

Protein Structure Determination

Dr. Albert A. Smith-Penzel

12 May, 2020

In the last two weeks, we have introduced basic concepts in magnetic resonance, and a strategy for assigning different resonances. Once we know where a peak in an NMR spectrum belongs in the protein's primary structure, we can use NMR to obtain information about the local structure and dynamics. We will discuss the following approaches to structure and dynamics:

- Determination of backbone torsional angles, using chemical shift and J-couplings (secondary structure)
- Estimation of distances, using ^1H - ^1H Nuclear Overhauser Effect (NOE)
- Structure calculation (tertiary structure)
- Dynamics characterization via relaxation measurements

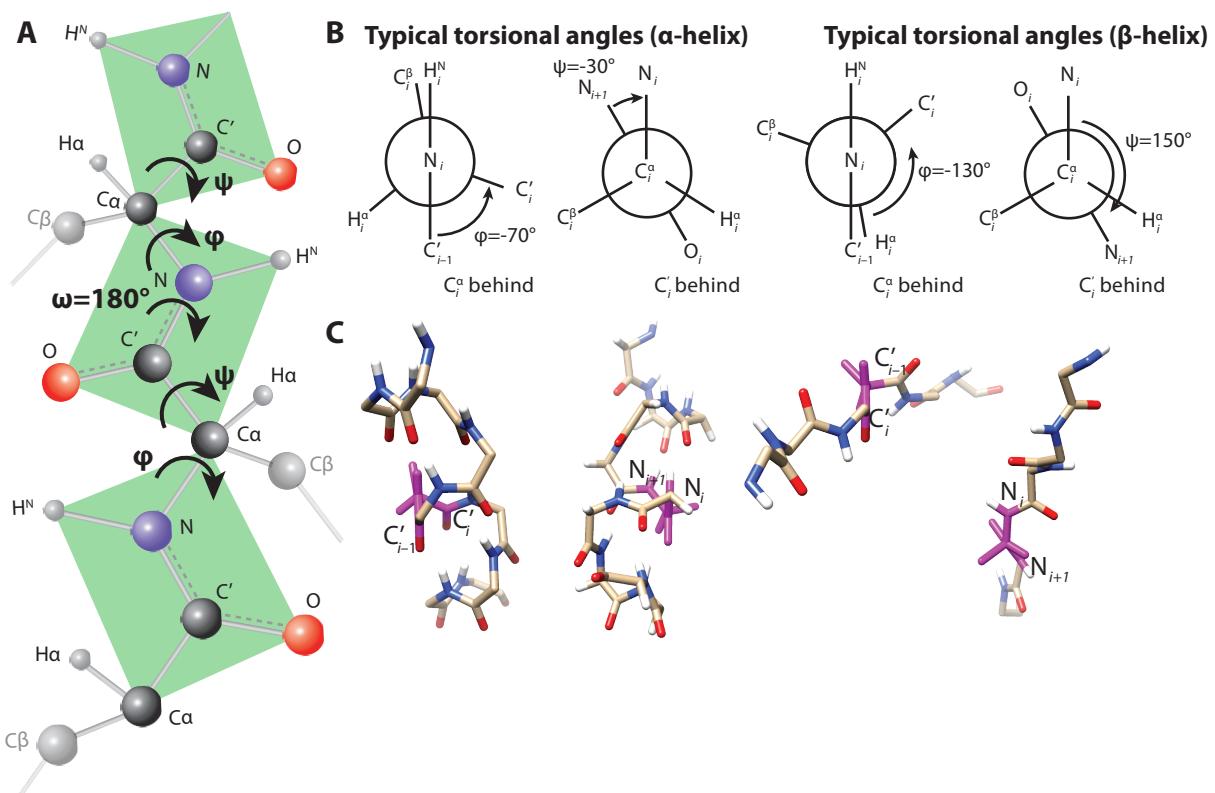


Figure 1: Torsional angles. **A** shows part of a protein backbone with the ϕ , ψ , and ω torsional angles indicated. Peptide planes and the partial double bonds between $\text{C}'-\text{O}$ and $\text{C}'-\text{N}$ are shown, which cause $\omega=180^\circ$. **B** shows the Newman projections for the ϕ and ψ torsional angles, for values typical for an α -helix and a β -sheet. **C** shows sections of an α -helix and a β -sheet. Atoms/bonds colored magenta are those shown in the Newman projection, where each molecule is rotated so the central bond of the Newman projection points out of the page ($\text{N}_i-\text{C}^\alpha_i$ for ϕ , $\text{C}^\alpha_i-\text{C}'_i$ for ψ).

1 Backbone Torsional Angles

Determination of backbone torsional angles plays a central role in structure determination using NMR. Torsional angles define the secondary structure of a protein (ex. α -helix v. β -sheet).

1.1 Definition of the torsional angles

Three torsional angles are defined (ϕ, ψ, ω , shown in Figure 1):

1. ϕ_i : For residue i , the angle between the carbonyl carbon (C') of the previous residue (C'_{i-1}) and the carbonyl carbon of the same residue (C'_i) viewed down the axis of the $N_i-C_i^\alpha$ bond.
2. ψ_i : For residue i , the angle between the nitrogen of the same residue (N_i) and the next residue (N_{i+1}), viewed down the axis of the $C_i^\alpha-C'_{i+1}$ bond.
3. ω_i : For residue i , the angle between $C\alpha$ of the previous residue (C_i^α) and $C\alpha$ of the same residue (C_{i+1}^α), viewed down the axis of the $C'_{i+1}-N_{i+1}$ bond. $\omega=180^\circ$ in almost all cases, because the partial double bond character of the $C-N$ bond requires that all atoms bonded to C' and N form a plane (the peptide plane). Then, ω is rarely reported with the ϕ and ψ torsional angles of a protein, since its value is already known.

1.2 Typical values of the torsional angles

Torsional angles cannot take on arbitrary values, but rather tend towards specific regions of values, depending on the secondary structure for the given residue. A Ramachandran plot is used to show what sets of values are typical for the ϕ and ψ angles (ω omitted since it is almost always 180°).

- The Ramachandran plot is a 2D plot of ϕ vs. ψ . An example Ramachandran plot is shown in Figure 2A.
- Colored regions indicate the likelihood of observing that pair of angles:
 1. Core regions: pairs of ϕ, ψ angles that are very common in proteins (**red**).
 2. Allowed regions: common, but less likely (**yellow**).
 3. Generous regions: uncommon, but still may occur occasionally in a protein structure (**tan**).
- Core, allowed, and generous regions are determined by evaluating a database of protein structures and extracting their torsional angles.
- The fraction of torsional angles in the database of protein structures found in the core, allowed, and generous regions are given in Table 1 [1].
- The three core regions observed in Figure 2 correspond primarily to β -sheet, α -helix, and left-handed helices.
- Typically, one uses a Ramachandran plot to evaluate the quality of a protein structure. For example, we plot torsional angles of ubiquitin onto the Ramachandran plot in Figure 2A. A structure having too many values in the less probable regions would be rejected as a bad structure (ProCheck [2] suggests 90% of values should fall in the allowed regions).

Table 1: Ramachandran Plot Regions

Region	Area (%)	Population (%)	Color
Core	11.0	81.9	red
Allowed	28.0	14.8	yellow
Generous	31.0	2.0	tan
Outside	30.0	1.3	—

- The distribution of torsional angles depends on the amino acid itself, so one may also compare torsional angles to amino-acid specific Ramachandran plots, as shown in 2B.

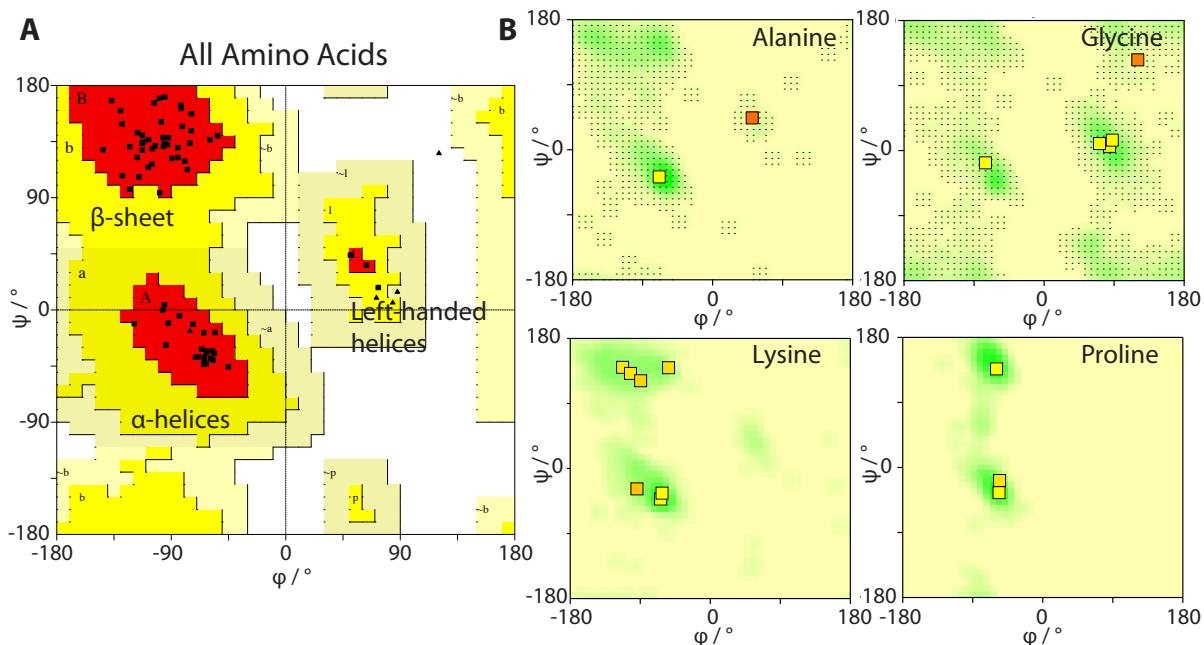


Figure 2: Ramachandran plots. **A** shows the Ramachandran plot for all amino acids, created using ProCheck (<https://servicesn.mbi.ucla.edu/PROCHECK/>, [2]). The different areas are defined in Table 1. Scatter points (black) are torsional angles determined for human Ubiquitin, from the structure "1ubi" (<http://www.rcsb.org/structure/1UBI>, [3]). Most torsional angles for Ubiquitin are found in the core regions. **B** shows the Ramachandran plots for specific amino acids, which have different distributions of torsional angles than the average distribution shown in **A**.

1.3 J-couplings and the Karplus Equation

One approach to extracting torsional angles using NMR is measurement of vicinal J-couplings. The relationship between the J-coupling of the backbone H^N – H^α and the ϕ torsional angle is given by the Karplus equation [4]:

$$J_{HN,H\alpha}^3 = A \cos^2(\phi - 60^\circ) + B \cos(\phi - 60^\circ) + C \quad (1)$$

- A , B , and C can be parameterized experimentally. Bax and co-workers have done so with SNase [5], where one observes a strong relationship between the ϕ torsional angle and the measured $J_{HN,H\alpha}^3$ coupling, as plotted in Figure 3B.

- Evolution under $J_{\text{HN},\text{H}\alpha}^3$ can be obtained with a variety of pulse sequences; one example is given in Figure 3C.
- However, measuring ϕ alone does not necessarily even determine whether a residue is located in an α -helix or a β -sheet. This can be seen in Figure 2A, where values of ϕ near -90° can easily be located in either an α -helix or a β -sheet. Furthermore, measuring $J_{\text{HN},\text{H}\alpha}^3$ does not uniquely determine ϕ , so vicinal J-couplings are best suited to supplement other structural data.

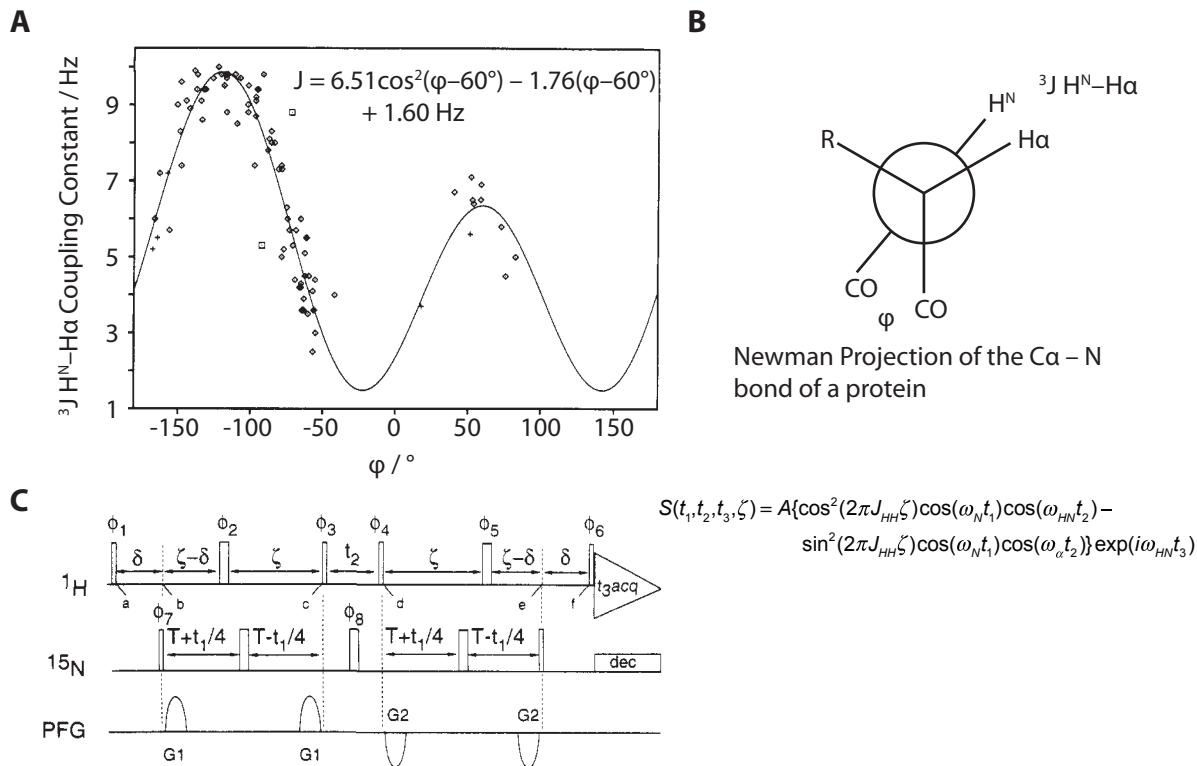


Figure 3: $J_{\text{HN},\text{H}\alpha}^3$ vs. ϕ . **A** plots experimentally measured J-couplings of SNase as scatter points against the ϕ torsional angle [5]. **B** shows the Newman projection for ϕ . **C** plots a 3D pulse sequence to measure $J_{\text{HN},\text{H}\alpha}^3$, where ζ can be varied, yielding a time dependence that depends on $J_{\text{HN},\text{H}\alpha}^3$.

1.4 Chemical Shifts vs. Secondary Structure

Another approach to determining secondary structure is the analysis of chemical shifts. In particular, the $\text{C}\alpha$ and $\text{C}\beta$ chemical shifts depend strongly on whether they are found in an α -helix or a β -sheet. This dependence is plotted in Figure 4 for the 20 amino acids.

- A simple approach to obtain secondary structure information (and therefore a rough estimate of ϕ and ψ) is calculating the secondary chemical shifts for $\text{C}\alpha$ and $\text{C}\beta$ resonances in a protein.
- The secondary chemical shift is defined as:

$$\begin{aligned} \Delta\delta(\text{C}\alpha) &= \delta(\text{C}\alpha) - \langle \delta(\text{C}\alpha) \rangle_{\text{AA}} \\ \Delta\delta(\text{C}\beta) &= \delta(\text{C}\beta) - \langle \delta(\text{C}\beta) \rangle_{\text{AA}} \end{aligned} \quad (2)$$

$\delta(\text{C}\alpha)$ and $\delta(\text{C}\beta)$ are the measured chemical shifts for the α and β carbons. $\langle \delta(\text{C}\alpha) \rangle_{\text{AA}}$ and $\langle \delta(\text{C}\beta) \rangle_{\text{AA}}$ are the average chemical shifts for the same amino

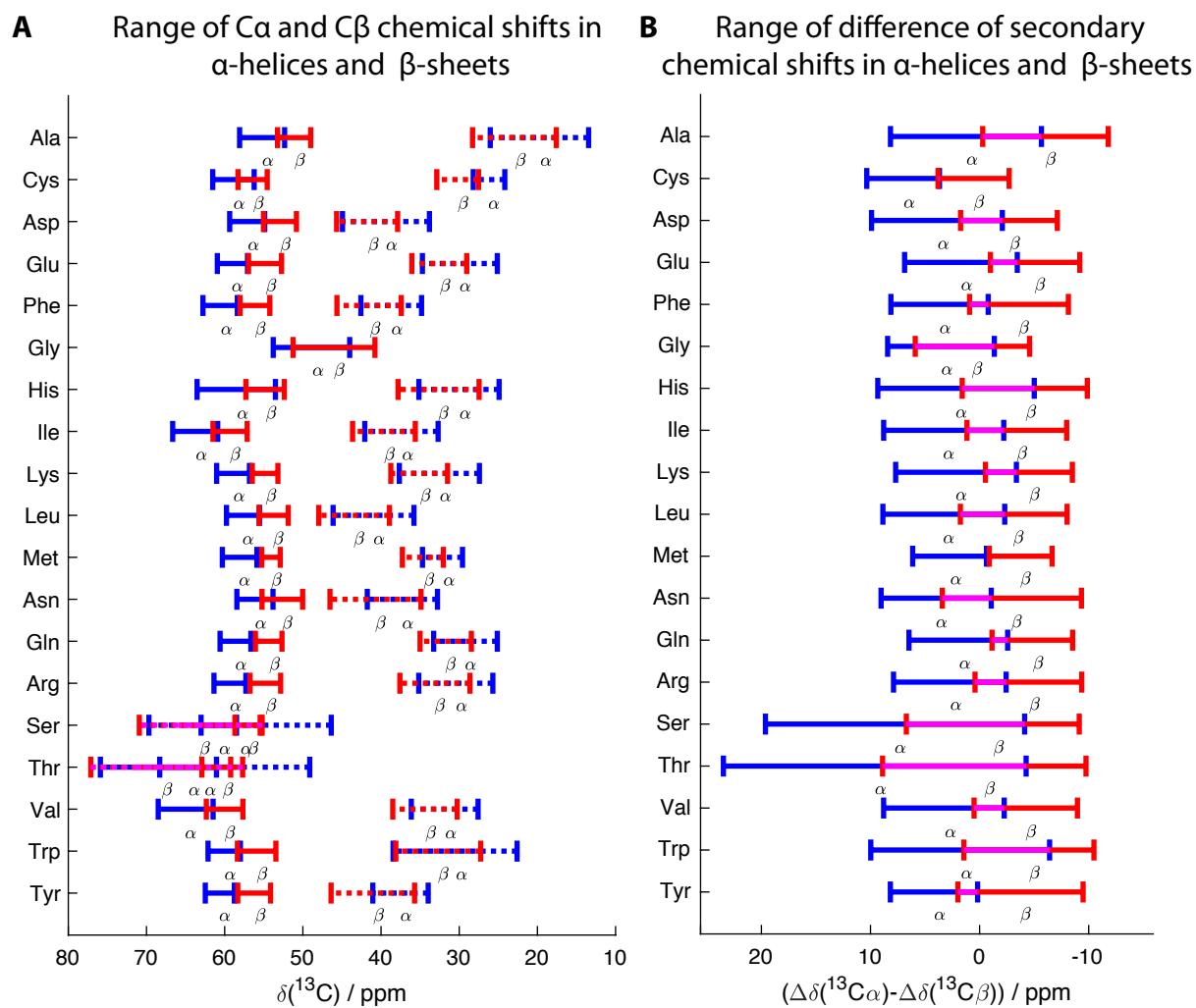


Figure 4: Secondary chemical shift. **A** Chemical shift ranges for different amino acids in α -helices (blue) and β -sheets (red) for $\delta(\text{C}\alpha)$ and $\delta(\text{C}\beta)$. **B** Range of $\Delta\delta(\text{C}\alpha) - \Delta\delta(\text{C}\beta)$ for α -helices and β -sheets.

acid type, averaged over a database of chemical shifts (here, we use data compiled in the Biological Magnetic Resonance Data Bank: http://www.bmrb.wisc.edu/ref_info/).

- Secondary chemical shifts are calculated for ubiquitin, where no clear correlation is seen between $\delta(\text{C}\alpha)$ or $\delta(\text{C}\beta)$ and the structure (Figure 5A, top). On the other hand, subtracting the average chemical shift for each amino acid type yields $\Delta\delta(\text{C}\alpha)$, which becomes positive for the α -helix and negative for β -sheets, whereas $\Delta\delta(\text{C}\beta)$ does the opposite (Figure 5A, middle).
- Since the secondary chemical shift has the opposite sign for $\text{C}\alpha$ and $\text{C}\beta$, we can get a better estimate of the secondary structure element by taking the difference $\Delta\delta(\text{C}\alpha) - \Delta\delta(\text{C}\beta)$ (Figure 5A, bottom).
- Color-coding this result onto the ubiquitin structure shows clearly the correlation between secondary chemical shift and secondary structure elements.
- More advanced approaches may be taken, to use chemical shift to extract torsional angles. For example, TALOS+ (Torsional Angle Likelihood Obtained from Shift and sequence similarity, [7]) considers not only $\Delta\delta(\text{C}\alpha)$ and $\Delta\delta(\text{C}\beta)$ for a given residue, but if available, utilizes $\Delta\delta(\text{N})$, $\Delta\delta(\text{H}^N)$, $\Delta\delta(\text{C}')$, and $\Delta\delta(\text{H}\alpha)$, and uses shifts from proceeding and following residues.

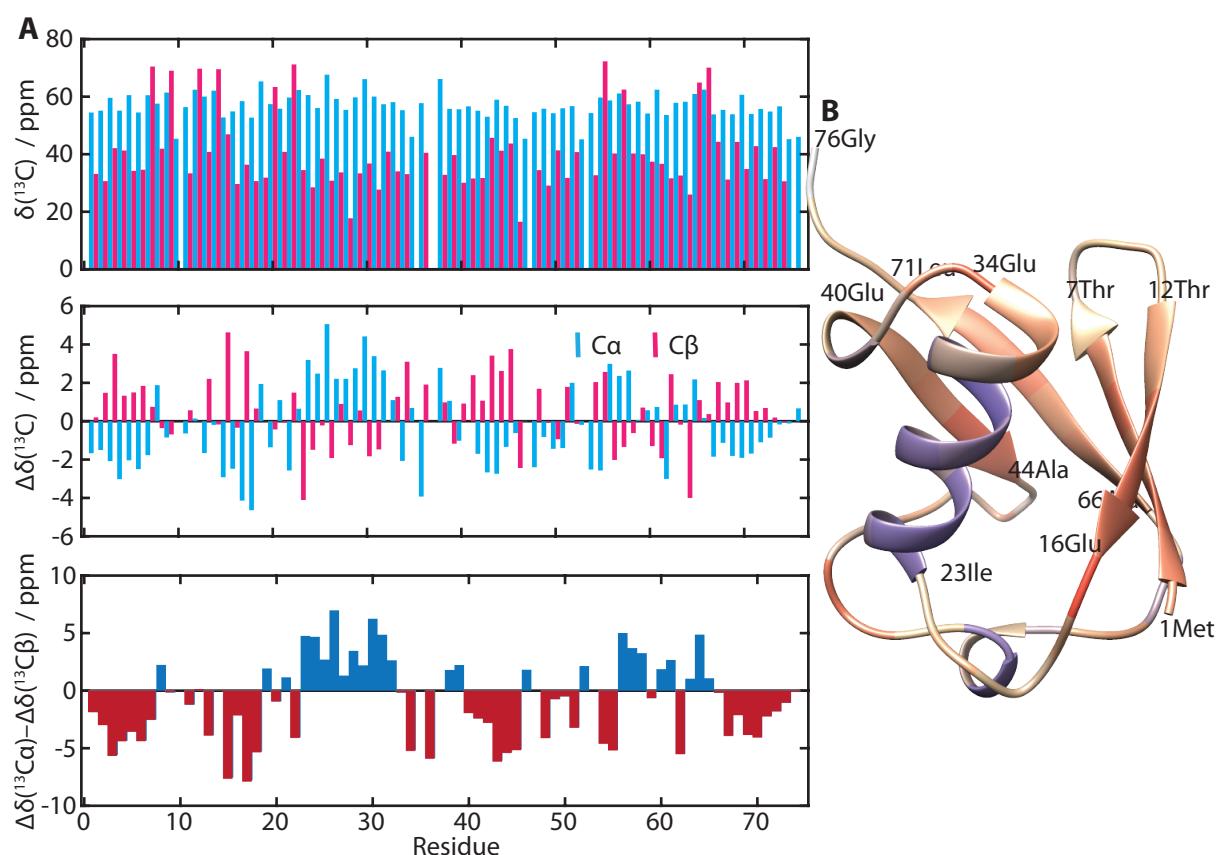
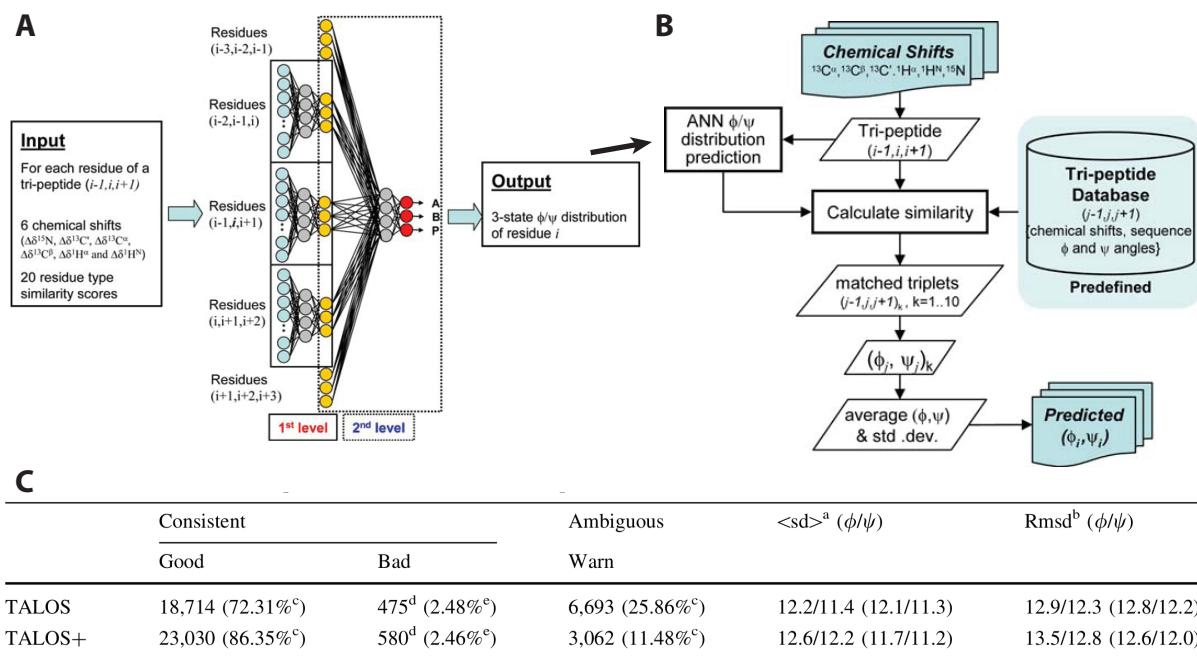


Figure 5: Secondary chemical shifts for Ubiquitin. **A** plots $\delta\text{C}\alpha$ and $\delta\text{C}\beta$ for each residue in ubiquitin (top) (available at <http://www.bmrb.wisc.edu/>, entry ID 15410, [6]). The second plot shows the secondary chemical shifts, calculated from eq. 2. $\Delta\delta(\text{C}\alpha)$ usually becomes positive in an α -helix and negative in a β -sheet. Conversely, $\Delta\delta(\text{C}\beta)$ usually becomes negative in an α -helix and positive in a β -sheet. The third plot takes the difference of $\Delta\delta(\text{C}\alpha)-\Delta\delta(\text{C}\beta)$, where positive values are colored blue and negative values are colored red. This result is color-coded onto the ubiquitin structure in **B**, where we see that the secondary chemical shift is highly correlated with the secondary structure elements of ubiquitin (ends of helices and β -sheets are labeled in **B**).

- To determine torsional angles for a given residue, these six secondary chemical shifts are provided in addition to the shifts for the two preceding and two following residues (30 secondary chemical shifts in total).
- This data is fed into a neural network, which estimates ϕ and ψ for a residue (Figure 6A).
- Result is compared to a database of 200 proteins. 10 nearest matches are found, where quality of match is evaluated based on the 3-residue sequence of amino acids, the measured chemical shifts, and the torsional angles compared with the results of the first step. These matches allow estimation of the mean and standard deviation of the ϕ and ψ angles (Figure 6B).
- Then, one has both an estimate of the torsional angles, but also an assessment of the quality of that estimate. Figure 6C shows the fraction of good (correct prediction), bad (wrong prediction), and ambiguous (no prediction due to high standard deviation) estimates of torsional angles.
- Further improvements of this methodology are implemented in TALOS-N [8].



TALOS and TALOS+ runs were performed for 200 proteins present in its reference database, with all residues from the protein under investigation excluded from the search

Figure 6: TALOS+ work flow and performance (figures from Reference [7]). **A** diagrams a two-layer neural network which provides initial predictions of ϕ and ψ . The first layer uses secondary chemical shifts of the residue of interest, and the prior and subsequent residue (18 shifts). The second layer uses shifts of two prior and two subsequent residues (30 shifts). **B** shows the complete TALOS+ workflow, where output of the neural network (ANN ϕ/ψ distribution prediction), along with chemical shifts and sequence similarity are compared to a database of 200 proteins. **C** tabulates performance of TALOS and TALOS+.

2 Structure Calculation (NOE)

So far, we have investigated how NMR may provide backbone torsional angles of proteins. However, one needs a continuous chain of highly accurate torsional angles to predict overall structure. This is usually not possible, even with advanced software such as TALOS. What we need is a way to connect nuclei that are close in space, but distant in the protein backbone sequence. Measurement of dipolar couplings would provide the required information, since the magnitude of a dipolar coupling is proportional to $1/r^3$, where r is the distance between two nuclei. However, as discussed in week 1, dipolar couplings are also scaled by $\frac{1}{2}(3\cos^2\theta - 1)$, where θ is the angle between the external magnetic field (B_0) and the vector connecting the two nuclei. Molecular tumbling in solution-state NMR, then, causes the average of this term to be zero, so that dipolar couplings are invisible in solution-state NMR spectra.

2.1 Principles of the Nuclear Overhauser Effect

However, dipolar couplings can still influence solution-state NMR spectra. Specifically, motion in the sample results in relaxation, mediated (in part) by the dipole coupling. In week 1, we introduced the Bloch equations describing magnetization of a single spin. This included terms T_1 , that is, the return of magnetization to the z -axis after excitation (longitudinal relaxation), and T_2 , the decay of magnetization in the xy -plane (transverse relaxation). Both these processes occur because of random motion of the sample— that motion can come from internal protein motion, and also from the molecular tumbling of the whole molecule in solution.

In case we have two spins (or more), and they are dipole coupled, longitudinal relaxation is no longer simply described by T_1 ; we can have transfer of magnetization between the two spins. This can be seen by considering the four possible states of a two spin- $1/2$ system: spin-up / spin-up, spin-up / spin-down, spin-down / spin-up, and spin-down / spin-down. ($|\alpha, \alpha\rangle$, $|\alpha, \beta\rangle$, etc.). These are illustrated in Figure 7A.

- The familiar T_1 relaxation results from transitions yielding the change of the state of one of the two spins, indicated by W_{1I} or W_{1S} .
- However, if there is a dipole coupling, both spins may flip simultaneously, either $|\alpha, \alpha\rangle \leftrightarrow |\beta, \beta\rangle$ or $|\alpha, \beta\rangle \leftrightarrow |\beta, \alpha\rangle$. These are known as double- or zero-quantum transitions, respectively, denoted in Figure 7A by W_2 and W_0 .
- The W_x indicate transition rates between the four possible states, where the resulting coupled differential equations are:

$$\begin{aligned} \frac{dN_{\alpha,\alpha}}{dt} &= W_{1I}(N_{\beta,\alpha} - N_{\alpha,\alpha}) + W_{1S}(N_{\alpha,\beta} - N_{\alpha,\alpha}) + W_2(N_{\beta,\beta} - N_{\alpha,\alpha}) \\ \frac{dN_{\alpha,\beta}}{dt} &= W_{1I}(N_{\beta,\beta} - N_{\alpha,\beta}) + W_{1S}(N_{\alpha,\alpha} - N_{\alpha,\beta}) + W_2(N_{\beta,\alpha} - N_{\alpha,\beta}) \\ \frac{dN_{\beta,\alpha}}{dt} &= W_{1I}(N_{\alpha,\alpha} - N_{\beta,\alpha}) + W_{1S}(N_{\beta,\beta} - N_{\beta,\alpha}) + W_2(N_{\alpha,\beta} - N_{\beta,\alpha}) \\ \frac{dN_{\beta,\beta}}{dt} &= W_{1I}(N_{\alpha,\beta} - N_{\beta,\beta}) + W_{1S}(N_{\beta,\alpha} - N_{\beta,\beta}) + W_2(N_{\alpha,\alpha} - N_{\beta,\beta}) \end{aligned} \quad (3)$$

where the $N_{\alpha,\beta}$ indicate the number of spins in each state. Note that what we can actually observe via NMR is the total magnetization on the I spin (the

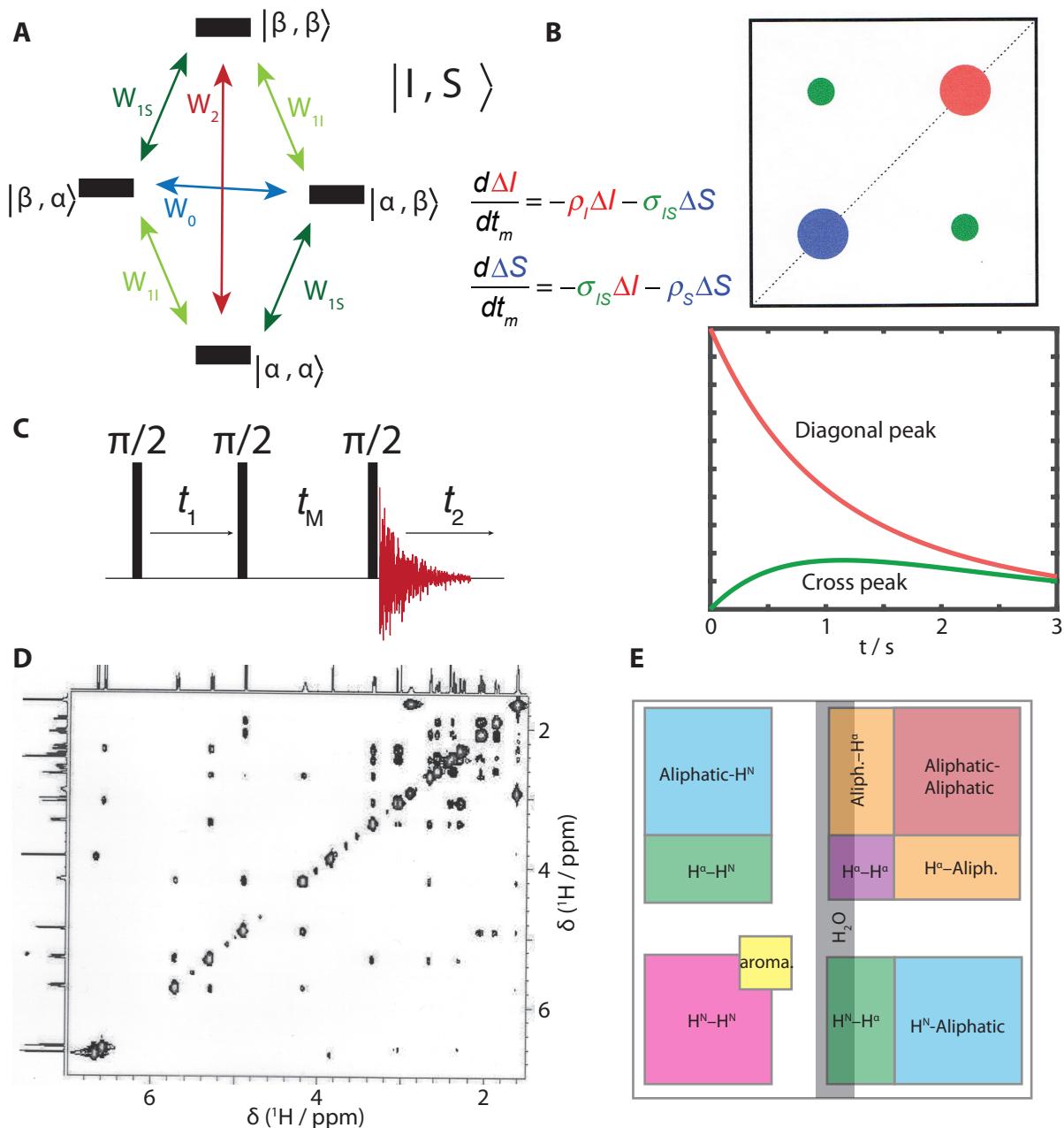


Figure 7: NOESY. **A** shows the energy levels of a two-spin $1/2$ system. Four states are possible, resulting in six total transitions between them. **B** plots diagonal and cross peaks of a NOESY spectrum, and plots typical decay behavior of the diagonal peak, and buildup/decay of the cross peak. **C** plots the basic NOESY pulse sequence. **D** Shows a NOESY spectrum. **E** is a schematic of regions of a ^1H - ^1H NOESY, where different regions correlate different types of protons.

first spin) or the total magnetization on the S spin, which are defined as the difference in magnetization in the α and β states for each spin, leading to

$$\begin{aligned}\Delta I &= N_{\alpha,\alpha} + N_{\alpha,\beta} - N_{\beta,\alpha} - N_{\beta,\beta} \\ \Delta S &= N_{\alpha,\alpha} + N_{\beta,\alpha} - N_{\alpha,\beta} - N_{\beta,\beta}\end{aligned}\quad (4)$$

- Then, we can find the differential equations describing the evolution of ΔI and ΔS .

$$\begin{aligned}\frac{d\Delta I}{dt} &= -\sigma_{IS}\Delta S - \rho_I\Delta I + \text{constant} \\ \frac{d\Delta S}{dt} &= -\sigma_{IS}\Delta I - \rho_S\Delta S + \text{constant}\end{aligned}\quad (5)$$

where ρ_I , ρ_S , and σ_{IS} are defined as:

$$\begin{aligned}\rho_I &= W_{1I} + W_0 + W_2 \\ \rho_S &= W_{1S} + W_0 + W_2 \\ \sigma_{IS} &= W_2 - W_0\end{aligned}\quad (6)$$

- The result is the Solomon equations [9]. Note, the use of I and S in these equations comes from the initials of their discover, Ionel Solomon.
- A solution of the Solomon equations results in transfer of magnetization between spins, as plotted in Figure 7B. The term σ_{IS} results in magnetization transfer between the two spins.
- For a homonuclear spin system (ex. two ^1H 's), this will result in equilibration of the magnetization on the two spins.
- Magnetization also will decay towards the thermal equilibrium of each spin, brought about the constant in the above equations. However, most implementations of the NOE pulse sequence will use phase-cycling to eliminate this term (the sign of the magnetization is inverted every other scan, so that the magnetization decays towards zero, Figure 7B, bottom).
- The transfer of magnetization between the spins is known as the Nuclear Overhauser Effect (NOE, where a similar electron-nuclear effect was previously described by Albert Overhauser [10] and verified by Charlie Schlichter [11]).
- Because this effect is mediated by the dipolar coupling, the rate of transfer, σ_{IS} is proportional to the square of the dipole coupling, such that

$$\begin{aligned}\sigma_{IS} &\propto \frac{1}{r^6} \\ \sigma_{IS} &\propto \tau_M.\end{aligned}\quad (7)$$

The second relationship results from the dependence of NOE on dynamics in the sample. For a relatively large molecule, such as a protein, and a homonuclear spin system, $W_0 \gg W_2$, and therefore W_0 determines σ_{IS} . The slowest motion in the system has the biggest influence on W_0 , which is the tumbling of the molecule in solution. τ_M is the correlation time of the tumbling, related to how long it takes for the molecule to make a full rotation in solution. Note that τ_M is larger for larger proteins, so that NOE becomes more efficient (but T_2 also depends similarly on τ_M , broadening signals and lowering resolution).

2.2 NOE Spectroscopy

The ^1H - ^1H NOE can be used as a means of measuring distance in solution-state NMR. Typically, this is implemented as a 2D ^1H - ^1H experiment as shown in 7C, although to obtain higher resolution, one may add additional dimensions. This experiment is known as Nuclear Overhauser Effect SpectroscopY (NOESY). An example 2D NOESY spectrum is shown in 7D, with regions of the ^1H - ^1H spectrum designated in 7E.

- Typically, one begins with ^1H magnetization on spin-1, which evolves during t_1 . This magnetization is then rotated to the z-axis (second π -pulse in 7C), where magnetization exchange occurs due to the NOE, as illustrated in 7A. After mixing (~ 100 ms), magnetization is rotated again to x, where signal is detected, resulting in a 2D spectrum.
- This yields cross-peaks, which indicate that the corresponding nuclei are nearby in space (usually <5 Å).
- The cross-peaks obtained depend strongly on the mixing time. For example, in Figure 8A, simulated NOESY spectra are shown for 10, 50, and 200 ms mixing times. One sees increasing number of peaks with increasing mixing time.

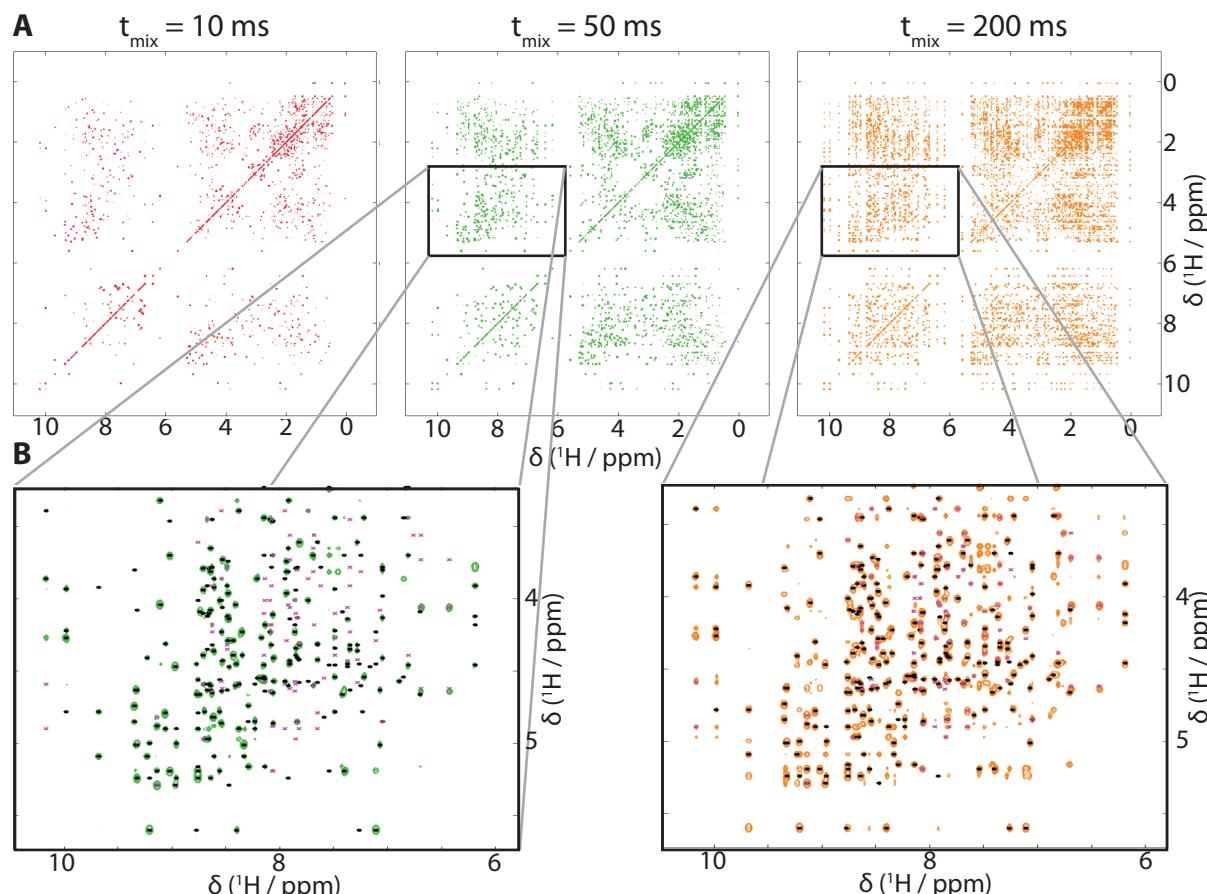


Figure 8: NOESY spectra of Ubiquitin, simulated for different mixing times. **A** shows the spectra for mixing times at 10, 50, and 200 ms (chemical shifts and structure from [12], shifts available at <http://bmrbb.wisc.edu/>, entry ID 4663, structure available at <http://www.rcsb.org>, PDB ID 1c3t). **B** expands regions and plots short-range contacts (black circles, same residue or 1 residue away) and medium range contacts (magenta x's, 2-4 residues away). Note that in all spectra, a real NOESY spectrum would have a significant water peak at around 4.8 ppm, running vertically through the spectrum, so that some peaks around this value cannot be seen in the direct dimension (see Figure 7E)

- When determining t_{mix} , one has to consider what information can be obtained from a given spectrum. For example, if a short mixing time is used, then one sees only the nearest distances. The problem, then, is that most of the cross-peaks result from nuclei that are nearby neighbors in the protein sequence.

Table 2: Short distances found in protein structures

Nuclei	α -helix	β_{10} helix	β -sheet (antipar.)	β -sheet (parallel)	Type-1* turn	Type-2* turn
$H_i^\alpha - H_i^N$	2.7	2.7	2.8	2.8	2.8/2.8	2.7/2.2
$H_i^\alpha - H_i^\beta$	2.2-2.9	2.2-2.8	2.2-2.9	2.2-2.9	2.2-2.9	2.2-2.9
$H_i^\beta - H_i^N$	2.0-3.4	2.0-3.4	2.4-3.7	2.6-3.8	2.0-3.5	2.0-3.4/3.2-4.0
$H_i^\alpha - H_{i+1}^N$	3.5	3.4	2.2	2.2	3.4/3.2	2.2/3.2
$H_i^N - H_{i+1}^N$	2.8	2.6	4.3	4.2	2.6/2.4	4.5/2.4
$H_i^\beta - H_{i+1}^N$	2.5-3.8	2.9-3.0	3.2-4.2	3.7-4.4	2.9-4.1/3.6-4.4	3.6/4.4
$H_i^\alpha - H_{i+2}^N$	4.4	3.8	—	—	3.6	3.3
$H_i^N - H_{i+2}^N$	4.2	4.1	—	—	3.8	4.3
$H_i^\alpha - H_{i+3}^N$	3.4	3.3	—	—	3.1-4.2	3.8-4.7
$H_i^\alpha - H_{i+3}^N$	2.5-4.4	3.1-5.1	—	—	—	—
$H_i^\alpha - H_{i+4}^N$ †	4.2	—	—	—	—	—
$H_i^\alpha - H_j^N$ †	—	—	3.2	3.0	—	—
$H_i^N - H_j^N$ †	—	—	3.3	4.0	—	—
$H_i^\alpha - H_j^\alpha$ †	—	—	2.3	4.8	—	—

all distances in Å

* includes both type I or II and I' or II'

† interstrand distances, so i and j are variable.

- For example, in Figure 8B, we zoom in on a region of the 50 ms spectrum and mark all peaks resulting from contacts within the same residue or one residue away with a black dot. Contacts resulting from nuclei that are 2-4 residues away are marked with a magenta 'x'. Almost all the peaks in this region are accounted for by these "short" and "medium"-range contacts.
- We learn very little from short-range contacts; we already know that these nuclei are nearby in space, so the appearance of cross-peaks in the spectrum is expected. The pattern of medium-range (2-4 residues) contacts, however, depends on the secondary structure. For example, in an α -helix, protons several residues away come closer than in a β -sheet, whereas interstrand contacts may occur in a β -sheet (see Table 2). Then, medium-range contacts can provide further information on secondary structure.
- One must mix sufficiently long to obtain long-range contacts. In Figure 8B, the short- and medium-range contacts are also indicated for the 200 ms mixing time, but we see many new peaks that do not correspond to these contacts. These additional contacts are long-range contacts, and provide information on the tertiary structure of the protein.
- Although we obtain more long-range contacts for the 200 ms mixing time, we should be careful not to mix for too long:
 1. The total signal decays during mixing, so we have poorer quality spectra for the longest mixing times
 2. Relayed transfers can occur, that is, of a set of three nuclei, 1 and 2 may be nearby, and 2 and 3 may be nearby, but 1 and 3 are far away. But, a

- transfer from 1-2, and 2-3 then appears in the spectrum as if 1 and 3 are close by, potentially confusing a subsequent structure determination.
- 3. If the measurable distance becomes too long, each peak actually provides less specific information. For example, if, from our spectrum, we know that two peaks are $<3 \text{ \AA}$ away (we might assume all cross-peaks come from contacts less than 3\AA if we use a relatively short mixing time), then we have a pretty good idea where the nuclei are relative to each other. On the other hand, if we use a longer mixing time, we may only know that their distance is $<5 \text{ \AA}$, such that there is a larger configurational space satisfying the distance restraint (although we will probably have more restraints in this case).
 - Solid-state NMR is also often used for structure determination. Because samples are not in solution, the dipolar coupling is not averaged by tumbling. Then, various pulse sequences exist to use the dipolar coupling directly for magnetization transfer (without NOE). Because solid-state NMR is lower resolution, one usually uses ^{13}C – ^{13}C correlation, instead of ^1H – ^1H , which results in narrower linewidths.
 - Common solid-state experiments are Proton-Driven Spin-Diffusion (PDSD), Dipolar-Assisted Rotary Resonance (DARR), Proton-Assisted Insensitive Nuclei cross-polarization (PAIN), and Proton Spin Diffusion (PSD, usually ^{13}C detected).

2.3 Structure determination software

As we can see in Figure 8, there is an enormous amount of information in an NOE spectrum. For example, in Figure 8A, $t_{mix}=200 \text{ ms}$, every peak provides some information, but there are about 10,000 peaks in the spectrum with amplitude at least 1% of the spectrum max. Many of those peaks are short-range, providing little structural information. Some are also medium range, providing primarily information on secondary structure. Finally, the most useful peaks, long-range peaks provide the critical information required to determine tertiary (and in some cases quaternary) structure. On top of the sheer number of resonances, it is often the case that the resonance frequencies are not known precisely enough to assign a 2D NOESY spectrum with enough certainty to obtain a structure. The solution to deal with this much information is structure determination software.

- A few common NMR structure determination programs are: ARIA [13], AutoStructure [14], CANDID [15], CYANA [16], Unio [17], and XPLOR-NIH [18].
- Typical input to NMR structure calculation are:
 1. A list of peaks from one or more 2D spectra (or higher dimensional spectra)
 2. A list of resonance assignments
 3. Torsional angle restraints (ex. TALOS output, optional)
 4. Vicinal J-couplings (H^N – $\text{H}\alpha$ couplings, optional)
- Note, the peaks themselves in the 2D spectra are not initially assigned; assignments are determined during structure calculation.
- The basic algorithm for structure determination from NOESY spectra (a more complete algorithm is given for the CANDID software in Figure 9):

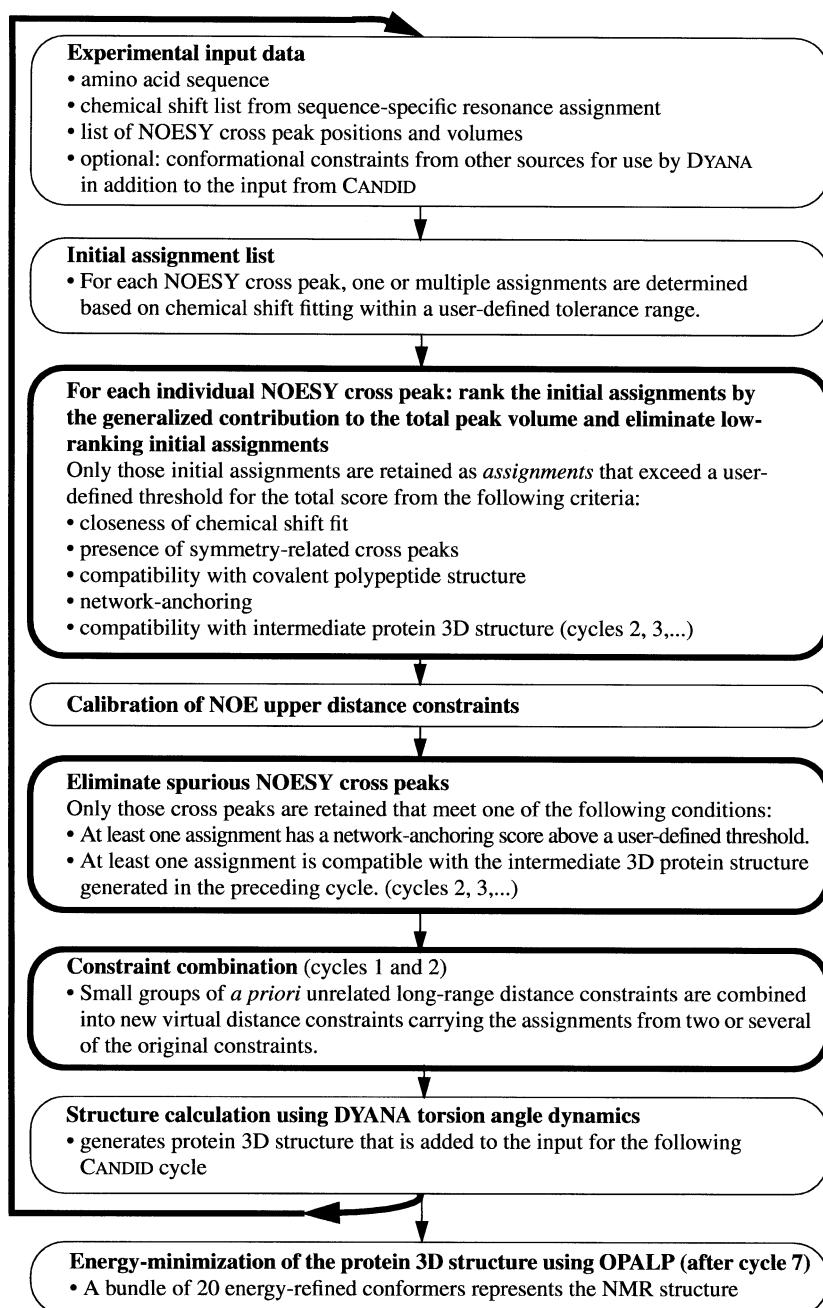


Figure 9: Flow-chart for the CANDID structure determination software (taken from Figure 1 in reference [15]).

1. Input chemical shift list (obtained via assignment, week 2), and NOESY peak list (extracted from NOESY spectrum). Also input additional structural restraints (J-couplings, torsional angles).
2. Perform initial assignment of peaks. Assignment of a peak depends on similarity of the peak's resonance frequencies and resonance frequencies in the chemical shift list. If an initial structure is available, software also estimates assignment likelihood for that structure (for example, a cross peak between two nuclei that are greater than 5 Å away is considered unlikely). Other considerations are also used for determining quality of a peak assignment (details in Figure 9).
3. Eliminate spurious peaks: require peaks to either be consistent with the current structure, OR be consistent with a network. Example: if a peak

connects nuclei a and b, then there should be other connections between nuclei in the two residues containing a and b, or in the residues immediately neighboring a and b (this stabilizes assignment performance).

4. Determine distance restraints from the current assignment.
 5. Calculate a structure from the distance restraints.
 6. Go back to step 2. Use the structure determined in 5 to perform the NOESY assignment and to eliminate spurious peaks. Repeat steps 2-6.
 7. After several iterations, finish by performing an energy minimization of structure.
- Then, a peak may not necessarily be assigned to the closest chemical shift in the list. Its assignment must also be consistent with the final structure. At the same time, the final structure relies on the assignment. Therefore, this process is iterative, in order to determine a consistent structure and assignment.
 - Algorithms usually add an element of randomness to this process, for example, by inputting a randomly generated initial structure. By varying the input, one can determine the stability of the output. This results in a "bundle" of structures, which is representative of how certain the resulting structure is (one may also see where in the protein the structure is less determined). For example, in Figure 10, 20 structures of ubiquitin are obtained from NOESY spectra. One sees that the C-terminus has considerable variation in its structure, so that a high degree of uncertainty exists for the structure in this region.

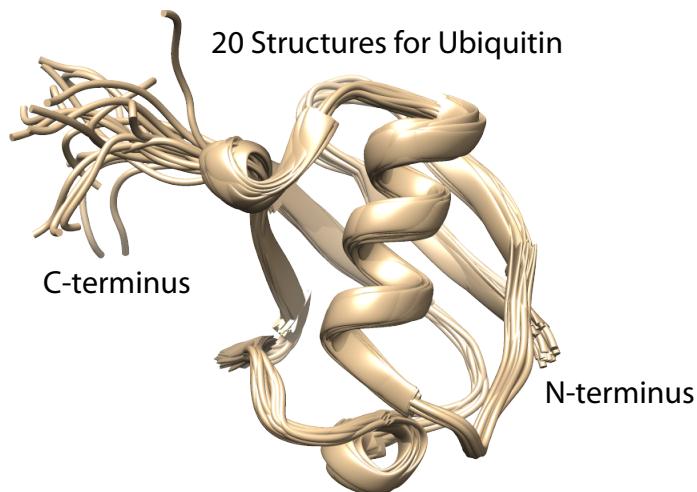


Figure 10: Bundle of 20 ubiquitin structures, from PDB ID 1c3t [12].

3 Dynamics Measurement with NMR

Obtaining protein structures is not the only function of NMR for protein studies. One may also determine the site-specific dynamics. One has a number of experiments useful for the characterization of amplitudes and rates of motion in a protein, including measurement of residual-dipole couplings, measurement of re-orientational relaxation, and measurement of chemical-exchange driven relaxation.

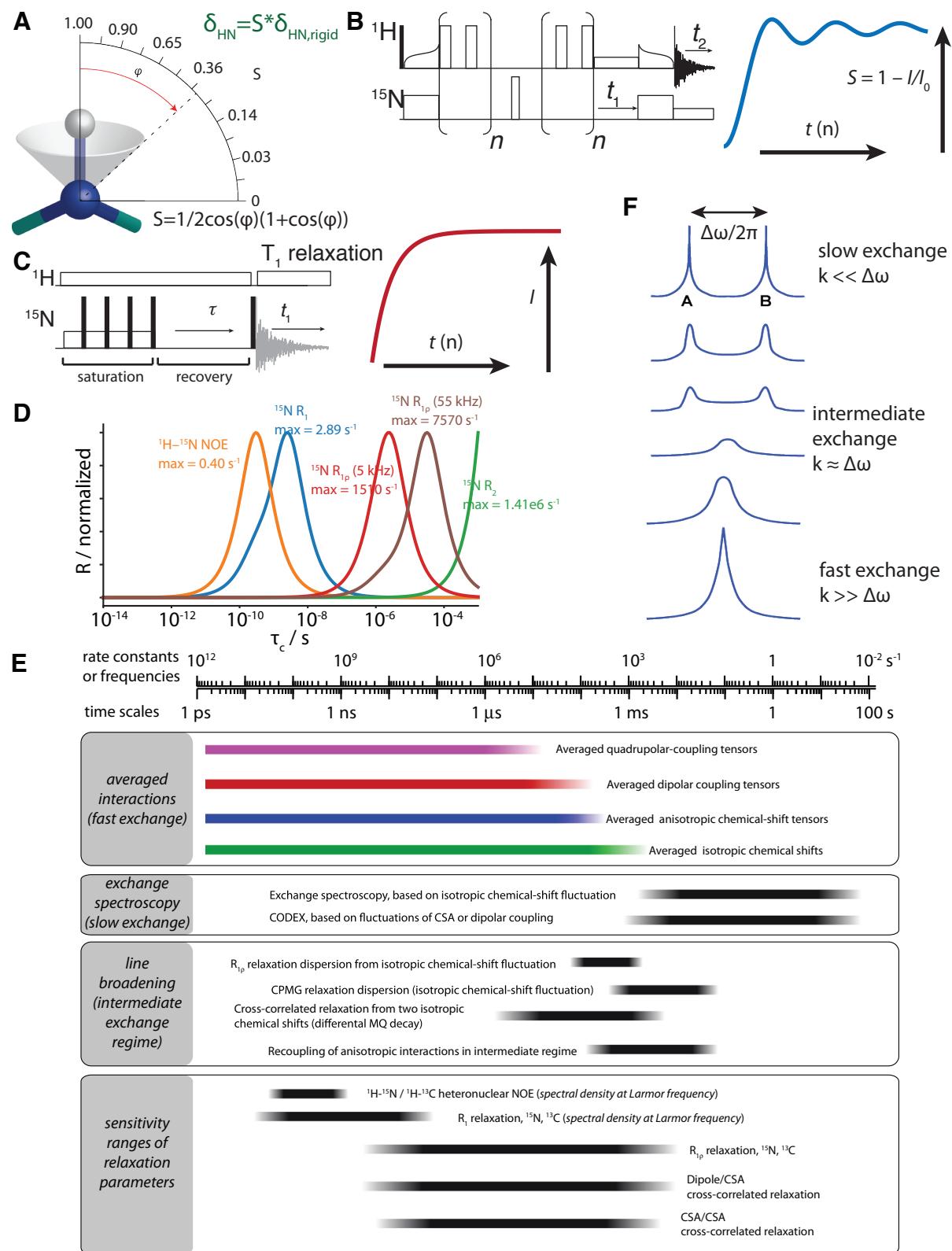


Figure 11: Dynamics measurements in NMR. **A** illustrates motion of a bond within a cone, where S , the order parameter depends on the opening angle of the cone. **B** shows the REDOR pulse sequence, and plots a typical REDOR curve, which can be fit to extract dipolar couplings. **C** shows a pulse sequence to measure $^{15}\text{N} T_1$ relaxation, via saturation-recovery and plots a T_1 recovery curve. **D** plots the normalized sensitivities of several ^{15}N relaxation experiments as a function of correlation time ($B_0=14.1$ Tesla). **E** shows various NMR-based dynamics measurements and the range of correlation times to which they are sensitive (plot from Schanda and Ernst [19]). **F** plots the lineshapes obtained from two-site chemical-exchange in the slow, intermediate, and fast exchange regimes.

3.1 Residual dipole couplings

The magnitude of a dipole coupling between two nuclei depends only on the types of nuclei, and the distance between them (an orientational dependence also exists). Then, for one-bond dipole couplings, we know the bond lengths and therefore can calculate the dipole coupling. However, if that bond moves within the protein, this results in orientational averaging and a reduction of the apparent dipole coupling strength.

- The reduction of dipole coupling is usually characterized with an order parameter, S . For example, if we measure dipole couplings of the backbone ^1H - ^{15}N dipolar coupling, we know the H-N bond length is 1.02 Å, and calculate the dipole coupling, $\delta_{\text{HN,rigid}} = 22.9$ kHz. Then, for an experimentally measured dipole coupling (referred to as the residual dipole coupling):

$$S = \frac{\delta_{\text{HN,exper.}}}{\delta_{\text{HN,rigid}}} \quad (8)$$

- S of 1 indicates an immobile bond, whereas lower values indicate increasing amplitude of motion. For specific models of motion, one can relate S to a particular amplitude of motion. For example, if a bond moves randomly within a cone (Figure 11A), then the opening of that cone is related to S , as

$$S = \frac{1}{2} \cos \phi (1 + \cos \phi) \quad (9)$$

- In solid-state NMR, one-bond dipole couplings may be measured using heteronuclear recoupling experiments, for example the DIPSHIFT experiment or the REDOR experiment. Figure 11B shows the REDOR pulse sequence, and the shape of a typical REDOR curve (this shape comes from an averaging of signals coming from dipolar couplings with the same magnitude but different orientations. Recall, $\omega_{IS} \propto \frac{1}{r^6} (3 \cos^2 \theta - 1)$).
- In solution-state NMR, residual-dipole couplings cannot be directly observed due to tumbling in solution (the dynamics of tumbling cover up the internal dynamics of the molecule). However, alignment media can be used to prevent isotropic tumbling, in order to extract S [20, 21].
- Measurement of quadrupolar couplings may also be used to obtain order parameters ($S = \delta_{\text{QC,exper.}}/\delta_{\text{QC,rigid}}$), for example, using deuterium NMR (spin-1).
- Residual couplings measurements are sensitive to motions with correlation times of up to ~ 100 μs , as shown in 11F (top).

3.2 Reorientation-driven relaxation

Reorientation of a given bond also re-orientates the dipole coupling between the two nuclei connected by that bond. This motion is "random", although can be described with amplitudes and correlation times. The correlation time can be thought of as the mean time between events for a motion (for a motion that "hops" between two orientations).

- These random motions occur at different times for different copies of the same molecule in a sample. Then, the trajectory of the magnetization for each copy of the same molecule takes a different path. This eventually destroys the net

magnetization in the sample (or causes it to return to thermal equilibrium). This is the origin of relaxation in NMR (ex. T_1 or T_2 decay. H–H NOESY is also a relaxation process).

- Most relaxation experiments are recorded by bringing magnetization away from thermal equilibrium, and observing its recovery. For example, the T_1 saturation-recovery experiment (Figure 11C) works by destroying (saturating) magnetization, often of ^{15}N or ^{13}C nuclei in proteins, and then varying a delay, τ , during which that magnetization recovers, as illustrated on the right of Figure 11C.
- Relaxation experiments are sensitive to correlation time (τ_c). This means that relaxation is faster if the inverse correlation time of motion matches a frequency present in the NMR spin system (ex. the nuclear Larmor frequency of one of the spins). See `wk3_videos.ppt` for a demonstration of the influence of correlation time sensitivity on the rate of a relaxation process.

$$\omega\tau_c \approx 1 \quad (10)$$

- Then, different types of experiments (and experiments performed at different B_0 fields) are sensitive to different correlation times. Normalized sensitivities of several ^{15}N relaxation experiments are shown in 11D. One may combine different relaxation experiments to be able to distinguish different timescales of motion.
- A set of relaxation experiments is often fitted to one or more correlation times and order parameters. This is known as the model-free approach [22], where internal motion of a bond in a protein is assumed to have a correlation function of the following form (in this case, the correlation function describes random re-orientation motions of a bond):

$$C(t) = S^2 + (1 - S^2) \exp(-t/\tau_c) \quad (11)$$

- We are also developing alternative approaches to model-free dynamics analysis in Leipzig, to improve joint-analysis of NMR data combined with other methods (esp. molecular dynamics simulation [23]).

3.3 Chemical exchange

Chemical exchange also relies on random changes of NMR interactions. However, in this case, random structural (or chemical) changes in the molecule result in a temporary change in the chemical shift itself ($\Delta\omega$).

- If the difference in chemical shift ($\Delta\omega$, expressed in radians/second, *not* ppm) between two states is very large compared to the rate of change ($k_{\text{ex}} \ll \Delta\omega$), we are in the slow-exchange limit (Figure 11F). In this case, one obtains a spectrum with two well-defined peaks (Figure 11F, top).
- On the other hand, if the chemical shift changes very rapidly ($k_{\text{ex}} \gg \Delta\omega$), the two peaks converge into one, and only the average chemical shift is obtained (Figure 11F, bottom).
- In between these extremes, we can see a gradual coalescence of the two peaks. Then, we can extract the difference of the chemical shift between the two states, *and* the rate of exchange, k_{ex} . $\Delta\omega$ is related to the degree of structural difference between the two states.

- The simple 1D spectrum is not always sufficient to extract $\Delta\omega$ and k_{ex} , in which case Carr-Purcell Meiboom-Gill (CPMG) experiments are used. These are trains of pulses, where the time between pulses may be varied. This variation can be used to separate two or more states in exchange from each other [24]. Another experiment for measuring chemical exchange is the Chemical Exchange Saturation Transfer (CEST) experiment [25].

4 Conclusions

This brings us to the end of the NMR-module. We see that we may use magnetic resonance to manipulate spins. Via assignment, we may associate particular chemical shifts with specific locations in a protein. In combination with the assignment, one may then use NOESY or other distance-dependent correlation experiments to determine protein structure. One may also use measurement of dipolar couplings or relaxation experiments to characterize protein dynamics.

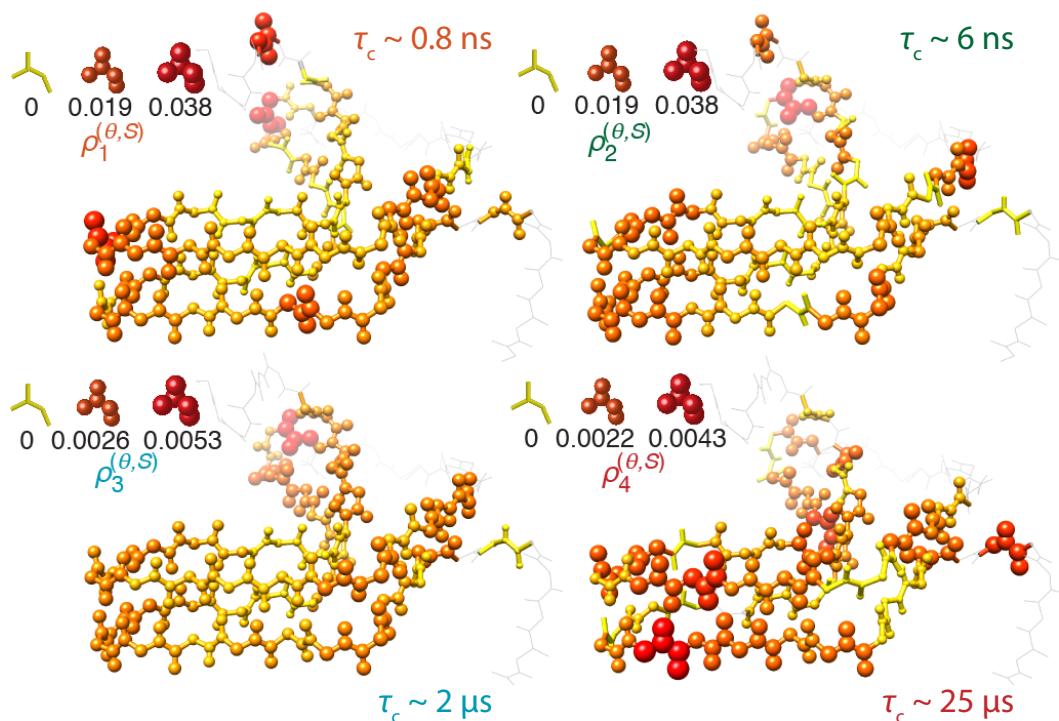


Figure 12: NMR-determined structure of the HET-s(218-289) fibril layer [26], where color and atom size encode the amplitude of dynamics measured by NMR relaxation for each backbone H-N bond [23].

References

- [1] A. L. Morris, M. W. MacArthur, E. G. Hutchinson, and J. M. Thornton, "Stereocchemical quality of protein structure coordinates," *Proteins: Structure, Function, and Bioinformatics*, vol. 12, no. 4, pp. 345–364, 1992.
- [2] J. Pontius, J. Richelle, and S. J. Wodak, "Deviations from standard atomic volumes as a quality measure for protein crystal structures," *Journal of Molecular Biology*, vol. 264, no. 1, pp. 121 – 136, 1996.
- [3] R. Ramage, J. Green, T. W. Muir, O. M. Ogunjobi, S. Love, and K. Shaw, "Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin," *Biochemical Journal*, vol. 299, pp. 151–158, 04 1994.
- [4] M. Karplus, "Vicinal proton coupling in nuclear magnetic resonance," *Journal of the American Chemical Society*, vol. 85, pp. 2870–2871, 09 1963.
- [5] G. W. Vuister and A. Bax, "Quantitative j correlation: a new approach for measuring homonuclear three-bond $j(\text{hnh}.\alpha.)$ coupling constants in ^{15}N -enriched proteins," *Journal of the American Chemical Society*, vol. 115, pp. 7772–7777, 08 1993.
- [6] V. A. Jaravine, A. V. Zhuravleva, P. Permi, I. Ibraghimov, and V. Y. Orekhov, "Hyperdimensional nmr spectroscopy with nonlinear sampling," *Journal of the American Chemical Society*, vol. 130, pp. 3927–3936, 03 2008.
- [7] Y. Shen, F. Delaglio, G. Cornilescu, and A. Bax, "Talos+: a hybrid method for predicting protein backbone torsion angles from nmr chemical shifts," *Journal of Biomolecular NMR*, vol. 44, no. 4, pp. 213–223, 2009.
- [8] Y. Shen and A. Bax, "Protein backbone and sidechain torsion angles predicted from nmr chemical shifts using artificial neural networks," *Journal of Biomolecular NMR*, vol. 56, no. 3, pp. 227–241, 2013.
- [9] I. Solomon, "Relaxation processes in a system of two spins," *Phys. Rev.*, vol. 99, pp. 559–565, Jul 1955.
- [10] A. W. Overhauser, "Polarization of nuclei in metals," *Phys. Rev.*, vol. 92, pp. 411–415, Oct 1953.
- [11] T. R. Carver and C. P. Slichter, "Polarization of nuclear spins in metals," *Phys. Rev.*, vol. 92, pp. 212–213, Oct 1953.
- [12] G. A. Lazar, T. M. Handel, E. C. Johnson, and J. R. Desjarlais, "Rotamer strain as a determinant of protein structural specificity," *Protein Science*, vol. 8, no. 12, pp. 2598–2610, 1999.
- [13] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T. E. Malliavin, and M. Nilges, "ARIA2: Automated NOE assignment and data integration in NMR structure calculation," *Bioinformatics*, vol. 23, pp. 381–382, 11 2006.
- [14] Y. J. Huang, R. Tejero, R. Powers, and G. T. Montelione, "A topology-constrained distance network algorithm for protein structure determination from noesy data," *Proteins: Structure, Function, and Bioinformatics*, vol. 62, no. 3, pp. 587–603, 2006.

- [15] T. Herrmann, P. Güntert, and K. Wüthrich, "Protein nmr structure determination with automated noe assignment using the new software candid and the torsion angle dynamics algorithm dyana," *Journal of Molecular Biology*, vol. 319, no. 1, pp. 209 – 227, 2002.
- [16] P. Güntert and L. Buchner, "Combined automated noe assignment and structure calculation with cyana," *Journal of Biomolecular NMR*, vol. 62, no. 4, pp. 453–471, 2015.
- [17] P. Serrano, B. Pedrini, B. Mohanty, M. Geralt, T. Herrmann, and K. Wüthrich, "The j-unio protocol for automated protein structure determination by nmr in solution," *Journal of Biomolecular NMR*, vol. 53, no. 4, pp. 341–354, 2012.
- [18] C. D. Schwieters, G. A. Bermejo, and G. M. Clore, "Xplor-nih for molecular structure determination from nmr and other data sources," *Protein Science*, vol. 27, no. 1, pp. 26–40, 2018.
- [19] P. Schanda and M. Ernst, "Studying dynamics by magic-angle spinning solid-state nmr spectroscopy: Principles and applications to biomolecules," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 96, pp. 1 – 46, 2016.
- [20] C. M. Thiele, "Residual dipolar couplings (rdcs) in organic structure determination," *European Journal of Organic Chemistry*, vol. 2008, no. 34, pp. 5673–5685, 2008.
- [21] N. A. Lakomek, C. Farès, S. Becker, T. Carlomagno, J. Meiler, and C. Griesinger, "Side-chain orientation and hydrogen-bonding imprint supra- τ_C motion on the protein backbone of ubiquitin," *Angewandte Chemie International Edition*, vol. 44, no. 47, pp. 7776–7778, 2005.
- [22] G. Lipari and A. Szabo, "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity," *Journal of the American Chemical Society*, vol. 104, pp. 4546–4559, 08 1982.
- [23] A. A. Smith, M. Ernst, S. Riniker, and B. H. Meier, "Localized and collective motions in het-s(218-289) fibrils from combined nmr relaxation and md simulation," *Angewandte Chemie International Edition*, vol. 58, no. 28, pp. 9383–9388, 2019.
- [24] A. L. Hansen, P. Lundström, A. Velyvis, and L. E. Kay, "Quantifying millisecond exchange dynamics in proteins by cpmg relaxation dispersion nmr using side-chain 1h probes," *Journal of the American Chemical Society*, vol. 134, pp. 3178–3189, 02 2012.
- [25] D. F. Hansen, P. Vallurupalli, P. Lundström, P. Neudecker, and L. E. Kay, "Probing chemical shifts of invisible states of proteins with relaxation dispersion nmr spectroscopy: How well can we do?," *Journal of the American Chemical Society*, vol. 130, no. 8, pp. 2667–2675, 2008. PMID: 18237174.
- [26] C. Wasmer, A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek, and B. H. Meier, "Amyloid fibrils of the het-s(218–289) prion form a β solenoid with a triangular hydrophobic core," *Science*, vol. 319, no. 5869, pp. 1523–1526, 2008.