

# ASSIGNMENT OF PROTEIN NMR SPECTRA USING HETERONUCLEAR NMR – A TUTORIAL

**Christina Redfield**

Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, U.K.

## 1. INTRODUCTION

The assignment of resonances in the complex NMR spectrum of a protein is the first step in any NMR study of protein structure, function or dynamics. Before ~1980, assignment was achieved using one-dimensional (1D) NMR and was based, to a large extent, on the assumption that the structure of the protein in solution was the same as that observed in an available X-ray structure. The introduction of two-dimensional (2D) NMR techniques such as correlated spectroscopy (COSY) and nuclear Overhauser effect spectroscopy (NOESY) in the late 1970's and early 1980's dramatically increased the resolution of protein NMR spectra [1,2]. This led to the development of a systematic method for the assignment of the 2D NMR spectra of proteins that relied only on information about the amino acid sequence of the protein; this is the sequential assignment method [3-5]. The availability of uniform  $^{15}\text{N}$  labelling and the development of three-dimensional (3D)  $^{15}\text{N}$ -edited NMR methods in the late 1980's led to improved resolution in NMR spectra and increased the size of proteins that could be assigned using the sequential assignment methodology [6,7]. Double labelling with  $^{13}\text{C}$  and  $^{15}\text{N}$  led, in the early 1990's, to the development of an alternative assignment approach based solely on through-bond scalar couplings; this again increased the molecular weight limit for complete resonance assignment [8,9]. This triple resonance approach in conjunction with deuteration and transverse relaxation optimised spectroscopy (TROSY) methods has increased further the molecular weight limit for full assignment to beyond 40 kD [10].

The triple resonance methodology is now the most commonly used approach for the assignment of protein NMR spectra. Because it relies solely on through-bond  $^1\text{J}$  and  $^2\text{J}$  spin-spin couplings, it is a more 'fool-proof' method than sequential assignment, which relies on through-space NOE data to establish connections between neighbouring amino acid residues. The nearly complete level of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  assignment that can be obtained by double labelling and the triple resonance assignment approach is also highly desirable, if structure determination is the ultimate goal of a project, because backbone and  $\beta$  position  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts provide information about secondary structure and backbone  $\phi/\psi$  torsion angles [11-14]. In addition, complete assignment ensures identification of the maximum number of NOEs in 3D and 4D  $^{15}\text{N}$  and/or  $^{13}\text{C}$ -edited NOESY spectra leading to better defined structures. Nevertheless, the sequential assignment approach is still useful when double labelling with  $^{13}\text{C}$  and  $^{15}\text{N}$  is too costly or not feasible. This method has been used in the past to assign spectra of proteins up to ~15 kD [15-17] and is straightforward to apply to peptides and small proteins. In this chapter, the triple resonance approach will be described first and then the sequential assignment methodology will be outlined.

## 2. SCOPE OF THIS CHAPTER

The aim of this chapter is to provide a tutorial on protein NMR resonance assignment. There are many review articles and books dedicated to explaining the theoretical basis of NMR pulse sequences but very little has been published explaining the steps involved in

analysing NMR spectra to obtain protein resonances assignments (a notable exception is the excellent web site by Victoria Higman: <http://www.protein-NMR.org.uk>). In this chapter, the available methodologies for assignment will be described. Examples of how sequence specific resonance assignments can be obtained using a suite of 2D and 3D NMR experiments will be presented and tips on how overlap problems can be overcome are included. The pulse sequences that are used to obtain the 2D and 3D NMR spectra will not be discussed; readers are encouraged to consult the original papers and also excellent books and review articles describing the pulse sequences [18,19]. Software packages for automated or semi-automated resonance assignment have been developed and can speed up the assignment process [20-25]. However, for a newcomer to protein NMR, attempting to assign spectra manually, at least once, can provide a better understanding of the methodology and an appreciation for the potential problems that arise.

### 3. THE $^1\text{H}$ - $^{15}\text{N}$ HSQC – A FINGERPRINT OF THE PROTEIN

The majority of the 3D experiments used in the triple resonance and sequential assignment approaches contain a  $^1\text{H}$ - $^{15}\text{N}$  HSQC (heteronuclear single quantum correlation) element [26,27]. Peaks in the 2D HSQC arise due to the one-bond  $^1\text{H}$ - $^{15}\text{N}$  coupling of ~90 Hz. The HSQC provides a ‘fingerprint’ of the protein because all amino acids, except proline, give rise to a peak in the spectrum (Figure 1). The HSQC also contains peaks for some side-chain nitrogens of Asn, Arg, Gln, His, Lys and Trp. Before attempting resonance assignment using 3D NMR methods it is advisable to determine whether the protein of interest will give NMR spectra of the required quality; as a starting point, the 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of the protein should be collected. The 2D HSQC can provide a good way to assess the quality of the protein sample, to permit optimisation of sample conditions (pH, temperature, ionic strength) and acquisition parameters, and to assess the likelihood of success in resonance assignment.

#### Optimising the $^1\text{H}$ - $^{15}\text{N}$ HSQC spectrum

1. The  $^1\text{H}$  carrier frequency (the frequency at the centre of the  $^1\text{H}$  spectrum) is usually chosen so that the water peak is at the centre of the spectrum (~4.8 ppm at 20 °C). The optimal  $^1\text{H}$  sweep width will vary from protein to protein. Initially collect a 1D  $^1\text{H}$  spectrum with a large sweep width (~15 ppm). Identify the most upfield and downfield shifted signals and then decrease the sweep width so that these are close to the edges of the spectrum.
2. The optimal  $^{15}\text{N}$  sweep width and  $^{15}\text{N}$  carrier frequency (the frequency at the centre of the  $^{15}\text{N}$  spectrum) will also vary from protein to protein. The HSQC spectrum can be collected initially using a large sweep width in the indirect  $^{15}\text{N}$  dimension (~35-40 ppm) centred at ~118 ppm. The HSQC spectrum may contain peaks that are shifted out of the main  $^{15}\text{N}$  envelope by several ppm. It may be desirable to fold, or alias, these outlying peaks in order to decrease the  $^{15}\text{N}$  sweep width and thereby increase resolution. Identify the desired upper and lower frequencies in the  $^{15}\text{N}$  dimension, these will determine the optimal sweep width. The  $^{15}\text{N}$  carrier frequency should be set to the average of the upper and lower frequencies selected. Collect the HSQC spectrum again with the optimised parameters (Figure 1).
3. The HSQC spectrum should contain one peak from the backbone amide of each residue, with the exception of proline and the N-terminal residue. In addition, pairs of cross peaks, sharing a common  $^{15}\text{N}$  shift, will be observed for each Asn and Gln side chain; these are found in the upfield region of both the  $^1\text{H}$  and  $^{15}\text{N}$  dimensions. A single peak will be observed for each Trp indole  $\text{H}_{\epsilon 1}$  in the downfield region of both the  $^1\text{H}$  and  $^{15}\text{N}$  dimensions. Depending on the pH and temperature of the sample, peaks may be observed from Arg, His and Lys side chains. Arg  $^{15}\text{N}_{\epsilon}-^1\text{H}_{\epsilon}$  peaks are often observed at pH values below ~6-7; Arg

side chain H $\eta$ , His H $\delta_1$ /H $\epsilon_2$  and Lys H $\zeta$  do not usually give observable peaks unless they are buried in the protein core or involved in hydrogen bonds or salt bridges at interfaces. The chemical shifts of these side-chain nitrogens differ substantially from those of the backbone amides and these peaks will be folded in the indirect dimension of the HSQC spectrum. Count the number of observed peaks and compare with the number expected on the basis of the amino acid sequence of the protein.

4. Pairs of peaks correlating a pair of  $^1\text{H}^{\text{N}}$  shifts with a single  $^{15}\text{N}$  shift arise from the Asn/Gln side chain N $\delta/\epsilon\text{H}_2$ . These can be distinguished from the peaks arising from backbone amides in two ways. In a decoupled HSQC [18], the peaks from the Asn/Gln NH $_2$  are absent. In a variant of the HSQC, in which one of the pair of  $\tau$  delays is set to  $1/8J$  rather than  $1/4J$ , only the Asn/Gln NH $_2$  peaks are observed (Figure 2) [28].

5. HSQC spectra are usually collected using pulsed-field gradients and water flip-back methods [29,30]. This eliminates the need for solvent presaturation and reduces the loss of  $\text{H}^{\text{N}}$  peaks as a result of solvent cross saturation. Exchange of  $\text{H}^{\text{N}}$  with  $\text{H}_2\text{O}$  is both acid and base catalysed and has a pH minimum at pH  $\sim 2$ -3. At pH values below  $\sim 6$ , HSQC peaks should be observed for all backbone amides. At pH values above  $\sim 7$ , the intrinsic  $\text{H}^{\text{N}}$  exchange rates may become so fast that  $\text{H}^{\text{N}}$  peaks are lost, particularly for disordered regions of the protein. The intrinsic rate of exchange of an  $\text{H}^{\text{N}}$  depends on the identity of that amino acid and of the previous residue [31]; intrinsic rates at a particular pH value can vary by 2-3 orders of magnitude along the sequence. The 2<sup>nd</sup> residue in the sequence has the fastest intrinsic exchange rate due to its proximity to the N-terminal NH $_3^+$  group. The C-terminal residue has the slowest intrinsic rate due to the terminal carboxyl group. Relatively high intrinsic rates are observed for sequences of His residues; as a result of this, peaks in the HSQC are often absent for His tags. If a significant number of peaks are missing from the HSQC spectrum then try decreasing the pH or lowering the temperature and repeating the HSQC experiment. The intrinsic exchange rates for a protein amino acid sequence at a particular pH can be predicted using the program SPHERE [31,32] (<https://www.fccc.edu/research/labs/roder/sphere/sphere.html>).

6. When the HSQC has been optimised give each peak in the spectrum a number to identify it. Numbering peaks from left to right in the spectrum (most downfield  $^1\text{H}^{\text{N}}$  to most upfield  $^1\text{H}^{\text{N}}$ ) is a convenient method.

7. Observation of more than the expected number of peaks in the HSQC could indicate the presence of impurities, sample degradation or multiple conformations in slow exchange. Protein impurities or sample degradation can be checked by SDS PAGE. Degradation of the protein may generate additional C-terminal residues which will appear in the spectrum as new cross peaks in the region around  $\sim 7.7$ -8.2 ppm and  $\sim 127$ -130 ppm. If the number of peaks in this region increases with time this is a sign of increasing amounts of degradation. Multiple conformations in slow exchange often arise from cis/trans isomerisation of an Xxx/Pro peptide bond. This may be an intrinsic property of the protein that cannot be altered.

8. If the protein does not give close to the expected number of  $^1\text{H}^{\text{N}}$ - $^{15}\text{N}$  peaks in the HSQC spectrum then the protein may not be a good candidate for triple resonance assignment using double labelling. Perdeuteration in conjunction with TROSY-based methods may give improved data (Section 4.9). Intrinsically disordered proteins give poorly resolved  $^1\text{H}$  NMR spectra and the standard triple resonance methods may not provide the required resolution. Higher dimensional methods and  $^{13}\text{C}$  rather than  $^1\text{H}$  detection have been introduced recently to address the problems associated with IDPs [33,34]. Line broadening is often observed for paramagnetic proteins and carbon detection has proved to be a useful tool in these systems as well [35,36].

## 4. THE TRIPLE RESONANCE ASSIGNMENT METHOD

Triple resonance assignment experiments, which require uniformly ( $> 90\%$ )  $^{15}\text{N}/^{13}\text{C}$  double-labelled protein, rely on  $^1\text{J}$  and  $^2\text{J}$  spin-spin couplings to identify neighbouring amino acid residues. Lower levels of  $^{15}\text{N}/^{13}\text{C}$  labelling lead to a significant reduction in signal-to-noise because these experiments rely on directly bonded  $^{15}\text{N}-^{13}\text{C}$  or  $^{13}\text{C}-^{13}\text{C}$  pairs. The one bond couplings between  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  in proteins vary in magnitude from  $\sim 7$  to  $\sim 140$  Hz (Figure 3). Of particular importance for this methodology is the one-bond coupling of  $\sim 15$  Hz between  $^{13}\text{CO}(i-1)$  and  $^{15}\text{N}(i)$  and the two-bond coupling between  $^{13}\text{C}\alpha(i-1)$  and  $^{15}\text{N}(i)$  which varies between 4 and 9 Hz; these allow the correlation of adjacent residues in the sequence. Such connections cannot be made using  $^1\text{H}-^1\text{H}$  couplings alone because the four-bond coupling between  $^1\text{H}\alpha(i-1)$  and  $^1\text{H}^{\text{N}}(i)$  is vanishingly small. The triple resonance methodology was first introduced in 1990 and was based on a set of four 3D triple resonance experiments and a 3D  $^1\text{H}-^{15}\text{N}$  TOCSY-HMQC [8,9]. Two of the triple resonance experiments were collected in  $\text{H}_2\text{O}$  (HNCO and HNCA) and two in  $\text{D}_2\text{O}$  (HCACO and HCA(CO)N) leading to differences in the measured  $^{13}\text{C}\alpha$  and  $^{15}\text{N}$  shifts due to H/D isotope effects. The methods currently used involve detection of  $^1\text{H}^{\text{N}}$  and all spectra are collected in  $\text{H}_2\text{O}$ , thereby removing the problem of isotope shifts. A large number of triple resonance pulse sequences have been developed since 1990 and it is not possible to discuss all of these in this chapter [18,19]. In the following sections several of the common approaches to assignment using triple resonance experiments are outlined.

Triple resonance spectra generally correlate a backbone  $^1\text{H}^{\text{N}}$  and  $^{15}\text{N}$  pair with one or more  $^{13}\text{C}$  chemical shifts. The naming convention for triple resonance experiments uses letters that denote the nuclei whose chemical shifts are correlated. For example, in the 3D HNCO experiment, backbone  $^1\text{H}^{\text{N}}-^{15}\text{N}$  pairs are correlated with the carbonyl carbon ( $^{13}\text{CO}$  or  $^{13}\text{C}'$ ) of the preceding residue. Nuclei involved in the coherence transfer pathway but whose chemical shift is not measured in the experiment are enclosed by parentheses ( ). For example, in the HN(CO)CA experiment, backbone  $^1\text{H}^{\text{N}}-^{15}\text{N}$  pairs are correlated with the  $^{13}\text{C}\alpha$  of the preceding residue with coherence transfer via the preceding  $^{13}\text{CO}$ ; the  $^{13}\text{C}\alpha$  chemical shift is measured but the  $^{13}\text{CO}$  shift is not.

For reasons of sensitivity, triple resonance experiments generally start with a  $^1\text{H}$  pulse and  $^1\text{H}$  signal is detected during the acquisition period.  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts are measured in the indirectly-detected dimensions. Many of these experiments are of the ‘out-and-back’ type where the proton spin that is initially excited, the  $^1\text{H}^{\text{N}}$ , is the same as the proton spin that is detected. So, in the HNCO, for example, the sequence starts with  $^1\text{H}^{\text{N}}$ , transfer occurs to  $^{15}\text{N}$  and then to  $^{13}\text{CO}$ , which is frequency labelled during  $t_1$ . This is followed by transfer back to  $^{15}\text{N}$ , which is frequency labelled during  $t_2$ , followed by a final transfer back to  $^1\text{H}^{\text{N}}$ ; the latter frequency is measured in  $t_3$ . The resulting 3D HNCO has the  $^1\text{H}^{\text{N}}$  chemical shifts in the  $F_3$  dimension,  $^{15}\text{N}$  in  $F_2$  and  $^{13}\text{CO}$  in  $F_1$ .

### 4.1 Optimising parameters for triple resonance experiments.

Before embarking on the collection of numerous 3D data sets it is advisable to optimise the sweep widths and carrier frequencies used in the aliphatic and carbonyl carbon dimensions. The range of aliphatic  $^{13}\text{C}$  chemical shifts observed for a protein can be determined by collecting a  $^1\text{H}-^{13}\text{C}$  HSQC; the most downfield shifted  $^{13}\text{C}$  usually arise from Thr  $\text{C}\beta$  while the most upfield  $^{13}\text{C}$  belong to Ile  $\text{C}\delta_1$  [37]. The sweep widths and  $^{13}\text{C}$  carrier frequencies for the indirect  $^{13}\text{C}\alpha$ ,  $^{13}\text{C}\alpha/^{13}\text{C}\beta$  and  $^{13}\text{CO}$  dimensions can be optimised by collecting 2D  $^1\text{H}^{\text{N}}-^{13}\text{C}$  planes of the HNCA, CBCA(CO)NH and HNCO experiments. The random coil shifts for aliphatic carbons in the 20 amino acids are shown in Figure 4 [37]; carbonyl carbons are centred around 175 ppm. For the 2D HNCA, an initial sweep width of

~35 ppm centred at ~55 ppm can be used. For the 2D CBCA(CO)NH, an initial sweep width of ~70 ppm centred at ~45 ppm can be used while for the HNCO an initial sweep width of ~20 ppm centred at ~175 ppm is appropriate. These sweep widths and carrier frequencies can then be optimised depending on the range of  $^{13}\text{C}$  chemical shifts observed in the 2D spectra. For the (H)CC(CO)NH experiment, a larger  $^{13}\text{C}$  sweep width than used in CBCA(CO)NH will be required because the Ile  $\text{C}\delta_1$  and  $\text{C}\gamma_2$  are upfield of Ala  $\text{C}\beta$ . The 2D  $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$  and/or  $^1\text{H}^{\text{N}}\text{-}^{13}\text{C}$  planes of most triple resonance experiments can be collected in a relatively short period of time. These 2D spectra allow the signal-to-noise ratio of various experiments to be assessed before embarking on the more time consuming collection of 3D data sets.

## 4.2 Viewing 3D data sets and generating strip plots and peak lists

3D triple resonance spectra usually contain a  $^1\text{H}^{\text{N}}$ , a  $^{15}\text{N}$  and a  $^{13}\text{C}$  dimension. The typical way of viewing these spectra is to plot 2D  $^1\text{H}^{\text{N}}\text{-}^{13}\text{C}$  planes that correspond to a particular  $^{15}\text{N}$  chemical shift. The plane will contain peaks for all residues with that particular  $^{15}\text{N}$  chemical shift, that is all peaks that fall along a horizontal line drawn through the  $^1\text{H}\text{-}^{15}\text{N}$  HSQC at that  $^{15}\text{N}$  chemical shift (Figures 1 and 5). The 3D triple resonance spectra contain a relatively small number of peaks, generally one or two peaks per residue; these correlate  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}$  and one or more  $^{13}\text{C}$  chemical shifts. The 3D matrices can be reduced to strip plots (Figure 5), as described below, or simply to a list of peaks containing the  $^1\text{H}^{\text{N}}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  chemical shifts of each peak observed in the 3D spectrum.

### Creating a 2D strip plot

1. Strips are extracted from  $^{15}\text{N}$  planes. Each strip is centred about a  $\text{H}^{\text{N}}$  chemical shift in  $\text{F}_3$  and the strip covers the full  $^{13}\text{C}$  sweep width in the indirect  $^{13}\text{C}$  dimension,  $\text{F}_1$  (or indirect  $^1\text{H}$  dimension for experiments such as HBHA(CBCA)NH). In order to make a 2D strip plot from a 3D matrix it is necessary to identify the  $^1\text{H}^{\text{N}}(\text{F}_3)\text{-}^{15}\text{N}(\text{F}_2)$  positions at which the strips parallel to  $\text{F}_1$  should be extracted. This can be done from peak positions in an HSQC spectrum collected prior to the 3D experiment or from peak positions in a 2D matrix created by projection of the 3D matrix along the indirect  $^{13}\text{C}$  dimension ( $\text{F}_1$ ).
2. The strips should be ordered according to the  $\text{H}^{\text{N}}$  chemical shift; the most downfield  $\text{H}^{\text{N}}$  is strip 1 and the most upfield  $\text{H}^{\text{N}}$  is the last strip. The numbering of the strips corresponds to the arbitrary peak numbers assigned in the HSQC spectrum (Figure 5).

## 4.3 Backbone assignments using $^1\text{H}^{\text{N}}$ , $^{15}\text{N}$ , $^{13}\text{C}\alpha$ and $^{13}\text{CO}$

Backbone assignments can be obtained from analysis of four triple resonance experiments, HNCO, HNCA, HN(CA)CO and HN(CO)CA as summarised in Figure 6. The HNCO experiment correlates  $\text{H}^{\text{N}}$  and N of residue  $i$  with CO of residue  $i-1$  [9,38]. HNCA correlates  $\text{H}^{\text{N}}$  and N of residue  $i$  with  $\text{C}\alpha$  of both residues  $i$  and  $i-1$  [9,38]. The pair of correlations is observed because  $^2\text{J}_{\text{N}(i)\text{-C}\alpha(i-1)}$  is of comparable magnitude to  $^1\text{J}_{\text{N}(i)\text{-C}\alpha(i)}$  (Figure 3). The HN(CA)CO experiment correlates the  $\text{H}^{\text{N}}$ , N and CO of both residues  $i$  and  $i-1$  as a result of the transfer from  $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$  of  $i$  via the  $\text{C}\alpha$  of residues  $i$  and  $i-1$  and subsequent transfer to  $\text{CO}(i)$  and  $\text{CO}(i-1)$  [39]. The HN(CO)CA experiment contains only the  $i-1$   $\text{C}\alpha$  connection because of the transfer via the CO of  $i-1$  [38,40]. Comparison of the chemical shifts of the observed peaks in these experiments should allow identification of backbone connectivities as described below.

### Analysis of HNCO, HNCA, HN(CA)CO and HN(CO)CA spectra

1. Prepare strip plots for each 3D dataset (Figure 7). The strips should be extracted from  $^{15}\text{N}$  planes. Each strip is centred about a  $\text{H}^{\text{N}}$  shift in  $\text{F}_3$  and covers the full  $^{13}\text{C}\alpha$  or  $^{13}\text{CO}$  sweep

width in the indirect dimension. Order the strips according to the  $H^N$  chemical shift. The numbering of the strips corresponds to the arbitrary peak numbers assigned in the  $^1H$ - $^{15}N$  HSQC.

2. For each  $H^N(i)$  compare the HNCA and HN(CO)CA strips; the former should contain 2 peaks and the latter a single peak. The  $C\alpha$  peak observed in both strips belongs to residue  $i-1$ . The  $C\alpha$  peak observed in only the HNCA strip belongs to residue  $i$ . It is usually, but not always, the case that in the HNCA the intraresidue ( $i$ )  $C\alpha$  peak is stronger than the interresidue ( $i-1$ ) peak. If only one  $C\alpha$  peak is observed in the HNCA strip then the two  $C\alpha$  have very similar chemical shifts (that is, they overlap) or the interresidue  $C\alpha$  connectivity is missing.

3. The CO peak observed in the HNCO strip for  $H^N(i)$  corresponds to  $CO(i-1)$ . The CO peak observed in the HN(CA)CO strip but not in the HNCO for  $H^N(i)$  corresponds to  $CO(i)$ . It is usually, but not always, the case that in the HN(CA)CO the intraresidue ( $i$ ) CO peak is stronger than the interresidue ( $i-1$ ) peak. If only one CO peak is observed in the HN(CA)CO strip then the two CO overlap or the interresidue CO connectivity is missing.

4. For each  $H^N(i)$  a list of 6 chemical shifts,  $C\alpha(i-1)$ ,  $CO(i-1)$ ,  $^{15}N(i)$ ,  $H^N(i)$ ,  $C\alpha(i)$  and  $CO(i)$  can be tabulated from the strip plots of the 4 experiments (Table 1).

5. Two  $H^N$ ,  $j$  and  $k$ , correspond to neighbouring residues in the sequence if their peak lists share a pair of  $C\alpha$  and a pair of CO shifts. Residue  $j$  precedes residue  $k$  if the shifts  $C\alpha(j)=C\alpha(k-1)$  and  $CO(j)=CO(k-1)$ . Residue  $k$  precedes residue  $j$  if the shifts  $C\alpha(k)=C\alpha(j-1)$  and  $CO(k)=CO(j-1)$  (Figure 8).

6. When two  $H^N$  have been identified as neighbours continue to extend the segment at each end using the procedure in step 5. Breaks in the segment will occur at the N and C-termini of the protein, at Pro residues and when two or more residues share the same  $C\alpha$  and CO shifts. In the case of Pro, the  $C\alpha$  and CO of the Pro will be observed in the  $H^N$  strip of the residue that follows it in the sequence (the  $i-1$  peaks) but because the Pro does not have an  $H^N$  there is no connection to be residue preceding the Pro.

In principle, assignments could be based on pairs of experiments, the HNCA and HN(CO)CA or the HNCO and HN(CA)CO. However, chemical shift overlap in the  $C\alpha$  or CO regions of the spectrum would lead to ambiguities in the analysis based on a single shared chemical shift. The approach described above, which relies on two shared chemical shifts, is more likely to give long segments of unambiguously assigned residues. Analysis of these four datasets results in segments of adjacent backbone spin systems. However, the measured  $H^N$ ,  $^{15}N$ , and CO chemical shifts do not vary significantly with amino acid type and therefore do not provide information that would allow these segments to be placed within the amino acid sequence of the protein. The  $C\alpha$  chemical shifts do have some dependence on amino acid type (Figure 4) [37]. Gly has a random coil  $C\alpha$  shift close to  $\sim 45$  ppm. In addition, the  $^{15}N$  of Gly are generally found to be upfield of  $\sim 110$  ppm. Ile, Pro, Thr and Val have a  $C\alpha$  shift downfield of  $\sim 60$  ppm. These are distinct from the  $C\alpha$  shifts of 50-60 ppm for the other residues. The distribution of Gly and Ile/Pro/Thr/Val is limited in protein sequences and therefore residue specific assignment is difficult on the basis of these 4 triple resonance experiments alone. In addition, secondary structure can lead to upfield or downfield shifts of  $C\alpha$  of  $\sim 3$ -4 ppm leading to a blurring of the 60 ppm cut off [41]. More detailed spin system information can be obtained from a variety of experiments. For each  $H^N$ - $^{15}N$  pair, side chain  $^1H$  chemical shifts can be obtained from a 3D  $^{15}N$ -edited TOCSY-HSQC, as described in section 5.2. In addition, triple resonance experiments that provide information about  $^{13}C\beta$  and/or  $^1H\beta$  can provide spin system information as described in the next section.

#### 4.4 Resonance assignments using $^1\text{H}^{\text{N}}$ , $^{15}\text{N}$ , $^{13}\text{C}\alpha$ , $^{13}\text{C}\beta$ , $^1\text{H}\alpha$ and $^1\text{H}\beta$ .

The triple resonance approach described in the previous section yields only backbone connectivities involving  $\text{H}^{\text{N}}$ ,  $^{15}\text{N}$ ,  $\text{C}\alpha$  and  $\text{CO}$ . Spin system information based on  $\text{C}\alpha$  chemical shift is limited to Gly and Ile/Pro/Thr/Val residues. In addition, overlap in the  $\text{C}\alpha$  region of the spectrum is common in proteins and can lead to ambiguities in the approach described above. An alternative set of triple resonance experiments can be used to obtain both the backbone connectivities and the spin system information needed for residue specific assignment. This methodology uses the CBCANH [42], CBCA(CO)NH [43] experiments and the closely related HBHA(CBCA)NH [44] and HBHA(CBCACO)NH [45] experiments (also abbreviated as HBHANH and HBHA(CO)NH (Figure 9)). These experiments are not of the ‘out-and-back’ variety. Instead, these ‘straight-through’ sequences start with a  $^1\text{H}$ - $^{13}\text{C}$  HSQC element in which either the  $^{13}\text{C}$  or  $^1\text{H}$  chemical shift is encoded in  $F_1$ . Spin system information is obtained from the chemical shifts of  $\text{C}\alpha$ ,  $\text{C}\beta$  and  $\text{H}\beta$  (Figures 4 and 10). The CBCA(CO)NH experiment correlates  $\text{C}\beta$  and  $\text{C}\alpha$  of residue  $i-1$  with the  $\text{N}$  and  $\text{H}^{\text{N}}$  of residue  $i$  while the CBCANH spectrum contains both the inter ( $i-1$ ) and intraresidue ( $i$ ) connections as seen above for the HNCA spectrum. In the CBCANH spectrum there is a phase difference for the  $\text{C}\alpha$  and  $\text{C}\beta$  peaks ( $\text{C}\alpha$  peaks of all residues, except Gly, are positive while  $\text{C}\beta$ , and  $\text{C}\alpha$  of Gly, are negative). Analysis of the CBCANH and CBCA(CO)NH spectra relies on matching of two shared chemical shifts; like the approach presented in section 4.3, this is more likely to give long segments of unambiguously assigned residues.

The HBHA(CO)NH experiment correlates  $\text{H}\beta$  and  $\text{H}\alpha$  of residue  $i-1$  with the  $\text{N}$  and  $\text{H}^{\text{N}}$  of residue  $i$  while the HBHANH spectrum contains both the inter and intraresidue connections. In this pair of experiments spin system information is obtained from the  $\text{H}\beta$  shifts (Figure 10) [37]. The HBHANH has the same phase differences for  $\text{H}\alpha$  and  $\text{H}\beta$  as observed for CBCANH. In principle, assignments can be based on the CBCA(CO)NH and CBCANH pair alone. However, chemical shift overlap in the  $\text{C}\alpha$  and  $\text{C}\beta$  regions of the spectrum may lead to ambiguities in the analysis. The use of the HBHA(CO)NH and HBHANH spectra in conjunction with CBCA(CO)NH and CBCANH can decrease the ambiguity in assignment and provide additional spin system information through the  $\text{H}\beta$  shift. When unambiguous connectivities can be established from the CBCA(CO)NH and CBCANH experiments alone, the HBHA(CO)NH experiment can be used to assign  $\text{H}\alpha$  and  $\text{H}\beta$  and to provide additional spin system information from the  $\text{H}\beta$  chemical shifts. These experiments can also be complemented by the HNCO/HN(CA)CO pair described above which provides the  $\text{CO}$  shifts. The assignment approach based on these 4 experiments is described below.

##### Analysis of CBCA(CO)NH and CBCANH spectra

1. Prepare strip plots for each 3D dataset as described above (Figure 11).
2. For each  $\text{H}^{\text{N}}(i)$  compare the CBCA(CO)NH and CBCANH strips. The  $\text{C}\alpha$  and  $\text{C}\beta$  peaks will have opposite sign in the CBCANH. The  $\text{C}\alpha$  and  $\text{C}\beta$  peaks observed in both strips belong to residue  $i-1$ . The  $\text{C}\alpha$  and  $\text{C}\beta$  peaks observed in only the CBCANH strip belong to residue  $i$ . Each CBCA(CO)NH strip should contain 2 peaks and each CBCANH strip should contain 4 peaks. Fewer peaks may indicate that residue  $i$  or  $i-1$  is a Gly or that  $\text{C}\alpha$  and/or  $\text{C}\beta$  peaks of  $i$  and  $i-1$  overlap.
3. The CBCA(CO)NH strip for a residue that follows a Gly will contain only one peak corresponding to the Gly  $\text{C}\alpha$  at  $\sim 45$  ppm. The CBCANH strip of a Gly residue or of a residue that follows a Gly will contain three rather than four peaks (note: the peak corresponding to Gly  $\text{C}\alpha$  will be negative).
4. For each  $\text{H}^{\text{N}}(i)$  a list of 4-6 chemical shifts,  $\text{C}\beta(i-1)$ ,  $\text{C}\alpha(i-1)$ ,  $\text{N}(i)$ ,  $\text{H}^{\text{N}}(i)$ ,  $\text{C}\beta(i)$  and  $\text{C}\alpha(i)$

can be tabulated from the strip plots of these 2 experiments. The number of chemical shifts depends on whether one or both residues are Gly (Table 2).

5. Two  $H^N$ ,  $j$  and  $k$ , correspond to neighbouring residues if their peak lists share a set of  $C\alpha$  and  $C\beta$  shifts. Residue  $j$  precedes residue  $k$  if the shifts  $C\alpha(j) = C\alpha(k-1)$  and  $C\beta(j) = C\beta(k-1)$ . Residue  $k$  precedes residue  $j$  if the shifts  $C\alpha(k) = C\alpha(j-1)$  and  $C\beta(k) = C\beta(j-1)$  (Figure 12).

6. When two  $H^N$  have been identified as neighbours continue to extend the segment at each end using the procedure in step 5. Breaks in the segments will occur at the N and C-termini, at Pro

residues and when two or more residues share the same  $C\alpha$  and  $C\beta$  shifts. In the case of Pro, the  $C\alpha$  and  $C\beta$  of the Pro will be observed in the  $H^N$  strip of the residue that follows it in the sequence (the  $i-1$  peaks) but because the Pro does not have an  $H^N$  there is no connection to be residue preceding the Pro.

7. Spin system information for residue  $i$  can be obtained from the chemical shifts of  $C\alpha(i)$  and  $C\beta(i)$ . Compare the spin system information derived from these shifts for each residue with the amino acid sequence of the protein and identify the residue specific assignment.

#### Analysis of HBHA(CO)NH and HBHANH spectra

1. Prepare strip plots for each 3D dataset as described above.

2. For each  $H^N(i)$  compare the HBHA(CO)NH and HBHANH strips. The  $H\alpha$  and  $H\beta$  peaks observed in both strips belong to residue  $i-1$ . The  $H\alpha$  and  $H\beta$  peaks observed in only the HBHANH strip belong to residue  $i$ . (note: the  $H\alpha$  and  $H\beta$  shifts for residue  $i$  can also be obtained from a 3D  $^{15}N$ -edited TOCSY-HSQC, section 5.2 below, so collection of the HBHANH may not be necessary).

3. The number of peaks observed in the HBHA(CO)NH will range from 1 to 3; a single peak will be observed if  $i-1$  is a Gly with overlapping  $H\alpha$ , two peaks will be observed if  $i-1$  is a Gly with two separate  $H\alpha$  or  $i-1$  is a residue with a single  $H\beta$  or  $i-1$  is a residue with two overlapping  $H\beta$ , three peaks will be observed if  $i-1$  is a residue with two resolved  $H\beta$ . In a similar way, the number of peaks observed in the HBHANH will range from 2 to 6.

4. If only the HBHA(CO)NH has been collected, then for each  $H^N(i)$  a list of 3-5 chemical shifts,  $H\beta(i-1)$ ,  $H\alpha(i-1)$ ,  $N(i)$ ,  $H^N(i)$  can be tabulated from the strip plots of this experiment. The number of chemical shifts depends on whether residue  $i-1$  is a Gly and on whether pairs of  $H\beta$  have unique chemical shifts (Table 2, Figure 12).

5. If both the HBHANH and HBHA(CO)NH have been collected, then for each  $H^N(i)$  a list of 4-8 chemical shifts,  $H\beta(i-1)$ ,  $H\alpha(i-1)$ ,  $N(i)$ ,  $H^N(i)$ ,  $H\alpha(i)$ , and  $H\beta(i)$  can be tabulated from the strip plots of these 2 experiments. The number of chemical shifts depends on whether one or both residues are Gly and on whether pairs of  $H\beta$  have unique chemical shifts.

6. Two  $H^N$ ,  $j$  and  $k$ , correspond to neighbouring residues if their peak lists share a set of  $H\alpha$  and  $H\beta$  shifts. Residue  $j$  precedes residue  $k$  if the shifts  $H\alpha(j) = H\alpha(k-1)$ , and  $H\beta(j) = H\beta(k-1)$ . Residue  $k$  precedes residue  $j$  if the shifts  $H\alpha(k) = H\alpha(j-1)$ ,  $H\beta(k) = H\beta(j-1)$ .

7. When two  $H^N$  have been identified as neighbours continue to extend the segment at each end using the procedure in step 5. Breaks in the segments will occur at the N and C-termini, at Pro

residues and when two or more residues share the same  $H\alpha$  and  $H\beta$  shifts.

8. Spin system information for residue  $i$  can be obtained from the chemical shifts of  $H\alpha(i)$  and  $H\beta(i)$ . Compare the spin system information derived from these shifts for each residue with the amino acid sequence of the protein and identify the residue specific assignment.

Spin system information can be derived from the CBCANH and CBCA(CO)NH



experiments because of unique  $C\alpha$  and  $C\beta$  chemical shifts of some amino acids (Figure 4) [37]. Gly has  $C\alpha$  at  $\sim 45$  ppm and no  $C\beta$ . A  $C\alpha$  downfield of 60 ppm is likely to arise from Ile, Pro, Thr or Val. Further distinction between these four residues can be based on the  $C\beta$  shift. Ile, Pro, Thr and Val have  $C\beta$  chemical shifts of  $\sim 38$  ppm,  $\sim 32$  ppm,  $\sim 70$  ppm and  $\sim 33$  ppm, respectively. Even with secondary structure dependent shifts, it is possible to identify Ile and Thr residues unambiguously. Ser, like Thr, has  $C\beta$  downfield of its  $C\alpha$ ; Ser can be distinguished from Thr because its  $C\alpha$  is generally further upfield ( $\sim 58$  ppm) than the Thr  $C\alpha$ . Ala can also be identified unambiguously due to its  $C\beta$  at  $\sim 18$  ppm. The other 13 residue types can be divided into two categories: residues with  $C\beta$  between 45 and 37 ppm (Asn, Asp, Cys, Leu, Phe, Tyr) and residues with  $C\beta$  upfield of  $\sim 35$  ppm (Arg, Gln, Glu, His, Lys, Met, Trp). Some of the spin-system ambiguity resulting from  $C\beta$  shifts can be resolved using  $H\beta$  shift information obtained from HBH $\alpha$ NH/HBHA(CO)NH (Figure 10). For example, a  $C\beta$  shift of  $\sim 37$ -45 ppm could correspond to Asn, Asp, Cys, Ile, Leu, Phe or Tyr. If this residues is found to have a pair of  $H\beta$  shifts at  $\sim 1.5$  ppm then it can be assigned unambiguously to Leu. A  $C\beta$  shift upfield of  $\sim 35$  ppm could correspond to Arg, Gln, Glu, His, Lys, Met, Trp. If this residues is found to have a pair of  $H\beta$  shifts at  $\sim 2.5$ -3.5 ppm then the spin system can be narrowed down to His or Trp.

#### 4.5 Obtaining further side-chain assignments using triple resonance experiments

Information about  $^{13}C\beta$  and  $^1H\beta$  chemical shifts is generally sufficient to enable residue specific assignment if a long stretch of connected residues has been identified. Further spin system information for each  $^1H^N$ - $^{15}N$  pair can be derived from experiments which utilise  $^{13}C$  TOCSY transfer along the side chain. Two widely used experiments are (H)CC(CO)NH and H(CCCO)NH [46-49]. The (H)CC(CO)NH experiment correlates aliphatic side chain  $^{13}C$  of residue  $i-1$  with the  $^{15}N$  and  $H^N$  of residue  $i$  while the H(CCCO)NH experiment correlates side chain  $^1H$  of residue  $i-1$  with the  $^{15}N$  and  $H^N$  of residue  $i$ . These experiments are more efficient than 3D  $^1H$ - $^{15}N$  TOCSY-HSQC because coherence transfer occurs via the larger  $^{13}C$ - $^{13}C$  coupling. The additional side chain  $^{13}C$  and  $^1H$  shifts are useful for distinguishing residues with similar  $C\beta$  and  $H\beta$  shifts. For example, Met, Lys and Val have similar  $C\beta$  and  $H\beta$  shifts of  $\sim 32$  ppm and  $\sim 2$  ppm but very different  $C\gamma$  and  $H\gamma$  shifts (Figure 4 and 10).

These experiments only correlate  $^{13}C$  which are bonded to  $^1H$  with the following  $^1H^N$ - $^{15}N$ . Therefore, they do not allow the correlation of the aromatic side chains of Tyr, Trp, Phe or His with the backbone nor do they allow the correlation of the amide side chain of Asn or Gln with the backbone. For the aromatic amino acids, experiments such as (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE [50] can be used to correlate the  $^{13}C\beta$  with the aromatic  $H\delta$  and  $H\epsilon$  resonances. Correlation between the aromatic side chains and the backbone can also be carried out on the basis of NOESY spectra (see section 5.3). For Asn/Gln, the side chain amides can often be correlated with the rest of the spin system from information available in the CBCA(CO)NH experiment. For the NHD components of the Asn/Gln side chain amides (Figure 2) correlations from  $H\delta$ - $N\delta$  to  $C\beta$  and  $C\alpha$  are observed for Asn and from  $H\epsilon$ - $N\epsilon$  to  $C\gamma$  and  $C\beta$  for Gln. The correlations are generally stronger for the more downfield peak of the  $NH_2$  pair, which corresponds to the  $E$  proton ( $H\delta_{21}$  or  $H\epsilon_{21}$ ), and this may provide an easy method for stereospecific assignment [51].

#### 4.6 A Worked Example

The examples used to illustrate the analysis of triple resonance spectra shown in Figures 7, 8, 11 and 12 are relatively simple and are not complicated by resonance overlap.

Of course this is not realistic when it comes to the analysis of protein spectra. A slightly more realistic example, which is complicated by overlap, is illustrated in Figures 13 and 14. Strips from the HNCA are shown for 10 residues. The strips are ordered according to the C $\alpha$  chemical shift. Some sequential links are unambiguous and simple to make on the basis of the HNCA alone. For example, residue 1 precedes residue 6 and residue 9 precedes 3 (Figure 13). The analysis becomes more complicated when considering links for some of the other residues including 6 and 7. As a result of overlap in the C $\alpha$  region, unambiguous links cannot be established from the HNCA alone. However, by using a second triple resonance experiment, the HN(CA)CO, unambiguous links can be established for all residues (Figure 14). This highlights the importance of using two or more  $^{13}\text{C}$  chemical shifts to establish and confirm connectivities between adjacent residues.

#### 4.7. Completing side chain $^{13}\text{C}/^1\text{H}$ assignments using $^{13}\text{C}$ HCCH-COSY/TOCSY.

The triple resonance experiments described above will yield nearly complete backbone and C $\beta$ /H $\beta$  assignments. If the (H)CC(CO)NH and H(CCCO)NH spectra are collected then further side chain  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts are available. However, ambiguity may still remain. For example, the HBHA(CO)NH experiment will give two  $^1\text{H}$  chemical shifts for a Thr at position *i-1*. However, depending on the secondary structure context, H $\beta$  may be upfield ( $\beta$ -sheet) or downfield ( $\alpha$ -helix) of H $\alpha$ ; further information is needed to distinguish H $\alpha$  from H $\beta$ . The (H)CC(CO)NH and H(CCCO)NH spectra will provide  $^{13}\text{C}$  and  $^1\text{H}$  chemical shifts for the  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$  positions. However, for residues such as Val, the spectra do not provide information about which of the two H $\gamma$  shifts is correlated with which of the two C $\gamma$  shifts. Nor does it distinguish between H $\beta$ /H $\gamma$ /H $\delta$  or C $\beta$ /C $\gamma$  with very similar chemical shifts (H $\beta$ /H $\gamma$  of Leu, H $\beta$ /H $\gamma$  of Pro, H $\beta$ /H $\gamma$  of Arg, C $\beta$ /C $\gamma$  of Lys). The  $^1\text{H}$ - $^{13}\text{C}$  HSQC combined with  $^{13}\text{C}$  HCCH-COSY and HCCH-TOCSY experiments, described below, can be used to complete  $^1\text{H}$  and  $^{13}\text{C}$  side chain assignments. These methods require  $^{13}\text{C}$  labelling only and the spectra may be collected in D $_2\text{O}$ .

The HCCH-COSY and HCCH-TOCSY methods for side-chain assignment depend on the one-bond  $^1\text{H}$ - $^{13}\text{C}$  and  $^{13}\text{C}$ - $^{13}\text{C}$  couplings (Figure 3); the  $^1\text{H}$ - $^{13}\text{C}$  one-bond coupling is typically 140 Hz and the  $^{13}\text{C}$ - $^{13}\text{C}$  one-bond couplings are 30-40 Hz. The HCCH-COSY [52] and HCCH-TOCSY [53,54] experiments differ in the mechanism of transfer between  $^{13}\text{C}$  nuclei. In the HCCH-COSY method a COSY-type mixing pulse is applied. In the HCCH-TOCSY method isotropic mixing is achieved between  $^{13}\text{C}$  nuclei using a DIPSI sequence. These experiments can be thought of as  $^1\text{H}$ - $^1\text{H}$  COSY and TOCSY spectra in which cross peaks are spread out in a third dimension according to the  $^{13}\text{C}$  chemical shift of one of the carbon nuclei. In the HCCH-COSY spectrum only 3-bond  $^1\text{H}$ - $^1\text{H}$  correlations are observed because of the COSY-type transfer between a pair of directly bonded carbons. In the HCCH-TOCSY spectrum longer-range  $^1\text{H}$ - $^1\text{H}$  correlations are observed; the number of  $^{13}\text{C}$ - $^{13}\text{C}$  transfers will depend on the length of the isotropic mixing time used. The peaks observed in the HCCH-COSY represent a subset of the peaks observed in the HCCH-TOCSY and often only the latter experiment is collected. Suitable sweep widths to use in the  $^{13}\text{C}$  dimension can be obtained by optimising the  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum.

##### Analysis of the 3D HCCH-TOCSY spectrum

1. Assignments of  $^{13}\text{C}\alpha/^{13}\text{C}\beta$  and  $^1\text{H}\alpha/^1\text{H}\beta$  obtained from experiments such as HNCA, CBCANH and HBHA(CO)NH are the starting point for the assignment of side chains using HCCH-TOCSY. For an assigned amino acid spin system look at the  $^{13}\text{C}$  planes of the HCCH-TOCSY corresponding to C $\alpha$  and C $\beta$  (Figure 15).
2. The  $^{13}\text{C}$  plane corresponding to C $\alpha$  will contain a diagonal peak at the  $^1\text{H}\alpha$  shift. There

will also be a cross peak corresponding to the correlation between  $H\alpha$  and  $H\beta$ . Similarly, the  $^{13}\text{C}$  plane corresponding to  $C\beta$  will contain one or two diagonal peaks at the  $^1\text{H}\beta$  shift(s) and cross peaks corresponding to the correlation with  $H\alpha$ . The observation of these peaks confirms that  $C\alpha$ ,  $C\beta$ ,  $H\alpha$  and  $H\beta$  belong to the same spin system, confirming the information obtained using the triple resonance experiments described above (Figures 15 and 16).

3. For some residue types, further cross peaks will be observed in the strips for  $C\alpha$  and  $C\beta$  identifying other protons belonging to the spin system. If the  $\text{H}(\text{CCCO})\text{NH}$  spectrum has been collected then these additional  $^1\text{H}$  shifts will correspond to peaks observed in this spectrum. If the  $(\text{H})\text{CC}(\text{CCO})\text{NH}$  spectrum has also been collected then the additional  $^{13}\text{C}$  peaks seen in this will correspond to  $^{13}\text{C}$  at the  $\gamma$ ,  $\delta$ ,  $\epsilon$  positions. Looking in the  $^{13}\text{C}$  planes corresponding to these additional  $^{13}\text{C}$  chemical shifts should identify further strips arising from the spin system (Figure 16).

4. If the  $(\text{H})\text{CC}(\text{CO})\text{NH}$  experiment has not been collected then it is possible to use the known random coil  $^{13}\text{C}$  chemical shift value for a particular carbon to act as a guide in the search for further correlations corresponding to the spin system. This is demonstrated in Figure 16.

5. Repeat this process starting with the  $C\alpha$  and  $C\beta$  planes for each amino acid spin system to obtain full side chain assignments.

6. During the triple resonance assignment process,  $C\alpha$ ,  $C\beta$ ,  $H\alpha$  and  $H\beta$  chemical shifts will have been identified for Pro residues from the residues that follow them in the sequence (the *i-1* peaks in strips for residues that follow Pro). These can be used as a starting point for assignment of the Pro spin systems in the  $\text{HCCH-TOCSY}$ .

#### 4.8. Triple resonance methods with deuteration

Triple resonance experiments permit the assignment of proteins that are significantly larger than those that can be assigned by  $^1\text{H-}^1\text{H}$  or  $^{15}\text{N}$ -edited methods because they rely on magnetisation transfer involving large one and two-bond scalar couplings. However, as the molecular weight increases above 30 kD, short transverse relaxation times will limit the transfer efficiency between coupled spins; this is particularly true for  $C\alpha$ . One solution to this problem involves 70-75% deuteration of the protein. Deuteration leads to significantly increased  $T_2$  values for carbons that normally have hydrogens attached and leads to improved transfer of magnetisation. A series of 3D experiments that takes advantage of the increased  $^{13}\text{C}$  relaxation times was developed [55]. The high level of deuteration precludes the use of experiments that involve correlations to  $H\alpha$  and  $H\beta$ . In these samples the only proton sites significantly populated correspond to the exchangeable  $\text{H}^{\text{N}}$ ; therefore, all experiments are of the out-and-back variety. The choice of a 70-75% deuteration level is a compromise between the advantage of increased relaxation times and the disadvantage of the significantly longer recycle delays that are required if the  $\text{H}^{\text{N}} T_1$  relaxation time is too long. The 3D experiments include constant time (CT)-HNCA, CT-HN(CO)CA, HN(CA)CB and HN(COCA)CB. The HN(CA)CB and HN(COCA)CB experiments differ from the previously described CACBNH and CACB(CO)NH experiments in that they are tuned to select for  $C\beta$  over  $C\alpha$ . Analysis of these 4 datasets can be carried out using the procedures described in sections 4.3 and 4.4. The improved relaxation properties of the deuterated protein makes backbone assignments of  $\text{H}^{\text{N}}$ ,  $\text{N}$ ,  $C\alpha$  and  $C\beta$  possible for significantly larger proteins (30-40 kD) than possible with the conventional experiments applied to non-deuterated samples. Side chain  $^1\text{H}$  assignments are not possible for the deuterated samples because of the low proton occupancy at aliphatic sites. These assignments can be attempted using the  $\text{HCCH-TOCSY}$  methods described in section 4.7 using non-deuterated protein.

#### 4.9 Triple resonance methods for proteins above 40kD

70-75% deuteration in combination with  $^{13}\text{C}$  and  $^{15}\text{N}$  double labelling can lead to significant improvements in 3D spectra as a result of significantly increased relaxation times for  $\text{C}\alpha$ ,  $\text{C}\beta$  and  $\text{H}^{\text{N}}$ . However, this approach does not significantly alter the relaxation properties of the backbone  $^{15}\text{N}$ ; this is dominated by the attached  $\text{H}^{\text{N}}$  which must be present for signal detection in all the experiments described above. Transverse relaxation optimised spectroscopy (TROSY) [10] has been developed to extend the molecular weight limit for resonance assignment. The experiment applies constructive use of interference between dipole-dipole and chemical shift anisotropy relaxation. This minimises  $^{15}\text{N}$  transverse relaxation during  $^{15}\text{N}$  evolution and  $\text{H}^{\text{N}}$  transverse relaxation during detection. Combined with high levels of deuteration the TROSY method can be used for assignment of backbone resonances of significantly larger proteins. The HSQC element in triple resonance experiments can be replaced by TROSY. A group of six 3D TROSY-based experiments has been introduced. These include TROSY-HNCA, TROSY-HNCO, TROSY-HN(CO)CA, TROSY-HN(CA)CO, TROSY-HNCACB and TROSY-HN(CO)CB [56,57]. Assignment based on these experiments is the same as described above in sections 4.3 and 4.4. A pair of 4D TROSY-based methods has also been developed [58]. The 4D HNCACO and HNCOCA experiments can be used for complete assignment of  $\text{H}^{\text{N}}$ ,  $\text{N}$ ,  $\text{C}^{\alpha}$  and  $\text{CO}$ . The HNCACO contains both intra and interresidue connectivities,  $\text{H}^{\text{N}}(\text{i})\text{-N}(\text{i})\text{-C}^{\alpha}(\text{i})\text{-CO}(\text{i})$  and  $\text{H}^{\text{N}}(\text{i})\text{-N}(\text{i})\text{-C}^{\alpha}(\text{i}-1)\text{-CO}(\text{i}-1)$  whereas the HNCOCA contains only interresidue connectivities.

The TROSY method in conjunction with deuteration makes possible the collection of triple resonance data for significantly larger proteins than possible before. However, the increase in the molecular weight leads to a proportional increase in the number of resonances in the spectrum for monomeric systems. Eventually overlap problems, particularly in the regions of the spectrum with a small intrinsic chemical shift range such as  $\text{C}\alpha$ , will make unambiguous assignment problematic. An increase in dimensionality will address this problem to some extent. There are several situations where this methodology will be particularly applicable. Multimeric proteins have the short relaxation times characteristic of large molecular weight proteins but the symmetry of the molecules leads to a reduction in the number of resonances in the spectrum. Protein complexes composed of a  $^{15}\text{N}$ - $^{13}\text{C}$ - $^2\text{H}$  labelled protein component and an unlabelled protein or nucleic acid will also be ideal systems for the TROSY-based methods. Selective labelling of a particular amino acid or group of amino acids can dramatically reduce the complexity of spectra. Another application will be to proteins in which only a portion of the sequence, corresponding to a single domain, has been labelled.

#### 5. THE SEQUENTIAL ASSIGNMENT METHOD

The sequential assignment method is fundamentally different from the triple resonance assignment strategy described above. It was developed in the early 1980's when uniform  $^{13}\text{C}/^{15}\text{N}$  labelling of proteins was not feasible and was originally based on analysis of 2D  $^1\text{H}$ - $^1\text{H}$  NMR spectra [3,4,59,60]. The sequential assignment method consists of two stages (Figure 17). The first involves identification of systems of spin-spin coupled resonances belonging to each amino acid residue. This was achieved using experiments such as COSY and TOCSY. During this process, for example, the spin systems of all the valine residues may be identified but there is nothing to distinguish one valine spin system from another. The second stage involves assignment of each spin system to a particular residue within the amino acid sequence of the protein. This cannot be achieved using COSY or TOCSY experiments because there is no resolved  $^1\text{H}$ - $^1\text{H}$  spin-spin coupling across the peptide bond.

Instead assignments are deduced from the through-space dipole-dipole interactions found in NOESY spectra. This method was used during the 1980's to assign the spectra of numerous proteins of up to ~10kD and in favourable cases for proteins up to ~15 kD.

Although originally developed for  $^1\text{H}$ - $^1\text{H}$  2D NMR spectra, the sequential assignment method can be adapted for use with  $^{15}\text{N}$ -labelled protein using 3D  $^{15}\text{N}$ -edited experiments such as TOCSY-HSQC, for stage one, and NOESY-HSQC, for stage two. Uniform labelling of recombinant proteins with  $^{15}\text{N}$  is relatively inexpensive (compared to uniform  $^{13}\text{C}/^{15}\text{N}$  labelling). The advantage of  $^{15}\text{N}$  labelling is that much improved resolution is achieved by spreading out the peaks that would be observed in 2D  $^1\text{H}$ - $^1\text{H}$  TOCSY and NOESY spectra into a third dimension using the  $^{15}\text{N}$  chemical shift of the backbone amide. The problems of overlap that make assignment of proteins of more than ~10 kD difficult using 2D  $^1\text{H}$ - $^1\text{H}$  methods are removed to a large extent using the  $^{15}\text{N}$ -edited TOCSY and NOESY methods. In favourable circumstances proteins of up to ~20 kD can be assigned using this approach. Here the sequential assignment approach will be outlined for  $^{15}\text{N}$ -labelled proteins using 3D NMR spectra.

### 5.1 2D and 3D experiments for $^{15}\text{N}$ -labelled proteins.

The 3D NMR experiments used in the  $^{15}\text{N}$ -based sequential assignment approach contain a  $^1\text{H}^{\text{N}}$ - $^{15}\text{N}$  HSQC (heteronuclear single quantum correlation) element. Before collecting 3D data, the acquisition parameters for the HSQC spectrum should be optimised in order to ensure good resolution in the 3D spectra (as described above in section 3).

There are four  $^{15}\text{N}$ -edited 3D experiments that are appropriate for sequential assignment. The most useful experiments are the  $^{15}\text{N}$ -edited TOCSY-HSQC and NOESY-HSQC [6,7]; these should be collected initially. The other two experiments, HSQC-NOESY-HSQC [61,62] and HNHA [63], may provide useful information but should only be collected after some analysis has been carried out using the first two data sets. The TOCSY-HSQC and the HNHA experiments provide through-bond spin system information useful in stage one of assignment. The NOESY-HSQC and HSQC-NOESY-HSQC experiments provide through-space NOE connectivities useful in stage two.

The TOCSY-HSQC and NOESY-HSQC spectra are similar to their 2D  $^1\text{H}$ - $^1\text{H}$  counterparts. These spectra contain a subset of the peaks in the  $^1\text{H}$  spectra; only TOCSY or NOESY effects involving at least one  $\text{H}^{\text{N}}$  are observed because the final HSQC component of the pulse sequences selects for  $^1\text{H}$  bonded to  $^{15}\text{N}$  (Figure 18). The  $^1\text{H}$ - $^1\text{H}$  TOCSY or NOESY peaks are separated into 'planes' on the basis of the  $^{15}\text{N}$  chemical shift of the backbone amide nitrogen. Thus, each  $^{15}\text{N}$  plane in the 3D spectrum contains a small subset of the information from the  $^1\text{H}$ - $^1\text{H}$  spectrum and overlap is reduced considerably.

Strip plots are a useful way to analyse these 3D data sets. For  $^{15}\text{N}$ -edited TOCSY-HSQC and NOESY-HSQC spectra strips are extracted from  $^{15}\text{N}$  planes. Each strip is centred about a  $\text{H}^{\text{N}}$  chemical shift in  $F_3$  and the strip covers the full  $^1\text{H}$  sweep width in the indirect  $^1\text{H}$  dimension,  $F_1$  (Figure 18). The strip plot can be created as described above. For the analysis of these spectra, it is convenient to order the strips according to the  $\text{H}^{\text{N}}$  chemical shift; the most downfield  $\text{H}^{\text{N}}$  is strip 1 and the most upfield  $\text{H}^{\text{N}}$  is the last strip. The numbering of the strips corresponds to the arbitrary peak numbers assigned in the HSQC spectrum.

### 5.2. Stage 1: Spin system identification from 3D TOCSY-HSQC and 3D HNHA

The first stage of sequential assignment involves the identification of systems of spin-spin coupled resonances that correspond to individual amino acid residues. It will be shown later that the analysis of NOE data for the second stage of sequential assignment relies most heavily on the  $\text{H}^{\text{N}}$  resonances. Therefore, the approach to spin system identification described here emphasises the  $^1\text{H}$ - $^{15}\text{N}$  HSQC. Information about the type of residue from which a

particular  $H^N$ - $^{15}N$  peak arises is collected using a 3D  $^{15}N$ -edited TOCSY-HSQC experiment. For many residues, such as glycine and alanine, complete spin system identification is quite straightforward even for large proteins. For other residues, including arginine and lysine, complete spin system identification is more difficult. However, the lack of complete spin systems for these residues does not necessarily hamper the sequential assignment process.

The TOCSY experiment uses isotropic mixing to transfer magnetisation between scalar coupled  $^1H$  spins within an amino acid residue. The amount of spin system information contained in a TOCSY spectrum will depend on the length of the TOCSY mixing time, the size of the coupling constants in the spin system, and resonance linewidths. With shorter TOCSY mixing times (20-25 ms) only  $H^N$ - $H\alpha$  peaks may be observed. As the mixing time is increased (up to ~100 ms) connectivities to  $H\beta$ ,  $H\gamma$  and other sidechain protons will appear. In general the intensity of  $H^N$ - $H\alpha$  peaks will depend on the  $^3J_{HNH\alpha}$  coupling constant; an intense peak is seen in the case of a large coupling constant and a weaker peak for a small coupling constant. Residues with a large  $H^N$ - $H\alpha$  coupling constant will give stronger intensity  $H^N$ - $H\beta$  connectivities than residues with small  $H^N$ - $H\alpha$  coupling constants. Therefore, the TOCSY-HSQC spectrum will usually contain more spin system information for a  $\beta$ -sheet protein than for an  $\alpha$ -helical one. Although the amount of spin system information will, in principle, increase as the TOCSY mixing time is increased, transverse relaxation will result in a decrease in signal at very long mixing times. As a general rule, longer-range connectivities, such as  $H^N$ - $H\delta$ , will only appear in TOCSY spectra once the shorter-range connectivities such as  $H^N$ - $H\beta$  have appeared. However, different behaviour may be observed for the same amino acid type at different positions in the sequence as a consequence of differences in coupling constants. For example, a Val residue with large  $H^N$ - $H\alpha$  and  $H\alpha$ - $H\beta$  coupling constants will show  $H^N$ - $H\gamma$  peaks in a TOCSY spectrum collected with a mixing time of 50 ms while a Val residue with small coupling constants may only show an  $H^N$ - $H\alpha$  peak in this spectrum. To optimise the mixing time prior to collection of 3D data, 2D  $^1H$ - $^1H$  TOCSY planes can be collected with different mixing times. If spectrometer time allows, the 3D experiment can be collected with two different mixing times. On the basis of the rules outlined below each  $H^N$ - $^{15}N$  pair can be assigned to a more or less specific spin system category.

#### Rules for analysing 3D $^{15}N$ -edited TOCSY-HSQC spectra

1. Collect one or more 3D  $^{15}N$ -edited TOCSY-HSQC spectra with mixing times ranging from ~25 to 100 ms. If a single experiment is being collected then a mixing time of ~40 ms is a good compromise. Create a strip plot as described above. This allows spin system information to be assigned for each peak in the HSQC.
2. Spin system assignment is based on the  $^1H$  random coil chemical shifts shown in Figure 10.
3. If a single peak is observed upfield of ~1.75 ppm (in addition to  $H\alpha$ ) at all mixing times then the spin system is likely to be an Ala. The peak should have a narrow shape in the  $CH_3$  dimension. The peak observed downfield of ~3.5 ppm will correspond to  $H\alpha$ .
4. If 2 peaks are observed downfield of 3.5 ppm then the spin system is a Gly, Ser or Thr. If 3 peaks are observed then the spin system is a Ser. If 2 peaks are observed downfield of 3.5 ppm and an additional peak is observed upfield of ~1.6 ppm then the spin system is a Thr.
5. If 1 or 2 peaks are observed in the region of 2.5-3.5 ppm and no peaks are observed further upfield then the spin system is either an Asp, Asn, Cys, His, Phe, Trp or Tyr. The peak observed downfield of ~3.5 ppm will correspond to  $H\alpha$ .
6. If at least 1 peak is observed upfield of ~2.25 ppm then the spin system is either Arg, Gln, Glu, Ile, Leu, Lys, Met or Val.

(a) If an additional peak is observed downfield of this peak at longer mixing times then the spin system is likely to be Glu, Gln or Met, which have  $H_\gamma$  downfield of  $H_\beta$ .

(b) If peaks are observed upfield of  $\sim 1.5$  ppm with a narrow shape then the spin system is likely to be Ile, Leu or Val. If more than 3 peaks upfield of 2.5 ppm are observed then the spin system is Ile or Leu.

(c) If more than 2 peaks are observed between 1.5 and 2.25 ppm the spin system is likely to be Arg or Lys. At long mixing times 1 or 2 peaks at  $\sim 3$  ppm may be observed. A total of 6 and 8 peaks in addition to the  $H^N$ - $H_\alpha$  peaks are expected for Arg and Lys, respectively.

7. If no peaks are observed other than the  $H^N$ - $H_\alpha$  peak then the spin system cannot be classified.

The 3D HNHA spectrum contains  $H^N$ - $H^N$  diagonal peaks and  $H^N$ - $H_\alpha$  cross peaks. The ratio of these peak intensities provides quantitative information about the  $^3J_{HNH_\alpha}$  coupling constant [63]. In addition, this data set can be used to identify, unambiguously,  $H_\alpha$  shifts for each residue. For example, comparison of HNHA and TOCSY-HSQC spectra may be useful for distinguishing Gly, which will give a pair of  $H_\alpha$  peaks in both spectra, from Ser or Thr, which give a single  $H_\alpha$  peak in the HNHA but may give 2 or 3 peaks downfield of 3.5 ppm in the TOCSY-HSQC spectrum due to their downfield shifted  $H_\beta$ .

### 5.3 Stage 2: Sequence specific assignment from 3D NOESY-HSQC and HSQC-NOESY-HSQC

The second stage of assignment involves the assignment of an amino acid spin system identified in stage 1 to a specific residue in the protein sequence. This is achieved by correlating an amino acid spin system with the spin systems of its neighbouring residues in the sequence, and relies on the through-space interactions observed in NOESY spectra.

In 1982 Wüthrich and coworkers showed that, for all sterically allowed values of  $\phi$ ,  $\psi$  and  $\chi_1$  at least one of the distances between  $H^N$ ,  $H_\alpha$  and  $H_\beta$  of residue  $i$  and  $H^N$  of residue  $i+1$  is short enough to give rise to an observable NOE effect (Figure 17) [3-5]. The most useful NOE effects for sequential assignment were found to involve the  $H_\alpha$  of residue  $i$  and the  $H^N$  of residue  $i+1$ , abbreviated as  $d_{\alpha N}(i, i+1)$ , the  $H^N$ 's of residues  $i$  and  $i+1$ , abbreviated as  $d_{NN}(i, i+1)$ , and the  $H_\beta$  of residue  $i$  and the  $H^N$  of residue  $i+1$ , abbreviated as  $d_{\beta N}(i, i+1)$ . The intensities of these three sequential NOEs depend on the torsion angles  $\psi$ ,  $\phi$  &  $\psi$ , and  $\chi_1$  &  $\psi$ , respectively. This torsion angle dependence means that specific types of secondary structure are characterised by specific sequential NOE effects. In the extended backbone structure, characteristic of  $\beta$ -sheet, the distance  $d_{\alpha N}(i, i+1)$  is short, 2.2 Å, whereas  $d_{NN}(i, i+1)$  is longer, 4.3 Å. In helical structure, the distance  $d_{NN}(i, i+1)$  is short, 2.8 Å, whereas  $d_{\alpha N}(i, i+1)$  is longer, 3.5 Å [4].

The problem with the sequential assignment approach is that NOE effects between  $H^N$ ,  $H_\alpha$  and  $H_\beta$  resonances are not restricted to adjacent residues of the sequence. NOE effects involving these protons also arise as a result of longer-range contacts in regions of secondary and tertiary structure. Wüthrich and coworkers showed that if both a  $d_{\alpha N}$  and a  $d_{NN}$  NOE or a  $d_{\alpha N}$  and a  $d_{\beta N}$  NOE effect are observed between two residues then in  $>90\%$  of cases these two residues are adjacent in the sequence if an upper limit of 3.8 Å is used for the observation of an NOE effect [3,4]. Thus, the identification of two of the three NOE effects is a more reliable criterion for sequential assignment.

If a strong  $d_{\alpha N}$  NOE is observed between two spin systems then these are likely to be adjacent in the sequence. If the resonances of the spin systems have been assigned to a specific amino acid type, such as Gly and Ala, then sequential assignment may be possible from this single NOE effect because the pair Gly-Ala may occur only once in the sequence.

However, as the size of the protein increases the occurrence of unique pairs of residues in the sequence decreases. Thus, it may be necessary to identify the residue adjacent to the Gly or Ala before a specific assignment can be made. If the resonances of the spin systems have been assigned to a more general class of amino acids, such as Asp/Asn/Cys/His/Phe/Trp/Tyr, or if no information about the spin system is known then sequential assignment of the basis of the single NOE is very unlikely. Additional NOE's to the adjacent residues will be needed before an assignment can be made. The procedure for carrying out the second stage of assignment for proteins of this kind follows.

The 3D NOESY-HSQC spectrum contains  $^1\text{H}$ - $^1\text{H}$  NOEs involving at least one  $\text{H}^{\text{N}}$  (Figure 18). The NOE effects that can be identified most unambiguously from this spectrum are the  $\text{H}^{\text{N}}$ - $\text{H}^{\text{N}}$  NOEs because they appear twice in the dataset. For example, an NOE between  $\text{H}^{\text{N}}(i)$  and  $\text{H}^{\text{N}}(j)$  will be observed in both the strip corresponding to  $\text{H}^{\text{N}}(i)$  and that of  $\text{H}^{\text{N}}(j)$ . NOEs that involve a  $\text{H}^{\text{N}}$  and an  $\text{H}\alpha$  or  $\text{H}\beta$  are only observed in the strip corresponding to the  $\text{H}^{\text{N}}$ . The methods for analysing the NOESY-HSQC spectrum are outlined below.

### Sequential assignment from 3D NOESY-HSQC

1. Identify intraresidue NOEs in the NOESY-HSQC strip plot by comparison with the TOCSY-HSQC strips; any peaks present in both the NOESY and TOCSY spectra are likely to be intraresidue connectivities (Figures 18 and 19). Label the  $\text{H}\alpha$  for each residue and any  $\text{H}\beta$  or other side chain resonances that have been identified in the TOCSY spectrum. Further analysis of the NOESY spectrum does not require overlaying of the NOESY and TOCSY spectra because the intraresidue NOE effects have now been labelled in the NOESY spectrum.

2. Starting with the most downfield  $\text{H}^{\text{N}}$  (i.e. the first strip) look in the region between 6 and 10 ppm in the indirect  $^1\text{H}$  dimension for NOE cross peaks. For each cross peak observed look for a strip which has its diagonal peak at this  $^1\text{H}$  frequency and a cross peak with a  $^1\text{H}$  frequency corresponding to the original diagonal peak. These two symmetric strips represent a pair of  $\text{H}^{\text{N}}$ s, which are correlated by a  $d_{\text{NN}}$  NOE (Figure 20). If the cross peaks are strong then the pair of  $\text{H}^{\text{N}}$ 's are likely to represent sequential residues located in a helix or turn. If the cross peaks are weak then the NOE may be an interstrand contact in a  $\beta$ -sheet. The sequential connectivity can be confirmed by identifying additional sequential NOEs,  $d_{\text{NN}}(i, i+1)$  and  $d_{\text{NN}}(i+1, i+2)$  (Figure 19). Place the two strips side by side and try to identify an NOE between the  $\text{H}^{\text{N}}$  of one residue and the  $\text{H}\alpha$  and/or  $\text{H}\beta$  of the other residue. Identification of such an NOE will confirm the assignment of the residues as neighbours in the sequence and will distinguish residue  $i$  from residue  $i+1$ . If appropriate  $d_{\text{NN}}(i, i+1)$  or  $d_{\text{NN}}(i+1, i+2)$  NOEs can not be seen for the two strips then the  $\text{H}^{\text{N}}$ - $\text{H}^{\text{N}}$  NOE may not be sequential.

3. Continue this procedure for all the NOEs observed in the region between 6 and 10 ppm. This will result in a list for each  $\text{H}^{\text{N}}$  resonance of 0, 1 or 2 neighbouring  $\text{H}^{\text{N}}$ s. Note that pairs of Asn and Gln side chain  $\text{H}^{\text{N}}$ s also give rise to a strong  $\text{H}^{\text{N}}$ - $\text{H}^{\text{N}}$  NOE cross peak.

4. If a strong cross peak is observed in the 6-10 ppm region but a reciprocal strip can not be identified then the cross peak involves an aromatic proton rather than a second  $\text{H}^{\text{N}}$ . NOEs between  $\text{H}^{\text{N}}$ 's and aromatic side chains are often intraresidue or sequential so observation of this NOE may provide additional spin system information.

5. From the sequential  $d_{\text{NN}}$  NOEs tabulated in 3 start to put together longer stretches of  $d_{\text{NN}}$  connectivities. Breaks in these segments will occur at Pro, at the end of helical regions or because of the degeneracy of neighbouring  $\text{H}^{\text{N}}$  shifts. However, because of the improved resolution compared to 2D  $^1\text{H}$ - $^1\text{H}$  methods longer segments of unambiguous connectivities should be seen. For each stretch of residues compare the spin system classification based on the TOCSY with the sequence of the protein. If the segment is long then even very limited



spin system information will result in a unique assignment.

6. When all possible  $d_{NN}$  NOEs have been identified the remaining assignments must be based on the strong sequential  $d_{\alpha N}$  NOEs that are expected in extended structure. Strips which did not show one or more strong  $d_{NN}$  NOEs are likely to contain a strong interresidue  $d_{\alpha N}$  NOE. Because of the lack of symmetry in these NOEs unambiguous assignment may be more difficult. Sequential assignment based on  $d_{\alpha N}$  NOEs requires that strips are matched up according to  $H\alpha$  shifts (Figure 19). A second strip plot in which strips are plotted in order of decreasing  $H\alpha$  shifts can aid in this stage of analysis.

7. For each strip containing a strong interresidue  $d_{\alpha N}$  NOE look for a strip with an intraresidue  $d_{\alpha N}$  NOE at this chemical shift; there may be several such strips. Confirm that this is a sequential NOE by identifying the appropriate intra and interresidue  $d_{\alpha N}$  NOEs (Figure 19).

8. From the sequential  $d_{\alpha N}$  NOEs tabulated in 6/7 start to put together longer stretches of  $d_{\alpha N}$  connectivities. Breaks in these segments will occur at Pro or because of the degeneracy of  $H\alpha$  and  $H\beta$  shifts. For each stretch of residues compare the spin system classification based on the TOCSY with the sequence of the protein. If the segment is long then even very limited spin system information will result in a unique assignment.

In proteins with a high helical content and poor chemical shift dispersion in the  $H^N$  region identification of sequential  $d_{NN}$  NOEs may be difficult. The 3D HSQC-NOESY-HSQC spectrum may be useful for these proteins. In this experiment only  $d_{NN}$  NOEs are observed. The indirect  $^1H$  dimension of the NOESY-HSQC is replaced by a second  $^{15}N$  dimension; the  $^{15}N$  frequency of one residue is correlated with the  $^{15}N$  and  $H^N$  frequencies of a second residue. This experiment allows  $d_{NN}$  NOEs to be observed between residues with degenerate  $H^N$  chemical shifts as long as they have distinct  $^{15}N$  shifts (Figure 21). This experiment is not useful for  $\beta$ -sheet proteins where very few  $H^N$ - $H^N$  NOEs are observed.

The assignment of the spectra of  $\beta$ -sheet proteins using  $^{15}N$ -edited methods can be difficult because of ambiguities in the identification of  $d_{\alpha N}$  NOEs. The  $d_{NN}$  NOEs that can be identified unambiguously often represent long-range contacts that may not be useful in the sequential assignment process. The main-chain directed (MCD) strategy, developed by Englander and Wand for analysis of  $^1H$  NOE spectra [64], can be useful for analysis of 3D NOESY-HSQC spectra of  $\beta$ -sheet proteins. This strategy uses cyclic patterns involving residues on adjacent strands of the  $\beta$ -sheet (Figure 22). The two MCD patterns for the antiparallel  $\beta$ -sheet are designated as the outer and inner loops (Figure 22a). The outer loop involves 2 TOCSY peaks (or intraresidue NOEs) and 3 interresidue NOEs:  $d_{\alpha N}(i,i)$ ,  $d_{\alpha N}(i+1,i+1)$ ,  $d_{\alpha N}(i,i+1)$ ,  $d_{\alpha N}(i+1,j)$  and  $d_{NN}(i,j)$ , where  $i$  and  $j$  are residues on opposite strands. This loop contains one interstrand  $d_{NN}$  NOE that can be identified unambiguously in the NOESY-HSQC. The inner loop involves 4 NOE effects:  $d_{\alpha N}(i+1,i+2)$ ,  $d_{\alpha N}(i+1,j)$ ,  $d_{\alpha N}(j-1,j)$ , and  $d_{\alpha N}(j-1,i+2)$ ; the  $d_{\alpha N}(i+1,j-1)$  NOE usually associated with this loop won't be observed in a  $^{15}N$ -edited experiment. The inner and outer loops share a common NOE,  $d_{\alpha N}(i+1,j)$ , and the  $H^N$  of residue  $i+2$ , identified in the inner loop, is the starting point for the next outer loop with a  $d_{NN}$  NOE to residue  $j-2$ . The MCD pattern for parallel  $\beta$ -sheet involves 2 TOCSY peaks (or intraresidue NOE's) and 4 interresidue NOEs:  $d_{\alpha N}(i,i)$ ,  $d_{\alpha N}(i+1,i+1)$ ,  $d_{\alpha N}(i,i+1)$ ,  $d_{\alpha N}(i+1,j+1)$ ,  $d_{\alpha N}(j,j+1)$  and  $d_{\alpha N}(j,i)$  (Figure 22b). Adjacent MCD loops in the parallel sheet share a common interstrand  $d_{\alpha N}$  NOE. The MCD pattern of the parallel sheet is less likely to give rise to unambiguous assignments because it does not involve a  $d_{NN}$  NOE as a starting point. Nevertheless, if assignment using the sequential approach proves difficult the MCD approach should be considered.

The  $^{15}N$ -based sequential assignment method can be used to assign proteins up to ~20

kD in size. The method works particularly well for helical proteins because of the strong unambiguous  $^1\text{H}^{\text{N}}\text{-}^1\text{H}^{\text{N}}$  NOEs that are observed. However, this method will not always be successful. In proteins with poor chemical shift dispersion, the degeneracy of  $^1\text{H}$  resonances will lead to ambiguities that may not be resolvable. In these circumstances double labelling with  $^{15}\text{N}$  and  $^{13}\text{C}$  and the triple resonance strategy described above may be required.

## 6. CONCLUSION

The assignment of resonances in the complex NMR spectrum of a protein is the first step in any NMR study of protein structure, function or dynamics. In this chapter the two widely used methods for resonance assignment have been presented. The sequential assignment methodology was introduced in the early 1980's and provided a systematic approach to resonance assignment that did not rely on information derived from an X-ray structure of the protein. 3D  $^{15}\text{N}$ -edited TOCSY-HSQC and NOESY-HSQC methods introduced in the late 1980's were suited to the sequential assignment method and extended the molecular weight limit of proteins that could be tackled by increasing spectra resolution. However, the reliance of this method on NOEs to make connections between neighbouring residues meant that it wasn't entirely 'fool proof'. The triple resonance assignment method introduced in the early 1990's uses  $^{13}\text{C}/^{15}\text{N}$  uniformly labelled protein and relies only on through-bond scalar couplings. This is now the most widely used approach and when combined with deuteration and TROSY, instead of HSQC, can be used to assign the spectra of proteins up to and beyond 40 kD.

## REFERENCES

1. Aue WP, Bartholdi E, Ernst RR. 1976. 2-Dimensional Spectroscopy - Application to Nuclear Magnetic-Resonance. *J Chem Phys* **64**:2229-2246.
2. Jeener J, Meier BH, Bachmann P, Ernst RR. 1979. Investigation of Exchange Processes by 2-Dimensional NMR-Spectroscopy. *J Chem Phys* **71**:4546-4553.
3. Wuthrich K, Wider G, Wagner G, Braun W. 1982. Sequential Resonance Assignments as a Basis for Determination of Spatial Protein Structures by High-Resolution Proton Nuclear Magnetic-Resonance. *J Mol Biol* **155**:311-319.
4. Billeter M, Braun W, Wuthrich K. 1982. Sequential Resonance Assignments in Protein H-1 Nuclear Magnetic-Resonance Spectra - Computation of Sterically Allowed Proton Proton Distances and Statistical-Analysis of Proton Proton Distances in Single-Crystal Protein Conformations. *J Mol Biol* **155**:321-346.
5. Wuthrich K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York
6. Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, Clore GM. 1989. Overcoming the Overlap Problem in the Assignment of H-1-NMR Spectra of Larger Proteins by Use of 3-Dimensional Heteronuclear H-1-N-15 Hartmann-Hahn Multiple Quantum Coherence and Nuclear Overhauser Multiple Quantum Coherence Spectroscopy - Application to Interleukin-1-Beta. *Biochemistry* **28**:6150-6156.
7. Messerle BA, Wider G, Otting G, Weber C, Wuthrich K. 1989. Solvent Suppression Using a Spin Lock in 2D and 3D NMR-Spectroscopy with H<sub>2</sub>O Solutions. *J Magn Reson* **85**:608-613.
8. Ikura M, Kay LE, Bax A. 1990. A Novel-Approach for Sequential Assignment of H-1, C-13, and N-15 Spectra of Larger Proteins - Heteronuclear Triple-Resonance 3-

- Dimensional NMR-Spectroscopy - Application to Calmodulin. *Biochemistry* **29**:4659-4667.
9. Kay LE, Ikura M, Tschudin R, Bax A. 1990. 3-Dimensional Triple-Resonance NMR-Spectroscopy of Isotopically Enriched Proteins. *J Magn Reson* **89**:496-514.
  10. Pervushin K, Riek R, Wider G, Wuthrich K. 1997. Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* **94**:12366-12371.
  11. Wishart DS, Sykes BD. 1994. The C-13 Chemical-Shift Index - a Simple Method for the Identification of Protein Secondary Structure Using C-13 Chemical-Shift Data. *J Biomol NMR* **4**:171-180.
  12. Cornilescu G, Delaglio F, Bax A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**:289-302.
  13. Cheung MS, Maguire ML, Stevens TJ, Broadhurst RW. 2010. DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J Magn Reson* **202**:223-233.
  14. Shen Y, Bax A. 2013. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* **56**:227-241.
  15. Driscoll PC, Hill HAO, Redfield C. 1987. H-1-NMR Sequential Assignments and Cation-Binding Studies of Spinach Plastocyanin. *Eur J Biochem* **170**:279-292.
  16. Redfield C, Dobson CM. 1988. Sequential H-1-NMR Assignments and Secondary Structure of Hen Egg-White Lysozyme in Solution. *Biochemistry* **27**:122-136.
  17. Redfield C, Smith LJ, Boyd J, Lawrence GMP, Edwards RG, Smith RAG, Dobson CM. 1991. Secondary Structure and Topology of Human Interleukin-4 in Solution. *Biochemistry* **30**:11029-11033.
  18. John Cavanagh WJF, Arthur G. Palmer, Mark Rance, Nicholas J. Skelton. (2007) *Protein NMR Spectroscopy - Principles and Practice*, Second Edition ed., Elsevier Academic Press
  19. Sattler M, Schleucher J, Griesinger C. 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Mag Res Sp* **34**:93-158.
  20. Bartels C, Guntert P, Billeter M, Wuthrich K. 1997. GARANT - A general algorithm for resonance assignment of multidimensional nuclear magnetic resonance spectra. *J Comput Chem* **18**:139-149.
  21. Moseley HNB, Monleon D, Montelione GT. 2001. Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. *Nuclear Magnetic Resonance of Biological Macromolecules, Pt B* **339**:91-108.
  22. Herrmann T, Guntert P, Wuthrich K. 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* **319**:209-227.
  23. Jung YS, Zweckstetter M. 2004. Mars - robust automatic backbone assignment of proteins. *J Biomol NMR* **30**:11-23.
  24. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas P, Ulrich EL, Markley JL, Ionides J, Laue ED. 2005. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins-Structure Function and Bioinformatics* **59**:687-696.

25. Serrano P, Pedrini B, Mohanty B, Geralt M, Herrmann T, Wuthrich K. 2012. The J-UNIO protocol for automated protein structure determination by NMR in solution. *J Biomol NMR* **53**:341-354.
26. Bodenhausen GR, D. 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett* **69**:185-189.
27. Akke M, Carr PA, Palmer AG, 3rd. 1994. Heteronuclear-correlation NMR spectroscopy with simultaneous isotope filtration, quadrature detection, and sensitivity enhancement using z rotations. *Journal of magnetic resonance. Series B* **104**:298-302.
28. Buck M, Boyd J, Redfield C, Mackenzie DA, Jeenes DJ, Archer DB, Dobson CM. 1995. Structural Determinants of Protein Dynamics - Analysis of N-15 NMR Relaxation Measurements for Main-Chain and Side-Chain Nuclei of Hen Egg-White Lysozyme. *Biochemistry* **34**:4041-4055.
29. Grzesiek S, Bax A. 1993. The Importance of Not Saturating H<sub>2</sub>O in Protein NMR - Application to Sensitivity Enhancement and Noe Measurements. *J Am Chem Soc* **115**:12593-12594.
30. Kay LE. 1995. Pulsed field gradient multi-dimensional NMR methods for the study of protein structure and dynamics in solution. *Prog Biophys Mol Bio* **63**:277-299.
31. Bai YW, Milne JS, Mayne L, Englander SW. 1993. Primary Structure Effects on Peptide Group Hydrogen-Exchange. *Proteins* **17**:75-86.
32. Connelly GP, Bai YW, Jeng MF, Englander SW. 1993. Isotope Effects in Peptide Group Hydrogen-Exchange. *Proteins* **17**:87-92.
33. Bermel W, Bruix M, Felli IC, Kumar MVV, Pierattelli R, Serrano S. 2013. Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins. *J Biomol NMR* **55**:231-237.
34. Bermel W, Felli IC, Gonnelli L, Kozminski W, Piai A, Pierattelli R, Zawadzka-Kazimierzczuk A. 2013. High-dimensionality C-13 direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. *J Biomol NMR* **57**:353-361.
35. Bertini I, Jimenez B, Pierattelli R, Wedd AG, Xiao Z. 2008. Protonless C-13 direct detection NMR: Characterization of the 37 kDa trimeric protein CutA1. *Proteins-Structure Function and Bioinformatics* **70**:1196-1205.
36. Bertini I, Luchinat C, Parigi G, Pierattelli R. 2008. Perspectives in paramagnetic NMR of metalloproteins. *Dalton T*:3782-3790.
37. Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. 1995. H-1, C-13 and N-15 Random Coil NMR Chemical-Shifts of the Common Amino-Acids .1. Investigations of Nearest-Neighbor Effects (Vol 5, Pg 67, 1995). *J Biomol NMR* **5**:332-332.
38. Grzesiek S, Bax A. 1992. Improved 3d Triple-Resonance NMR Techniques Applied to a 31-Kda Protein. *J Magn Reson* **96**:432-440.
39. Clubb RT, Thanabal V, Wagner G. 1992. A Constant-Time 3-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue H-1(N), N-15, and C-13(1) Chemical-Shifts in N-15-C-13-Labeled Proteins. *J Magn Reson* **97**:213-217.
40. Bax A, Ikura M. 1991. An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins. *J Biomol NMR* **1**:99-104.
41. Spera SB, A. 1991. Empirical Correlation between Protein Backbone Conformation and Ca and Cb 13C Nuclear Magnetic Resonance Chemical Shifts. *J Am Chem Soc* **113**:5490-5492.

42. Grzesiek S, Bax A. 1992. An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins. *J Magn Reson* **99**:201-207.
43. Grzesiek S, Bax A. 1992. Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. *J Am Chem Soc* **114**:6291-6293.
44. Wang AC, Lodi PJ, Qin J, Vuister GW, Gronenborn AM, Clore GM. 1994. An Efficient Triple-Resonance Experiment for Proton-Directed Sequential Backbone Assignment of Medium-Sized Proteins. *J Magn Reson Ser B* **105**:196-198.
45. Grzesiek S, Bax A. 1993. Amino-Acid Type Determination in the Sequential Assignment Procedure of Uniformly C-13/N-15-Enriched Proteins. *J Biomol NMR* **3**:185-204.
46. Grzesiek S, Anglister J, Bax A. 1993. Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in C-13/N-15-Enriched Proteins by Isotropic Mixing of C-13 Magnetization. *J Magn Reson Ser B* **101**:114-119.
47. Montelione GT, Lyons BA, Emerson SD, Tashiro M. 1992. An Efficient Triple Resonance Experiment Using C-13 Isotropic Mixing for Determining Sequence-Specific Resonance Assignments of Isotopically-Enriched Proteins. *J Am Chem Soc* **114**:10974-10975.
48. Lyons BA, Montelione GT. 1993. An HcncH Triple-Resonance Experiment Using C-13 Isotropic Mixing for Correlating Backbone Amide and Side-Chain Aliphatic Resonances in Isotopically Enriched Proteins. *J Magn Reson Ser B* **101**:206-209.
49. Logan TM, Olejniczak ET, Xu RX, Fesik SW. 1993. A General-Method for Assigning NMR-Spectra of Denatured Proteins Using 3d Hc(Co)Nh-Tocsy Triple Resonance Experiments. *J Biomol NMR* **3**:225-231.
50. Yamazaki T, Formankay JD, Kay LE. 1993. 2-Dimensional NMR Experiments for Correlating C-13-Beta and H-1-Delta/Epsilon Chemical-Shifts of Aromatic Residues in C-13-Labeled Proteins Via Scalar Couplings. *J Am Chem Soc* **115**:11054-11055.
51. McIntosh LPB, E.;Kay, L. E. 1997. Stereospecific assignments of the NH2 resonances from the primary amides of asparagine and glutamine side chains in isotopically labeled proteins. *J Biomol NMR* **9**:306-312.
52. Kay LE, Ikura M, Bax A. 1990. Proton Proton Correlation Via Carbon Carbon Couplings - a 3-Dimensional NMR Approach for the Assignment of Aliphatic Resonances in Proteins Labeled with C-13. *J Am Chem Soc* **112**:888-889.
53. Fesik SW, Eaton HL, Olejniczak ET, Zuiderweg ERP, McIntosh LP, Dahlquist FW. 1990. 2d and 3d NMR-Spectroscopy Employing C-13-C-13 Magnetization Transfer by Isotropic Mixing - Spin System-Identification in Large Proteins. *J Am Chem Soc* **112**:886-888.
54. Bax A, Clore GM, Gronenborn AM. 1990. H-1-H-1 Correlation Via Isotropic Mixing of C-13 Magnetization, a New 3-Dimensional Approach for Assigning H-1 and C-13 Spectra of C-13-Enriched Proteins. *J Magn Reson* **88**:425-431.
55. Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR, Kay LE. 1994. A Suite of Triple-Resonance NMR Experiments for the Backbone Assignment of N-15, C-13, H-2 Labeled Proteins with High-Sensitivity. *J Am Chem Soc* **116**:11655-11666.
56. Salzmann M, Pervushin K, Wider G, Senn H, Wuthrich K. 1998. TROSY in triple-resonance experiments: New perspectives for sequential NMR assignment of large proteins. *P Natl Acad Sci USA* **95**:13585-13590.
57. Salzmann M, Wider G, Pervushin K, Senn H, Wuthrich K. 1999. TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *J Am Chem Soc* **121**:844-848.

58. Yang DW, Kay LE. 1999. TROSY triple-resonance four-dimensional NMR spectroscopy of a 46 ns tumbling protein. *J Am Chem Soc* **121**:2571-2575.
59. Arseniev AS, Wider G, Joubert FJ, Wuthrich K. 1982. Assignment of the H-1 Nuclear Magnetic-Resonance Spectrum of the Trypsin Inhibitor-E from Dendroaspis-Polylepis-Polylepis Two-Dimensional Nuclear Magnetic-Resonance at 500 MHz. *J Mol Biol* **159**:323-351.
60. Wider G, Lee KH, Wuthrich K. 1982. Sequential Resonance Assignments in Protein H-1 Nuclear Magnetic-Resonance Spectra - Glucagon Bound to Perdeuterated Dodecylphosphocholine Micelles. *J Mol Biol* **155**:367-388.
61. Frenkiel T, Bauer C, Carr MD, Birdsall B, Feeney J. 1990. Hmqc-Noesy-Hmqc, a 3-Dimensional NMR Experiment Which Allows Detection of Nuclear Overhauser Effects between Protons with Overlapping Signals. *J Magn Reson* **90**:420-425.
62. Ikura M, Bax A, Clore GM, Gronenborn AM. 1990. Detection of Nuclear Overhauser Effects between Degenerate Amide Proton Resonances by Heteronuclear 3-Dimensional Nuclear-Magnetic-Resonance Spectroscopy. *J Am Chem Soc* **112**:9020-9022.
63. Vuister GW, Bax A. 1993. Quantitative J Correlation - a New Approach for Measuring Homonuclear 3-Bond J(H(N)H(Alpha) Coupling-Constants in N-15-Enriched Proteins. *J Am Chem Soc* **115**:7772-7777.
64. Englander SW, Wand AJ. 1987. Main-Chain-Directed Strategy for the Assignment of H-1-NMR Spectra of Proteins. *Biochemistry* **26**:5953-5958.

Peak Number	$H^N(i)$ (ppm)	$^{15}N(i)$ (ppm)	$^{13}C\alpha(i-1)$ (ppm)	$^{13}C\alpha(i)$ (ppm)	$^{13}CO(i-1)$ (ppm)	$^{13}CO(i)$ (ppm)
1	9.63	109.47	53.04	48.76	175.07	176.78
2	9.41	128.93	62.27	60.18	174.13	173.43
3	9.39	123.06	48.76	58.97	176.78	178.96
4	9.03	131.76	52.93	50.19	172.54	175.13
5	8.26	128.80	60.18	52.93	173.43	172.54

**Table 1.** Analysis of chemical shifts obtained from the HNCA, HNCO, HN(CO)CA and HN(CA)CO strip plots. Two  $H^N$ - $^{15}N$  peaks,  $j$  and  $k$ , correspond to neighbouring residues if their peak lists share a pair of  $C\alpha$  and a pair of CO shifts. Residue  $j$  precedes residue  $k$  if the shifts  $C\alpha(j) = C\alpha(k-1)$  and  $CO(j) = CO(k-1)$ . Residue  $k$  precedes residue  $j$  if the shifts  $C\alpha(k) = C\alpha(j-1)$  and  $CO(k) = CO(j-1)$ . Peaks 1 and 3 share the  $C\alpha$  chemical shift value of 48.76 ppm and the CO value of 176.78 ppm. Peak 3 has  $C\alpha(i-1)$  and  $CO(i-1)$  values of 48.76 and 176.78 ppm; therefore, residue 3 is preceded by residue 1. Using a similar logic, the other three peaks can be arranged in the order 2-5-4.

Peak Number	$H^N(i)$ (ppm)	$^{15}N(i)$ (ppm)	$^{13}C\alpha(i-1)$ (ppm)	$^{13}C\alpha(i)$ (ppm)	$^{13}C\beta(i-1)$ (ppm)	$^{13}C\beta(i)$ (ppm)	$H\alpha(i-1)$	$H\beta(i-1)$
1	9.63	109.47	53.04	48.76	40.67	-	5.05	3.61/2.98
2	9.41	128.93	62.27	60.18	70.70	34.62	5.26	4.10
3	9.39	123.06	48.76	58.97	-	42.07	4.83/3.87	
4	9.03	131.76	52.93	50.19	44.39	20.42	5.07	1.81/1.10
5	8.26	128.80	60.18	52.93	34.62	44.39	4.58	1.70

**Table 2.** Analysis of chemical shifts obtained from the CBCANH, CBCA(CO)NH and HBHA(CO)NH strip plots. Two  $H^N$ - $^{15}N$  peaks,  $j$  and  $k$ , correspond to neighbouring residues if their peak lists share a pair of  $C\alpha$  and a pair of  $C\beta$  shifts. Residue  $j$  precedes residue  $k$  if the shifts  $C\alpha(j) = C\alpha(k-1)$  and  $C\beta(j) = C\beta(k-1)$ . Residue  $k$  precedes residue  $j$  if the shifts  $C\alpha(k) = C\alpha(j-1)$  and  $C\beta(k) = C\beta(j-1)$ . Peaks 2 and 5 share the  $C\alpha$  chemical shift value of 60.18 ppm and the  $C\beta$  value of 34.62 ppm. Peak 5 has  $C\alpha(i-1)$  and  $C\beta(i-1)$  values of 60.18 and 34.62 ppm; therefore, residue 5 is preceded by residue 2. Using a similar logic, peak 4 follows peak 5 and peak 3 follows peak 1.

## FIGURE LEGENDS

**Figure 1.** 500 MHz  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of CheY3 at pH 4.5 and 20 °C. CheY3 is a 14 kDa globular protein composed of 122 residues. The HSQC spectrum contains peaks from backbone amides of all amino acids except proline. Side-chain nitrogens of some amino acid also give observable peaks. Trp indole  $^1\text{H}\epsilon_1$ - $^{15}\text{N}\epsilon_1$  peaks are generally observed in the downfield region in both  $^1\text{H}$  and  $^{15}\text{N}$ ; the peak from the side chain of the single Trp residue in CheY3 is indicated. The region enclosed in the box contains peaks from Asn/Gln side chains in addition to backbone amides (expanded in Figure 2). The horizontal dotted grey lines indicate the two  $^{15}\text{N}$  chemical shift ranges at which the planes shown in Figure 5 were extracted. The peaks labelled a to g are the peaks for which strips are extracted in Figure 5. The peaks numbered 1 to 5 are the peaks for which strips are shown in Figures 7, 8, 11 and 12.

**Figure 2.** 500 MHz  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of CheY3. (a) Expanded region of the HSQC spectrum highlighted with the box in Figure 1. Peaks from backbone amides and side chain  $\text{NH}_2$  of Asn/Gln are observed. (b) The same region of an HSQC spectrum collected with a pair of the  $\tau$  values set to  $1/8J$  rather than  $1/4J$ . This sub-spectrum contains only the pairs of peaks arising from Asn and Gln side-chain  $\text{NH}_2$ ; these are indicated in (a) with horizontal dashed lines connecting the pair of peaks. In (a) note that above the  $\text{NH}_2$  peaks there is a much weaker peak that arises from NHD (indicated with \* for one pair at  $\sim 112$  ppm). The intensity of these peaks depends on the amount of  $\text{D}_2\text{O}$  added to the sample. These peaks are offset in  $F_1$  from the strong  $\text{NH}_2$  peaks because of the one-bond deuterium isotope effect on the  $^{15}\text{N}$  chemical shift; a smaller isotope effect on the  $\text{H}^{\text{N}}$  can be observed in  $F_2$ . The broad nature of the peak in the  $F_1$  dimension arises from the quadrupolar coupling of the  $^{15}\text{N}$  to deuterium. It is these NHD peaks that give rise to the correlations observed for Asn and Gln side chains in the CBCA(CO)NH experiment (section 4.5).

**Figure 3.** Summary of the one- and two-bond couplings important for triple resonance experiments used for assignment. The one-bond couplings are indicated next to the relevant bond. The dashed line indicates the  $^2J$  coupling between  $\text{C}\alpha(i-1)$  and  $\text{N}(i)$ .

**Figure 4.** Random-coil aliphatic  $^{13}\text{C}$  chemical shifts for the 20 common amino acids [37]. The symbols used are defined in the legend in the upper right corner.

**Figure 5.** (a) Two  $^1\text{H}^{\text{N}}$ - $^{13}\text{C}$  planes extracted from a 3D CBCA(CO)NH data set. The planes were extracted at  $^{15}\text{N}$  chemical shift values of 104.73 ppm (top) and at 129.04 ppm (bottom). These  $^{15}\text{N}$  chemical shifts are indicated in Figure 1. The  $^1\text{H}^{\text{N}}$ - $^{15}\text{N}$  peaks giving rise to the peaks in these planes are labelled a to e from the most downfield to the most upfield  $^1\text{H}^{\text{N}}$ . (b) Strip plot generated for the 7 peaks observed in these two planes.

**Figure 6.** Schematic representation of the  $^1\text{H}$ - $^{13}\text{C}$ - $^{15}\text{N}$  correlations observed in some 3D triple resonance experiments. The four experiments shown are all of the ‘out-and-back’ variety. Nuclei whose chemical shift is measured in one of the three dimensions are shown in bold black typeface. Nuclei involved in the coherence transfer pathway but whose chemical shift is not measured are shown in italic black typeface. (a) The HNCO experiment correlates  $\text{H}^{\text{N}}$  and  $\text{N}$  of residue  $i$  with CO of residue  $i-1$ . (b) HNCA correlates  $\text{H}^{\text{N}}$  and  $\text{N}$  of residue  $i$  with  $\text{C}\alpha$  of both residues  $i$  and  $i-1$ . (c) HN(CA)CO correlates the  $\text{H}^{\text{N}}$ ,  $\text{N}$  of residues  $i$  with CO of



both residues  $i$  and  $i-1$  as a result of the transfer from  $^1\text{H}^{\text{N}}\text{-}^{15}\text{N}$  of  $i$  via the  $\text{C}\alpha$  of residues  $i$  and  $i-1$  and subsequent transfer to  $\text{CO}(i)$  and  $\text{CO}(i-1)$ . (d)  $\text{HN}(\text{CO})\text{CA}$  correlates  $\text{H}^{\text{N}}$  and  $\text{N}$  of residue  $i$  with  $\text{C}\alpha$  of only residue  $i-1$  because of the transfer via the  $\text{CO}$  of  $i-1$ .

**Figure 7.** Analysis of HNCA, HN(CO)CA, HNCO and HN(CA)CO spectra. Strip plots are shown for five residues (labelled 1 to 5 in Figure 1). Each strip is extracted from a  $^{15}\text{N}$  plane and centred about the  $\text{H}^{\text{N}}$  shift. For each  $\text{H}^{\text{N}}(i)$  the HNCA strip contains 2 peaks and the HN(CO)CA strip contains a single peak. The  $\text{C}\alpha$  peak observed in both strips belongs to residue  $i-1$ ; this peak is circled in the HNCA strip and labelled  $i-1$  in the HN(CO)CA strip. The  $\text{C}\alpha$  peak observed in only the HNCA strip belongs to residue  $i$  and is labelled. For each  $\text{H}^{\text{N}}(i)$  the HNCO strip contains 1 peak and the HN(CA)CO strip contains 2 peaks. The  $\text{CO}$  peak observed in both strips corresponds to  $\text{CO}(i-1)$ ; this peak is circled in the HN(CA)CO strip and labelled  $i-1$  in the HNCO strip. The  $\text{CO}$  peak observed in only the HN(CA)CO strip corresponds to  $\text{CO}(i)$  and is labelled. For each  $\text{H}^{\text{N}}(i)$  a list of 6 chemical shifts,  $\text{C}\alpha(i-1)$ ,  $\text{CO}(i-1)$ ,  $^{15}\text{N}(i)$ ,  $\text{H}^{\text{N}}(i)$ ,  $\text{C}\alpha(i)$  and  $\text{CO}(i)$  can be tabulated from the strip plots of the 4 experiments (Table 1).

**Figure 8.** Strips from the HNCA and HN(CA)CO spectra shown in Figure 7 are plotted as a function of sequence. Connectivities between neighbouring residues, established using the rules outlined in section 4.3 and the chemical shifts shown in Table 1, are indicated by horizontal lines. A blank strip has been inserted between the strips for residues 4 and 1 to indicate that these residues are not neighbours in the sequence.

**Figure 9.** Schematic representation of the  $^1\text{H}\text{-}^{13}\text{C}\text{-}^{15}\text{N}$  correlations observed in some 3D triple resonance experiments. Nuclei whose chemical shift is measured in one of the three dimensions are shown in bold black typeface. Nuclei involved in the coherence transfer pathway but whose chemical shift is not measured are shown in italic black typeface. (a) The CBCANH experiment correlates  $\text{H}^{\text{N}}$  and  $\text{N}$  of residue  $i$  with  $\text{C}\alpha$  and  $\text{C}\beta$  of both residues  $i$  and  $i-1$ . (b) CBCA(CO)NH correlates the  $\text{H}^{\text{N}}$ ,  $\text{N}$  of residue  $i$  with  $\text{C}\alpha$  and  $\text{C}\beta$  of only residue  $i-1$  because of the transfer via the  $\text{CO}$  of  $i-1$ . (c) The HBHA(CBCA)NH (also abbreviated HBHANH) experiment correlates  $\text{H}^{\text{N}}$  and  $\text{N}$  of residue  $i$  with  $\text{H}\alpha$  and  $\text{H}\beta$  of both residues  $i$  and  $i-1$ . (b) HBHA(CBCACO)NH (also abbreviated HBHA(CO)NH) correlates the  $\text{H}^{\text{N}}$ ,  $\text{N}$  of residue  $i$  with  $\text{H}\alpha$  and  $\text{H}\beta$  of only residue  $i-1$  because of the transfer via the  $\text{CO}$  of  $i-1$ .

**Figure 10.** Random-coil aliphatic  $^1\text{H}$  chemical shifts for the 20 common amino acids [37]. The symbols used are defined in the legend in the upper right corner.

**Figure 11.** Analysis of CBCANH and CBCA(CO)NH spectra. Strip plots are shown for five residues (labelled 1 to 5 in Figure 1). Each strip is extracted from a  $^{15}\text{N}$  plane and centred about the  $\text{H}^{\text{N}}$  shift. For each  $\text{H}^{\text{N}}(i)$  the CBCANH strip contains 2-4 peaks and the CBCA(CO)NH strip contains 1-2 peaks. In the CBCANH, all  $\text{C}\alpha$  peaks (except Gly) are positive (shown in black) while Gly  $\text{C}\alpha$  and all  $\text{C}\beta$  peaks are negative and shown in red. The  $\text{C}\alpha$  and  $\text{C}\beta$  peaks observed in both strips belongs to residue  $i-1$ ; these peaks are circled in the CBCANH strip and labelled  $i-1$  in the CBCA(CO)NH strip. The  $\text{C}\alpha$  and  $\text{C}\beta$  peaks observed in only the CBCANH strip belong to residue  $i$  and are labelled. For each  $\text{H}^{\text{N}}(i)$  a list of 4-6 chemical shifts,  $\text{C}\alpha(i-1)$ ,  $\text{C}\beta(i-1)$ ,  $^{15}\text{N}(i)$ ,  $\text{H}^{\text{N}}(i)$ ,  $\text{C}\alpha(i)$  and  $\text{C}\beta(i)$  can be tabulated from the strip plots of the 2 experiments (Table 2).

**Figure 12.** Strips from the CBCANH spectra shown in Figure 11 are plotted as a function of sequence. Connectivities between neighbouring residues, established using the rules outlined in section 4.4 and the chemical shifts shown in Table 2, are indicated by horizontal lines. In addition, strips from the HBHA(CO)NH for these 5 residues are shown. A blank strip has been inserted between the strips for residues 4 and 1 to indicate that these residues are not neighbours in the sequence. Spin system information can be obtained from the  $C\alpha$ ,  $C\beta$  and  $H\beta$  chemical shifts. The peak 4  $C\beta$  shift of 20.42 ppm identifies this as Ala. The peak 5  $C\beta$  of 44.39 could arise from Asn, Asp, Cys, Ile, Leu, Phe or Tyr; the  $C\alpha$  of 52.93 ppm suggests this is not Ile. The  $H\beta$  (*i-I*) shifts of 1.10 and 1.81 ppm in strip 4 of the HBHA(CO)NH identify peak 5 as Leu. The peak 2  $C\alpha$  of 62.27 ppm suggests this is Ile, Thr or Val. The  $C\beta$  of 34.62 ppm suggests that peak 2 is Val and the  $H\beta$  of 1.70 ppm from strip 5 is consistent with this. The downfield shifted  $C\beta$  (70.70 ppm) observed for residue *i-I* in strip 2 suggests that the residue preceding peak 2 is a Thr or Ser; the downfield  $C\alpha$  of 62.27 ppm would favour assignment to Thr. Therefore, 2-5-4 is identified as Val-Leu-Ala with the Val preceded by Thr; this Thr-Val-Leu-Ala sequence can be assigned to residues 4-7 of CheY3. Peak 3 has a single peak in the CBCA(CO)NH at 48.76 ppm suggesting that peak 1 that precedes it is a Gly; this is confirmed by the negative peak observed at this position in the CBCANH and the two  $H\alpha$  peaks observed in the HBHA(CO)NH at 4.83 and 3.87 ppm. The residue preceding peak 1 has  $C\beta$  at 40.67 ppm and two  $H\beta$  at 3.61 and 2.98 ppm; this residue can be identified as an Asn, Asp, Cys, Phe or Tyr.

**Figure 13.** Strips from the HNCA for 10 residues are shown. Each strip contains two peaks corresponding to  $C\alpha$  of *i* and *i-I*. These can be assigned to *i* and *i-I* by comparison with the HN(CO)CA which only contains the *i-I* peak (not shown); the *i* peaks identified in this way are circled in red. (Note: there are two overlapping  $^1H^N$  giving rise to 4 peaks instead of 2 in strip 10). The residue that follows 1 in the sequence should have a  $C\alpha$  *i-I* peak at the same position as the *i* peak circled in strip 1; this peak would fall along a horizontal line drawn at the position of the *i* peak (shown in blue). On this basis, residue 6 can be identified as the residue that follows 1. In a similar way, 7 follows 5, 3 follows 9 and 4 follows one of the residues in strip 10 (shown by green lines). (b) These assignments can be confirmed using data from the HN(CA)CO experiment. Each strip contains two peaks corresponding to CO of *i* and *i-I*. These can be assigned to *i* and *i-I* by comparison with the HNCO which only contains the *i-I* peak (not shown); the *i* peaks identified in this way are circled in red. The expected connectivities involving the CO peaks are shown for 1 and 6, 5 and 7, 9 and 3, and 10 and 4; these are shown with horizontal lines.

**Figure 14.** (a) Strips from the HNCA for 10 residues are shown. Each strip contains two peaks corresponding to  $C\alpha$  of *i* and *i-I*. These can be assigned to *i* and *i-I* by comparison with the HN(CO)CA which only contains the *i-I* peak (not shown); the *i* peaks identified in this way are circled in red. (Note: there are two overlapping  $^1H^N$  giving rise to 4 peaks instead of 2 in strip 10). Ambiguity arises in trying to identify the residues that follow 2, 3, 4, 6 and 7. Residues 2, 3 and 4 share similar  $C\alpha$  *i* chemical shifts. The blue horizontal line indicates that either 2, 3 or 4 could be followed by residue 5. Similarly, 6 and 7 share the same  $C\alpha$  *i* chemical shifts. The green horizontal line indicates that either 6 or 7 could be followed by residue 8 or 9. (b) Ambiguity can be overcome using data from the HN(CA)CO experiment. Each strip contains two peaks corresponding to CO of *i* and *i-I*. These can be assigned to *i* and *i-I* by comparison with the HNCO which only contains the *i-I* peak (not shown); the *i* peaks identified in this way are circled in red. It is clear from the HN(CA)CO that residue 2 is

followed by residue 5, 6 is followed by 8 and 7 is followed by 9; these connectivities are shown with horizontal lines.

**Figure 15.** Analysis of the HCCH-TOCSY spectrum for Asp 34. Strips extracted at  $^{13}\text{C}$  chemical shifts of 53.93 ppm (top) and 41.28 ppm (bottom) are shown; these correspond to  $\text{C}\alpha$  and  $\text{C}\beta$  of Asp 34 identified from the CBCANH and CBCA(CO)NH experiments. The horizontal line drawn at 5.24 ppm in the top strip corresponds to the  $\text{H}\alpha$  identified for Asp 34 in the HBHA(CO)NH. Two cross peaks can be seen at 3.73 and 2.96 ppm; these correspond to the two  $\text{H}\beta$  of Asp 34. In the strip extracted at 41.28 ppm, cross peaks between both  $\text{H}\beta$  and the  $\text{H}\alpha$  at 5.24 ppm are observed. The observed connectivities confirm the assignments made for  $\text{C}\alpha$ ,  $\text{C}\beta$ ,  $\text{H}\alpha$  and  $\text{H}\beta$  of Asp 34.

**Figure 16.** Analysis of the HCCH-TOCSY spectrum for Ile 65. The top two strips are extracted at  $^{13}\text{C}$  chemical shifts of 62.81 and 35.36 ppm; these correspond to  $\text{C}\alpha$  and  $\text{C}\beta$  of Ile 65 identified from the CBCANH and CBCA(CO)NH experiments. The horizontal line drawn at 3.48 ppm in the top strip corresponds to the  $\text{H}\alpha$  identified for Ile 65 in the HBHA(CO)NH. A strong cross peak can be seen at 2.16 ppm; this corresponds to the  $\text{H}\beta$  of Ile 65. In the strip extracted at 35.36 ppm, a cross peak between  $\text{H}\beta$  and the  $\text{H}\alpha$  at 3.48 ppm is observed. These connectivities confirm the assignments made for  $\text{C}\alpha$ ,  $\text{C}\beta$ ,  $\text{H}\alpha$  and  $\text{H}\beta$  of Ile 65. In the strip extracted at 62.81 ppm (top) additional cross peaks can be seen at 1.67, 1.16, 0.97 and -0.04 ppm; these correspond to other side chain  $^1\text{H}$  of Ile 65. Two additional side chain  $^{13}\text{C}$  chemical shifts of 27.68 and 19.08 ppm were identified for Ile 65 in the (H)CC(CO)NH spectrum. Inspection of the  $^{13}\text{C}$  plane at 27.68 ppm shows diagonal peaks at both 1.67 and 1.16 ppm which show strong cross peaks to -0.04 ppm and weaker cross peaks to 3.48 ppm. The  $^{13}\text{C}$  chemical shift and the pair of  $^1\text{H}$ , at 1.67 and 1.16 ppm, linked to this  $^{13}\text{C}$  suggest that this is the  $\text{C}\gamma_1$  methylene group of Ile 65. The two remaining peaks at 0.97 and -0.04 ppm must arise from the two methyl groups of Ile. Inspection of the  $^{13}\text{C}$  plane at 19.08 ppm shows a diagonal peak at 0.97 ppm with cross peaks to 3.48, 2.16, 1.67, 1.16 and -0.04 ppm; this  $^{13}\text{C}$  chemical shift and the stronger cross peak to 0.97 ppm (than to -0.04 ppm) from the  $\text{H}\alpha$  (top strip) suggest that this is the  $\gamma_2$  methyl group. The peak at -0.04 ppm is resolved in the  $^1\text{H}$ - $^{13}\text{C}$  HSQC and is linked to a  $^{13}\text{C}$  shift of 9.57 ppm; this value is typical of the Ile  $\delta_1$  methyl group. The strip 9.57 ppm, shown at the bottom of the figure, has cross peaks to all the other protons of Ile 65.

**Figure 17.** Schematic representation of the (a) through-bond and (b) through-space connectivities used in the sequential assignment method. (a) In the TOCSY-HSQC spectrum each backbone  $\text{H}^{\text{N}}$  will be correlated with  $\text{H}\alpha$ ,  $\text{H}\beta$  and other sidechain protons. (b) Identification of adjacent residues is based on the sequential  $d_{\text{NN}}(i,i+1)$ ,  $d_{\alpha\text{N}}(i,i+1)$ , and  $d_{\beta\text{N}}(i,i+1)$  NOE effects.

**Figure 18.** Schematic representation of a  $^{15}\text{N}$  plane from the  $^{15}\text{N}$ -edited TOCSY-HSQC and the  $^{15}\text{N}$ -edited NOESY-HSQC. The plane contains peaks from two  $\text{H}^{\text{N}}$  shown in blue and red. (a) The TOCSY-HSQC contains a strong diagonal peak ( $F_1=F_3$ , represented by the large circle) and cross peaks corresponding to  $\text{H}\alpha$  and  $\text{H}\beta$  of each  $\text{H}^{\text{N}}$ . Analysis of the chemical shifts of observed side chain  $^1\text{H}$  provides spin system information. (b) The NOESY-HSQC contains a strong diagonal peak ( $F_1=F_3$ ) and cross peaks corresponding to both intra- and interresidue through-space correlations. Comparison of the TOCSY-HSQC and NOESY-HSQC strip for each  $\text{H}^{\text{N}}$  allows the intra- and interresidue NOEs to be distinguished.

**Figure 19.** Schematic representation of the NOE connectivities used in the sequential assignment method (described in section 5.3). The Squares, circles and diamonds represent  $d_{NN}$ ,  $d_{\alpha N}$ , and  $d_{\beta N}$  NOE connectivities, respectively. The filled and open symbols represent intra- and interresidue NOE effects, respectively; these can be distinguished by comparing the TOCSY-HSQC and NOESY-HSQC spectra. The size of the symbols represents the intensity of the NOE effect. (left) Sequential assignment based on  $d_{NN}$  NOEs usually observed in helices. (right) Sequential assignment based on  $d_{\alpha N}$  NOEs usually observed in  $\beta$ -sheet.

**Figure 20.** Strips from a  $^{15}\text{N}$ -edited NOESY-HSQC spectrum. The strips show the  $\text{H}^{\text{N}}\text{-H}^{\text{N}}$  NOE region for 11 residues. (a) Strips are plotted according to the  $\text{H}^{\text{N}}$  chemical shift with the most downfield  $\text{H}^{\text{N}}$  in strip 1 and the most upfield  $\text{H}^{\text{N}}$  in strip 11. The diagonal peaks are indicated by the darker contours. Strip 1 contains a diagonal peak at 8.9 ppm and two strong  $\text{H}^{\text{N}}$  cross peaks at 7.7 and 8.3 ppm. A horizontal line drawn at 7.7 ppm shows a single diagonal peak with the same chemical shift. Inspection of the strip for residue 8 shows a cross peak at 8.9 ppm that lines up with the diagonal peak of residue 1. This confirms an  $\text{H}^{\text{N}}\text{-H}^{\text{N}}$  NOE between residues 1 and 8. A horizontal line drawn at 8.3 ppm shows two diagonal peaks with the same chemical shift (residues 3 and 4). Inspection of the strip for residue 4 shows a cross peak at 8.9 ppm that lines up with the diagonal peak of residue 1 while strip 3 does not contain a cross peak at this position. This confirms an  $\text{H}^{\text{N}}\text{-H}^{\text{N}}$  NOE between residues 1 and 4. (b) The strips from the 11 residues are plotted in their correct order in the protein sequence. Sequential  $\text{H}^{\text{N}}\text{-H}^{\text{N}}$  NOEs are indicated by the green and red lines. These residues are located in an  $\alpha$ -helix within the protein.

**Figure 21.** Strips from  $^{15}\text{N}$ -edited NOESY-HSQC and HSQC-NOESY-HSQC spectra. (a) The strips from the NOESY-HSQC show the  $\text{H}^{\text{N}}\text{-H}^{\text{N}}$  NOE region for residues 77 to 80; these residues are located in an  $\alpha$ -helix. The sequential  $\text{H}^{\text{N}}\text{-H}^{\text{N}}$  NOEs between residues 77 and 78 and between residues 79 and 80 are indicated by horizontal dashed lines. No NOE is observed between residues 78 and 79 because these two residues share the same  $\text{H}^{\text{N}}$  chemical shift of 7.9 ppm; the NOE for this pair of residues overlaps with the strong diagonal peaks (indicated by the oval). (b) The strips from the HSQC-NOESY-HSQC for the same residues clearly show a sequential  $\text{H}^{\text{N}}\text{-H}^{\text{N}}$  NOE between residues 78 and 79; this is observed because these residues have different  $^{15}\text{N}$  chemical shifts of 118.5 and 121.7 ppm, respectively.

**Figure 22.** Cyclic NOE patterns involving residues on adjacent strands of the  $\beta$ -sheet used in the main-chain directed (MCD) strategy for assignment of (a) antiparallel and (b) parallel  $\beta$ -sheet [64]. (a) The two MCD patterns for the antiparallel  $\beta$ -sheet are designated as the outer (dotted oval) and inner (dashed oval) loops. the  $d_{\alpha\alpha}(i+1,j-1)$  NOE usually associated with this loop won't be observed in a  $^{15}\text{N}$ -edited experiment. The inner and outer loops share a common NOE,  $d_{\alpha N}(i+1,j)$ , and the  $\text{H}^{\text{N}}$  of residue  $i+2$ , identified in the inner loop, is the starting point for the next outer loop with a  $d_{NN}$  NOE to residue  $j-2$ . (b) The MCD pattern for parallel  $\beta$ -sheet is indicated with the dotted oval; adjacent MCD loops share a common interstrand  $d_{\alpha N}$  NOE. Residue labels are placed above/below the nitrogen. Adapted with permission from reference 64. Copyright 1987 American Chemical Society.

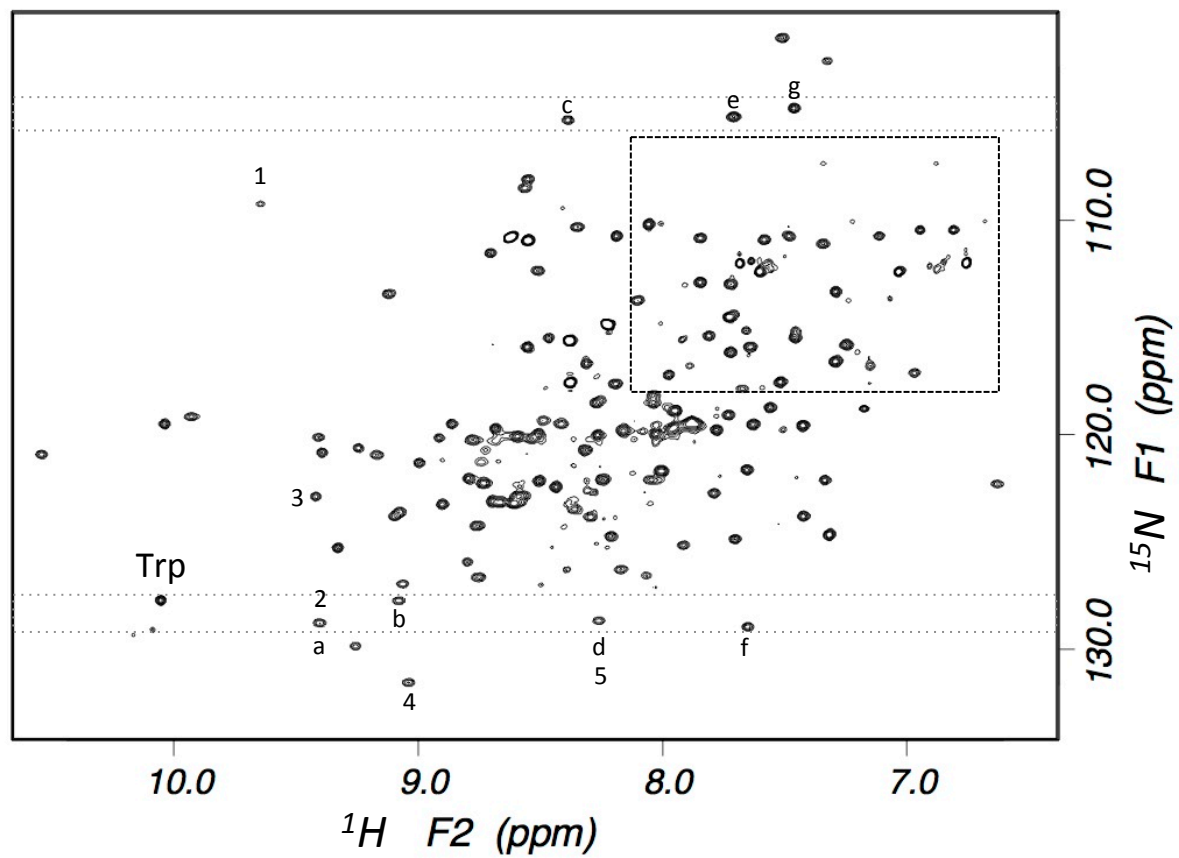


Figure 1

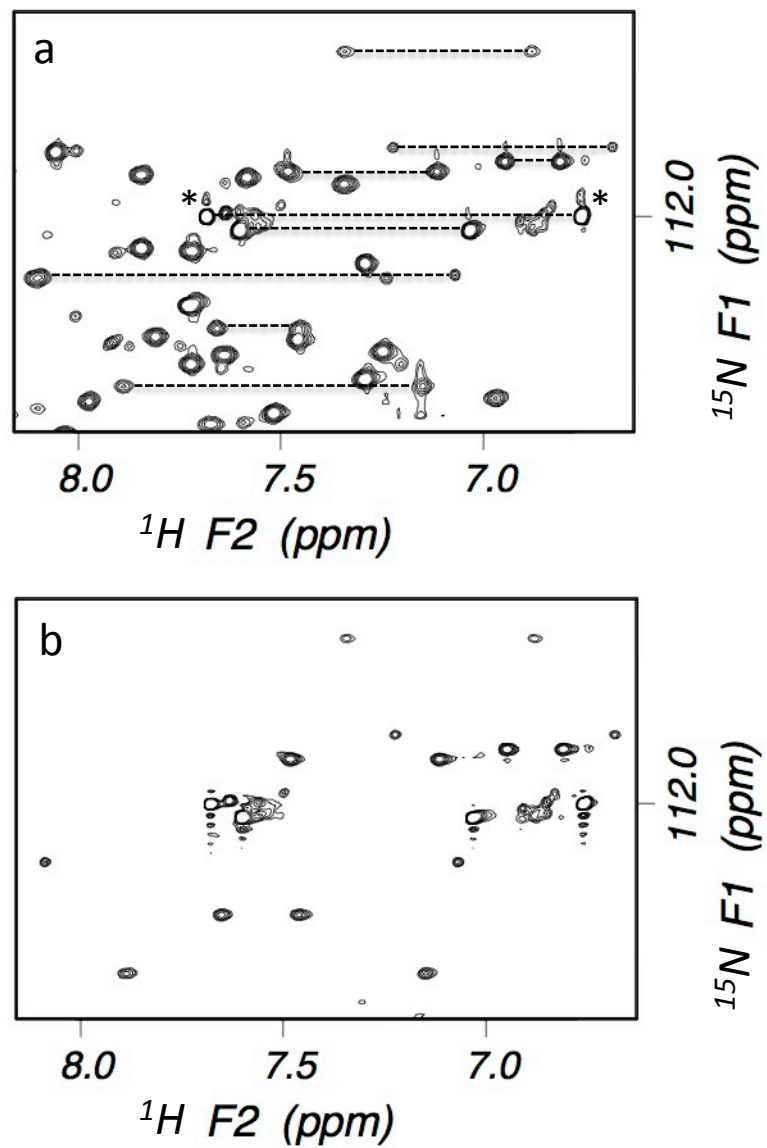


Figure 2

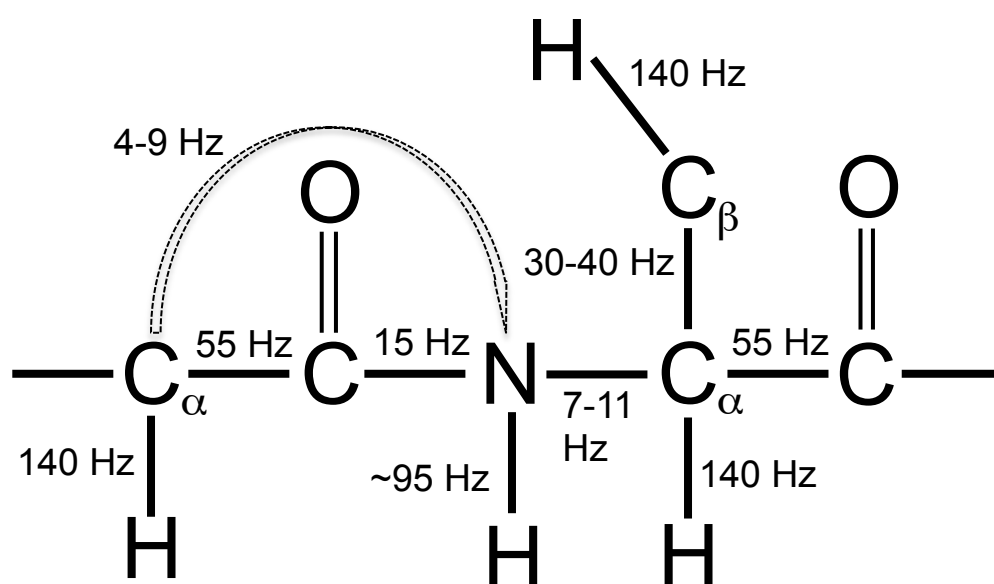


Figure 3

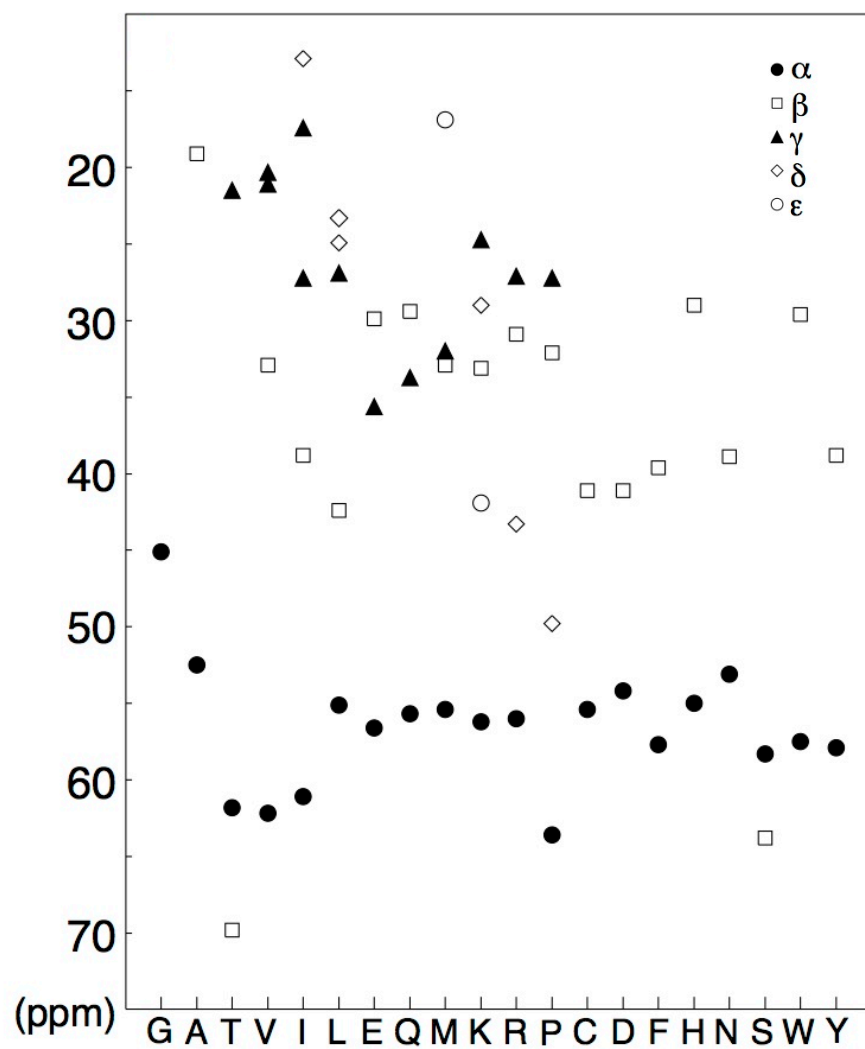


Figure 4



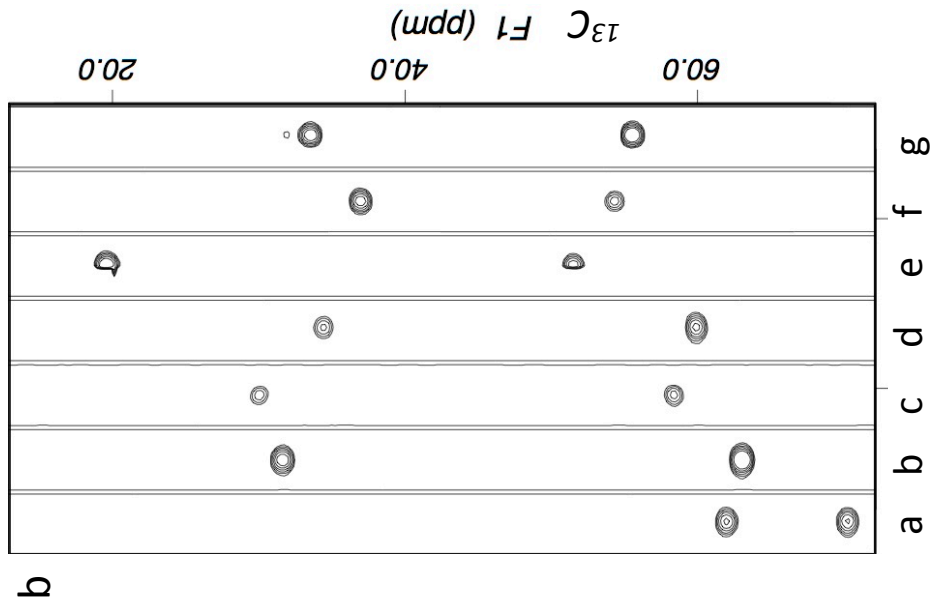
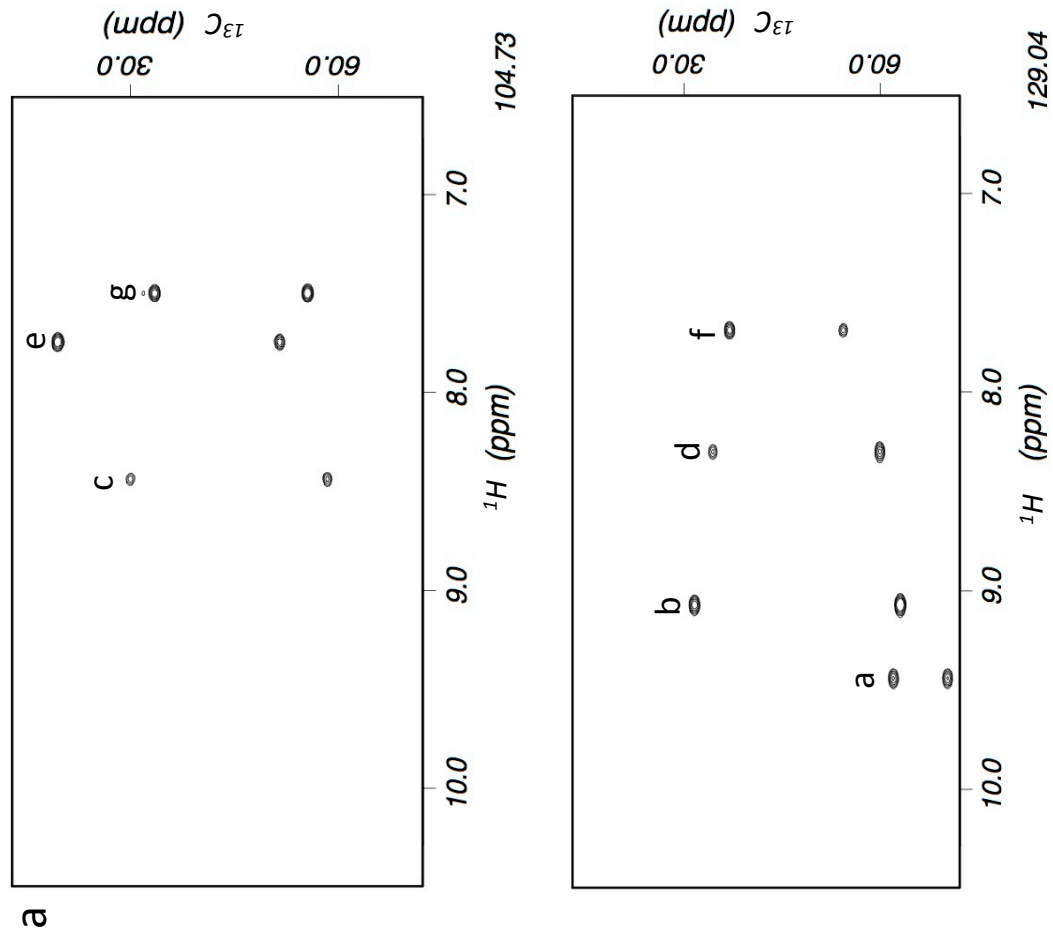
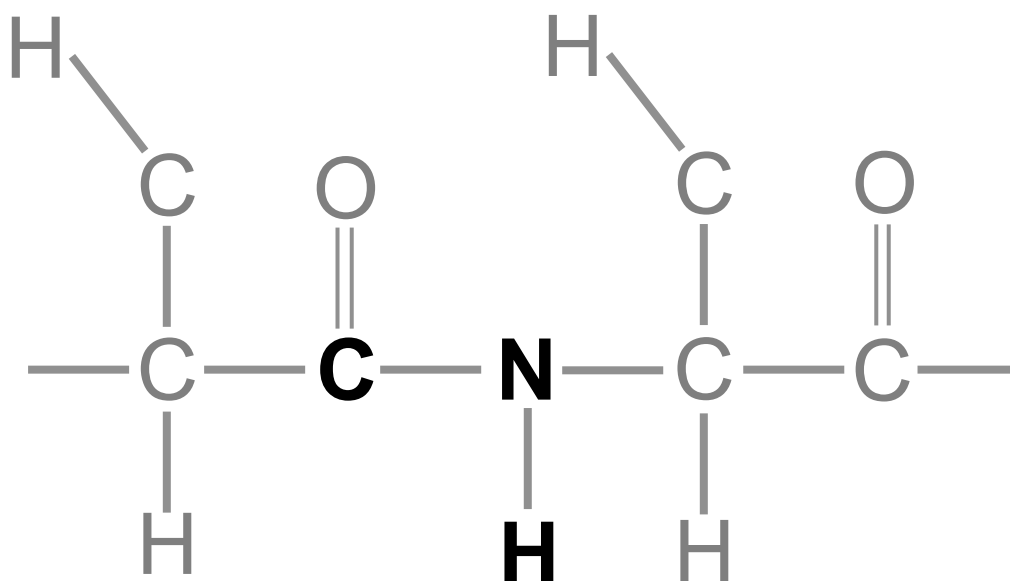
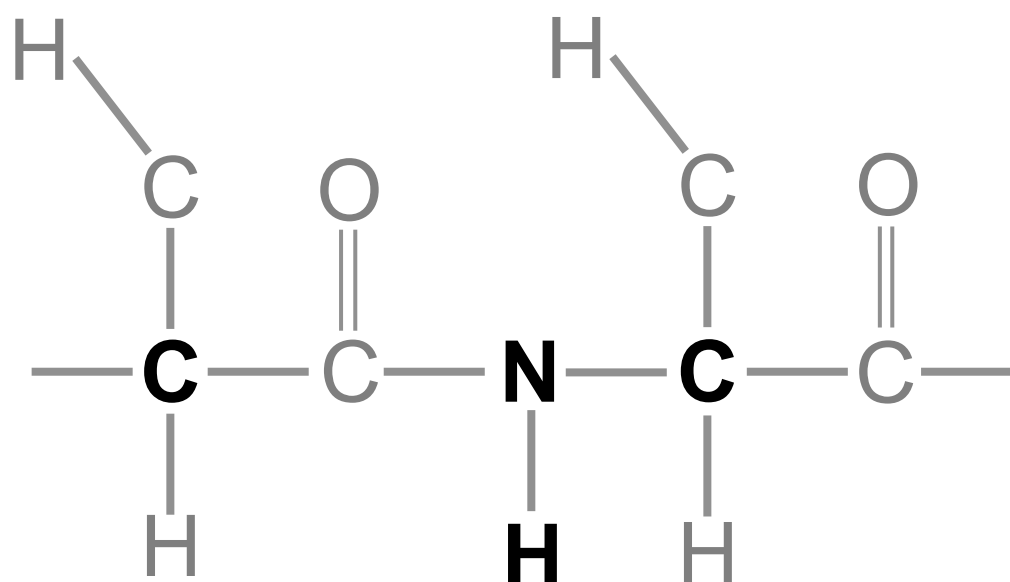


Figure 5

Figure 6

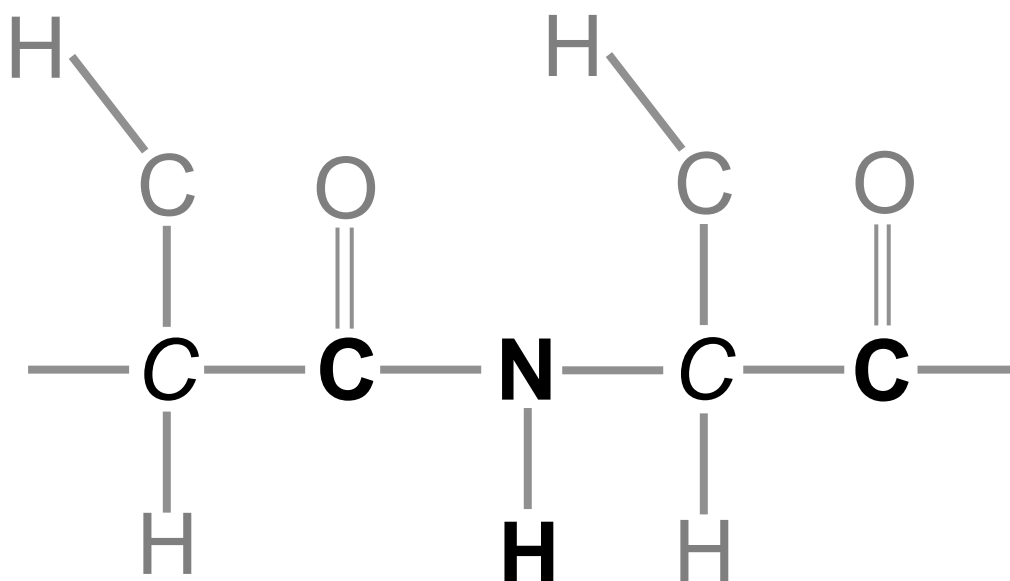


(a) HNCNCO

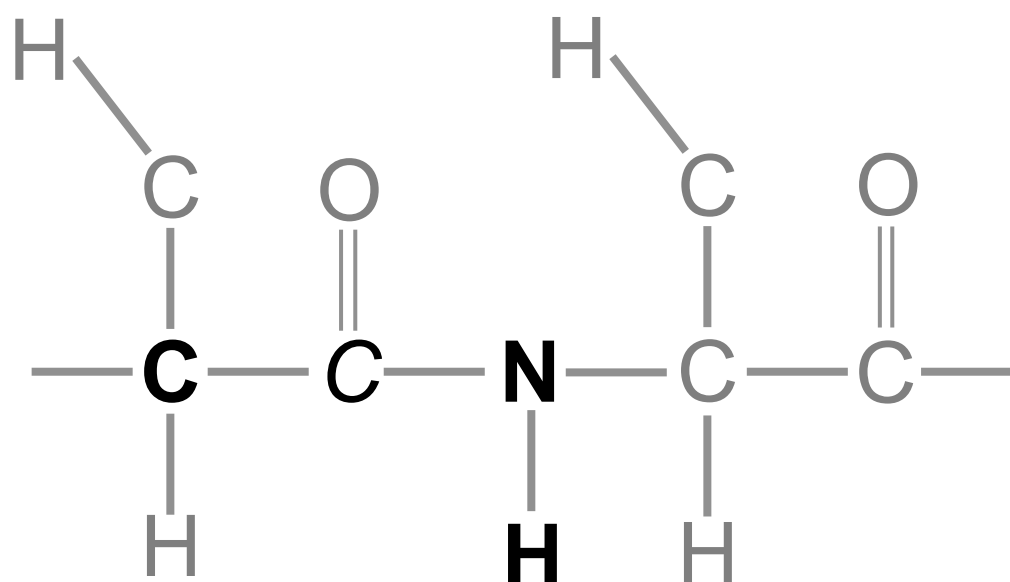


(b) HNCNCA

Figure 6 cont.



(c) HN(CA)CO



(d) HN(CO)CA

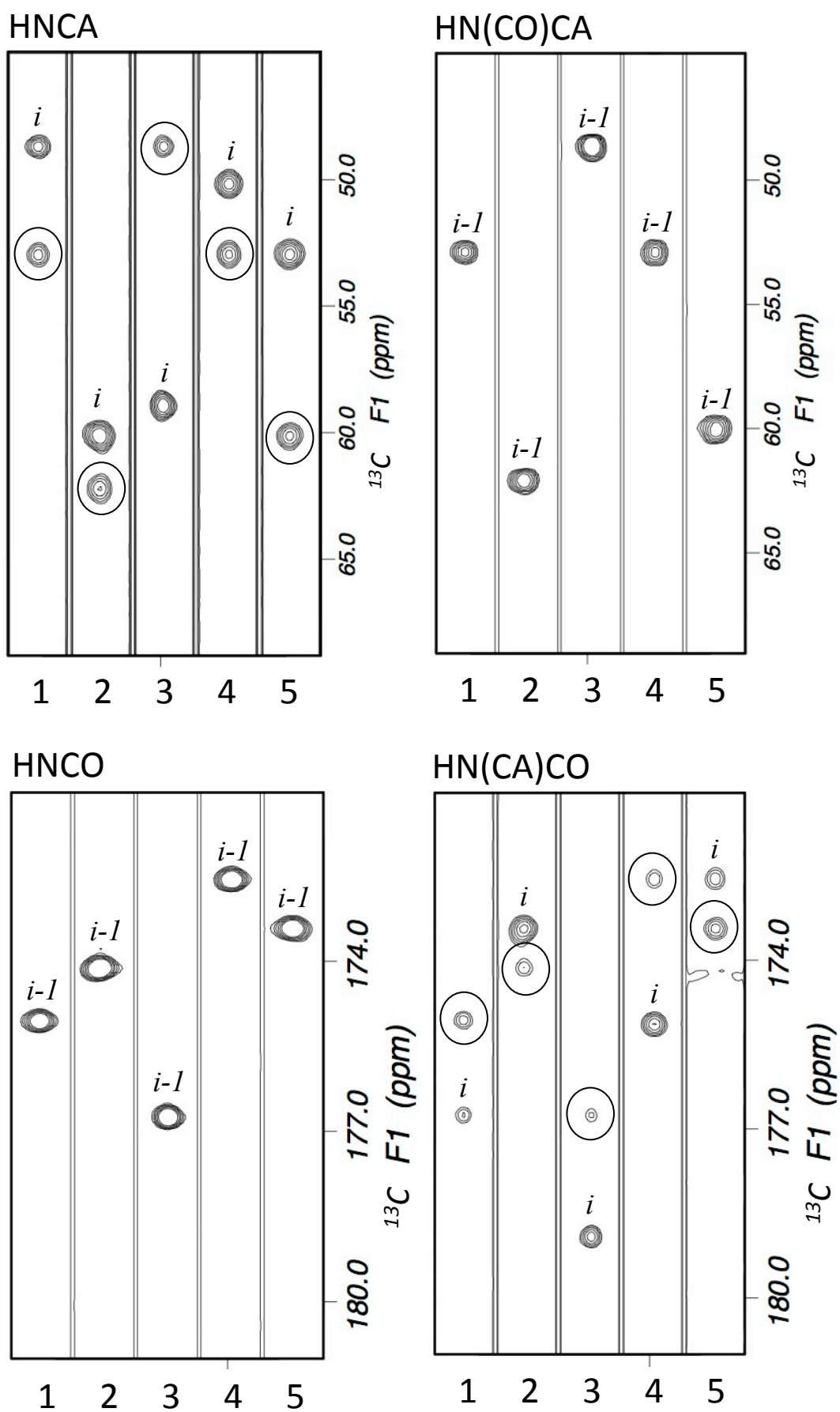
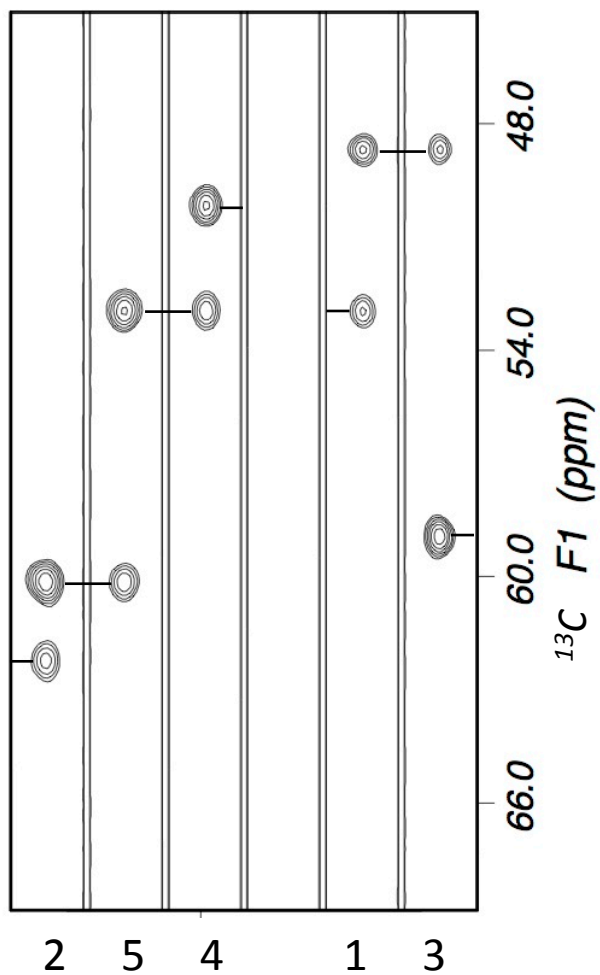


Figure 7

HNCA



HN(CA)CO

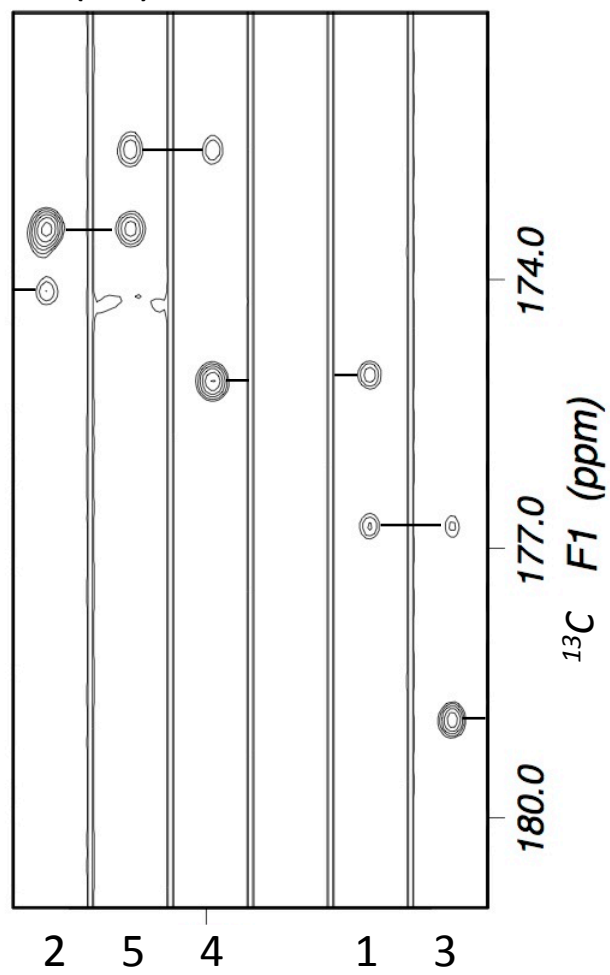
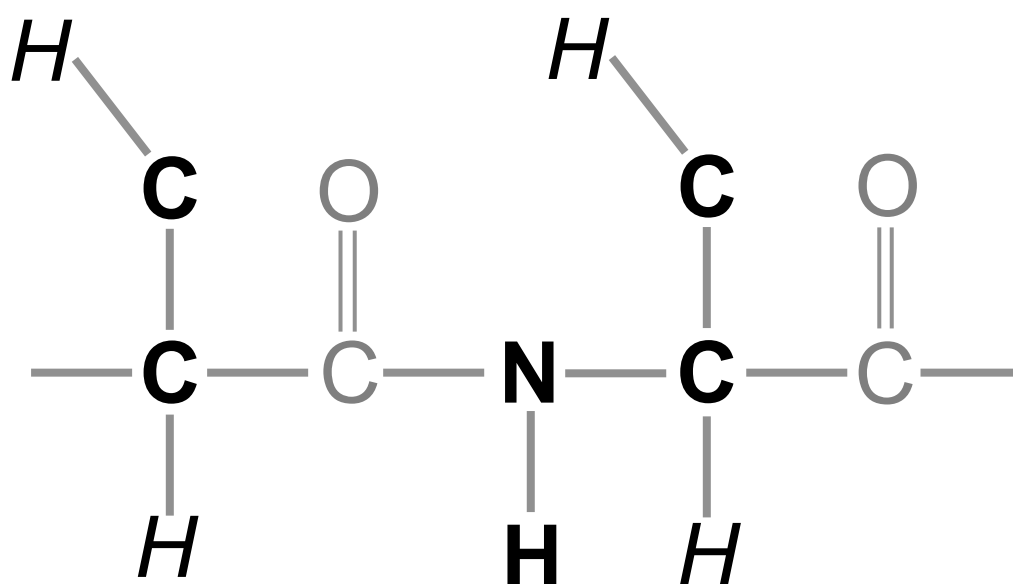
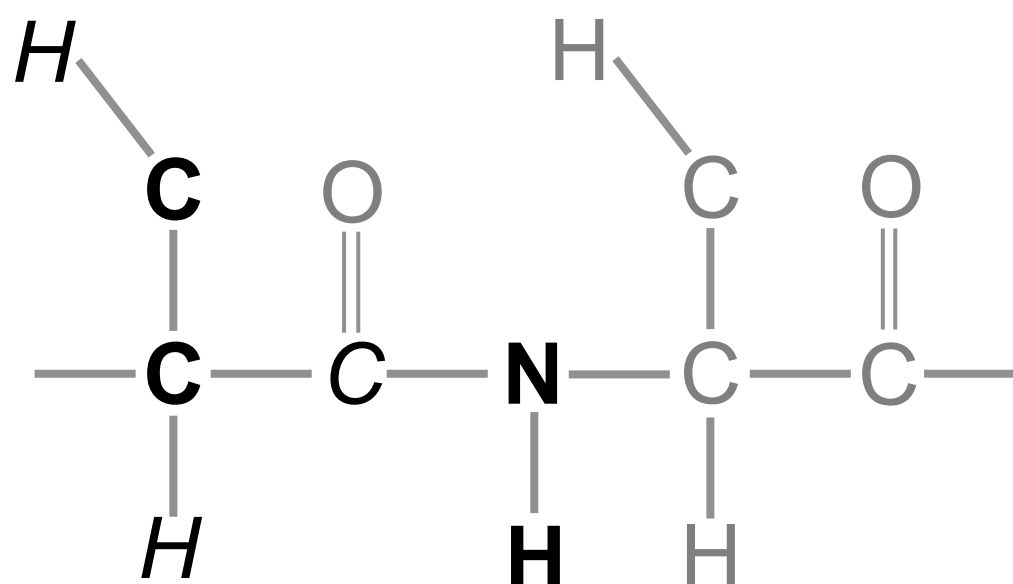


Figure 8

Figure 9

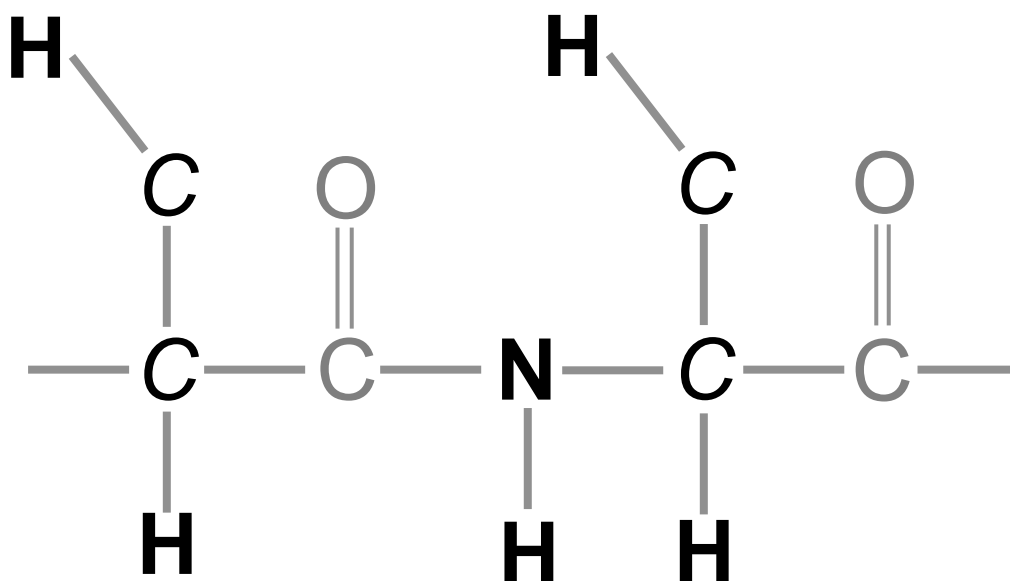


(a) CBCANH

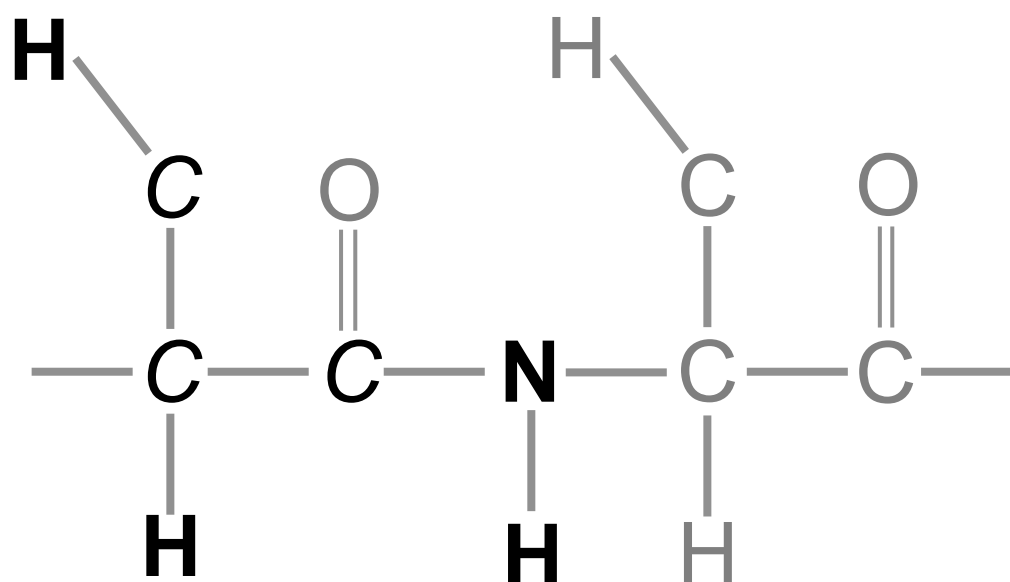


(b) CBCA(CO)NH

Figure 9 cont.



(c) HBHA(CBCA)NH



(d) HBHA(CBCACO)NH

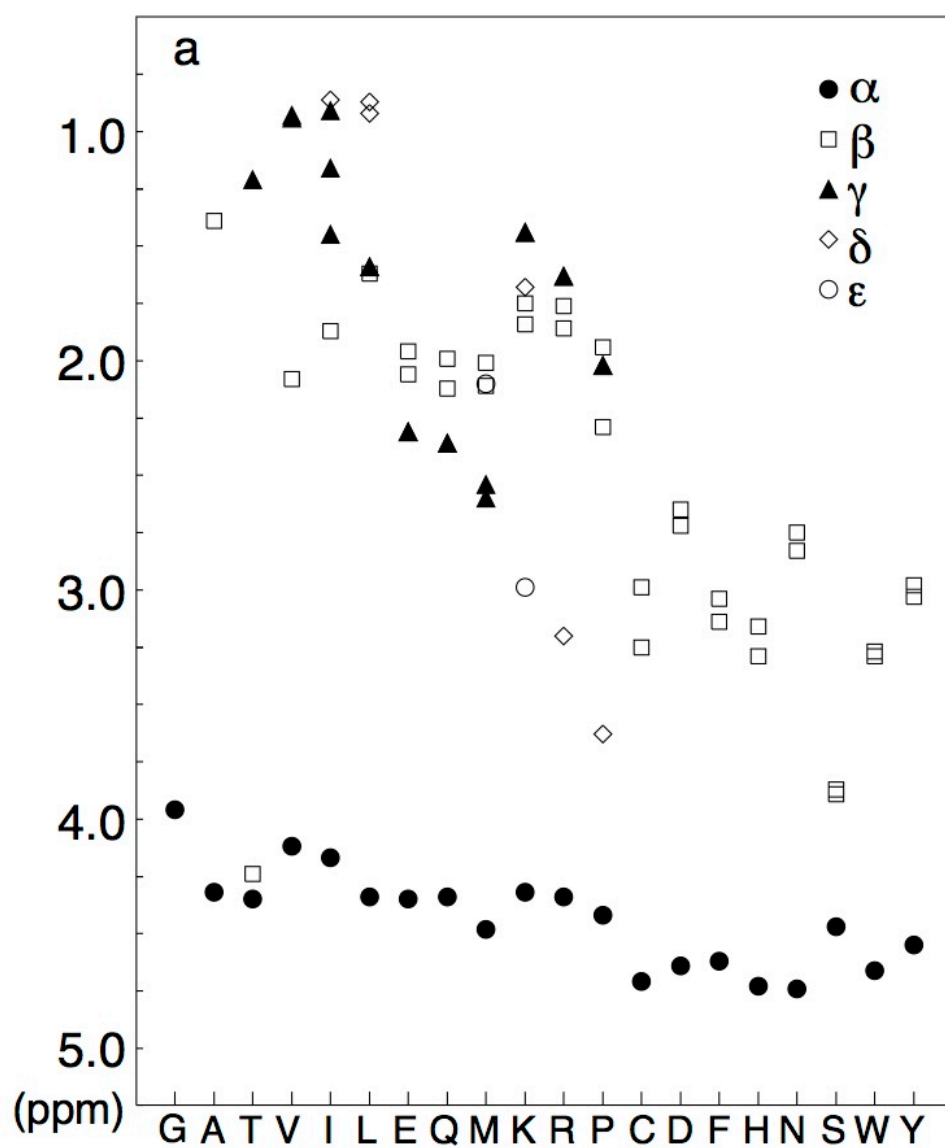
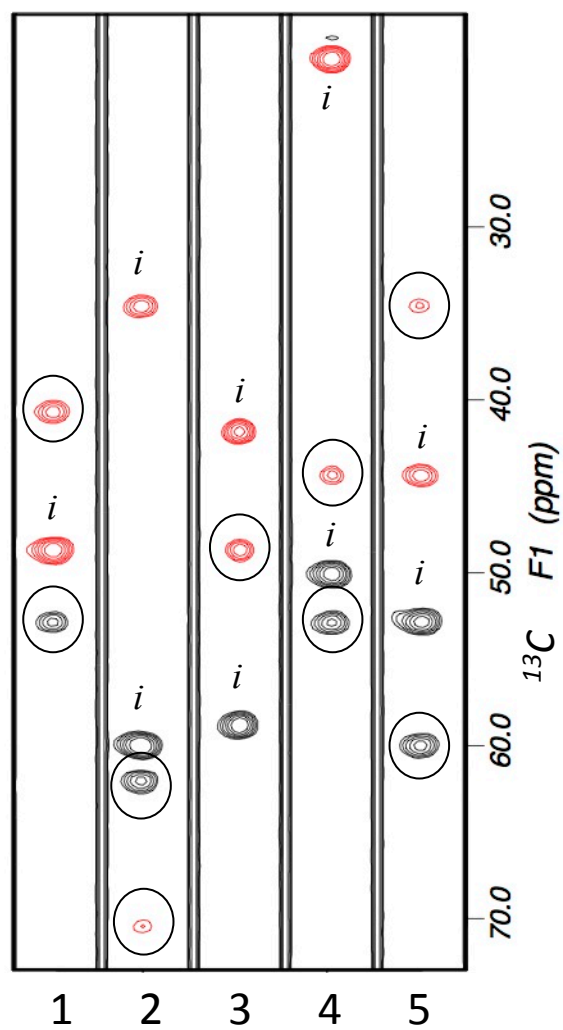


Figure 10



CBCANH



CBCA(CO)NH

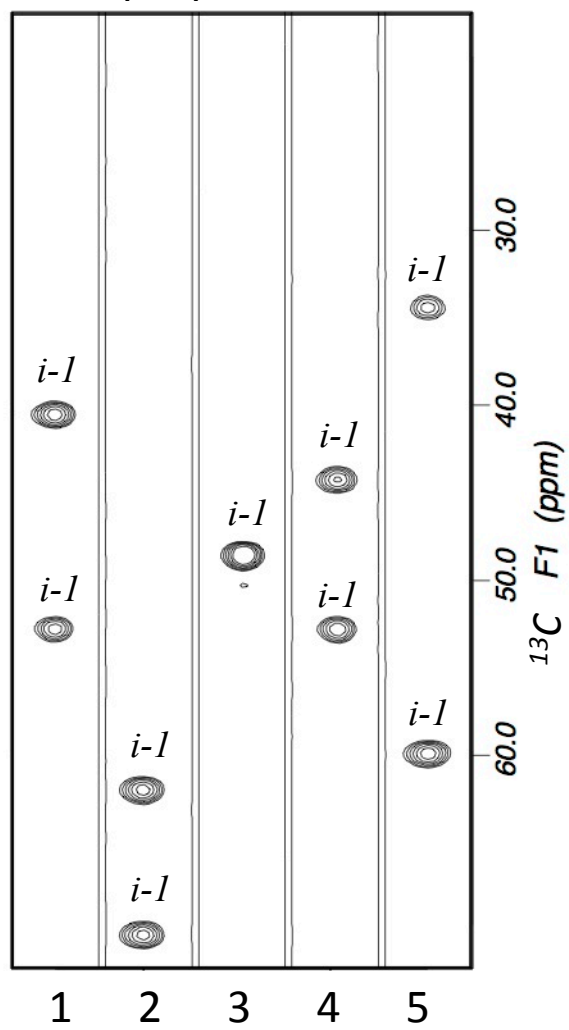
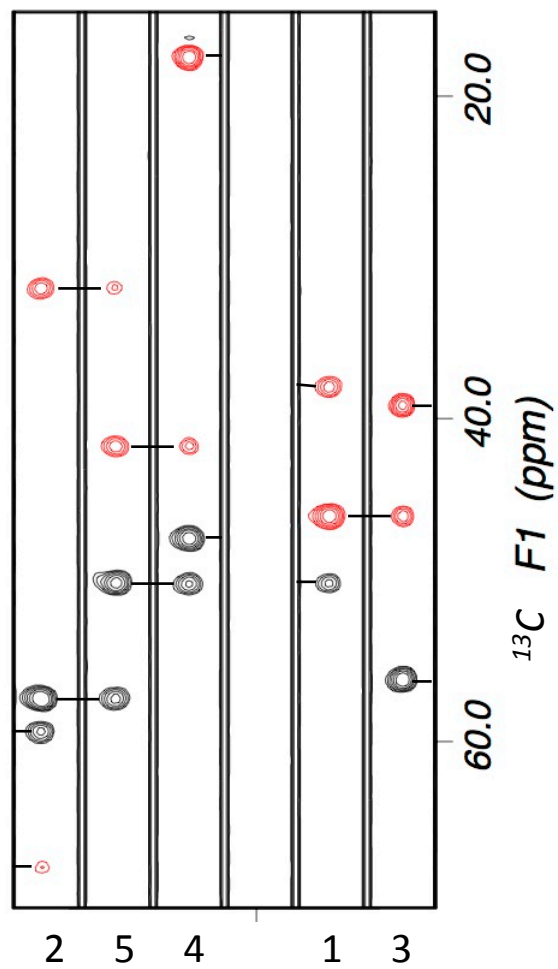


Figure 11

CBCANH



HBHA(CO)NH

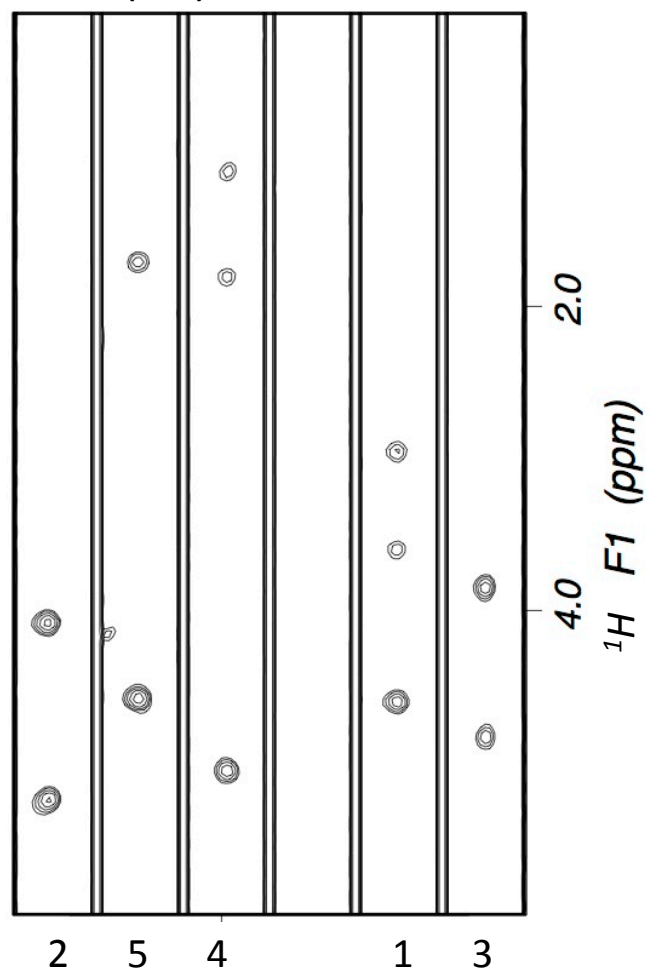


Figure 12

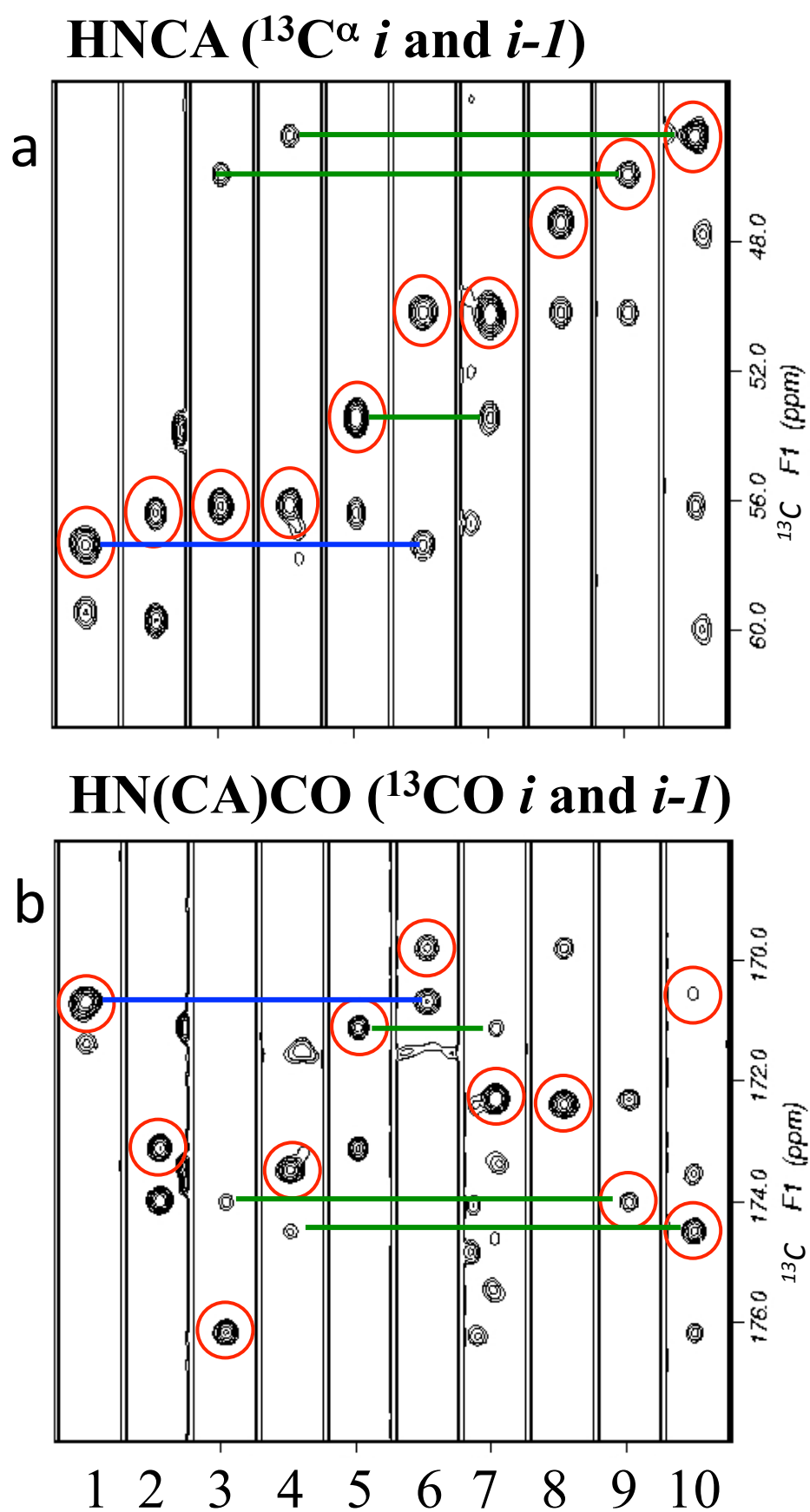


Figure 13

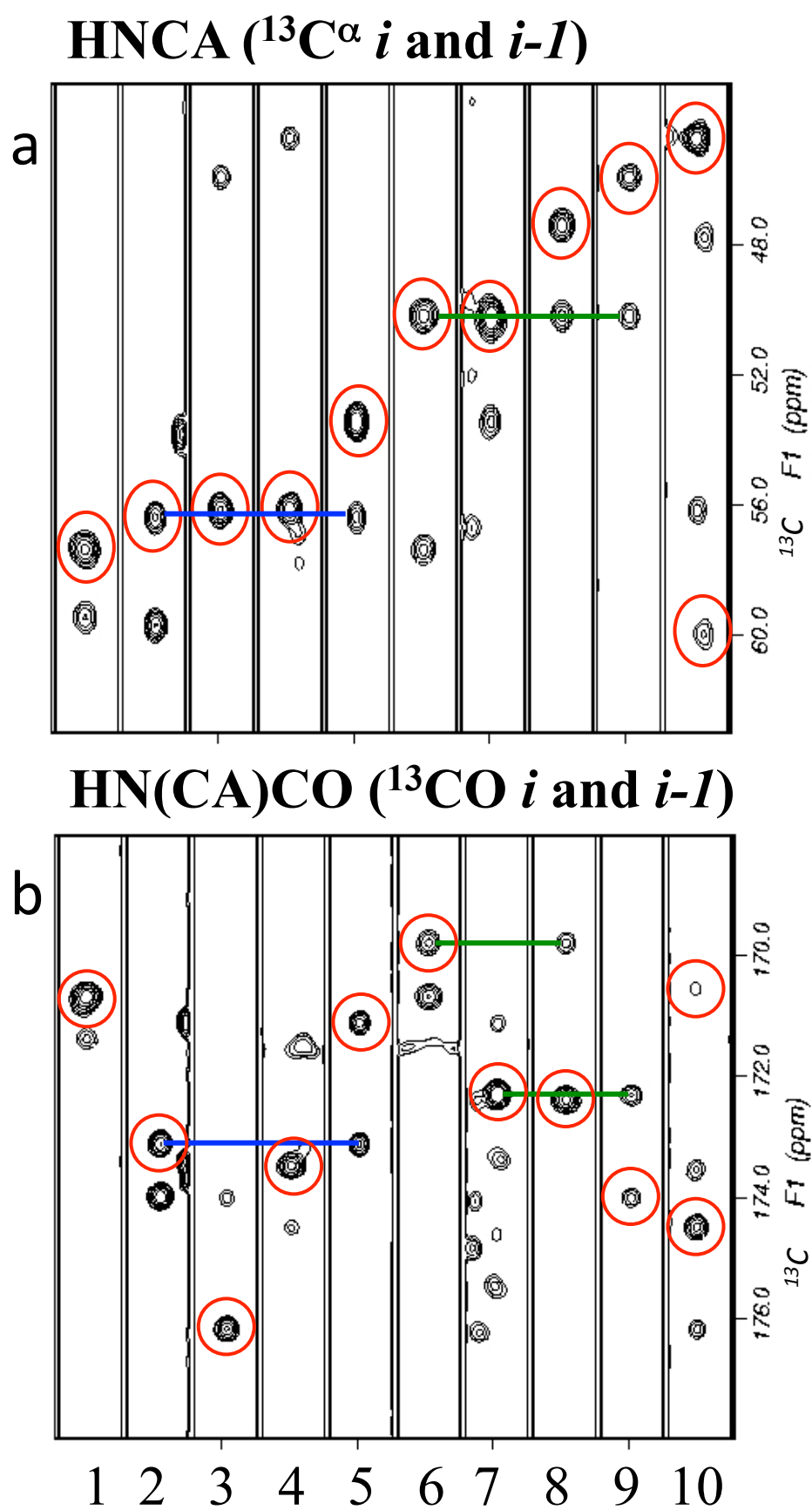


Figure 14

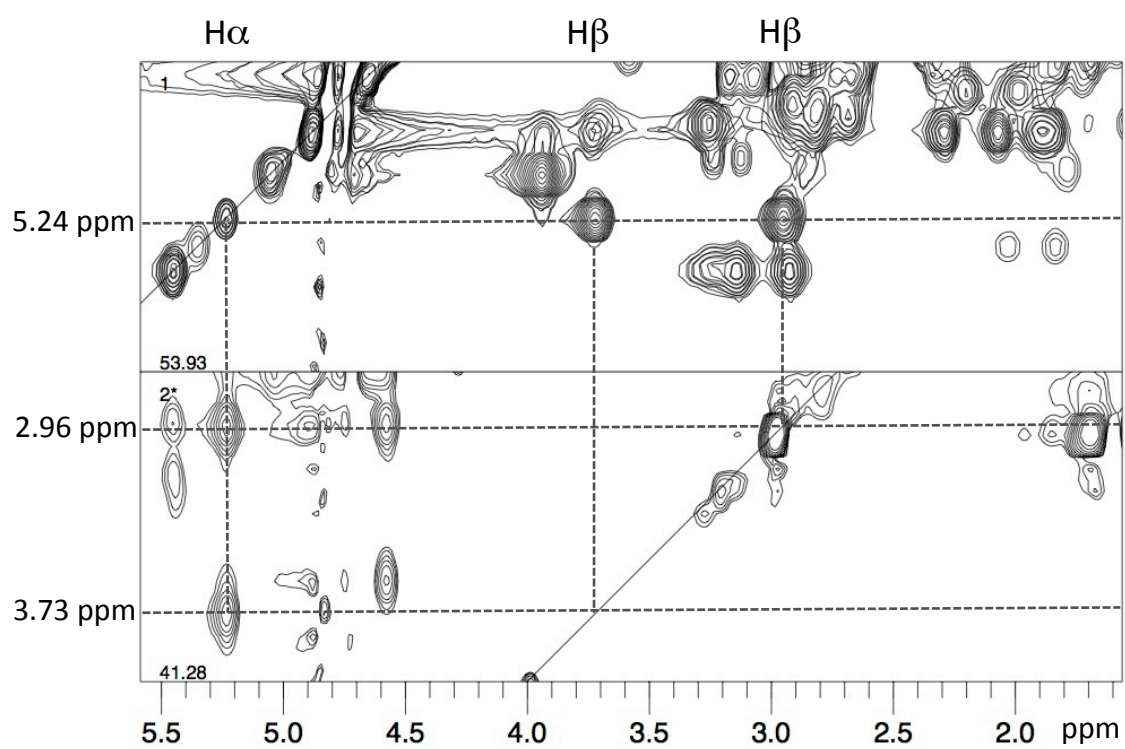


Figure 15

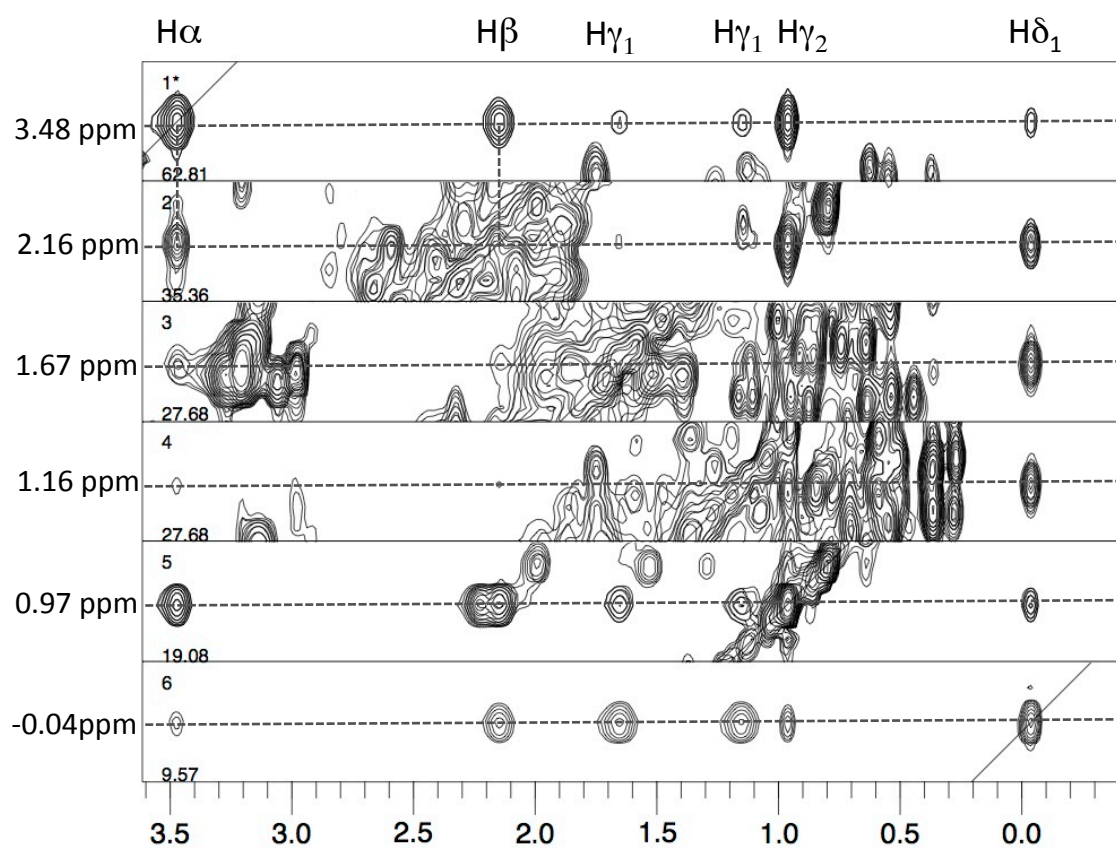


Figure 16

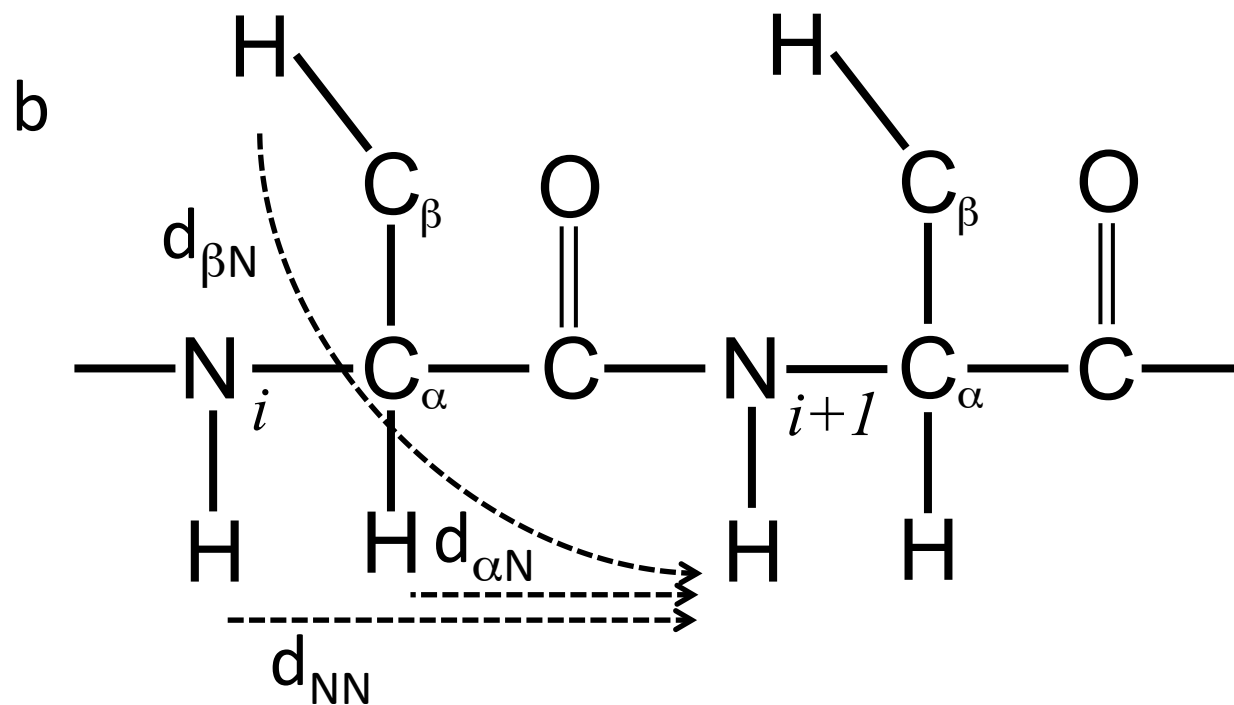
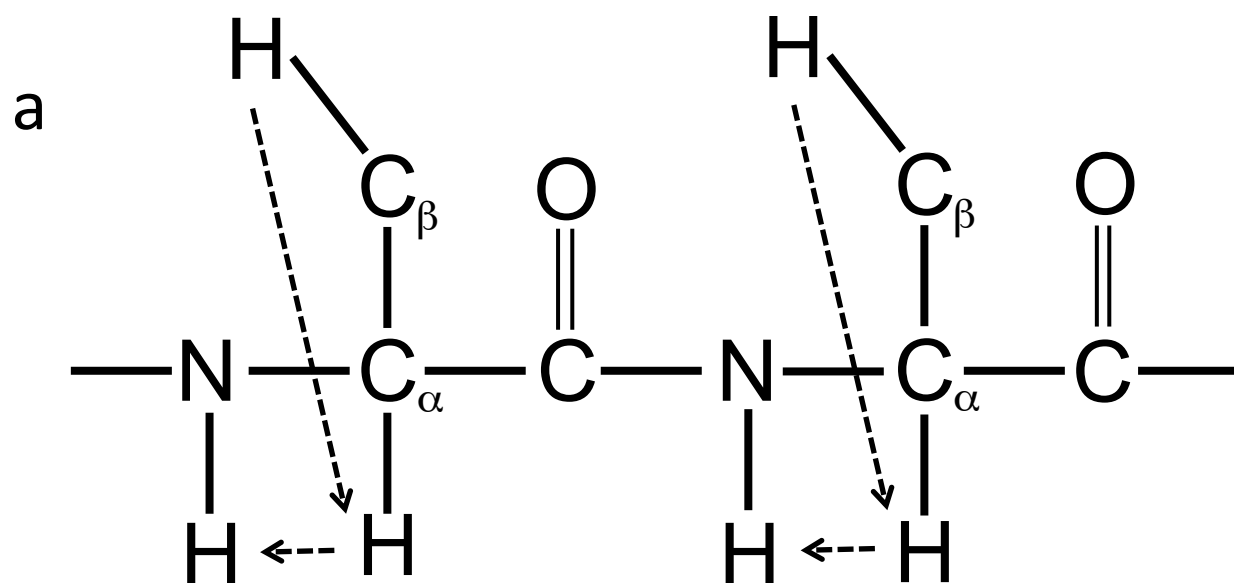


Figure 17

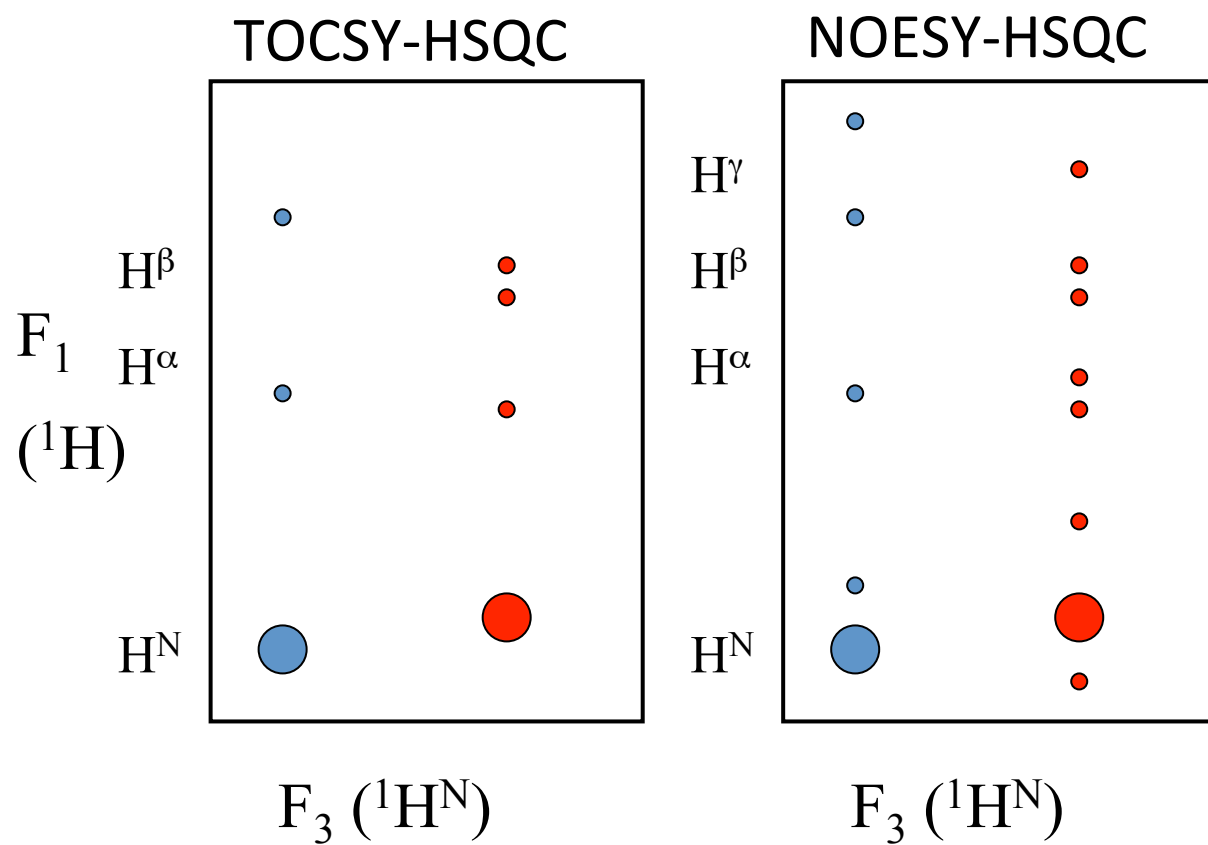


Figure 18



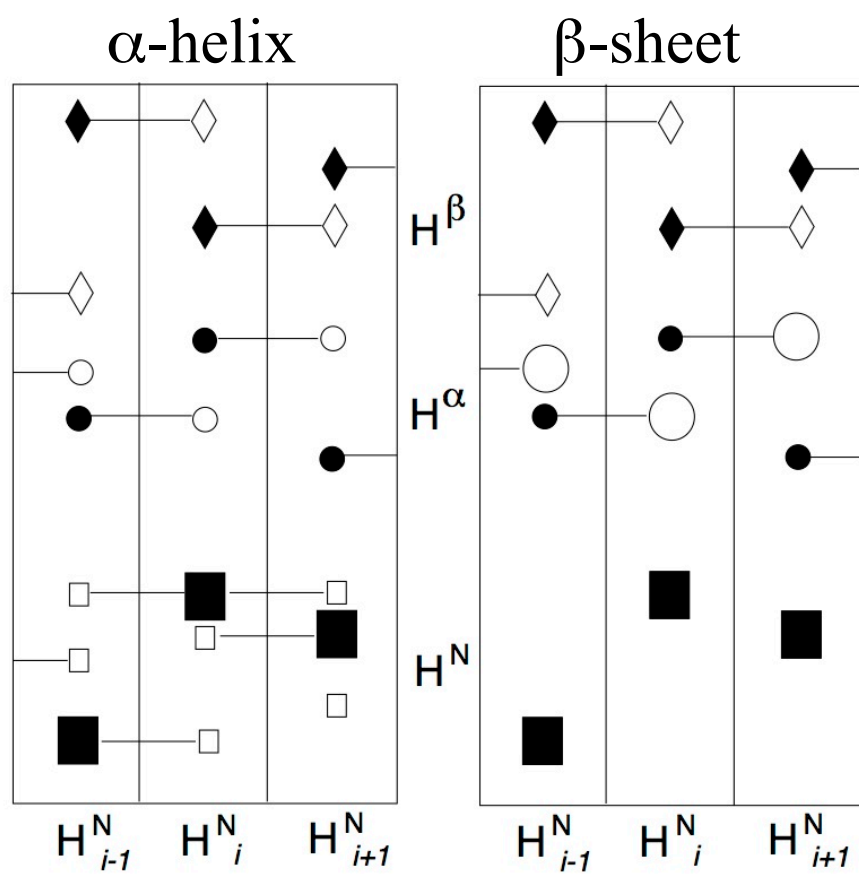


Figure 19

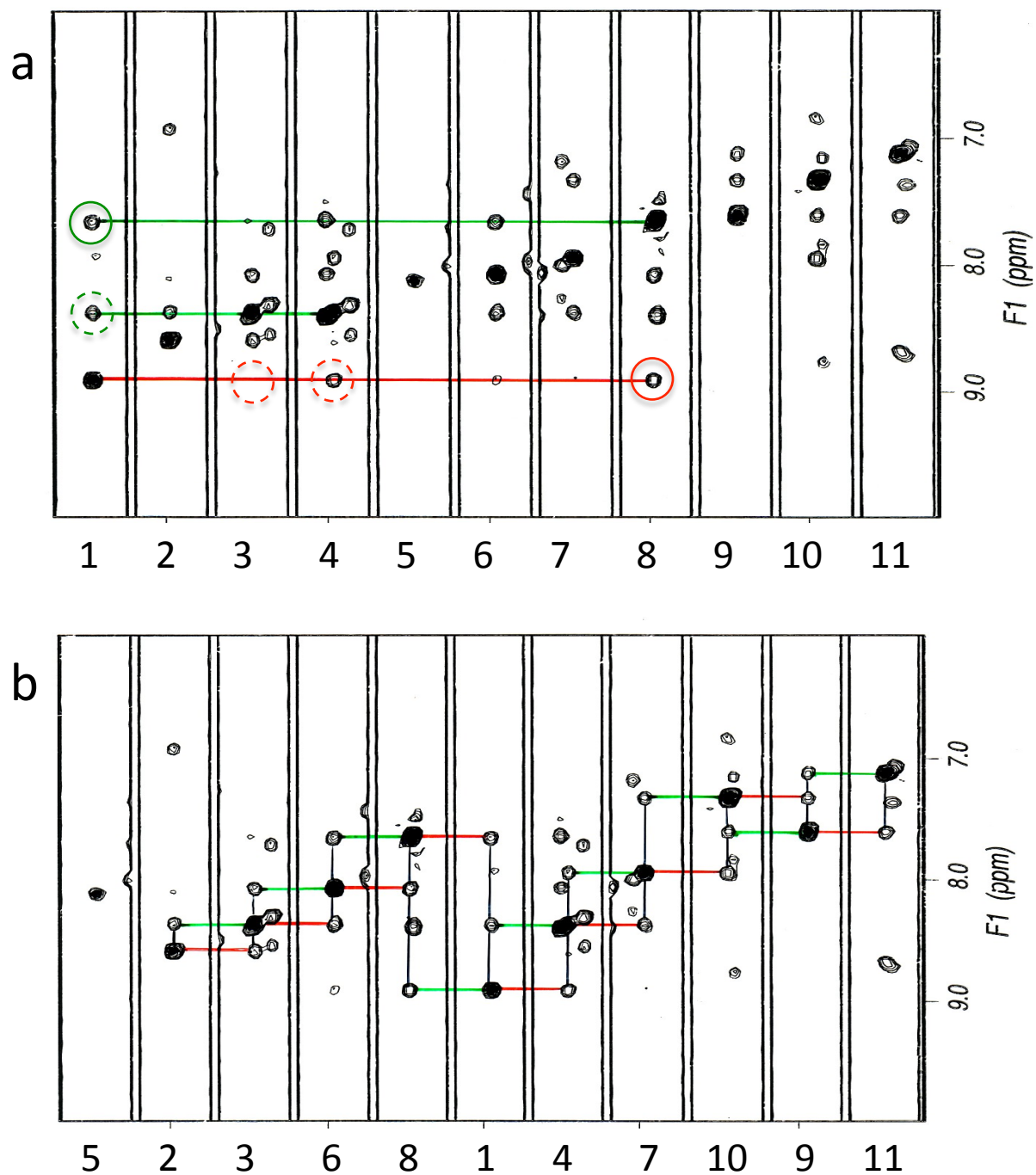


Figure 20

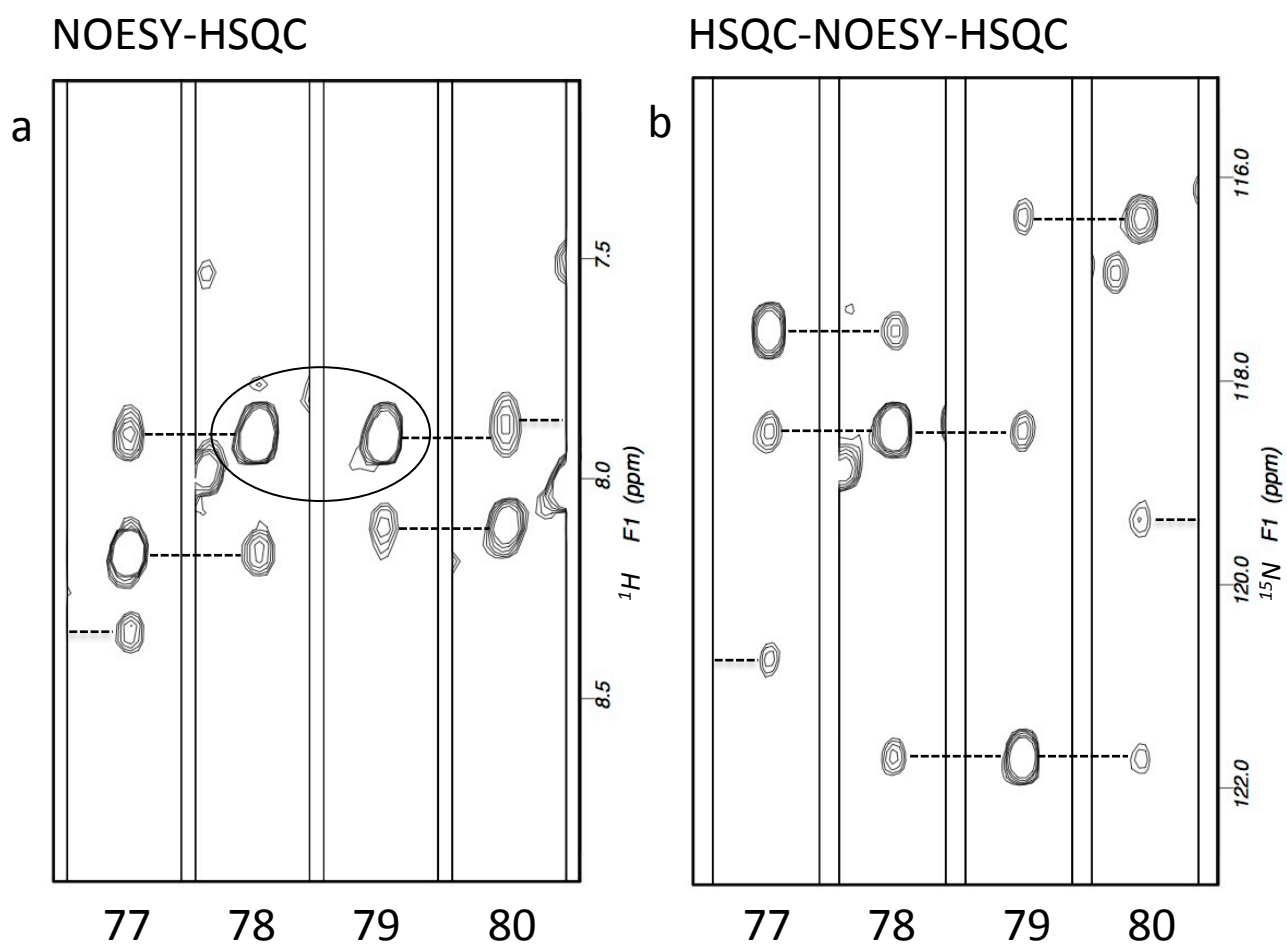


Figure 21

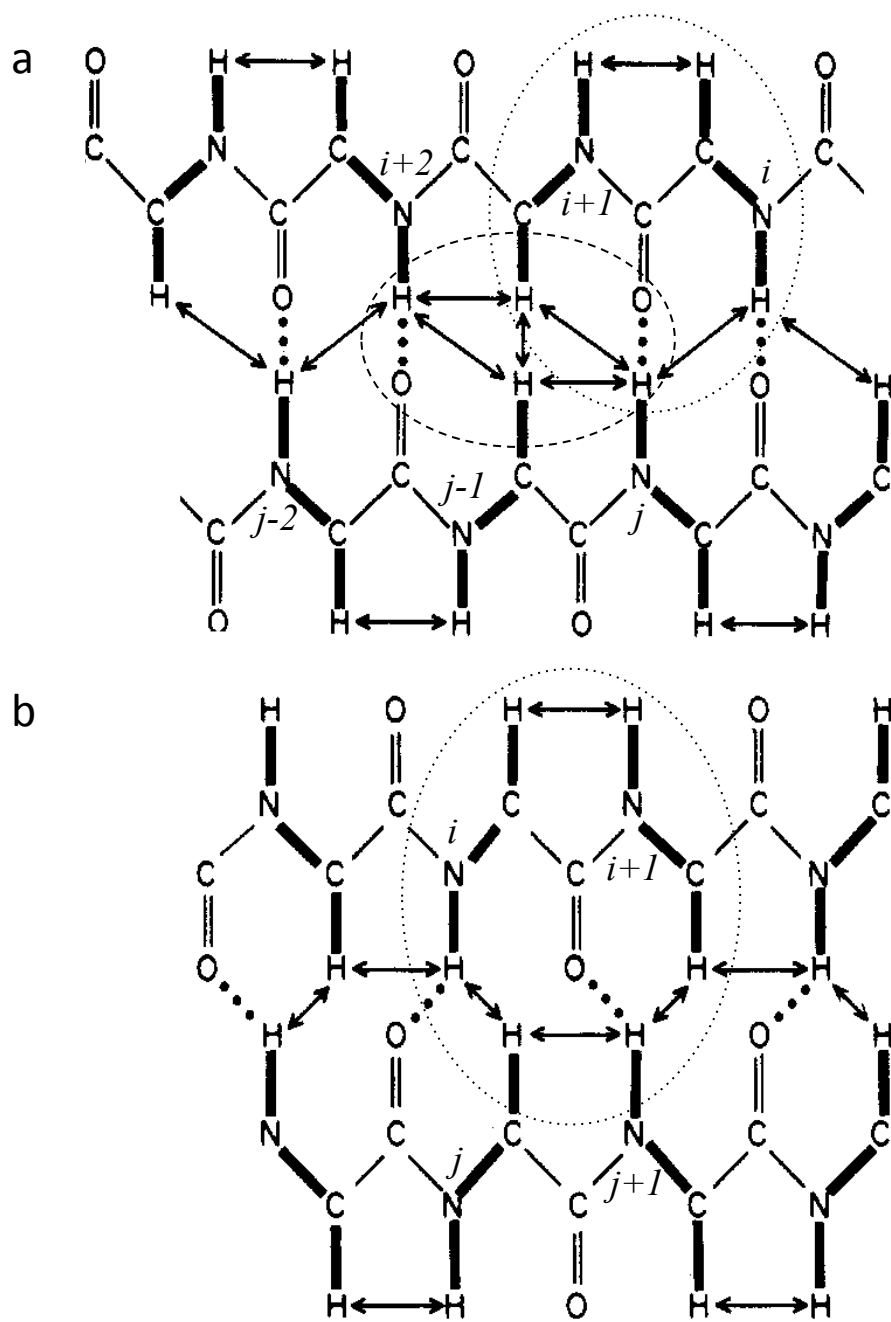


Figure 22