Proc. Indian Acad. Sci. (Chem. Sci.), Vol. 106, No. 7, December 1994, pp. 1537–1549.

Heteronuclear two- and three-dimensional NMR studies on the R1–R2–R3 domain of *Drosophila melanogaster* c-myb protein: Spin system identifications

P K RADHA, ANUP MADAN, L C PADHY and R V HOSUR*
Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India

Abstract. Advantages of heteronuclear two- and three-dimensional NMR experiments in obtaining better dispersion of peaks in spectra of large protein molecules have been described. The basic experimental techniques have been qualitatively presented and their application to a protein of 160 amino acid residues has been described. Several residue-type specific signals have been identified. The analysis of three-dimensional ^{13}C resolved ^{1}H–^{1}H TOCSY spectra for spin system identifications has been described in some detail.

Keywords. Multidimensional NMR; c-myb protein; heteronuclear NMR.

1. Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy has emerged during the last two decades as the single most invaluable spectroscopic technique for atomic level structure determination of biological macromolecules in aqueous solutions, under conditions similar to those prevalent in real systems (Wüthrich 1986). More than a hundred protein structures and a similar number of DNA structures, in the molecular weight range of 5–10 kDa have been determined within a short span of 10 years (see reviews, van de Ven and Hilbers 1988, Hosur *et al* 1988, Clore and Gronenborn 1989). Several methodological advances in the NMR technique have contributed to this success and the most important of these is the development of two-dimensional NMR (Jeener 1971; Aue *et al* 1976; Jeener *et al* 1979; Ernst *et al* 1987; Ernst 1992). The 3-D structure determinations have been based on identification and quantification of specific interactions between protons in a given molecule, which are vividly displayed on a plane in 2-D NMR spectra (Wüthrich 1986; Kaptein *et al* 1988; Kessler *et al* 1988; Bax 1989; Clore and Gronenborn 1989; Wüthrich 1989).

The emphasis today is to extend the applications of NMR to still larger molecules, 20–30 kDa, so that many of the enzymes and DNA-protein complexes become amenable to detailed investigations and thus lead to important biological insights. It has been recognized that for such large systems, ^{1}H NMR alone is not adequate because of the smaller chemical shift ranges resulting in heavy overlaps of peaks in 2-D NMR spectra. Individual specific correlations can no longer be identified, let alone their quantification. These difficulties have promoted developments in two directions: (a) extending NMR dimensions to 3 and 4 using heteronuclei such as ^{15}N

*For correspondence

and/or ^{13}C (Fesik and Zuiderweg 1988; Kessler *et al* 1988; Kay *et al* 1990; Bax and Grezesick 1993; Oschkinat *et al* 1994). These nuclei have larger chemical shift ranges and thus the correlations can now be spread with good dispersions in a 3- or 4-dimensional space, (b) use of genetic engineering techniques for overproducing and specifically labelling the molecules with heteronuclei (Miller 1972; Bachmann 1983; McIntosh and Dahlquist 1990; Tate *et al* 1992).

We embarked on such an adventure two years ago, with myb, an oncogene product, as our system of interest. Myb proteins are transcriptional factors, playing important roles in cell differentiation and growth (Beidenkepp *et al* 1988; Nishina *et al* 1989; Sakura *et al* 1989; Howe *et al* 1990; Oehler *et al* 1990). They bind DNA in a sequence-specific manner, the specific recognition sequence being pyAACG/TG (Beidenkepp *et al* 1988). The proteins are typically more than 600 residues long and are divided into several independent domains responsible for specific functions. Near their N-terminii, there is a highly conserved stretch of 155–160 residues, made of three imperfect repeats R1, R2, R3 each 52–53 residues long (Klempenauer and Sippel 1987; Ibanez *et al* 1988). Each of these repeats has three conserved tryptophans at regular intervals and this whole domain is known to be responsible for the specific DNA binding activity of the protein. The three-dimensional structure of the protein is not yet known and from experimental data there is a consensus that the conserved tryptophans have important roles to play in DNA-binding directly or indirectly.

By using gene cloning and bacterial expression technology, we have overproduced the 160 amino acid long DNA binding domain, R1–R2–R3, of the *Drosophila* c-myb protein and have carried out several spectroscopic investigations to study its stability, folding and DNA binding properties (Madan *et al* 1994; Radha *et al* 1994). We have found that the protein distinguishes specific and non-specific DNA targets very efficiently and the specific DNA binds to a hydrophobic domain of the protein, to which the tryptophans contribute in a significant manner (Madan *et al* 1994). We have also partially characterized the protein by ^{1}H NMR in two dimensions and have identified some residue-specific signals in the spectra (Radha *et al* 1994). In this paper we focus on results from heteronuclear NMR in 2- and 3-dimensions to obtain additional spin-system identifications.

2. Resolution enhancement by heteronucleus in two- and three-dimensions

Figure 1 illustrates how the use of heteronucleus allows resolution enhancements and removes ambiguities. In the ^{1}H–^{1}H 2-D-correlated spectrum, peak labelled 'p' is an overlap of two diagonal peaks correlated to two different protons. On the face of it, it appears that a single proton represented by 'p' is correlated to two other protons. In the 2-D ^{1}H–^{13}C correlated spectrum shown above, we see that, at the ^{1}H chemical shift of p, there are two peaks corresponding to different ^{13}C chemical shifts. This clearly indicates that p is an overlap of two different diagonal peaks and thus represents two different protons on different carbons. This spectrum however does not tell which proton in 'p' is connected to which of the two ^{1}H correlations seen in the ^{1}H–^{1}H spectrum.

The 2-D ^{1}H–^{1}H and ^{1}H–^{13}C spectra shown above also illustrate an opposite situation. The two protons represented by q and r diagonal peaks are well resolved in the ^{1}H spectrum but the carbon chemical shifts for the two are the same (peaks

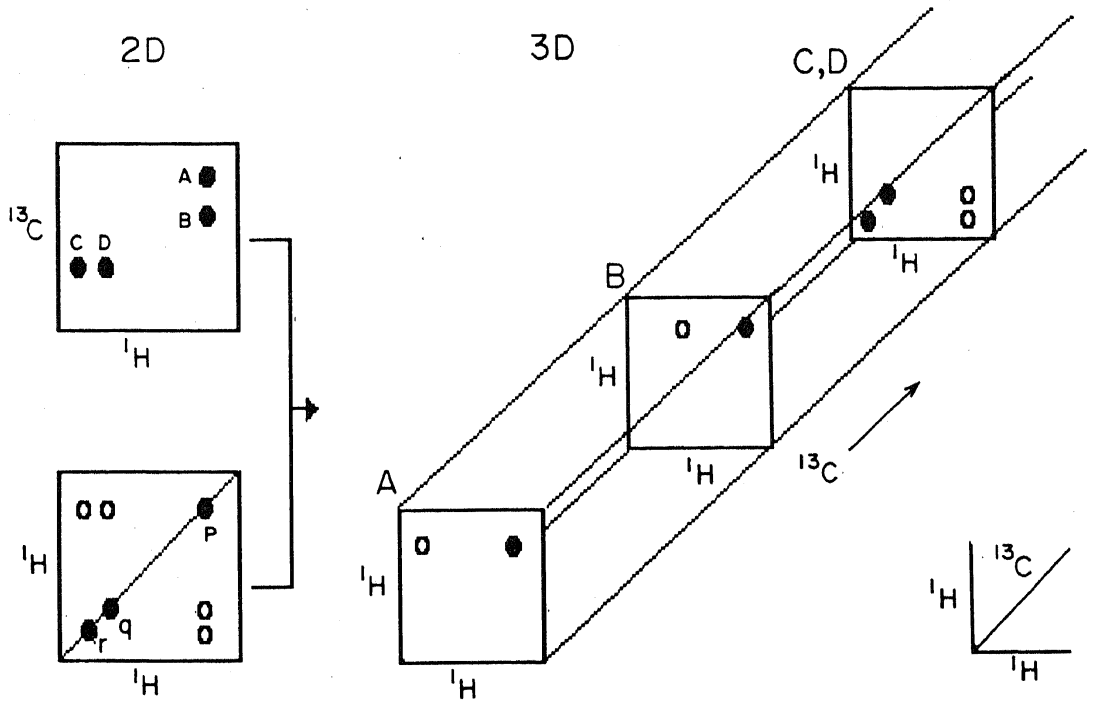

Figure 1. Illustration of the use of heteronucleus (^{13}C) for enhancement of resolution in two- and three-dimensional NMR spectra. Left side of the figure shows schematically ^{1}H–^{1}H (bottom) and ^{1}H–^{13}C (top) correlated spectra. The peak 'p' in the bottom spectrum represents two unresolved protons on different carbons but these are resolved in the top spectrum as peaks A and B. These protons are *J*-correlated to protons represented by peaks r and q in the bottom spectrum and the peaks C and D in the top spectrum. The correspondence is however not clear. The correlations can be clearly discerned from the 3-D spectrum shown in the right side of the figure. The ^{1}H–^{1}H correlations are separated along the carbon axis in the third dimension. Filled circles which lie on the diagonal and the open circles identify the correlations.

C and D in the ^{1}H–^{13}C spectrum). This actually represents a loss of information on the ^{13}C dimension.

A 3-D spectrum takes the best of the two, taking advantage of both ^{1}H and ^{13}C shift separations and results in an overall resolution enhancement as depicted in the schematic spectrum shown in the right panel of figure 1. This spectrum has both ^{1}H–^{1}H and ^{1}H–^{13}C correlations and the ^{1}H–^{1}H correlations corresponding to each ^{13}C shift are displayed in a separate plane. Thus it is now clear which ^{13}C is connected to which ^{1}H etc. and the ambiguities get resolved.

Each plane of a 3-D spectrum clearly contains much fewer peaks than a typical 2-D spectrum carrying similar information and thus the interpretations also become easier. An experimental illustration of the simplifications is shown in figure 2. Details of these spectra will be discussed later and it is enough to note here that the slice of the 3-D spectrum shown in figure 2b represents a filter through the 2-D spectrum shown in figure 2a.

3. Experimental techniques for heteronuclear correlations in two and three-dimensions

A number of experimental techniques for heteronuclear correlations in 2, 3 and 4 dimensions have been described and these have been elegantly reviewed recently by

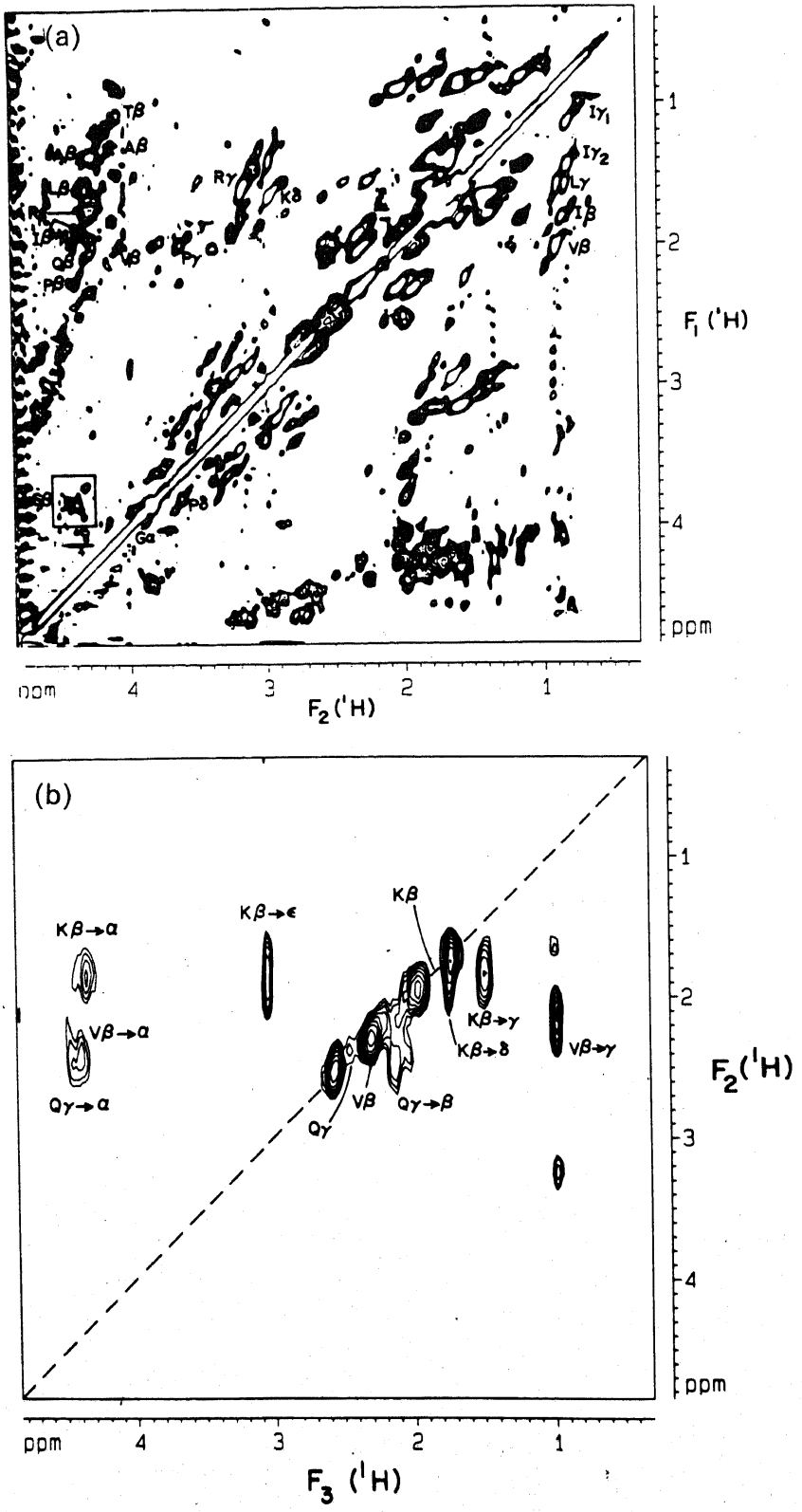

Figure 2. Experimental illustration of the spectral simplifications achieved in the 3-D spectrum as compared to the 2-D spectrum (**a**). The spectrum in (**b**) is a ^{13}C plane through the full 3-D spectrum shown in figure 7, at a ^{13}C shift of 33·4 ppm. A number of correlations which are hard to find out clearly in the top spectrum are readily identified in the lower spectrum and these have been indicated. The spin systems of K, V and Q residues can be clearly seen.

several authors (Clore and Gronenborn 1991; Bax and Grezesick 1993; Zuiderweg *et al* 1993; Oschkinat *et al* 1994). We refer the reader to these articles for extensive surveys and we shall consider only the few very basic techniques, which have been explicitly used in the present paper. Almost all the techniques which rely on these basic techniques, employ ^{1}H detection (Bodenhausen and Freeman 1977; Maudsley and Ernst 1977; Bodenhausen and Reuben 1980; Kay *et al* 1989) which gives substantially higher sensitivity, and rely on large one-bond X–H coupling constants (X represents a heteronucleus) for efficient coherence transfers. At some stages ^{1}H–^{1}H transfers (Braunschweiler and Ernst 1983; Davis and Bax 1985) also may be involved and these could be *J*-transfers or NOE-transfers (Jeener *et al* 1979; Anil Kumar *et al* 1980; Macura and Ernst 1980) depending on the needs.

Figure 3 shows the few experimental pulse sequences which have been used in the present work, described in detail in the following section. Pulse sequence A is the usual HMQC (Heteronuclear Multiple Quantum Coherence) transfer based correlation (Bodenhausen and Freeman 1977; Maudsley and Ernst 1977). The experiment works in the following fashion. The first ^{1}H pulse creates ^{1}H magnetization which evolves for a time τ under ^{1}H–^{13}C coupling and is converted into heteronuclear multiple quantum coherence by the first ^{13}C pulse. This evolves during t_1 with ^{13}C frequency, since the ^{1}H frequencies are refocussed by the 180° ^{1}H pulse in the middle of the t_1 period. At the end of t_1, the magnetization is transferred back to ^{1}H by the 90° carbon pulse and allowed to refocus during the period τ. The ^{1}H magnetization is then detected as a function of t_2, during which ^{1}H–^{13}C coupling is removed by ^{13}C decoupling by GARP (Shaka *et al* 1985) sequence. Although we have explicitly mentioned ^{13}C here, the same pulse sequence holds good for any other heteronucleus, the most useful for proteins being ^{15}N. Two-dimensional Fourier transformation of the data results in 2-D ^{1}H–^{13}C correlated spectrum with ^{13}C appearing along the F_1 axis and ^{1}H along the F_2 axis. The sensitivity of such an experiment would be comparable to that of a ^{1}H–^{1}H *J*-correlation experiment if all the carbons to which the protons are attached are labelled 100% by ^{13}C isotope. However, as we will see later, if the protein concentration is 5–6 mM, reasonably good spectra can be obtained even at natural abundance of ^{13}C, in an affordable period of time.

Pulse sequence B is similar to A, but for the ^{1}H–^{1}H TOCSY (Braunschweiler and Ernst 1983) mixing introduced prior to detection along t_2. Thus the spectrum reflects ^{1}H–^{13}C correlations in a direct manner but reflects ^{1}H–^{1}H correlations in an indirect manner. The GARP (Shaka *et al* 1985) decoupling employed during detection can also be removed in this case which will allow discrimination between ^{1}H–^{13}C transfer peaks and ^{1}H–^{1}H TOCSY relayed peaks at particular carbon frequencies. The latter do not show ^{1}H–^{13}C coupling and thus appear as single peaks, while the former appear as doublets due to the one-bond X–H coupling. These will be exemplified in the following section.

Pulse sequence C represents a 3-D version of the magnetization transfer process employed in B. The t_1, t_2 and t_3 periods have ^{13}C, ^{1}H and ^{1}H frequencies respectively. The t_1, t_2 periods are connected by ^{1}H–^{13}C coherence transfers and t_2 and t_3 periods are connected by ^{1}H–^{1}H coherence transfers. The data have to be Fourier-transformed along the three dimensions t_1, t_2 and t_3 to generate a 3-D spectrum $S(F_1, F_2, F_3)$

$$S(t_1, t_2, t_3) \xrightarrow{\text{3DFT}} S(F_1, F_2, F_3) \tag{1}$$

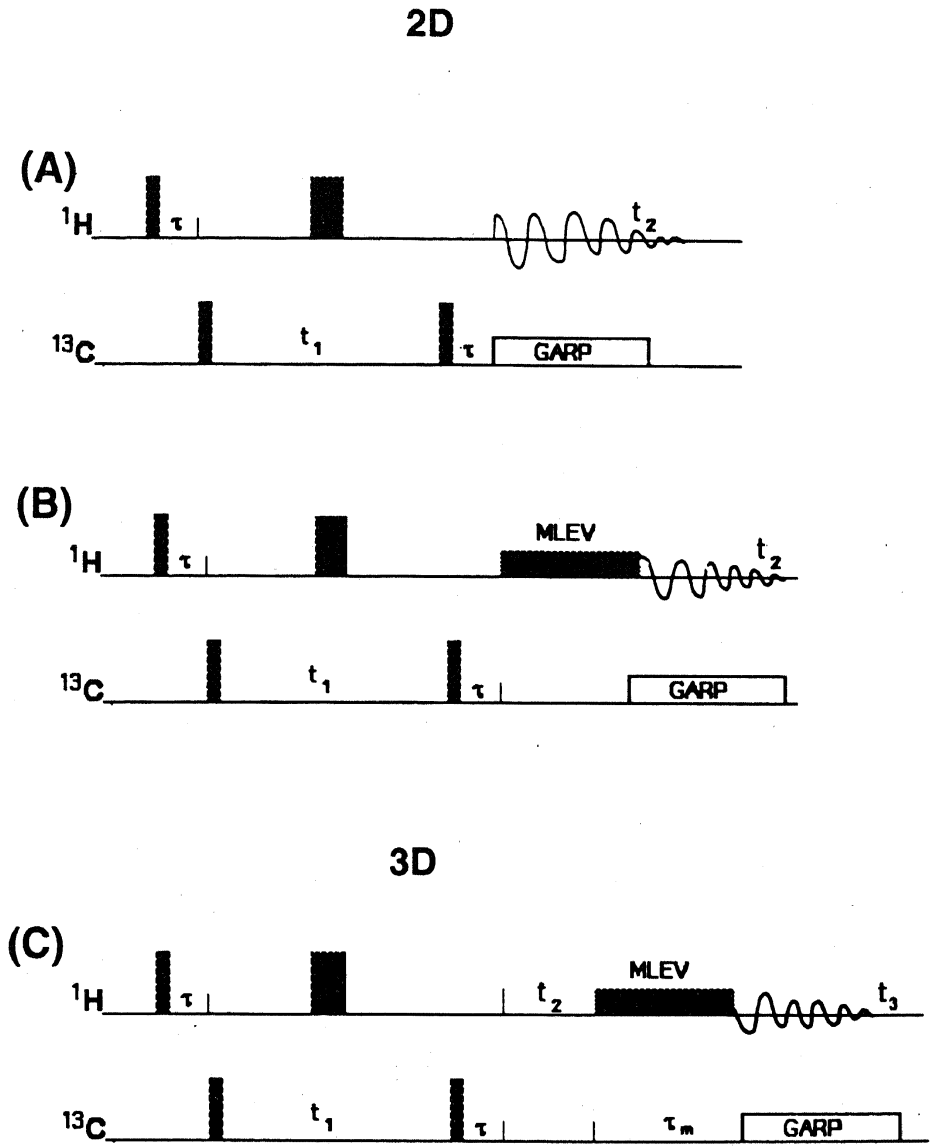

Figure 3. Experimental pulse sequences for 2-D HMQC (**A**), HMQC–TOCSY (**B**) and 3-D ^{13}C resolved HMQC–TOCSY (**C**) experiments. MLEV is a ^{1}H–^{1}H mixing sequence and GARP represents a ^{13}C-decoupling sequence. τ is a short delay equivalent to 1/2 $^1J_{CH}$ where $^1J_{CH}$ is the one-bond proton–carbon coupling constant, and t_1, t_2, t_3 are the usual incrementable time variables. The thin pulse is a 90° pulse and the thick pulse represents a 180° pulse.

The (F_2, F_3) planes in such a spectrum contain the normal TOCSY peaks observed in 2-D spectra, but separated by the ^{13}C chemical shifts. The spectrum shown in figure 2a is one of such planes.

4. Spin system identifications in R1–R2–R3 protein

Spin system identifications of the individual amino acid residues in the NMR spectra of proteins constitutes an important step for both sequence specific resonance

assignments and subsequent 3-D structure determinations. In large proteins, such as the present one, proton homonuclear experiments alone are not adequate and a combination of $^{1}H-^{1}H$ and $^{1}H-^{13}C$ 2-D and ^{13}C edited 3-D experiments have to be employed for unambiguous identifications. A particular advantage of ^{13}C that has been observed from the large number of assignments obtained so far is that ^{13}C chemical shifts are less sensitive (compared to ^{1}H) to secondary and tertiary structures of the protein and thus provide useful guides for identifying residue type specific signals.

Figure 4 shows a schematic 2-D $^{1}H-^{13}C$ correlated spectrum indicating the expected positions of different cross-peaks for the different amino acid residues. Several

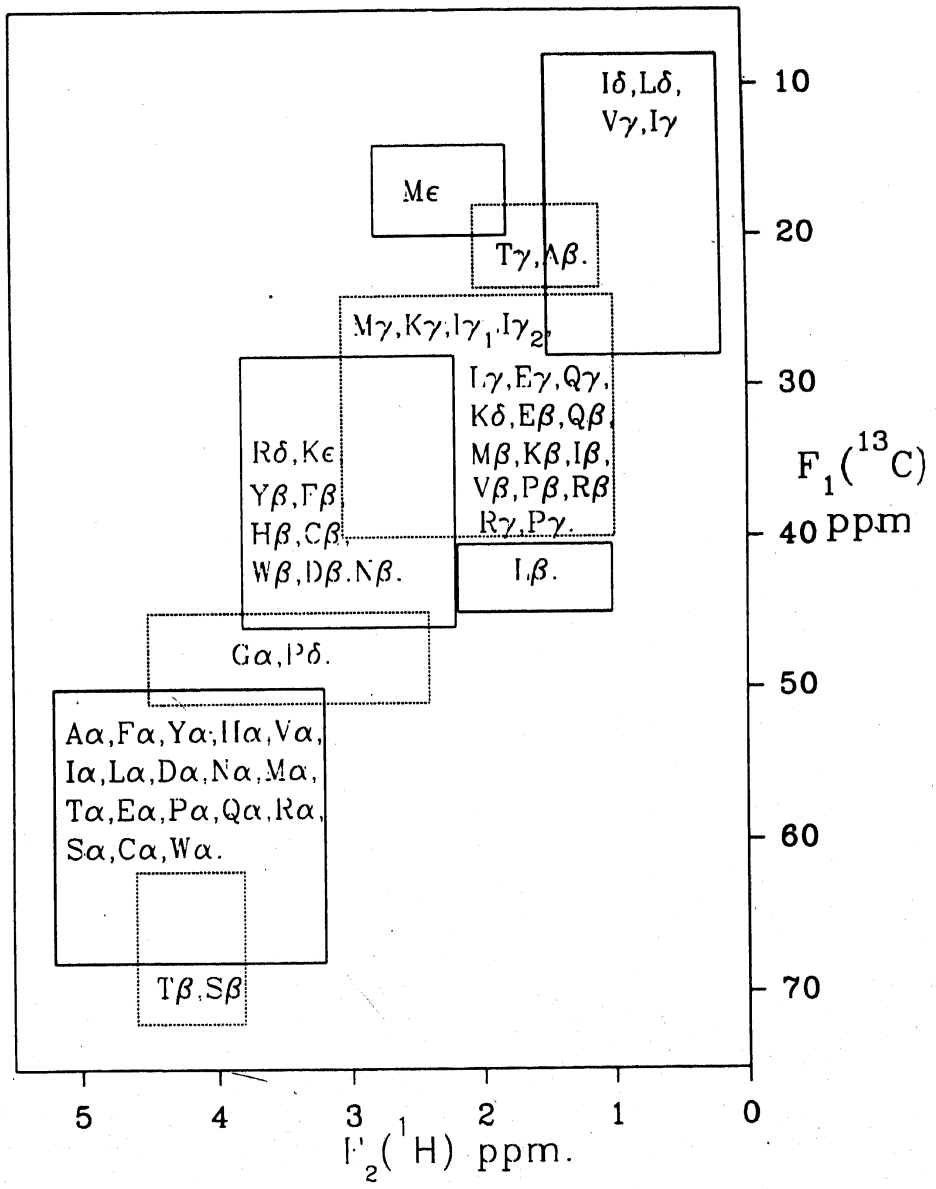

Figure 4. Schematic 2-D $^{1}H-^{13}C$ correlated spectrum of a protein indicating the generally expected positions of cross-peaks of different protons. The residues have been identified by one letter symbols and the Greek letters identify the carbons to which the protons are attached.

distinctive features are seen in the spectrum which help a great deal in specific identifications. First of all, Thr and Ser C^{β} carbons resonate furthest downfield in characteristic positions. Second, the various methyl carbons are well separated which is in contrast to the methyl protons of different residues. Gly C^{α} and Pro C^{δ} carbons resonate in distinctive regions. It is also noticeable that the C^{α} carbons have a distinct chemical shift range as compared to the C^{β}, C^{γ} or other carbons in the side chains.

Figures 5, 6 and 7 display the HMQC (pulse sequence A of figure 3), HMQC–TOCSY (pulse sequence B of figure 3) and 3-D ^{13}C resolved ^{1}H–^{1}H TOCSY (pulse sequence C in figure 3) spectra of the R1–R2–R3 protein respectively. The HMQC–TOCSY spectrum has been recorded without GARP decoupling in the t_2 period and this allows discrimination of ^{13}C to ^{1}H transfers and ^{1}H to ^{1}H TOCSY transfers.

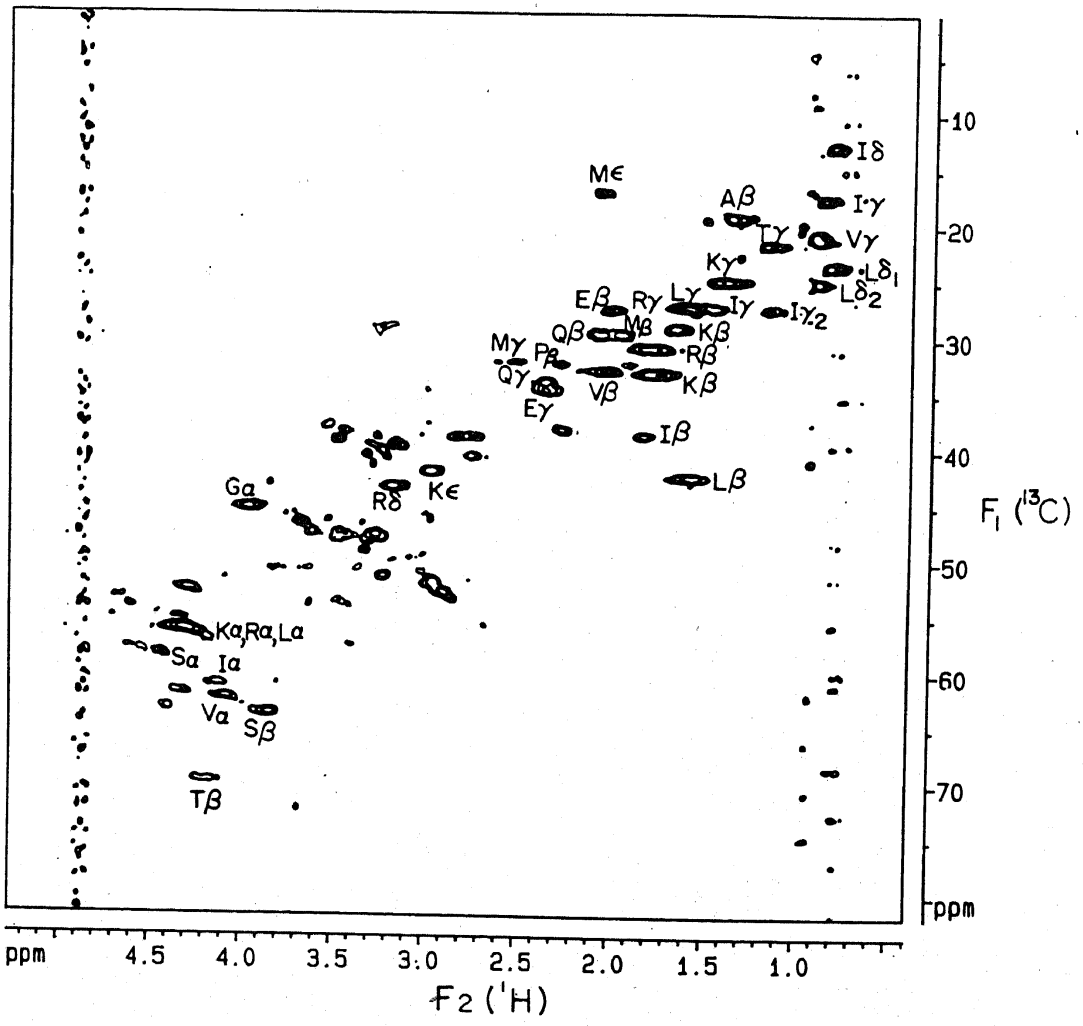

Figure 5. ^{1}H–^{13}C HMQC spectrum at natural abundance of R1–R2–R3 protein recorded with pulse sequence A of figure 3 on BRUKER AMX 500 NMR spectrometer. Sample conditions: 5 mM protein, *p*H 4·2, temperature 16°C. 2048 t_2 points and 512 t_1 points were used with carbon offset placed in the middle of the C^{β} region. The carbon spectral width was 90 ppm. Quadrature detection along F_1 was achieved by the standard TPPI procedure. The various peak assignments indicated have been obtained by analysis of different 2-D spectra and the 3-D spectrum shown in figure 7. Total experimental time was 12 h.

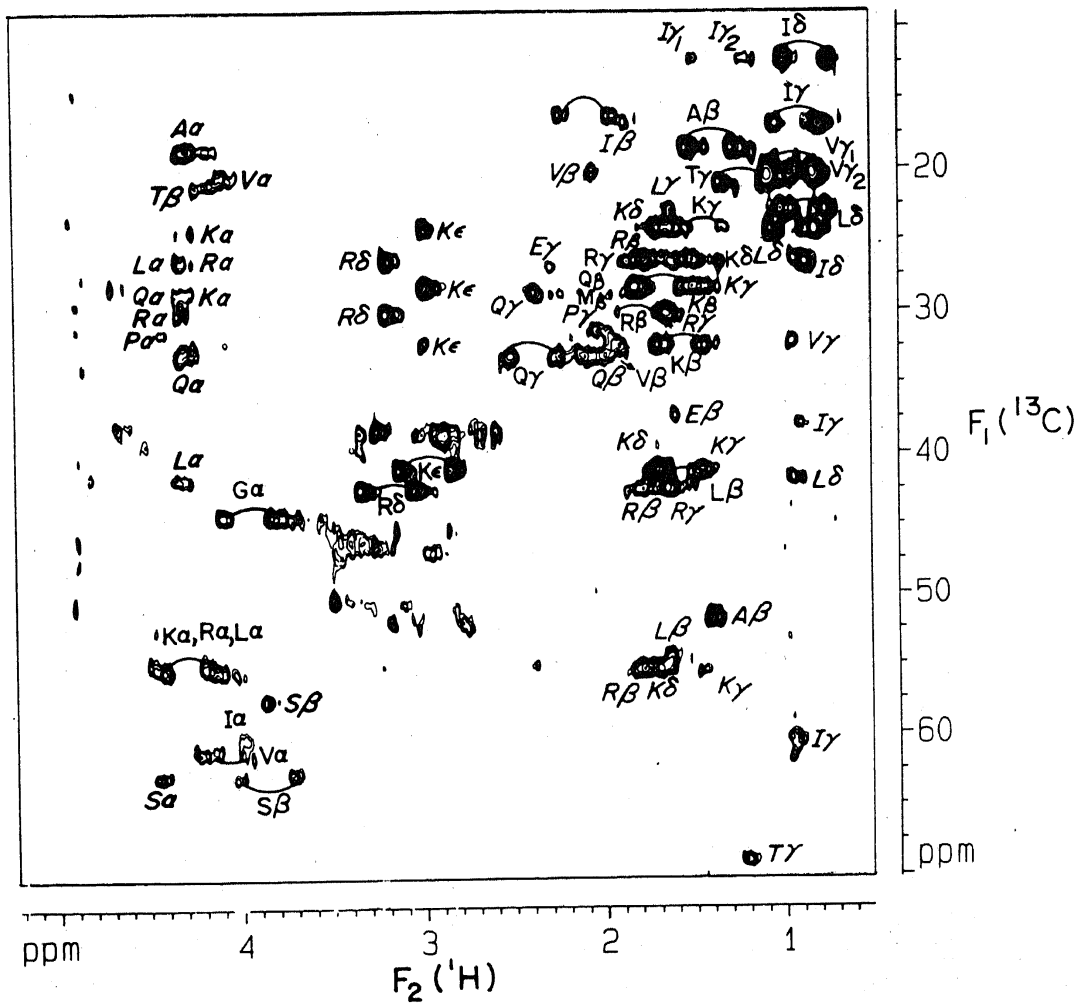

Figure 6. ^{1}H–^{13}C HMQC–TOCSY spectrum at natural abundance of the R1–R2–R3 protein under the same conditions as in figure 5. The MLEV mixing time was 51 ms. Other experimental parameters were the same as in figure 5. No ^{13}C decoupling was employed during detection and thus the direct peaks in the spectrum retain the one bond $^{1}J_{CH}$ coupling (they have been joined). The relay peaks are however single peaks. We notice that, here as well as in figure 5, dispersion does not exist within each residue type. Almost all the peaks except some in the C^{α} region have been assigned.

The former peaks will be doublets retaining the large one bond $^{1}J_{CH}$ couplings, while the latter will be single peaks. The 2-D spectra have been analysed in conjunction with the ^{1}H–^{1}H 2-D TOCSY spectra of the protein in $^{1}H_2O$ and $^{2}H_2O$ solutions. The $^{2}H_2O$ spectrum used is the same as in figure 2b. Many of the ambiguities originating from the equivalence of carbon chemical shifts were removed from the analysis of the 3-D spectrum shown in figure 7 which was also recorded without GARP decoupling in the t_3 period. In such a spectrum, each carbon plane will have pairs of peaks separated by $^{1}J_{CH}$ on the diagonal for each proton and the chemical shift of the proton corresponds to the centre of the doublet.

Spin system identifications from 3-D spectra involves examination of various ^{13}C planes and monitoring connectivities in the individual planes. The procedure is schematically indicated in figure 8 for a few residues. As an example consider the

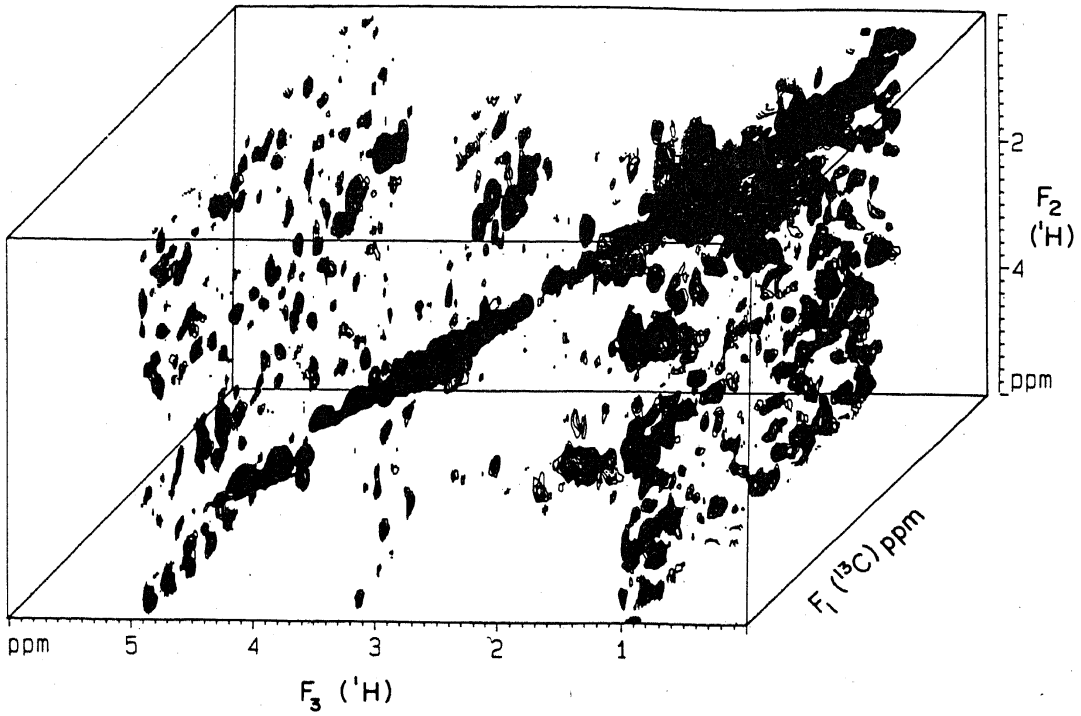

Figure 7. Three-dimensional ^{13}C resolved ^{1}H–^{1}H TOCSY spectrum of the R1–R2–R3 protein at natural abundance under the same sample conditions as in figures 5 and 6. MLEV mixing time was 30 ms and the experimental parameters of spectral width were same as in figure 6. 16 transients were used for each increment and the data set consisted of 512 t_3, 64 t_2 and 64 t_1 points. Total experimental time was approximately 36 h. GARP decoupling was not employed during t_3 period.

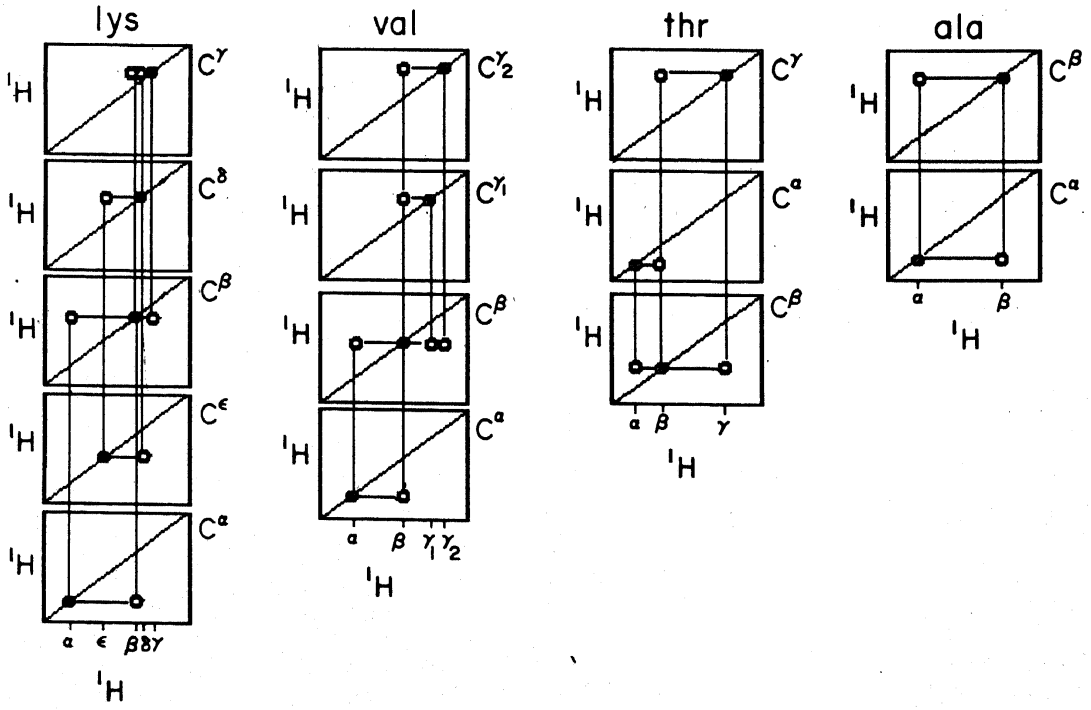

Figure 8. Schematic representation of spin system identification of different residues (4 are shown) from the 3-D spectrum shown in figure 7. In each lane, each rectangle represents a ^{1}H–^{1}H TOCSY spectrum (F_2–F_3 planes) at a particular carbon chemical shift (F_1 axis). The pattern of peaks expected in different carbon planes and their correlations are indicated for each residue type. Filled circles on diagonals originate from ^{1}H–> ^{13}C–> ^{1}H magnetisation transfer and open circles representing relays originate from ^{1}H–> ^{1}H TOCSY transfers.

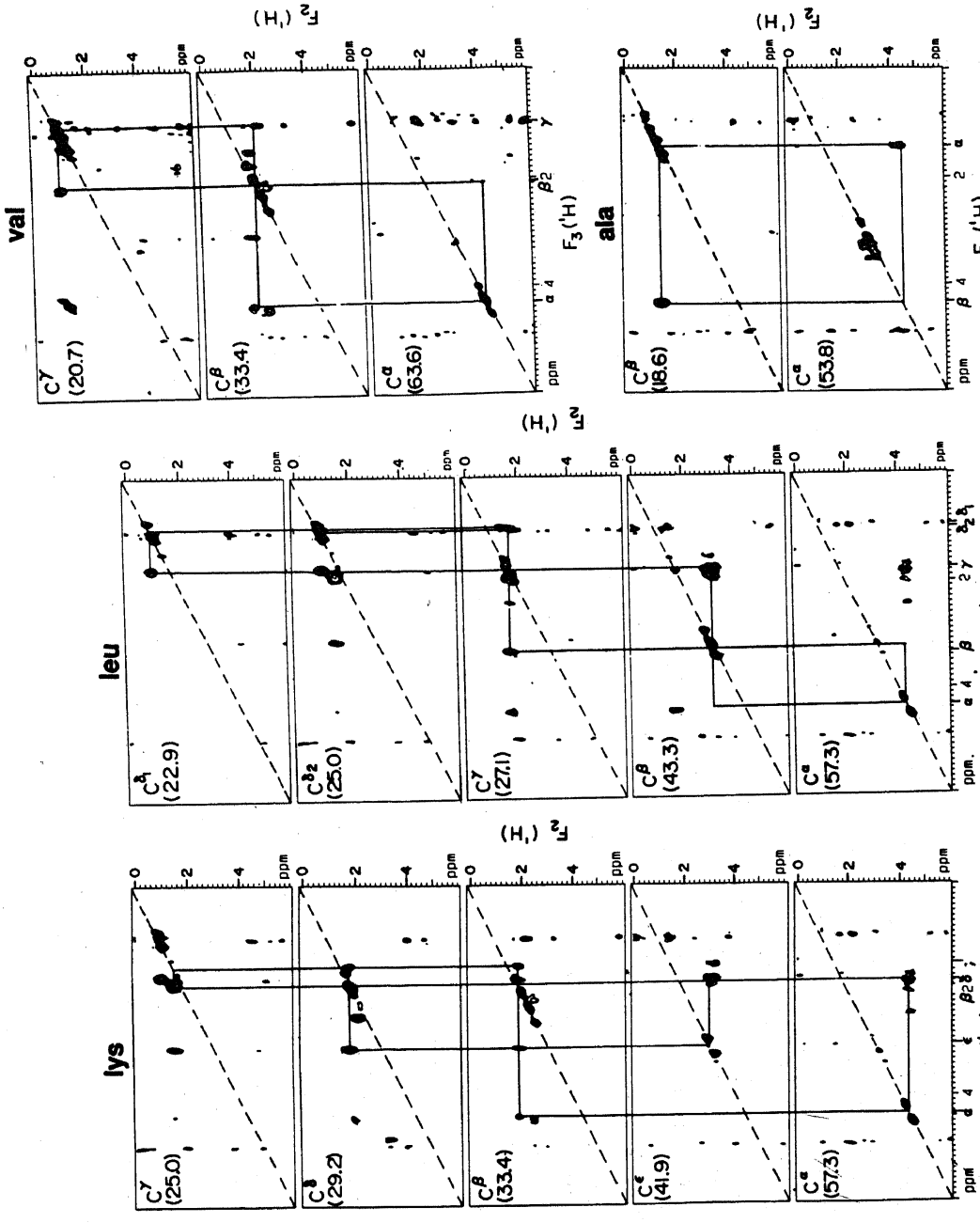

Figure 9. Illustrative spin system connectivities in the 3-D spectral planes for the residues Ala, Val, Lys and Leu. Since no GARP decoupling was employed, the diagonal peaks are doublets and the relay occurs from these doublets. Therefore the ^{1}H chemical shift actually corresponds to the midpoint of a doublet on the diagonal. The chemical shifts of the various carbon planes have been indicated and the connections drawn include only the first step of TOCSY relay. In many planes additional peaks are seen due to second and third steps of TOCSY relays.

simple case of alanine residue. Two carbon planes namely C^{β} and C^{α} need to be identified, which contain two peaks each. In the C^{β} plane the diagonal peak occurs at the methyl proton chemical shift and a cross peak appears at the α proton chemical shift. The reverse occurs in the C^{α} plane. Similarly for Thr, three carbon planes belonging to C^{γ}, C^{β} and C^{α} need to be connected as indicated in the figure. Figure 9 shows experimental illustrations of the above connectivities for a few spin systems. All the assignments thus obtained have been indicated in figures 5 and 6.

We notice from the experimental spectra and the identifications discussed above that, for each residue type, the chemical shift dispersion for both ^{1}H and ^{13}C is very small. We observed a similar situation in the case of ^{15}N chemical shifts as well in ^{15}N–^{1}H correlated spectra. This is known to be a common feature of proteins with high helical contents. All the experiments described here have been performed on unlabelled protein samples. This obviously has limitations of sensitivity and resolution achievable in individual dimensions. Further experiments with selective excitation schemes and with singly and doubly labelled proteins for obtaining greater dispersions and consequently residue-wise identifications are in progress.

Acknowledgement

We thank the National Facility for High Field NMR supported by Government of India and located at Tata Institute of Fundamental Research, Bombay for providing NMR facilities.

References

Anil Kumar, Ernst R R and Wüthrich K 1980 *Biochem. Biophys. Res. Commun.* **95** 1
Aue P, Bartholdi E and Ernst R R 1976 *J. Chem. Phys.* **64** 2229
Bachmann B J 1983 *Microbiol Rev.* **47** 180
Bax A 1989 *Annu. Rev. Biochem.* **58** 223
Bax A and Grezesick S 1993 *Acc. Chem. Res.* **26** 131
Biedenkepp H, Borgmeyer U, Sippel A E and Klempenauer K H 1988 *Nature (London)* **335** 835
Bodenhausen G and Freeman R 1977 *J. Magn. Reson.* **28** 471
Bodenhausen G and Reuben D J 1980 *Chem. Phys. Lett.* **69** 185
Braunschweiler L and Ernst R R 1983 *J. Magn. Reson.* **53** 521
Clore G M and Gronenborn A M 1989 *Crit. Rev. Biochem.* **24** 479
Clore G M and Gronenborn A M 1991 *Prog. NMR Spectrosc.* **23** 43
Davis D G and Bax A 1985 *J. Am. Chem. Soc.* **107** 2821
Ernst R R 1992 *Angew. Chem., Int. Ed. Eng.* **31** 805
Ernst R R, Bodenhausen G and Wokaun A 1987 *Principles of nuclear magnetic resonance in one and two dimensions* (Oxford: Clarendon)
Fesik S W and Zuiderweg E R P 1988 *J. Magn. Reson.* **78** 588
Hosur R V, Govil G and Miles H T 1988 *Magn. Reson. Chem.* **26** 927
Howe K M, Reakes C F L and Watson R J 1990 *Embo J.* **9** 161
Ibanez C E, Gracia A, Strobber-Grasser U and Lipsick J S 1988 *J. Virol.* **62** 4398
Jeener J 1971 *Ampere International Summer School*, Bosko Polje, Yugoslavia (unpublished)
Jeener J, Meier B H, Bachmann P and Ernst R R 1979 *J. Chem. Phys.* **71** 4546
Kaptein R, Boelens R, Scheek R and van Gunsteren W F 1988 *Biochemistry* **27** 5389
Kay L E, Clore G M, Bax A and Groenenborn A M 1990 *Science* **249** 411
Kay L E, Marion D and Bax A 1989 *J. Magn. Reson.* **84** 72
Kessler H, Gehrke M and Griesinger C 1988 *Angew. Chem., Int. Ed. Eng.* **27** 490

Klempenauer K H and Sippel A E 1987 *Embo J.* **6** 2719
Macura S and Ernst R R 1980 *Mol. Phys.* **41** 95
Madan A, Hosur R V and Padhy L C 1994a *Biochemistry* **33** 7120
Madan A, Radha P K, Hosur R V and Padhy L C 1994b *Eur. J. Biochem.* (in press)
Maudsley A A and Ernst R R 1977 *Chem. Phys. Lett.* **50** 368
McIntosh L P and Dahlquist F W 1990 *Q. Rev. Biophys.* **23** 1
Miller J H 1972 In *Experiments in molecular genetics* (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory)
Nishina Y, Nakagoshi H, Imamoto H, Gonda T J and Ishii S 1989 *Nucleic Acids Res.* **17** 107
Oehler T, Hannelore A, Beidenkepp H and Klempenauer K H 1990 *Nucleic Acid Res.* **18** 1703
Oschkinat H, Muller T and Dickmann T 1994 *Angew. Chem., Int. Ed. Eng.* **33** 277
Radha P K, Madan A, Padhy L C and Hosur R V 1994 *Curr. Sci.* **66** 287
Sakura H, Kanei-Ishii C, Nagase T, Nakagoshi H, Gonda T J and Ishii S 1989 *Proc. Natl. Acad. Sci., USA* **86** 5758
Shaka A J, Barker P B and Freeman R 1985 *J. Magn. Reson.* **64** 547
Tate S-I, Tate N U, Ravesa M W, Jaye M and Inagaki F 1992 *FEBS Lett.* **297** 39
van de Ven F and Hilbers C W 1988 *Eur. J. Biochem.* **178** 1
Wüthrich K 1986 *NMR of proteins and nucleic acids* (New York: Wiley)
Wüthrich K 1989 *Acc. Chem. Res.* **22** 36
Zuiderweg E R P, van Boren S R, Kurochkin A V, Neubig R R and Majumdar A 1993 *Perspectives in Drug Discovery and Design* **1** 391