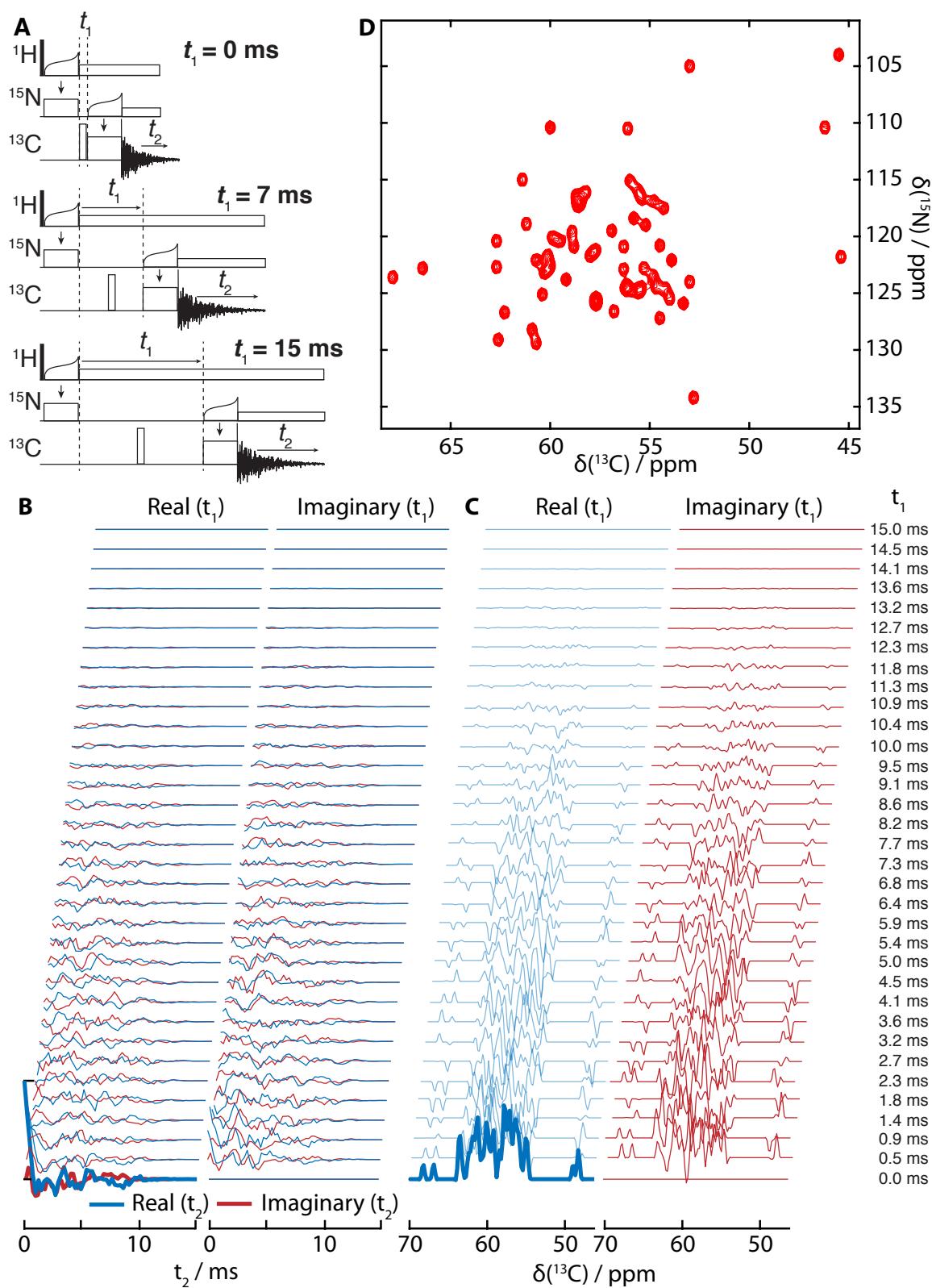
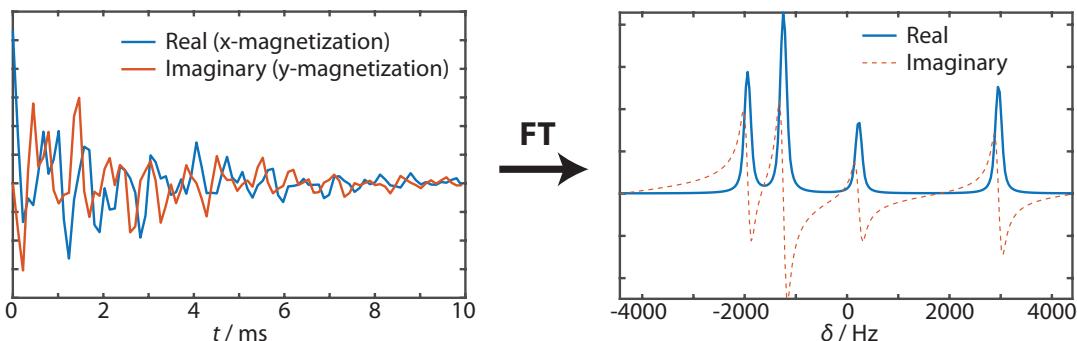


Figure 7: Acquisition of a 2D experiment. **A** Pulse sequence (solid-state cross-polarization based transfers, where transfer indicated with arrows). The sequence is repeated, with t_1 incremented from 0 to 15 ms. **B** At each t_1 , signal is acquired in the direct dimension (real+imaginary). **C** Each of these signals is Fourier transformed (t_2) to obtain a series of 1Ds. The 1D peaks correspond to the ^{13}C chemical shifts, but the signals oscillate at the frequency of the bonded ^{15}N resonance with t_1 incrementation. **D** Fourier transformation is performed along t_1 , to obtain the 2D spectrum, which correlates $^{13}\text{C}\alpha$ chemical shift to ^{15}N chemical shift.

Signal Acquisition and Processing (Fourier Transforms)

Here we briefly discuss the details of acquiring and processing the signal. The discrete Fourier transform allows us to decompose a signal into its frequency components. For example, in the following, we plot a time dependent signal that is the sum of four frequencies.



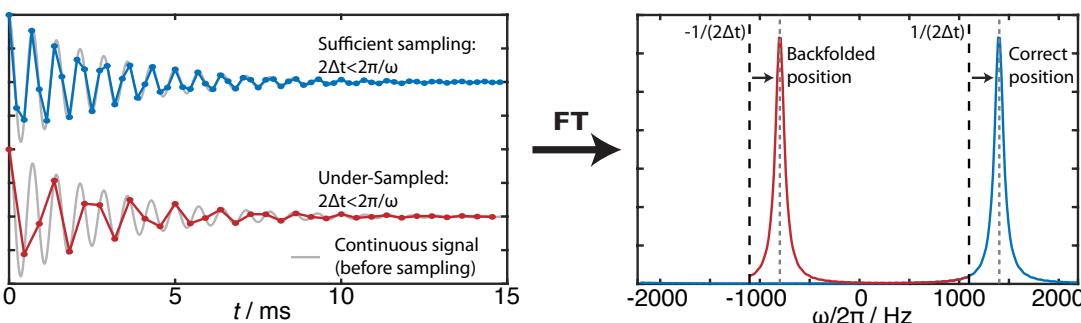
The Fourier transform yields a peak for each frequency. Note, the input has a "real" part and an "imaginary" part. The real part is the component measured on the x-axis (of the rotating frame) and the imaginary part is the component measured along the y-axis (rot. frame). Although the coil detects signal in the lab frame, the spectrometer multiples that signal by a reference signal with ω_{ref} , using two different phases:

$$\begin{aligned} \cos(\omega_{ref}t)\cos(\omega t) &= \frac{1}{2}(\cos((\omega - \omega_{ref})t) + \cos((\omega + \omega_{ref})t)) \\ \sin(\omega_{ref}t)\cos(\omega t) &= \frac{1}{2}(\sin((\omega - \omega_{ref})t) + \sin((\omega + \omega_{ref})t)) \end{aligned} \quad (1)$$

The sum term may be removed using a filter, so we are left with the difference term, which is in the rotating frame of the reference frequency. By mixing the signal with two different phases of the reference, we obtain an 'x' and 'y' component (practically, these two signals let us distinguish frequencies faster than ω_{ref} from frequencies slower than ω_{ref}).

Useful Fourier Transform properties

- The FT of a real signal (x-only) is symmetric around 0 (can't distinguish positive and negative frequencies).
- The FT of an imaginary signal (y-only) is anti-symmetric around 0.
- The maximum resolution obtained with a time trajectory (digital resolution) is $1/t_{acquisition}$, that is, the inverse of the longest time point acquired.
- Frequencies larger than $\omega/2\pi = 1/(2\Delta t)$ are backfolded (Nyquist). Consider the signal plotted below (grey). The second sampling scheme is too slow to capture the frequency, resulting in the peak appearing in the wrong place in the Fourier transformed spectrum (backfolding).



Three-dimensional (and higher dimensional) experiments may be acquired in the same way, but additional delays must be inserted and incremented sequentially, making the length of these experiments increasingly long.

- The length of one scan of an experiment is the length of the pulse sequence (~100 ms) plus a delay inserted between scans (this allows magnetization to recover to thermal equilibrium, and its length depends on T_1 of the nucleus that provides magnetization (usually ^1H , typical delays 1-2s)).
- For a 1D experiment, the total length is the #scans x (pulse sequence length + delay length)
- For a 2D experiment, the total length is #time points in indirect dimension (t_1) x #scans x (pulse sequence length + delay length)
- For higher dimensions, we have to take the product of time points in all indirect dimensions multiplied by #scans x (pulse sequence length + delay length)
- To reduce experimental time, 3D and higher dimensional experiments often use fewer time points in the indirect dimensions, sacrificing resolution and sweep width (see box: Signal Acquisition and Processing). One also uses "sparse-sampling" where many time points are omitted and later reconstructed.

Note that projections of higher dimensional experiments (that is, summing signal across one dimension) yield the corresponding lower dimensional spectra.

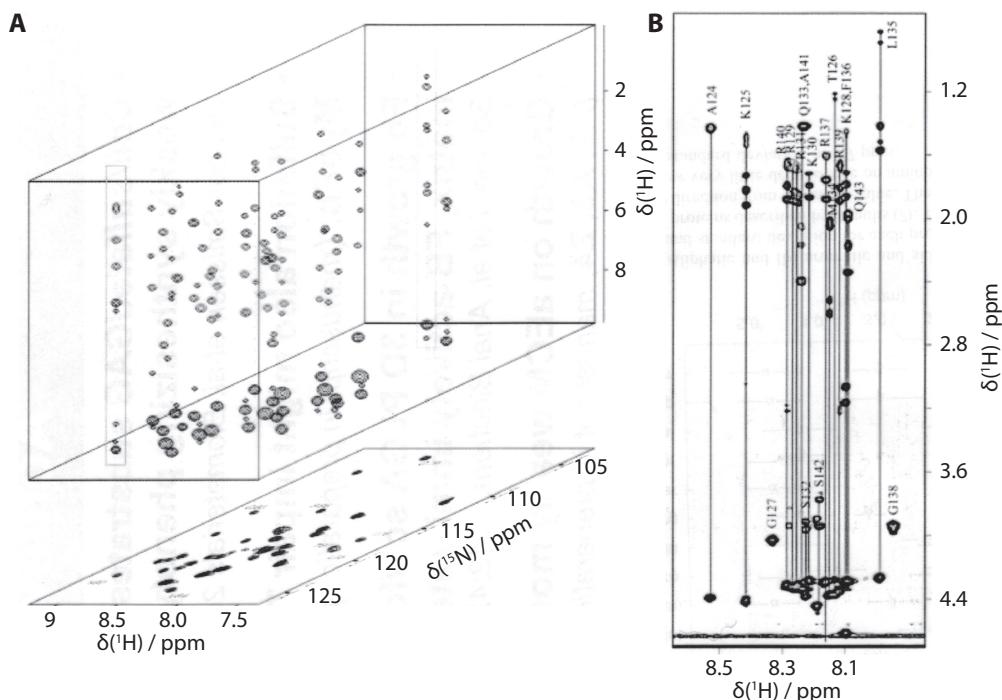
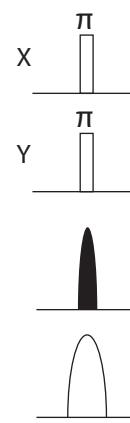
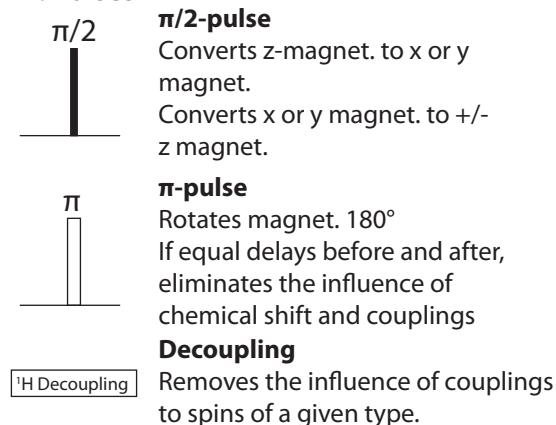


Figure 8: HN(CA)H TOCSY spectrum. The 3D spectrum in **A** correlates sidechain protons ($\text{H}\alpha$, $\text{H}\beta$, $\text{H}\gamma$, etc.) with backbone N and H^N . The projection through the indirect ^1H dimension yields the usual H-N HSQC spectrum (**A**, bottom). **B** shows the projection through the ^{15}N dimension, which yields a section of a TOCSY (H-H) spectrum, which correlates backbone H^N with $\text{H}\alpha$, $\text{H}\beta$, etc.

2.2 Two-Dimensional Correlation Experiments

We introduce several basic 2D experiments used for correlating chemical shifts in proteins.

A : Pulses **π -pulse (2 channels)**

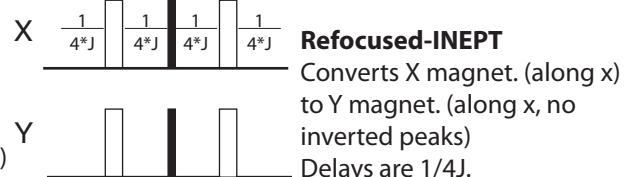
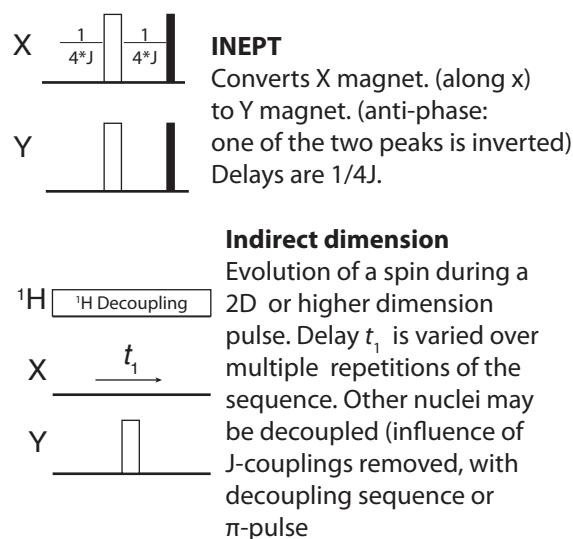
Rotates magnet. on both spins 180°
Eliminates influence of chemical shift
Does not eliminate influence of J-coupling

Selective $\pi/2$ -pulse

Same as $\pi/2$ -pulse, but acts only on some spins of a given nucleus (usually CO but not Ca/C β or vice versa).

Selective π -pulse

Same as π -pulse, but acts only on some spins of a given nucleus.

B : Sub-sequences**Spin-lock**

Magnetization is held along the x or y axis. On ^1H , this results in exchange of magnet. among systems of coupled spins (TOCSY)

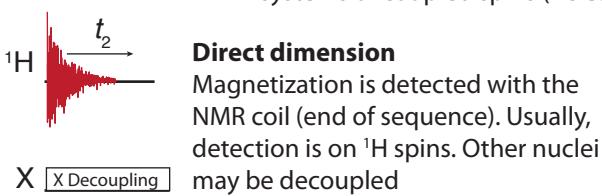
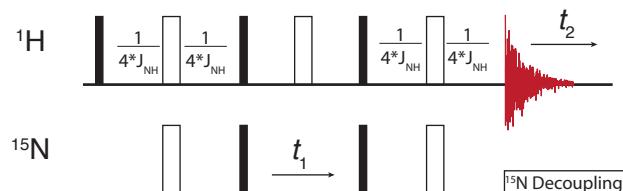
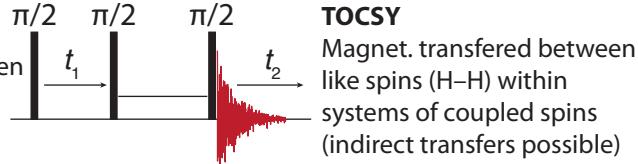
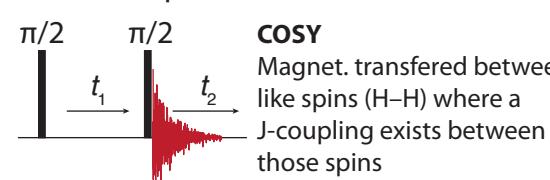
**C : Basic sequences**

Figure 9: Components of pulse sequences and basic pulse sequences. **A** shows the basic pulses used, and states their effect on magnetization.

- The simplest 2D experiment is the COSY (CORrelation SpectroscopY). COSY was the first 2D NMR experiment [3], and correlates like spins (ex. H-H) that are coupled via J-couplings (pulse sequence: Figure 9C). One can use COSY to identify spin-systems, that is, groups of protons which are coupled to each other. However, cross-peaks only appear for directly J-coupled ^1H . For example, in Figure 10B, we can start at the diagonal peak appearing at $\delta(^1\text{H}) = 5.3$ ppm. This connects to the next peak at 4.1 ppm, which then connects to the next

peak at 3.7 ppm. This continues until 6 resonances are connected, whereas other resonances belong to the second spin system.

- While one can in principle identify amino acids this way using the COSY experiment, connecting all resonances through the diagonal of the spectrum (as done in Figure 10B, left) can be challenging. An alternative approach is the TOCSY (TOtal Correlation SpectroscopY) experiment (pulse sequence: Figure 9C). In this experiment, a spin-lock is applied to achieve mixing via J-couplings. The results is that within a systems of spins that are J-coupled, magnetization is transferred between all spins, yielding cross-peaks between all spins (also between spins that are not directly coupled, ex. if AB and BC are J-coupled, then AC will have a cross-peak). The comparison of cross-peak patterns is shown in Figure 10A.
- TOCSY is generally easier to identify which resonance belong in a spin-system, and spectra are also somewhat cleaner (although less total signal) than COSY. However, one may not determine from TOCSY the exact connectivity within a spin system, only that all spins are either directly or indirectly coupled (compare Figure 10B).
- The number of resonances in a spin system and the range of the chemical shifts (see Figure 3A) are indicative of which amino acid one observes. Then, one may use either TOCSY or COSY to identify spin-systems.
- We may also obtain correlations between heteronuclear spin pairs, for example, between ^1H and ^{15}N nuclei. INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) is the basic sequence for polarization transfer between heteronuclei (Figure 9B), where two INEPT transfers are combined to form the HSQC (Heteronuclear Single Quantum Coherence spectroscopy) experiment (H–N transfer, ^{15}N evolution as the indirect dimension, and finally N–H transfer with direct detection of ^1H (Figure 9C).
- HSQC is one of the most commonly used pulse sequences in protein NMR. Most often, one may obtain H–N correlations for a protein, as shown in Figure 11 (although any pair of heteronuclei works in principle). Most of the resonances in these spectra result from correlations between the backbone H^N and N nuclei (some may result from H–N pairs found in amino acid sidechains, especially in case the ^{15}N resonance is backfolded into the spectrum).
- INEPT transfers magnetization via evolution of the J-couplings in the XY-plane. See [wk2_videos.pptx](#), slides 5-8 for a better explanation of how this works.
- HSQC spectra are often well-resolved, for proteins having ~ 100 residues, and the resulting spectra are often called the "protein fingerprint"..
- 3D spectra, used for assignment, use HSQC as the basic element of the pulse sequence, with additional transfers and a second indirect dimension.

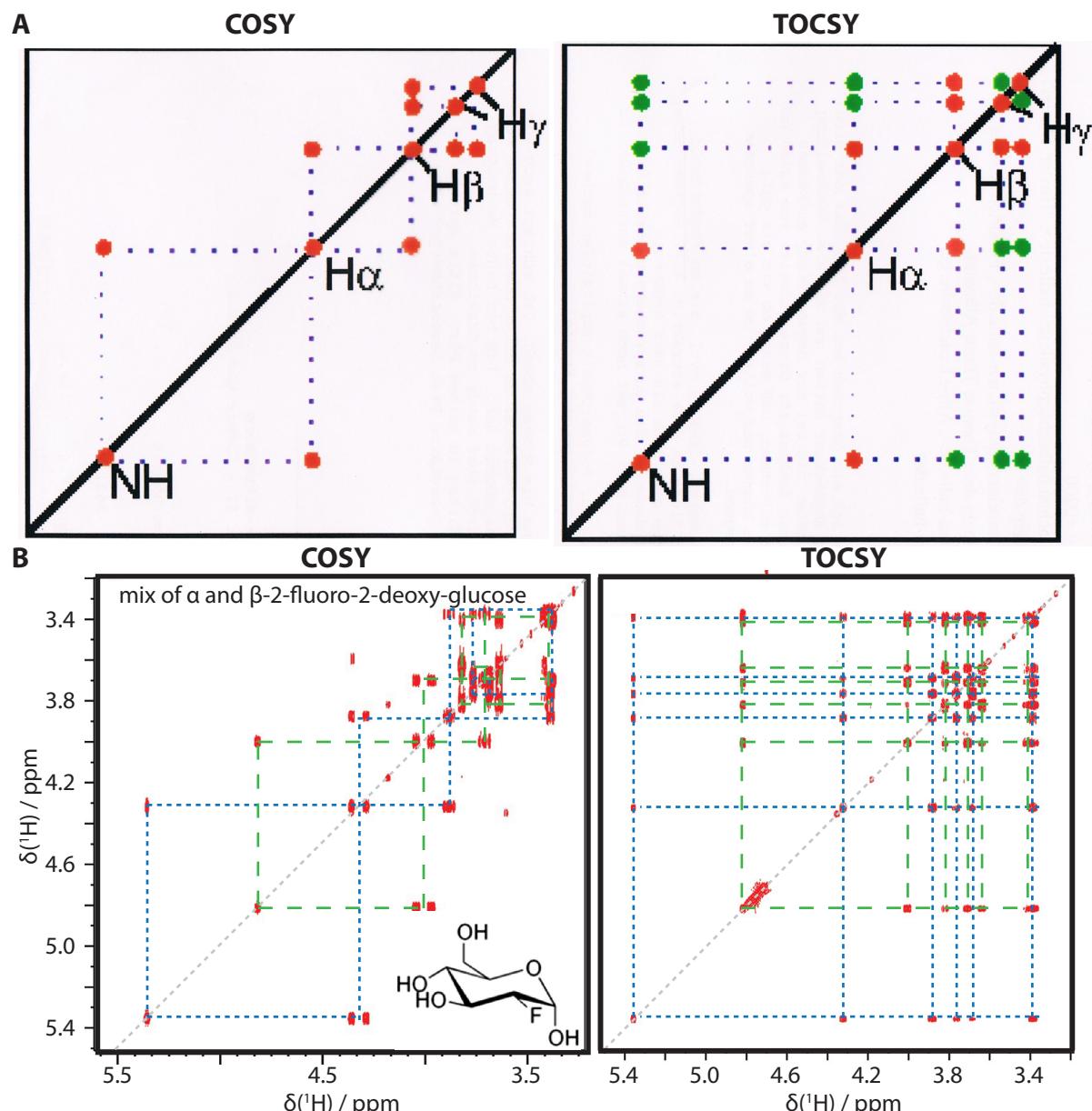


Figure 10: COSY vs. TOCSY. **A** shows the connectivity obtained via a COSY 2D experiment. Correlations appear between resonances that are directly J-coupled, so one can walk through the correlations to identify the full spin-system. **B** shows two example spectra, of a mix of α - and β -2-fluoro-2-deoxy-glucose. Then, there are two spin-systems present (couplings exist between spins of each system, but not between the two systems). In the COSY spectrum, we see cross-peaks between directly coupled spins, whereas in the TOCSY, we see cross-peaks between all spins in the same spin-system (COSY and TOCSY spectrum obtained from [4]).

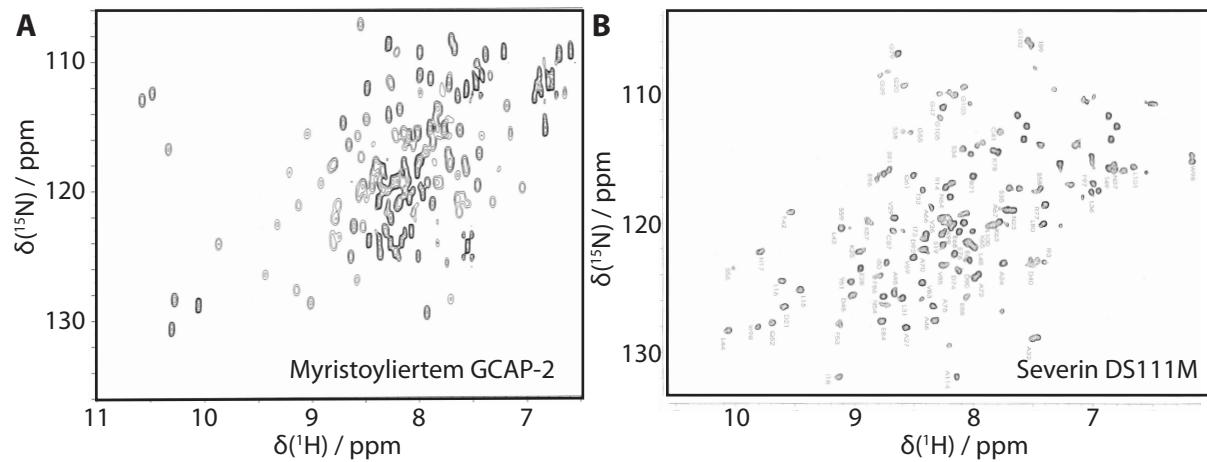


Figure 11: ^1H – ^{15}N HSQC spectra of **A** Myristoyliertem GCAP-2 and **B** Severin DS111M.

2.3 Backbone walk experiments (3D)

In the previous section, we introduce basic 2D homonuclear (ex. H–H) and heteronuclear (ex. H–N) correlation experiments. Homonuclear (COSY and TOCSY) correlation may be used in protein NMR to identify, for example, type of amino acid. Heteronuclear (HSQC) correlation may be used to obtain "fingerprint" spectra, that is, the backbone H–N correlation, as well as other heteronuclear correlations. Still, in complex proteins, we may be able to identify amino acid type for a set of resonances, but we still do not know where in the protein sequence that amino acid is found. This brings us to 3D spectroscopy.

- 3D spectra allow one to "walk" down the backbone of a protein, in order to assign all backbone chemical shifts (and possibly side chains). The third dimension is inserted in a similar fashion to the second dimension, as was illustrated in Figure 6C. An example 3D spectrum is shown in Figure 12.

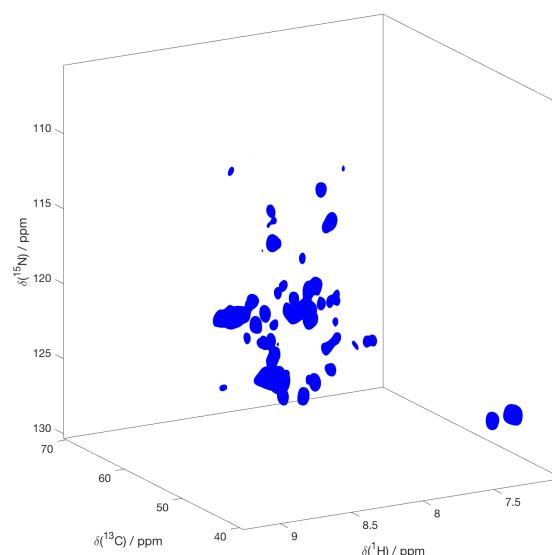


Figure 12: 3D HNCA spectrum

- The ability to assign a 3rd resonance allows us to perform a backbone walk. For example, suppose we have an HNCA 3D experiment. HNCA experiments transfer magnetization from the backbone H_i^N (i indicates the index of the amino acid), to backbone N_i (N of the same amino acid), and finally to *both* C_i^α and C_{i-1}^α . We can use this for assignment as follows (refer to Figure 13):
 1. To "walk" through the backbone, we start at 1Glycine (first strip). A single resonance is found, where the H, N, and C^α shifts are the result of an $H_i^N - N_i - C_i^\alpha$ transfer on 1-Glycine.
 2. We can find another resonance (second strip), have the same C^α chemical shift, but different H and N shifts. This peak results from the $H_i^N - N_i - C_{i-1}^\alpha$ transfer on 2-Asparagine.
 3. Within the same strip, we find a second peak. This has the same H, N chemical shifts as the first peak, but a new C^α shift, corresponding to the $H_i^N - N_i - C_i^\alpha$ transfer on 2-Asparagine.
 4. As with 1-Glycine, we can find another peak with the same C^α shift, but new H, N shifts (third strip). Proceeding in the same manner, we may walk through the whole backbone, thus assigning all H^N , N, and C^α resonances.

5. This assignment is performed as a video, using the ccpNMR software [5] in GNNQQNY_assign.mp4.

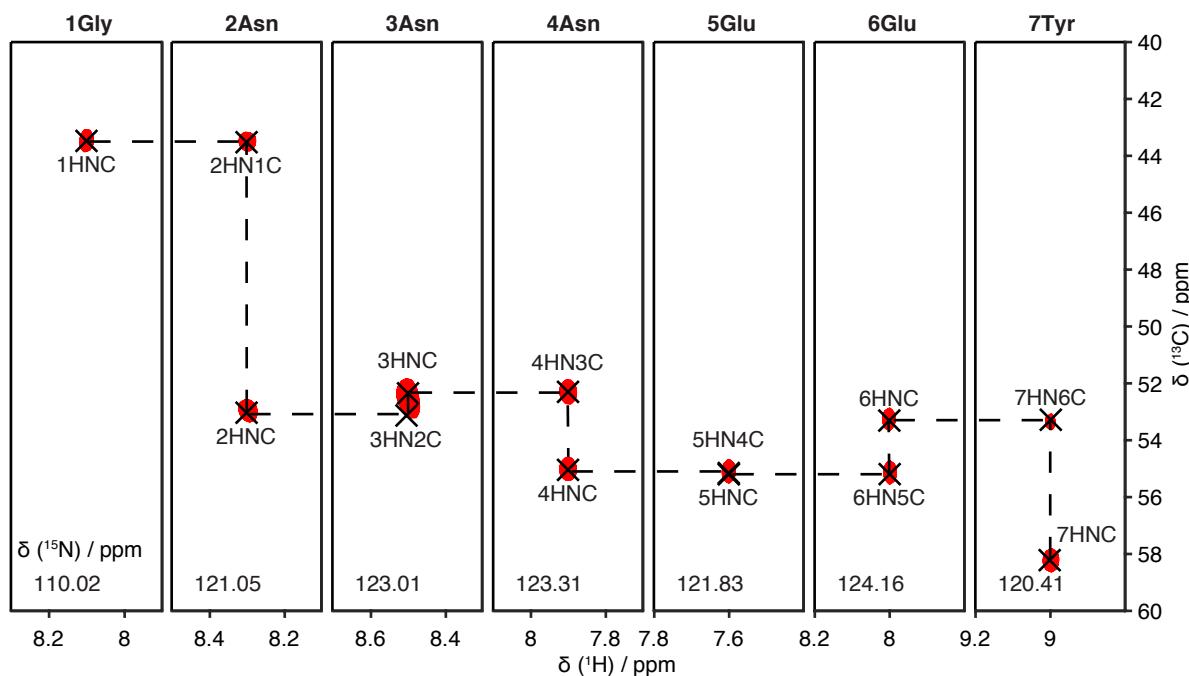


Figure 13: Strip plots of HNCA of GNNQQNY (simulated data). Each plot corresponds to a particular chemical shift on the ^{15}N axis (z-axis of the 3D), as indicated in the lower left. Then, each strip is zoomed in to the ^1H and ^{15}N chemical shifts of a given amino acid. Except for the first residue (which has no previous residue), each shows two peaks (sometimes these overlap, and so appear to be one peak). The backbone walk is indicated. Assignment of this peptide is shown using ccpNMR in GNNQQNY_assign.mp4 (separate file from wk2_videos.ppt).

- In principle, we may obtain an entire backbone assignment (without CO resonances) using only the HNCA experiment, since H and N resonances are correlated with both C_α^i and C_α^{i-1} resonances, allowing a continuous walk. However, peak overlap and missing resonances usually prevent such a simple assignment. Instead, we use a collection of experiments, some of which are summarized in 2. Connectivity obtained with each type of experiment is shown in Figure 14A. The usefulness of each experiment type is discussed below:

1. The CO_{i-1} shift for a given H_i^N and N_i resonance can be obtained with an HNCO experiment. An HN(CA)CO experiment can provide the CO shift for both the i and $i-1$ residues, although with reduced signal.
2. The $\text{H}_i^N-\text{N}_i-\text{C}_i^\alpha$ and $\text{H}_i^N-\text{N}_i-\text{C}_{i-1}^\alpha$ shifts are often not distinguishable, so that we do not know whether we should walk to the next or previous residue from a given peak. An HN(CO)CA experiment provides a $\text{H}_i^N-\text{N}_i-\text{C}_{i-1}^\alpha$ transfer, but not a $\text{H}_i^N-\text{N}_i-\text{C}_i^\alpha$ transfer. Then, if both HNCA and HN(CO)CA experiments are acquired, one can differentiate the $\text{H}_i^N-\text{N}_i-\text{C}_i^\alpha$ and $\text{H}_i^N-\text{N}_i-\text{C}_{i-1}^\alpha$ resonances, since only $\text{H}_i^N-\text{N}_i-\text{C}_{i-1}^\alpha$ resonances will appear in the HN(CO)CA spectrum
3. Walking through the backbone using an $\text{H}_i^N-\text{N}_i-\text{C}_i^\alpha$ experiment requires that the C_α dimension is sufficiently resolved. If we find multiple H,N resonances with approximately the same C_α chemical shift, we cannot

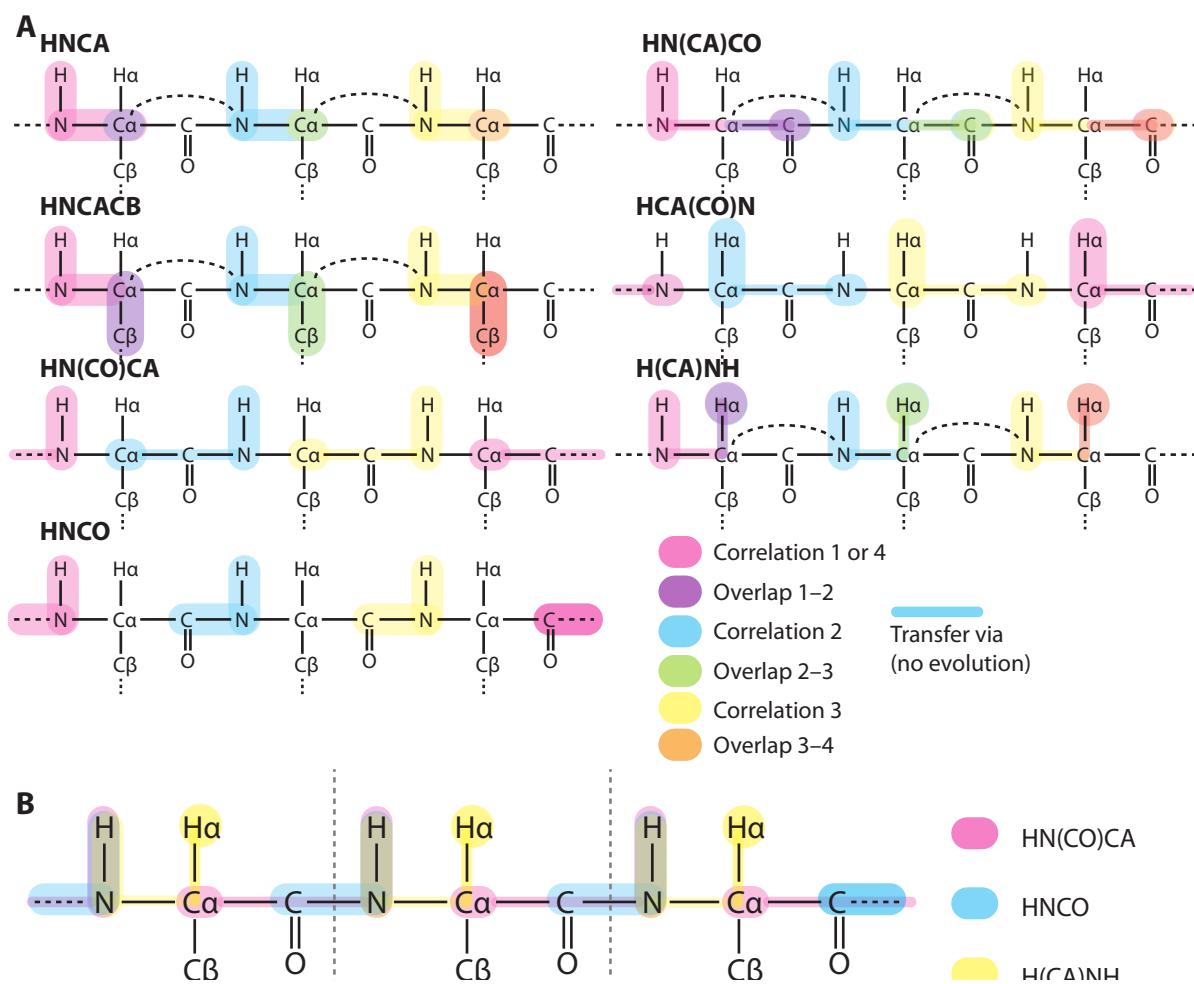


Figure 14: Correlations of different experiments. Several 3D experiments are selected, and the connectivity between nuclei is highlighted. **A** Connectivity of experiments. A single color (cyan, magenta, yellow) indicates what nuclei are correlated by the experiment (dotted arc indicates the $N_i-C_{i-1}^\alpha$ transfer). HNCA, HNCACB, H(CA)NH have overlapping correlations, allowing a backbone walk based on these spectra alone. Other experiments can be combined to obtain a backbone walk, as illustrated in **B**. Note that experiments contain CACB are run as 3D experiments, where magnetization evolves simultaneously on $C\alpha$ and $C\beta$ nuclei. A more complete list of experiments, and further details are found in Table 2.

determine which is the correct choice for a backbone walk. CBCANH and CBCA(CO)NH experiments can be used to resolve this.

- In this case, when searching for the next residue's ^{13}C shifts, we have to find a resonance for which both the $C\alpha$ and $C\beta$ chemical shifts match. This makes assignment less ambiguous, since it is less likely that both $C\alpha$ and $C\beta$ are the same for two different residues.
- These experiments are often run as 3Ds (a 4D version is also possible), where the $C\alpha$ and $C\beta$ shifts appear in the same dimension.
- This experiment also helps determine the amino acid type, since the $C\beta$ chemical shift also depends more strongly on amino acid type than $C\alpha$ (Figure 3B).
- As with the HNCA and the HN(CO)CA experiment, the CBCANH experiment connects the H and N shifts to the $C\alpha$ and $C\beta$ shifts of both the current (i) and previous ($i-1$) residue. CBCA(CO)NH, however, only connects to the $i-1$ residue.

4. $\text{H}(\text{CA})\text{NH}$, HCACO , and $\text{HCA}(\text{CO})\text{N}$ experiments allow assignment of the $\text{H}\alpha$ resonance, and provide further pathways for the backbone walk. In case a protein has been deuterated and back-exchanged (see Isotope Labeling box), these experiments will not work.
5. Note that, ideally, we would simply walk through the whole backbone without interruption. This is not always possible, for example, because two residues have the same chemical shifts (if both H and N match, or both $\text{C}\alpha$ and $\text{C}\beta$ match for two residues, it is not usually possible to complete the backbone walk).
 - In this case, we still have some residues connected via the backbone walk. Since we know the primary sequence of the protein, and typical values for $\text{C}\alpha$ and $\text{C}\beta$ shifts (Figure 3B), we can guess what part of the amino acid sequence that series of $\text{C}\alpha$ and $\text{C}\beta$ shifts belong to.
 - We may also use other experiments, for example, to obtain the ^1H shifts of the side chain, to determine the amino acid type. For example, the 3D experiment depicted in Figure 8A correlates the backbone H^N and N resonances to the sidechain ^1H resonances via a TOCSY transfer. Such a spectrum is very useful in determining amino acid type, since the number of peaks in the ^1H dimension and their shifts (Figure 3A) provide significant information on the amino acid.

Table 2: Triple-Resonance Experiments for Sequential Resonance Assignment

Experiment	Correlations Observed	Magnetization Transfer	J Couplings
HNCA	$^1\text{H}_i^N - ^{15}\text{N}_i - ^{13}\text{C}_i^\alpha$ $^1\text{H}_i^N - ^{15}\text{N}_i - ^{13}\text{C}_{i-1}^\alpha$		$^1\text{J}_{\text{NH}}$ $^1\text{J}_{\text{NCA}}$ $^2\text{J}_{\text{NCA}}$
HN(CO)CA	$^1\text{H}_i^N - ^{15}\text{N}_i - ^{13}\text{C}_{i-1}^\alpha$		$^1\text{J}_{\text{NH}}$ $^1\text{J}_{\text{NCO}}$ $^1\text{J}_{\text{CaCO}}$
H(CA)NH	$^1\text{H}_i^\alpha - ^{15}\text{N}_i - ^1\text{H}_i^N$ $^1\text{H}_i^\alpha - ^{15}\text{N}_{i+1} - ^1\text{H}_{i+1}^N$		$^1\text{J}_{\text{CaHa}}$ $^1\text{J}_{\text{NCA}}$ $^2\text{J}_{\text{NCA}}$ $^1\text{J}_{\text{NH}}$
HNCO	$^1\text{H}_i^N - ^{15}\text{N}_i - ^{13}\text{CO}_{i-1}$		$^1\text{J}_{\text{NH}}$ $^1\text{J}_{\text{NCO}}$
HN(CA)CO	$^1\text{H}_i^N - ^{15}\text{N}_i - ^{13}\text{CO}_i$ $^1\text{H}_i^N - ^{15}\text{N}_i - ^{13}\text{CO}_{i-1}$		$^1\text{J}_{\text{NH}}$ $^1\text{J}_{\text{NCA}}$ $^2\text{J}_{\text{NCA}}$ $^1\text{J}_{\text{CaCO}}$
HCACO	$^1\text{H}_i^\alpha - ^{13}\text{C}_i^\alpha - ^{13}\text{CO}_i$		$^1\text{J}_{\text{CaHa}}$ $^1\text{J}_{\text{CaCO}}$
HCA(CO)N	$^1\text{H}_i^\alpha - ^{13}\text{C}_i^\alpha - ^{15}\text{N}_{i+1}$		$^1\text{J}_{\text{CaHa}}$ $^1\text{J}_{\text{CaCO}}$ $^1\text{J}_{\text{NCO}}$
CBCA(CO)NH	$^{13}\text{C}_i^\beta / ^{13}\text{C}_i^\alpha - ^{15}\text{N}_{i+1}^\alpha - ^1\text{H}_{i+1}^N$		$^1\text{J}_{\text{CH}}$ $^1\text{J}_{\text{CaCb}}$ $^1\text{J}_{\text{CaCO}}$ $^1\text{J}_{\text{NCO}}$ $^1\text{J}_{\text{NH}}$
CBCANH	$^{13}\text{C}_i^\beta / ^{13}\text{C}_i^\alpha - ^{15}\text{N}_i^\alpha - ^1\text{H}_i^N$ $^{13}\text{C}_i^\beta / ^{13}\text{C}_i^\alpha - ^{15}\text{N}_{i+1}^\alpha - ^1\text{H}_{i+1}^N$		$^1\text{J}_{\text{CH}}$ $^1\text{J}_{\text{CaCb}}$ $^1\text{J}_{\text{NCA}}$ $^2\text{J}_{\text{NCA}}$ $^1\text{J}_{\text{NH}}$

Isotope Labeling

This week, we have seen that assignment of protein resonances often relies on a backbone walk, that is, magnetization transfers between 3 (or more) spins, to correlate resonances. However, for these transfers to work, one needs all spins involved in the transfer to have spin 1/2. Yet, 99% of carbon nuclei are ^{12}C , which has spin of 0, and 99.63 % of nitrogen are ^{14}N , which has a spin of 1 (in principle, we can detect spin-1 nuclei, but these have a large quadrupole coupling which shortens relaxation times and broadens the spectrum, limiting resolution).

- In a 2D experiment correlating H and C spins, only 1 % of molecules will yield signal due to the low abundance of ^{13}C
- IN a 3D HNCA experiment, correlating H^N , N, and $\text{C}\alpha$ spins, only $(1-0.99)\times(1-0.9963)=0.0037$ % of molecules will yield signal, making the experiment impossible under usual conditions
- This requires isotope labeling: we prepare the protein in such a way that we have ^{13}C and ^{15}N nuclei at the majority (~90%) of positions in the molecule.
- The most common scheme uses expression of proteins in E.Coli, where the growth medium provided has $^{15}\text{NH}_4\text{Cl}$ as the sole nitrogen source and uniformly ^{13}C labeled glycerol or glucose. This scheme results in uniformly labeled samples, with ^{13}C and ^{15}N at all positions (residual ^{12}C and ^{14}N is present, but randomly distributed).
- One also often uses D_2O and glycerol or glucose that is also deuterated during expression, resulting in a protein that is also uniformly deuterated. After such a preparation, exchangeable protons (bonded to O or N) are switched back to ^1H by putting the unfolded protein in H_2O (a folded protein, that is, in its native structure, may not exchange all protons). Then, the exchangeable ^1H can still be used in NMR experiments. This results in narrower linewidths, yielding better resolution. However, experiments where exchangeable H's are not the source of magnetization are no longer possible (ex. $\text{H}(\text{CA})\text{NH}$, HCACO , $\text{HCA}(\text{CO})\text{N}$, $\text{CBCA}(\text{CO})\text{NH}$, and CBCANH experiments in Table 2). This technique is referred to as deuterating the protein and "back-exchanging" the protons.
- Numerous other labeling schemes exist for "selective" labeling: One may want to label certain types of amino acids, or labeling certain carbons within the amino acids. These may be used to reduce signal and regain resolution on larger proteins, or limit magnetization transfer (esp. in relaxation experiments). Cell-free expression in particular is a powerful method for selectively labeling particular amino acids (discussed later by Peter Schmidt).

Next Week

This week, we have established an approach to assign resonances. Once we know what resonances corresponds to which positions in a protein's primary structure (i.e. sequence), we can use additional experiments to extract structural and dynamic information using NMR.

References

- [1] G. Bodenhausen and D. J. Ruben, "Natural abundance nitrogen-15 nmr by enhanced heteronuclear spectroscopy," *Chemical Physics Letters*, vol. 69, no. 1, pp. 185 – 189, 1980.
- [2] T. Yamazaki, W. Lee, M. Revington, D. L. Mattiello, F. W. Dahlquist, C. H. Arrowsmith, and L. E. Kay, "An hnca pulse scheme for the backbone assignment of ¹⁵n,¹³c,²h-labeled proteins: Application to a 37-kda trp repressor-dna complex," *Journal of the American Chemical Society*, vol. 116, no. 14, pp. 6464–6465, 1994.
- [3] W. P. Aue, E. Bartholdi, and R. R. Ernst, "Two-dimensional spectroscopy. application to nuclear magnetic resonance," *The Journal of Chemical Physics*, vol. 64, no. 5, pp. 2229–2246, 1976.
- [4] H. Kovacs and E. Kupce, "Parallel nmr spectroscopy with simultaneous detection of ¹h and ¹⁹f nuclei," *Magnetic Resonance in Chemistry*, vol. 54, no. 7, pp. 544–560, 2016.
- [5] S. P. Skinner, R. H. Fogh, W. Boucher, T. J. Ragan, L. G. Mureddu, and G. W. Vuisster, "Ccpnmr analysisassign: a flexible platform for integrated nmr analysis," *Journal of Biomolecular NMR*, vol. 66, no. 2, pp. 111–124, 2016.