Lecture 10. Assignment and Structure Determination in Proteins.

We have presented several experiments over the past few lectures, and haven't spent any time really discussing how they are used in chemical shift assignment and structure determination. We'll correct that error in this lecture. I'll break this up into three categories: 1. natural products, 2. peptides, and 3. proteins. Nucleic acids will be discussed separately, but brief comments about how nucleic acids would be treated differently than proteins will be made where appropriate.

10.1. Assignment and Structure Determination of Natural Products.

The objective of assigning natural (and non-natural) products is to obtain some sort of picture of the molecule you are studying. In most cases, an accurate three-dimensional structure will not be determined. Instead, you will be interested in determining a “two dimensional structure”, similar to what can be drawn on paper. In more sophisticated examples, we will need three dimensional information to shed light on a particular aspect of the molecules chemistry. First we'll cover chemical shift assignments, then briefly discuss “structure determination”. Many of the rules for chemical shift assignment in natural products are the same for amino and nucleic acids, and the common rules/strategies will only be covered here. As I am not an expert in natural products, much of what I say in the next few paragraphs is anecdotal, but since that hasn’t stopped me in the past, here goes.

Chemical shift assignment consists of three conceptual steps: 1. identifying unique spin systems; 2. assigning the “structure”, or topology of the individual spin systems (or networks); and 3. connecting the different spin systems together to identify the molecule. This is, in general, a non-trivial process, as you might have gathered from the exam problem you had to solve, and these steps are called conceptual because you generally work on steps 1 and 2 simultaneously. The first step, identifying individual spin systems, is based in part on identifying coupled nuclei through COSY-type experiments. The problem can be approached by finding “starting” points. Singlets, of course, aren't scalar coupled to anything (otherwise they wouldn't be singlets), and can be assigned by inspection. Assigning singlets doesn't help with identifying spin system topologies, but these resonances must be assigned eventually, and they make convenient starting positions. Other good starting points are up-field or down-field shifted resonances. These arise from $^1$Hs next to electron donating or withdrawing functional groups, respectively. If these resonances are well separated from the others, then their couplings and connectivities are easily identified. The scalar-coupled resonances are identified by:

1. identifying collapsed couplings in selective decoupling experiments, as was done in your exam,

2. identifying cross peaks in a 2D COSY experiment. Of course, other experiments can, and should, be used to confirm the correlations identified in the COSY (what might those be?), and

3. measuring the scalar coupling constants observed in the self-resolved resonances. This is because the J values must be symmetric; scalar coupling of $^1$Hs A and B gives the same interaction whether you look at nucleus A or B. A trivial example is shown in Figure 10.1; it is often not this simple to identify the couplings when the coupling patterns become complicated
due to multiple coupling partners. In the example shown, the couplings between resonances A and C, and B and D, are the same, demonstrating that nuclei A and C, and B and D, are scalar-coupled, but that A is not coupled to B or D, etc. Specific problems to be aware of in identifying spin coupling networks:

1. chemical shift degeneracy (overlap). Several $^1$Hs can have similar magnetic environments that contribute to “accidental” degeneracy; other $^1$Hs can be related by symmetry. This degeneracy can lead to fewer than expected resonances. Accidental degeneracy can generally be identified by complicated multiplet structures (different multiplets are overlapped) and by the use of relayed-COSY and TOCSY experiments because different correlation peaks will be identified. These overlapped resonances might not be assigned initially as overlapped; many overlapped peaks are identified only after you have enough information to say that it doesn’t make sense that this many cross peaks can be associated with a single $^1$H. Symmetry-related degeneracy can be distinguished by simplified multiplet structures, and by higher intensity (the resonance contains signals from more than one $^1$H). A rough estimate of the number of $^1$Hs involved in a single resonance can be obtained by integrating the spectrum. This is only approximate because of relaxation differences between the different nuclei.

2. odd local geometry. Scalar coupling constants are dependent upon the dihedral angle between the two $^1$Hs; it is largest for anti conformations; smallest for eclipsed. In some structures, there will be only weak coupling between two nuclei that are part of the same spin coupling network because of this. In extreme cases, the entire spin system won’t be identified without NOESY data. On the other hand, you can attempt to assign these $^1$Hs to the same spin network using heteronuclear correlations. Assume you’ve assigned the majority of $^1$H spins. If you perform an HMQC to correlate directly-bound $^1$H-$^{13}$C spin pairs, then you can run an HMBC experiment to identify correlations between $^1$H and other $^{13}$Cs. This experiment will be discussed again below.

3. Poly aromatic spin systems. These are particularly difficult to assign because they are especially poor in $^1$Hs, and the $^1$Hs that are present are often overlapped. In this case, there can be too many carbons between the $^1$Hs to use the HMQC / HMBC trick I just mentioned, and NOESY experiments don’t work because the $^1$Hs are too far away.

4. $^1$H chemical shifts are strongly affected by neighboring functional groups. For instance, the $\text{CH}_3$ resonance frequency shifts when bound to an oxygen or nitrogen; it also shifts when bound to sp$^3$ vs. sp$^2$ carbons. The same is true for $\text{CH}_2$ and CH groups. There are published tables that indicate commonly observed chemical shift ranges for $^1$Hs near important organic functional groups; these should be consulted when assigning natural products. However, I can

![Figure 10.1. Stick spectrum showing four sets of doublets.](attachment:figure101.png)
guarantee that just about each interesting system will have one or two $^1$Hs that don’t “fit” into these categories and will require more thought in assignment.

5. The types of coupling networks encountered are, generally, not known at the beginning of the assignment process. This is in contrast to proteins and nucleic acids, where the basic spin systems are known ahead of time.

Connecting assigned spin coupling networks in natural products can be difficult; many times the different spin networks are separated by several bonds lacking $^1$Hs, and / or by functional groups. There are two approaches to bridging these gaps: using NOE interactions, or using heteronuclear couplings. The NOE is limited to $^1$Hs that are about 5-7 Å apart (the exact distances will depend on the molecular tumbling rate, and the local “proton density”). Another way to correlate the different spin systems is to use HMQC / HMBC experiments. This can be accomplished in the following manner. Consider the spin system indicated in Figure 10.2, which indicates the $^1$H and $^{13}$C nuclei around the glycosidic bond in polysaccharides. From the COSY / DQF-COSY, and TOCSY experiments, the $^1$H spin systems of each individual sugar can be assigned; furthermore, the specific carbohydrate can be identified through scalar coupling constants between different $^1$Hs (all hexoses are epimers of each other). The problem is in connecting the individual sugar spin systems. This is done using a combination of HMQC and HMBC experiments. Recall that the HMQC is used to provide $^1$H-$^{13}$C correlations, and provides the C1 and C4 carbon chemical shifts. Once these are obtained, an HMBC experiment is performed that correlates H4 of sugar A with C1 of sugar B (arrows in the figure). In the same experiment, you get a correlation between H1 of sugar B and C4 of sugar A to confirm the assignments.

This particular example was chosen to represent a general solution to connecting spin systems separated by an ether linkage; similar approaches can be taken to correlate spin systems separated by a limited number of bonds; the limitation arises from the size of the $^1$H-X nucleus coupling constant relative to the $T_2$ relaxation time of the molecule.

To summarize, COSY-type experiments are used to identify and assign individual scalar coupling networks. The different networks are correlated using either NOE information or heteronuclear correlations. In general, assignment and structure determination in these molecules is complicated by the variety of spin systems that can be encountered, by the relative lack of 1Hs, and by an absence of a defined, three dimensional structure.


In this section, I will discuss assignment and structure determination of peptides and small proteins (< 100 amino acids). These methods discussed in this section are based on homonuclear, $^1$H-$^1$H correlations. The approach and limitations will be discussed. In the next
section, I’ll discuss specific methods for dealing with larger proteins.

Assignments of these molecules follow the same conceptualized steps identified for natural products; many of the steps are the same, some are different. Peptides and proteins offer some simplifications over natural products, and some additional complications. These are:

1. The individual spin systems are well-known, and the relative chemical shifts and coupling topologies are documented (Wüthrich’s book). This is also true of nucleic acids. The fact that the spin systems are known in proteins and nucleic acids represents a substantial simplification over natural products where the spin systems are not known a priori.

2. Peptides are generally flexible and don’t have a fixed secondary or tertiary structure in solution. Assignments are simplified in peptides because $T_2$ is generally long, which results in nearly maximal cross peak intensity. On the other hand, the flexibility increases the likelihood of accidental chemical shift overlap, e.g., all valines have similar chemical shifts for the $H_\alpha$, $H_\beta$ and $H_\gamma$’s. On the other hand, short DNA/RNA fragments can have a defined secondary and tertiary structure through intramolecular hydrogen bonding.

3. As peptides get longer, they tumble more slowly and adopt a more compact, well-defined structure. This has the following implications: as $T_2$ decreases, cross peak intensities decrease because the time needed to evolve antiphase states approaches $T_2$. This is true in nucleic acids, as well. Second, as the structure becomes more compact, there are many more unique chemical environments, which removes the accidental overlap problem, but complicates the spectral appearance.

In the following paragraphs, I’ll discuss the analysis of 2D spectra and assignment of amino acid spin systems. A complete and thorough discussion of this can be found in the book by Wüthrich. This book covers peptide and nucleic acid assignment strategies. It is a bit out-date in that there is no discussion of heteronuclear 3D methods, but it nevertheless is a solid book that belongs on the shelf of any serious biomolecular NMR spectroscopist.

**Peptide assignment strategy: Example.** I previously showed you TOCSY and NOESY spectra from a 17 residue peptide that we are studying (Figure 7.19 and 20). In general, the TOCSY and NOESY spectra can be broadly classified in two different “parts” - the amide part and the sidechain part. There are two approaches to assigning the spectrum: one is to start with the amide $^1$Hs; the other is to start with the sidechain $^1$Hs. There are pros and cons to either approach.

**Starting with the HN.** This method generally allows you to identify the individual amino acids with ease because each amino acid contains a single backbone amide $^1$H (the sidechain amides of glutamine and asparagine are easily identified and distinct from the backbone amides). Figure 10.X shows an expansion of the TOCSY spectrum shown in Figure 7.20, emphasizing the ability to identify discrete amino acid spin systems from the correlations detected between aliphatic sidechain and amide protons Using the standard chemical shift ranges and known coupling topologies for the different amino acids, the data in this TOCSY spectrum allowed us to 1. identify the individual spin systems (indicated by the lines drawn in the figure) and 2. assign them to amino acid type in a relatively short amount of time. In many cases (e.g., when a particular amino acid occurred once in the peptide), this information was sufficient to obtain the sequential assignment; in cases where there were two amino acids of the same kind, we couldn’t
tell specifically which amino acid it was.

In my opinion, this is the easiest and most direct method for assignment. The limitations to this approach, though, are that 1. only HN-Hα correlations are observed in a COSY, which doesn’t provide enough information to assign the amino acid type; 2. since we need to observed correlations with HN, we need to use H2O solvent, which presents severe problems with solvent suppression; and 3. if the HN chemical shift dispersion is not high, then there can be significant problems due to chemical shift overlap.

**Starting from the Sidechain 1H.** An alternative approach is to start from the sidechain resonances (see Figure 7.20). The advantages of this approach are that 1. there are no exchangeable resonances so the spectra can be run in D2O solvent and solvent suppression becomes trivial; 2. many correlations are used to assign the individual amino acids, so overlap is not as much of a problem; 3. the cross-peak patterns and chemical shifts of the different amino acids are well-known, and easily assigned in 2D spectra of peptides. The disadvantages of this approach are 1. certain regions of the spectrum exhibit significant chemical shift degeneracy (e.g., methyls); 2. it is sometimes hard to identify individual amino acids because of overlapping or missing correlations; 3. in D2O there are no correlations with backbone amides. This last point is important because the primary way to identify secondary structure involves backbone amide resonances.

In fact, you generally need to assign all cross peaks in every spectrum you collect. The approach I use is to 1. identify individual spin systems from the amides; 2. use the correlations from the HN protons to assign amino acid type where possible; 3. confirm and extend these assignments by assigning the sidechain correlations. Using this approach, overlapping resonances in one region of the spectrum can usually be worked around because there may be no (or at least different) overlap in the other region. Also, the assignments are obtained in H2O, which will facilitate assigning the NOE spectra used in structure characterization.

**Sequential assignments.** This approach doesn’t provide complete, sequence-specific chemical shift assignments; we’ve only gone through the first two conceptual assignment steps.
We still need to sequentially link the individual amino acids. This is invariably done using a 2D NOESY experiment (Figure 10.5). The amino acids are linked together in what's called a NOESY walk. The NOESY experiment, of course, contains correlations between $^1$Hs nearby in space. This always includes an intra-residue HN-Hα correlation (fixed by bonded geometry to be less than ~3.5 Å away, independent of secondary structure). The NOESY experiment also contains other HN-Hα correlations, and, in small peptides, these are generally "sequential" NOEs to the preceding amino acid residue (Figure 10.4, 10.5). Therefore, from one amide resonance, we get information on (at least) two spin systems: the intraresidue spin system, and the preceding spin system. In the absence of assignments, it would be nearly impossible to determine which resonances belong to the intra- versus sequential residues from the NOESY.

**Figure 10.4.** Schematic indicating NOEs involved in sequential assignments of peptides.

**Figure 10.5.** Fingerprint region of a NOESY spectrum indicating sequential assignments identified in a NOESY walk. The NOESY spectrum is shown in black; the TOCSY spectrum is shown in red. Assignments from Takahashi and Logan, unpublished.
spectrum alone. This distinction is easily made by comparing the data in the TOCSY and NOESY spectra because the TOCSY experiment only provides correlations with $^1$Hs of the intra-residue amino acid (Figure 10.5).

This assignment strategy (using TOCSY and NOESY spectra) provides unambiguous sidechain assignments, especially for aliphatic resonances, but there are two problematic groups of residues: aromatics and prolines. Aromatic sidechain $^1$Hs are difficult to assign because the aromatic and aliphatic protons are not scalar coupled. Aromatic resonances are assigned by 1. chemical shift patterns in the TOCSY (to distinguish Phe, Tyr, and Trp residues); correlations observed in the TOCSY; and $H_\delta \leftrightarrow H_\beta$ crosspeaks in NOESY spectra to connect the aromatic and aliphatic portions (Figure 10.6). Also, assigning Trp aromatic sidechains generally requires NOE correlations between the indole HN and the other aromatic $^1$Hs. Prolines are difficult to assign because they don’t have an amide $^1$H. The correlations expected in a proline residue are well-known, and distinct from any other residue. Sequential assignments are made via NOE interactions, but from the HN on the residue preceding the proline (Figure 10.7) to either $H_\alpha$ or $H_\delta$ of the proline for trans and cis conformations, respectively.

10.3. Assigning Proteins using Heteronuclear Methods.

The homonuclear methods described in the previous section become useless when proteins approach sizes > 100 amino acid residues. There are two reasons for this: first, the number of resonances becomes larger and overlap becomes increasingly severe; but, second, the tumbling rate slows and $T_2$ relaxation rates increase, preventing the buildup of correlations through the relatively small $^1$H-$^1$H coupling constants. Therefore, for proteins larger than ~100 amino acid residues, the double- and triple-resonance 3D methods discussed in the previous lecture are used for making chemical shift assignments. Again, the same three conceptual steps are
involved in making chemical shift assignments, but a completely different set of experiments are used.

**Identifying Individual Spin Systems.** There are two experiments used to identify the individual spin systems (e.g., count the number of amide resonances): the HSQC and HNCO. The HSQC provides a high resolution 2D spectrum correlating $^1$H bound to $^{15}$N; the HNCO is a very sensitive, (relatively) high resolution 3D triple resonance experiment that correlates CO$_{i-1}$, N$_i$, and HN$_i$, where i and i-1 refer to the nuclei being of the intra-residue and preceding residue, respectively. A single correlation is expected for each residue in these two experiments (glutamine and asparagine also give amide resonances, but, again, these are distinct in chemical shift and spectral appearance from the backbone resonances, and are readily identified). A protein of n residues, m of which are prolines, should have n-m-1 correlations in these spectra (often a few less than that are observed due to chemical shift overlap or protein dynamics). Figure 10.8 shows a TROSY spectrum of arginine kinase kindly provided by Omar Davulcu, Jack Skalicky and Michael Chapman.

**Assigning Spin Systems to Amino Acid Type.** There are several basic types of amino acids that are classified according to their topology: These groups are 1. glycine (1 methylene); 2. proline (five membered ring); 3. AMX (asn, asp, cys, his, ser, phe, tyr, trp; AMX refers to relative $^1$H chemical shifts of the one H$_\alpha$ and two H$_\beta$ signals); 4. long spin systems (gln, glu, lys, met, & arg); 5. branched chain (val, ile, & leu); and 6. other (ala, thr). Most of these can be identified based on the $^{13}$C chemical shift of the $\alpha$ and $\beta$ nuclei (see Spera and Bax reference, end of Lecture 2). For instance, Val C$_\alpha$ and C$_\beta$ have chemical shifts of 60 & 33 ppm,

---

**Figure 10.8.** TROSY spectrum or arginine kinase. For larger proteins, TROSY spectra replace the HSQC due to the more favorable relaxation properties of the TROSY correlation compared to the HSQC. From this spectrum one obtains information about the total number of individual spin systems. Davulcu, Skalicky, and Chapman, unpublished work.
respectively, while leu has 56 & 42 ppm. The values for glu, gln, & met are distinct enough to allow a strong tentative assignment based on the Cβ chemical shift alone; the aromatic amino acids can generally be distinguished from the asn/asp/cys spin systems in a similar fashion.

Thus, the individual amino acid spin systems are assigned to type based on their 13C chemical shifts. These are generally obtained from the HNCACB experiment (see section 9.10 for a discussion and sample spectra).

**Linking Amino Acids into the Amino Acid Sequence.** Of course, the final stage in backbone assignments is to link together the individual amino acids to map the amino acid sequence of the protein being studied. This process provides the sequence-specific chemical shift assignments. Again, this is generally done using the HNCACB in combination with the CBCA(CO)NH experiment (see Lecture 9 for description of these experiments). As discussed in section 9.10, one often detects correlations that indicate the Cα/Cβ chemical shift of the residue preceding the amide on which you are detecting. This information can be used to make the sequential assignments by finding the 13Cα/13Cβ frequencies that match best. Sample data indicating the use of the HNCACB and CBCA(CO)NH spectra for sequential assignments are shown in Figure 10.10. In this specific example of arginine kinase, the high molecular weight required sidechain deuteration which precludes using the CBCA(CO)NH experiment (because the polarization transfer starts on sidechain protons). In this case, the HN(CO)CACB experiment was used.

A summary of the triple resonance experiments commonly used in assigning protein backbone resonances is shown in Figure 10.12. This list is by no means exhaustive. For a more complete list, see Edison et al. 1994 *Methods Enzymol.* 239, 3-79.

**Sidechain assignments.** In the final step, the side chain 1H and 13C resonances must be assigned; this is done using the HCCCH-TOCSY experiment, as discussed in Lecture 9, although there are other experiments that provide very specific information. Among the more important of these is the HNHA experiment, providing correlations between HN, N, and Hα atoms and the 15N-separated TOCSY-HSQC, which frequently only provides correlations between HN and Ha, but sometimes with sidechain methyl signals in larger proteins.

In summary, the assignment of proteins is accomplished as follows: 1. identify the individual spin systems in HSQC / HNCO spectra; 2. assign to type using HNCACB; 3. link sequentially
using CBCA(CO)NH; assign all sidechains using HCCH-TOCSY. In general, one collects many experiments that provide redundant correlations to make sure that the chemical shift assignments are as complete as possible. In the end, your structures are only as good as your assignments!

Figure 10.10. Strip plot indicating combined analysis of HNCACB / CBCA(CO)NH type spectra. The figure shows side-by-side strips from an HNCACB and an HN(CO)CACB spectrum collected on arginine kinase prepared with uniform $^{15}$N, $^{13}$C, and sidechain $^2$H labeling. The HN(CO)CACB experiment was run rather than the CBCA(CO)NH because the latter experiment requires polarization transfer from sidechain $^1$H to $^{13}$C, which is impossible when the sidechains are deuterated. The analysis starts by identifying the intramolecular and sequential (i-1) correlations in the right-most panel. The residue that identically matches the C$^\alpha_{i-1}$ and C$^\beta_{i-1}$ resonances is selected as the next residue in the polypeptide chain and the process is repeated (Davulcu, Skalicky, and Chapman, unpublished).
### Determining Structure in Proteins.

In X-ray crystallography, one observes scattering reflections, which is FT’d to present regions of electron density. The structure of the protein is determined by fitting a polypeptide chain into this electron density. Protein structure determination by NMR is not as direct in that there is no structure into which nuclei are fit. Instead, NMR structures are determined by providing a different set of structural constraints; NOE interactions provide pair-wise distance information, and scalar couplings are used to provide pair-wise dihedral angle information. Other information, such as $^{13}$C chemical shift and amide hydrogen exchange rates, are used to identify secondary structures. All of this data is used to calculate structures that are consistent with the

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Correlations Observed</th>
<th>Magnetization Flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>HNCA</td>
<td>H(i), N(i), Cα(i, i-1)</td>
<td><img src="HNCA.png" alt="HNCA Diagram" /></td>
</tr>
<tr>
<td>HN(CO)CA</td>
<td>H(i), N(i), Cα(i-1)</td>
<td><img src="HN(CO)CA.png" alt="HN(CO)CA Diagram" /></td>
</tr>
<tr>
<td>HNCO</td>
<td>H(i), N(i), CO(i-1)</td>
<td><img src="HNCO.png" alt="HNCO Diagram" /></td>
</tr>
<tr>
<td>HNCACB</td>
<td>HN(i), N(i), Cα/Cβ(i, i-1)</td>
<td><img src="HNCACB.png" alt="HNCACB Diagram" /></td>
</tr>
<tr>
<td>CBCA(CO)NH</td>
<td>HN(i), N(i), Cα/Cβ(i-1)</td>
<td><img src="CBCA(CO)NH.png" alt="CBCA(CO)NH Diagram" /></td>
</tr>
</tbody>
</table>

Figure 10.11.
observed structural constraints. We’ll discuss the calculation and evaluation of structures in the next lecture; here we’ll discuss experiments to collect and provide structural constraints.

10.4. NOE and Protein Structure.

Again, the best reference for this section is the Wüthrich book, chapters 7-9. Recall that the NOE interaction between two nuclei can be represented as

\[ \text{NOE} \propto \left( \frac{1}{r^6} \right) f(\tau_c)p_A \]  

where \( f(\tau_c) \) represents a function describing the local motional properties of the spin and \( p_A \) represents the population of the molecule in a given conformation (typically assumed to be 100%). While in principle, the NOE can provide precise internuclear distances; in practice, the data are much less precise. There are three reasons for this:

1. The NOE equations are only valid for two isolated spins. This situation is not obtained in proteins where there are on the order of 1000 or more 1Hs.
2. Spin diffusion effects cause problems in quantifying NOEs (spin diffusion is non-direct dipolar interactions between two protons; it makes two 1Hs appear closer in space than they really are).
3. Motion during the NOE mixing period (~ 80-100 ms) averages the distance of the detected interactions.

For these reasons, the NOE is used in a semi-quantitative or qualitative manner, depending on the particular application. In the semi-quantitative mode, we evaluate NOE intensity as indicating distance within a certain range; in the qualitative mode, we simply determine whether there is an interaction between two nuclei. One application of the qualitative mode is secondary structure identification. This is discussed below.

10.5. Identifying Secondary Structure from NOEs.

There are four basic types of secondary structure in proteins: helix (\( \alpha \) or 310), strand (parallel or anti-parallel sheets), turn (several types), and random coil, each with its own unique NOE signature. These are presented in the following paragraphs, but first, some nomenclature, Figure 10.12. The standard nomenclature is to indicate the two nuclei participating in the NOE;
the N-terminal nucleus is indicated first, and the relative residue positions are indicated in parenthesis, e.g., αN(i,j). If the NOE is between atoms on sequential residues, then there is no numerical index. For instance, an NOE between Hα of residue i and the HN of residue j (j=i+2) is indicated as αN (i, i+2), say, or αβ (i, i+3) for an NOE between the Hα and Hβ protons of two residues separated by two intervening amino acid residues. NOEs are classified as sequential if |i-j| = 1; medium-range NOEs are for 1< |i-j| ≤ 5, and long range for |i-j| > 5.

**β-Strands and β-Sheets.** These are the easiest secondary structures to identify. In the extended conformation, the characteristic NOEs are strong NOEs between HN of residue i and Hα of residue i-1 (denoted as αN(i,i+1) or αN), and weak NOEs between HN on adjacent residues (NN (i, i+1) or NN), Figure 10.13. In β-sheets, individual strands can be oriented in a parallel or anti-parallel fashion; the distinguishing feature here being strong NH-NH and NH-Hα

**Figure 10.13.**

**Figure 10.14.**
NOEs between residues that are distant in the amino acid sequence. Parallel vs. anti-parallel are distinguished by the N → C sense of the strands, and by the presence of Hα-Hα NOEs in the antiparallel sheets.

**Turns.** There are several types of turns found in proteins; I will concentrate only on hairpin turns of type I and type II, Figure 10.14. These are characterized by short NN distances between residues 2,3 and 3,4, and by a short αN(i, i+2) distance between residues 2 and 4 for Type I turns.

**Helices.** There are two commonly encountered helices in proteins: α-helix and 310-helix, which differ in their characteristic NOE patterns. In α-helices, the characteristic NOEs are strong NN, weak αN, strong αN(i,i+3) and αN(i, i+4), with some αN(i, i+2) observed, but being weaker than the "i+4" NOEs (Figure 10.15). In 310 helices, there are no αN(i, i+4) NOEs and the αN(i, i+2) are much more intense. Another characteristic NOE observed in α-helices is the αβ(i, i+3).

### 10.6. Scalar Couplings and Protein Structure.

Scalar couplings between two nuclei separated by 3 bonds are important indicators of dihedral angles in proteins, and currently an area of active interest in the development of methods to provide a range of precise J coupling measurements. The idea is shown in Figure 10.16. The angle between HN & Hα (as shown in the Figure) provides a measure of the ϕ angle; χ angles are obtained by measuring either 1H-1H, 1H-13C, or 13C-13C couplings at various points along the sidechains. From the plot shown in this figure, we can identify scalar couplings consistent with β-strand (ϕ = -150° → -90°) and α-helix (ϕ = -60° → -30°) for $^3J_{HNH\alpha} > 8$ Hz and $^3J_{HNH\alpha} < 4$ Hz, respectively. However, it is also obvious that the Karplus curve does not give a unique ϕ value for $^3J_{HNH\alpha}$ between 4 and 9 Hz, and this is the limitation of scalar coupling constants.

The most commonly measured coupling constant is the $^3J_{HNH\alpha}$, and the experiment most commonly used to measure it is the HNHA-J experiment. The original HNHA sequence was published by Vuister and Bax (1993 JACS 115, 7772-7777) with a significant improvement published by Kuboniwa (also in Bax’s group; 1994 JBNMR 4, 871-878). The experiment is shown in Figure 10.17, and a brief product operator analysis is presented after that. Essentially this experiment is an HMOC correlation between 1H and 15N, coupled with a COSY-type HN-Hα coherence transfer (t2 period) in the middle of the experiment. This results in two peaks per spin.
system: the HN “diagonal” peak and the H_a “cross peak”. Each has intensity given by the following expressions immediately prior to detection:

\[
\text{diagonal peaks} = \cos(2\pi J\delta_2)^2, \quad [10.2]
\]

\[
\text{cross peaks} = -\sin(2\pi J\delta_2)^2 \quad [10.3]
\]

where J is the \(3J_{\text{HNH}_\alpha}\) and \(\delta_2 = 13 \text{ ms}\). The value of J is obtained from the ratio of the cross peak and diagonal peak intensity,

\[
J = -\tan(2\pi J\delta_2)^2. \quad [10.4]
\]

Similar experiments are used to measure coupling constants between other nuclei, and Karplus curves are used to analyze them, as well, but with different coefficients.

### 10.7. Amide Hydrogen Exchange and Protein Structure.

A third experimental measure of protein structure is to determine the amide hydrogen exchange rates; the exchange rate is slowed when the amide hydrogen is involved in H-bonded secondary structure. Note that this data does not tell us what type of secondary structure a residue is involved in, but can be used to identify hydrogen bonds. The use of amide hydrogen exchange to identify hydrogen bonds has been supplanted to some extent by the ability to directly measure hydrogen bonds via scalar couplings, as described in Lecture 2.
10.8. Chemical Shift and Protein Structure.

This discussion was presented some time ago, in the second lecture. Again, the idea is that the chemical shift of a particular amino acid type is shifted upfield or downfield if it is involved in secondary structure formation. For $\alpha$-helices, $\text{C}_\alpha$ resonances are shifted downfield and upfield, respectively; while the opposite shift is observed in $\beta$-strand. Secondary chemical shifts of CO and H$_\alpha$ are also useful.