Finding Bonds, H-bonds…

• A hydrogen bond (HB) allows chunks of peptide relatively far away from each other to come close together. They are all over the place in globular proteins, so if we could identify were they are (donor and acceptor atoms), we have a huge constraint in the structure.

• In a protein the most interesting HBs are those formed between the peptide backbone amide protons and carbonyls, as the ones we see in $\alpha$-helixes and $\beta$-sheets. We can also see some from side the chains (Asn, Asp, Gln, Glu) to the backbone amides or carbonyls:

• To find them we study the exchange rates of amide protons. The idea is that labile protons (NHs) that are protected from the solvent will not exchange as fast with solvent protons as things that are solvent exposed.
Amide exchange rates

• Therefore, if we add D$_2$O to our H$_2$O solution and take spectra at different times, we'll see that signals from different amide protons will decrease in size at different rates.

• Since the amide region of a 1D is way too crowded in proteins, we normally use a quick 2D experiment, as a DQF-COSY. We look at the NH to H$_\alpha$ fingerprint at different times.
Amide exchange rates

• From this data we can tell which amide is H-bonded strongly, which one weakly, and which ones not at all. Since we also have NOE and $^3J$ coupling data, we can try to see if these hydrogen bonded amides match with regions that we identified previously as $\alpha$-helices, $\beta$-sheets, or $\beta$-turns.

• If we can do this, then, and ONLY then, we can use a H-bond constraint during the generation of our 3D model.

• Why the ONLY? We only now the H-bond donor, but there is (or there was until a while ago) no way we can tell who the acceptor atom is (the C=O oxygen). If we miss-place one of these we screw up big time. Since we are basically cyclizing the peptide, there is no way we can get the right structure.

• If we decide that it’s reasonable to use a H-bonding energy penalty, we can put it into the force field more or less as a distance constraint:

$$E_{\text{HB}} = K_{\text{HB}} \cdot (r_i - r_{\text{HB-ideal}})^2$$

• $r_{\text{HB-ideal}}$ is ~2.5 Å (depending on the reference). Since H-bonds also have an angular requirement (the N-H\cdots O angle has to be between 135 and -135), we can make a more complex term to reflect this.
Amide temperature gradients

• Studying exchange rates works OK in proteins, because the time in which the amides turnover is long (globular). In small peptides this ain’t true.

• Since we have a lot more flexibility in a peptide (a lot more contact with solvent), everything usually exchanges in relatively short times (minutes as opposed to hours). By the time you put some D₂O in the tube, brought it to the NMR lab, placed it in the magnet, and shimmed the sample, there are no amide protons…

• For peptides, instead of studying the exchange rates, we analyze the change in chemical shift of the amide protons with change in sample temperature (temperature gradients).

• This is because the more the proton is exposed, the more it’ll interact with solvent as we increase temperature, moving it upfield towards water…

• We measure T gradients in parts per billion (ppb). Values below -2 indicate shielding from the solvent (water), between 4 and 5 ppb we have partial shielding (some of the time we have a H-bond, some of the time we don’t), and above 6 or 7 ppb indicates complete exposure.

• As with proteins, we cannot tell which one is the acceptor (the oxygen). Therefore, we have to be really really careful using these data…
An example of amide temperature gradients

- For the peptide **Ala-Arg-Pro-Tyr-Asn-Aic-Cpa-Leu-NH$_2$**: 

- **Leu** NH is partially H-bonded (shielded from solvent)...
Using ERs and TGs

• Knowing that you have a H-bond and not being able to use it as a constraint in the model is painful.

• If we want to be safe, we can just do the whole calculation of structures with NOEs and $^3$J$s$ as we saw last time, and then discard structures in which the NH $^1$H we know is H-bonded does not appear H-bonded (use it as a check).

• The other way is to have some other sources to corroborate that the H-bond exists (NOEs and couplings). This works better in proteins because we have sizable $\alpha$-helices and $\beta$-sheets. In peptides we may have a $\beta$-turn, which is very tiny, and may not have decent NOEs and $^3$J couplings.

• Or, we may do it the hard way - If we have 3 or 4 possible H-bond acceptors, we can try each one of them in different simulations and see at the end which one gives us the lowest energy structures:
Isotopic labeling

• The only nuclei that we can look in a protein are usually the $^1$H. In small proteins (up to 10 KDa, ~80 amino acids) this is OK. We can identify all residues and study all NOEs, and measure most of the $^3J$ couplings.

• As we go to larger proteins (>10 KDa), things start getting more and more crowded. We start losing too many residues to overlap, and we cannot assign the whole backbone chain.

• What we need is more NMR sensitive nuclei in the sample. That way, we can edit the spectra by looking at those, or, for example, add a third (and maybe fourth) dimension.

• To do this we need several things:

a) We need to know the gene (DNA chunk) that is responsible for the synthesis of our proteins.

b) We need a molecular biologist to get a plasmid that overexpresses this gene, possibly in an *E. coli* vector system in minimal media, so that even Joe Blow, the new (and clueless) undergrad in the lab, can grow lots of it.

c) The overexpressed protein has to be functional (85% of the time you get a beautiful band in the SDS-gel that is all inclusion bodies…).

d) A good/simple purification procedure.
Isotopic labeling (continued)

• 10 to 1 that your particular protein will fail one of these requirements in real life. But most of the time, we can work around either overexpression, activity and purification problems. Getting the gene is the toughest one to overcome.

• In any case, now that we have the plasmid, we grow it in isotopically enriched media. This usually means M9 (minimal media), which only has NH$_4$Ac and glucose as sources of N and C. No cell homogenates or yeast extracts.

• So, if we want a $^{15}$N labeled protein, we use $^{15}$NH$_4$Ac (that is dirt-cheap). Glucose-U-$^{13}$C is a lot more expensive, but it is sometimes necessary.

• In that way we get partially- or fully-labeled protein, in which all nuclei are NMR-sensitive ($^{13}$C=O, $^{13}$C$\alpha$, and $^{15}$Ns). All the protein backbone is NMR-sensitive.

![Diagram of amino acid with isotopic labels](image)

• Another cool thing is that we now have new $^3$$J$-couplings to use for dihedral angles: H-C-C-$^{15}$N, H-N-C-$^{13}$C. These have their own Karplus parameters.
Isotopic labeling (…)

• One of the most common experiments performed in $^{15}$N-labeled proteins is a $^{15}$N-$^1$H hetero-correlation. Instead of doing the normal HETCOR which detects $^{15}$N (low sensitivity), we do an HSQC or HMQC, which gives us the same data but using $^1$H for detection.

• This experiment is great, because we can spread the signals using the chemical shift range of $^{15}$N:

![Diagram showing spread of signals using $^{15}$N-$^1$H correlation]

• It is ideal for several things. One of them is measurement of amide exchange rates.

• It is also good to do spin system identification: If we don’t have good resolution in the COSY or TOCSY and some signals are overlapped, we can use the $^1$H-$^{15}$N correlation to spread the TOCSY correlations in a third dimension…
3D NMR spectroscopy

• …which brings us to 3D spectroscopy. There is nothing to be afraid of. The principles behind 3D NMR are the same as those behind 2D NMR.

• Basically we can think of them this way: In the same fashion that an evolution time \( t_1 \) gave us the second dimension \( f_1 \), we can add another evolution time (which will be in the end \( t_1 \)), and obtain a third frequency axis after some sort of math transformation.

• For a 2D we had:

\[
\begin{align*}
\text{Preparation} & \quad \text{Evolution} \quad t_1 & \quad \text{Mixing} & \quad \text{Acquisition} \quad t_2 \\
& \quad f_1 & \quad f_2 \\
\end{align*}
\]

• For a 3D we will have:

\[
\begin{align*}
\text{Preparation} & \quad \text{Evolution} \quad t_1 & \quad \text{Mixing} \quad (1) & \quad \text{Evolution} \quad t_2 & \quad \text{Mixing} \quad 2 & \quad \text{Acquisition} \quad t_3 \\
& \quad f_1 & \quad f_2 & \quad f_3 \\
\end{align*}
\]

• As in the 2D experiments, depending on the type of mixing we use in each chunk, the type of data we’ll get…
3D NMR spectroscopy (continued)

- We will not try to go pulse by pulse seeing how they work, but just mention (and write down) some of the sequences, and understand how they are analyzed.

- We first have to separate into different categories depending on the type of mixing:

  - **3D separation spectra**: We take a spin system and separate different parameters (chemical shift, couplings) in different dimensions. A conceivable example would be a 3D version of an HOMO2DJ experiment. They are not used that frequently, at least for proteins/DNA.

  - **3D transfer spectra**: In these ones we have some sort of transfer process, such as scalar J-couplings or NOE enhancements, for passing information between the different dimensions. They are an extension of the 2D experiments we have seen, and the most used ones.

- The way we build them is nothing else that putting two 2D experiments one after the other (at least we can understand them this way).

- According to this, we will have things like **TOCSY-HSQC**, **NOESY-HSQC**, **COSY-COSY**, **COSY-NOESY**, **TOCSY-NOESY**, etc., etc..

- We’ll see how one of them works…
3D NMR spectroscopy (...)

- Say that we could somehow selectively tickle only certain amide protons in the sample (we’ll see more on selective pulses today, but this is only an example). Only protons attached to this amide proton will give us cross-peaks:

- So, we do this selective amide excitation followed by a 2D TOCSY experiment. Our 2D plot will only have the line that corresponds to the amide proton we selected. For a **Leu**:

- If we changed the frequency of the selective pulse to another NH we would get another spin system and so forth.
3D NMR spectroscopy (…)

• Now, we could put all the 2D experiments stacked like if they were posters in a rack, and each slice would have the connectivities of a particular spin system:

![Resolved NH frequencies](image)

![Aliphatic H frequencies](image)

• This would be a ‘pseudo’ 3D experiment. The problem here is the way we do the selection of the NHs. Usually, we isolate each NH (or whatever we want to isolate) by doing a 2D experiment that resolves it.

• From what we’ve seen, an NH correlation would be good for this purpose, because most of the cross-peaks are well resolved.

• Additionally, in the way we do 3Ds we don’t usually collect all the 2Ds into a 3D, but get the 3D which has all the cross-peaks and then analyze the slices at different frequencies.
3D NMR spectroscopy (…)

• Furthermore, peaks cross-peaks appearing in the cube arise due to a transfer of polarization between the nuclei that we look at in the 3 dimensions.

• A 3D using a $^{15}\text{N}-^{1}\text{H}$ correlation and TOCSY combination will look like this (hope you like it - it took me forever…):

  ![Diagram](image)

  - We have each of the individual (hopefully) TOCSY cross-peaks in a single line, which starts in the $^{15}\text{N}-^{1}\text{H}$ cross-peak. The amide protons are separated by the $^{15}\text{N} \delta$s.

  - Looking at the cube is kind of hard when you have 200 amides. We usually take slices at different 15N frequencies.
3D NMR spectroscopy (…)

- Depending on the slice (plane) we chose, we’ll have TOCSY spectra corresponding to different NHs:
3D NMR spectroscopy (...)

• Some real data:

• This is an NOESY-HSQC of a dimeric GpA mutant (taken from bioc.rice.edu/~mev/spectra3.html)
TOCSY-HSQC pulse sequence

• The experiment that we used in this explanation is one of the most employed ones when doing 3D spectroscopy. The pulse sequence looks like this:

![Diagram of TOCSY-HSQC pulse sequence]

• Briefly, the first chunk is a TOCSY in $^1$H, in which we have to decouple either $^{13}$C or $^{15}$N (yes, that means saturating these nuclei). Here we have $t_1$ ($f_1$), meaning that we’ll have a $^1$H-TOCSY like spectrum in this dimension.

• The second part is an inversely detected HETCOR, in which we will have $^1$H frequencies in one axis ($f_1$), and $^{13}$C or $^{15}$N frequencies in $f_2$.

• Finally, detection in $t_3$ is in $^1$H, so we have $^1$H shifts and couplings in $f_3$. 
• In a similar way we can combine a NOESY with the HSQC. The sequence is this one below:

• Now instead of a TOCSY on the first part we have a NOESY-type deal. The second part is identical to the other one.

• In the 3D we will have a normal $^1$H-$^{15}$N (or $^{13}$C) correlation in the $f_1$-$f_2$ plane, and NOESY planes for the $f_1$-$f_3$ planes. Each of these NOESY planes will have only a few protons. If the resolution is good, each proton has its own $f_1$-$f_3$ plane.

• In the same way we can combine any other type of 2D experiment to obtain a 3D experiment. These two are the most used ones in proteins, which are the molecules for which 3Ds are used the most…
Selective pulses

• Many other 3D sequences are used to identify spin systems at the beginning of the assignment process. Most of these rely in some sort of selective excitation of part of the spin systems present in the peptide.

• For example, we may want to see what is linked to the $H_\alpha$ but not to the NHs. Also, upon labeling the peptide completely we may want to select how the transfer of magnetization goes through the peptide backbone.

• In order to do such a thing we need selective pulses, which we have mentioned before, but never described in detail.

• A non-selective pulse is very short and square, which in turn makes it affect frequencies to the side of the carrier due to the frequency components it has (we saw all this…). On the other hand, a selective pulse is a lot longer in time, which makes the range of frequencies a lot narrower:

![Time:](image1)

![Frequency:](image2)
Selective pulses

• A problem of using longer square pulses as selective pulses is that we still have the wobbles to the sides (remember the FT of a square pulse). Therefore we are still tickling more frequencies than what we would like.

• What we have to do is figure a pulse in the time domain that will almost exclusively affect only certain frequencies. This means that our pulse will have a certain intensity profile vs. time different from a square, and is therefore shaped.

• The quick-and-dirty way to obtain a shaped pulse is to see how it would look in the frequency domain (what frequencies we need it to affect), and then do a inverse Fourier Transform to the time domain (FT⁻¹):

• To do it properly we have to do a complete analysis of the Bloch equations. We usually use Gaussian and squared Gaussian shaped pulses. By changing both the shape, power, and length we can tune them to affect, for example, only the $H_\alpha$, $NH$, $^{13}C_\alpha$, or $^{13}C=O$ regions of the spectrum…
3D experiments using selective pulses

• …which brings us back to the use of selective pulses in 3D spectroscopy. Now we can fine tune what we want to see in our 3D even more. Instead of, say, exciting all $^{13}$C atoms in the sample, we can just flip $^{13}$C$_{\alpha}$ atoms of the backbone. In this way, we will only see polarization transfer processes which involve these type of carbons.

• This is really useful, because we can follow atoms along the peptide backbone in a selective way. The experiments are named according to how the polarization is followed. For example, we have the **HNCA** experiment:
3D using selective pulses (continued)

• Here we have a transfer of polarization from the \(^1\text{H}\) (to the \(^{15}\text{N}\) (we are both enhancing the \(^{15}\text{N}\) signal and obtaining a correlation between both nuclei), then passing it only to the \(^{13}\text{C}_\alpha\) atoms (we use selective \(\pi/2\) and \(\pi\) pulses).

• This is a transfer of polarization between the \(^{15}\text{N}\) and the \(^{13}\text{C}_\alpha\). Since the \(^{15}\text{N}\) had information on the \(^1\text{H}\) it was attached to, the \(^{13}\text{C}_\alpha\) will ‘know’ this too.

• In the end we will get a cross-peak at the \(^{15}\text{N}-^{1}\text{H}-^{13}\text{C}_\alpha\) frequency (a blob in space with those coordinates). Using other selective pulses we get other correlations. For example, we have \(\text{HCA(CO)N}\), which actually ‘jumps’ the \(^{13}\text{C}=\text{O}\):