Journal of Biomolecular NMR, 2 (1992) 323–334
ESCOM



J-Bio NMR 064

# Support of $^{1}H$ NMR assignments in proteins by biosynthetically directed fractional $^{13}C$-labeling

Thomas Szyperski, Dario Neri, Barbara Leiting, Gottfried Otting and Kurt Wüthrich

*Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland*

Received 6 February 1992
Accepted 25 March 1992

*Keywords:* Protein structure; NMR assignment; Isotope labeling; Biosynthetically directed fractional $^{13}C$-labeling; Stereospecific NMR assignment

## SUMMARY

Biosynthetically directed fractional incorporation of $^{13}C$ into proteins results in nonrandom $^{13}C$-labeling patterns that can be investigated by analysis of the $^{13}C$-$^{13}C$ scalar coupling fine structures in heteronuclear $^{13}C$-$^{1}H$ or homonuclear $^{13}C$-$^{13}C$ correlation experiments. Previously this approach was used for obtaining stereospecific $^{1}H$ and $^{13}C$ assignments of the diastereotopic methyl groups of valine and leucine. In the present paper we investigate to what extent the labeling patterns are characteristic for other individual amino acids or groups of amino acids, and can thus be used to support the $^{1}H$ spin-system identifications. Studies of the hydrolysates of fractionally $^{13}C$-labeled proteins showed that the 59 aliphatic carbon positions in the 20 proteinogenic amino acids exhibit 16 different types of $^{13}C$-$^{13}C$ coupling fine structures. These provide support for the assignment of the resonances of all methyl groups in a protein, which are otherwise often poorly resolved in homonuclear $^{1}H$ NMR spectra. In particular, besides the individual methyl assignments in Val and Leu, unambiguous distinctions are obtained between the methyl groups of Ala and Thr, and between the γ- and δ-methyl groups of Ile. In addition to the methyl resonances, the $\gamma CH_2$ groups of Glu and Gln can be uniquely assigned because of the large coupling constant with the δ-carbon, and the identification of most of the other spin systems can be supported on the basis of coupling patterns that are common to small groups of amino acid residues.

## INTRODUCTION

Biosynthetically directed fractional $^{13}C$-labeling has recently been used for obtaining stereospecific $^{1}H$ and $^{13}C$ NMR assignments for the diastereotopic methyl groups of valine and leucine in

*Abbreviations:* NOE, nuclear Overhauser effect; fractional $^{13}C$ labeling, biosynthetically directed fractional $^{13}C$-labeling; TOCSY, total correlation spectroscopy; ROESY, rotating frame Overhauser enhancement spectroscopy; [$^{13}C$,$^{1}H$]-COSY, two-dimensional $^{13}C$-$^{1}H$ correlation spectroscopy; isotopomer, isotope isomer; P22 c2 repressor, c2 repressor of the salmonella phage P22 consisting of a polypeptide chain with 216 residues; P22 c2(1–76), N-terminal domain of the P22 c2 repressor with residues 1–76.

proteins and peptides (Senn et al., 1989; Neri et al., 1989,1990). Since the isopropyl moieties of valine and leucine are usually involved in a large number of NOE distance constraints, and because outstandingly large pseudoatom corrections must be applied in the absence of stereospecific assignments (Wüthrich, 1983,1986), the resulting individual methyl assignments are an important factor for obtaining high-quality protein structures (Güntert et al., 1989; Driscoll et al., 1989). In the present paper we exploit the fact that the NMR spectra recorded for obtaining these stereospecific assignments contain additional information on nonrandom labeling in the other amino acid residues. This information, which is obtained without additional sample preparations or NMR experiments, can be used to support the spin-system identification in protein $^1$H NMR spectra.

## MATERIALS AND METHODS

We have so far obtained biosynthetically directed fractional $^{13}$C-labeling of proteins or peptides by growing microorganisms on minimal media containing a mixture of roughly 10% [$^{13}C_6$]glucose and 90% glucose at natural isotope abundance as the sole carbon source (Neri et al., 1989; Senn et al., 1989). (In principle, other $^{13}$C-labeled source molecules could be used in the place of glucose.) Since contiguous fragments of two or more carbon atoms originating from one source molecule of glucose are embedded into the carbon skeleton of the amino acids during biosynthesis (e.g., Umbarger, 1978; Neidhardt, 1987; Szyperski et al., 1992) each amino acid is represented by a pool of nonrandomly labeled isotope isomers (isotopomers). Therefore, the $^{13}$C-$^{13}$C scalar coupling fine structure observed for a given $^{13}$C resonance depends not only on the magnitude of the scalar coupling constants to the adjacent carbons, but also on the relative abundance of the different isotopomers. In proton-detected [$^{13}$C-$^1$H]-COSY experiments (Bodenhausen and Ruben,

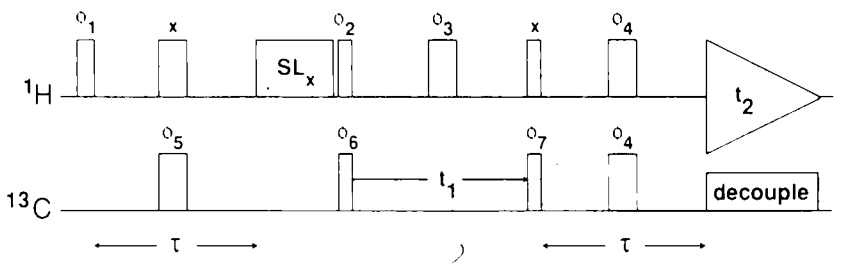


Fig. 1. Experimental scheme of [$^{13}$C,$^1$H]-COSY (Bodenhausen and Ruben, 1980) which provides decoupling of the $^1$H-$^{13}$C scalar couplings in both dimensions and ensures that the $^1$H-$^1$H couplings do not affect the $^{13}$C-$^{13}$C scalar coupling fine structure along $\omega_1$. 90-degree pulses are indicated by thin bars, and 180-degree pulses by broad bars. $SL_x$ denotes a spin-lock pulse of 2 ms duration with phase x, which purges the magnetization from $^{12}$C-bound protons (Otting and Wüthrich, 1988). The delay $\tau$ is set to $1/2^1J(^1H,^{13}C)$, which corresponds to $\tau$ = 3.6 ms in the presently used experiments. The following 64-step phase cycle was used to record the spectrum of P22 c2(1-76) in Fig. 3. $\varphi_1 = (x,x,-x,-x)_{16}$, $\varphi_2 = [(y)_4(-y)_4]_8$, $\varphi_3 = [(x)_8(-x)_8]_4$, $\varphi_4 = [(x)_{16}(-x)_{16}]_2$, $\varphi_5 = (x)_{64}$, $\varphi_6 = [(x)_4(-x)_4]_8$, $\varphi_7 = (x,-x)_{16}(-x,x)_{16}$, receiver = $(x,-x,-x,x)_8(-x,x,x,-x)_8$. The spectra of the hydrolysates with 7000 $t_1$ values were recorded with a two-step phase cycle: $\varphi_3 = (x,-x)$, $\varphi_2 = (y)_2$, $\varphi_1 = \varphi_4 = \varphi_5 = \varphi_6 = \varphi_7$ = receiver = $(x)_2$. The phases $\varphi_5$ and $\varphi_6$ were subjected to TPPI for quadrature detection in $\omega_1$ (Marion and Wüthrich, 1983). $^{13}$C broadband decoupling during acquisition was obtained using WALTZ-16 with suppression of cycling sidebands (Shaka et al., 1986).

1980) the carbon fine structure in the $^{13}C$ dimension can be related with the resonances of the directly bound protons, so that the information on the spectral properties of distinct carbon atom positions can be used to support the assignment of the $^{1}H$ spin system of the amino acid residue.

For the studies presented in this paper we used 3 different samples which were obtained using the *Escherichia coli* overexpression system W3110 *lac* $1^{Q}/_{p}$TP125 for high-level production of P22 c2 repressor, which consists of a polypeptide chain with 216 amino acid residues, including all 20 proteinogenic amino acids (De Anda et al., 1983). Details of the growth conditions, the protein purification and further steps of the NMR sample preparation have been described elsewhere (Neri et al., 1990; Szyperski et al., 1992; Wüthrich et al., 1992). The first sample was a 5 mM $D_2O$ solution of the N-terminal DNA-binding domain P22 c2(1–76) obtained by chymotrypsin cleavage of P22 c2 repressor produced in a fermenter with vigorous aeration during the growth of the culture. The level of $^{13}C$-labeling was 8%. The second sample was a solution in 0.1 M DCl of a mixture of amino acids obtained by hydrolysis of the P22 c2(1–76) preparation of sample one. The third sample was a 0.1 M DCl solution of a mixture of amino acids obtained by hydrolysis of the complete P22 c2 repressor isolated from a culture grown in contact with air but without forced aeration (for details see Szyperski et al., 1992). The level of $^{13}C$-labeling was 12%.

When using the fine-structure patterns arising from fractional $^{13}C$-labeling in support of the assignment of the $^{1}H$ NMR spectrum, the measurements must be made on the intact protein. For the spectral analysis one needs to take into account that the isotopomeric composition of the individual amino acid types, which is reflected in the $^{13}C$ NMR multiplets, may be dependent on the growth conditions. Thereby, with otherwise well-defined medium conditions, the oxygen supply is the major variable to be considered (Neidhardt, 1987). To investigate the influence of oxygen supply we made use of the improved spectral resolution achieved in the protein hydrolysates.

The experimental scheme of Fig. 1 was used to record the [$^{13}C$,$^{1}H$]-COSY spectra. The measurements were done at a $^{1}H$ frequency of 600 MHz using a Bruker AM600 spectrometer. For the amino acid mixtures the acquisition parameters were $t_{1max}$ = 336 ms, $t_{2max}$ = 102 ms, data size before zero-filling 7000 points in $t_1$ and 1024 points in $t_2$. The total recording time per spectrum was about 16 h. Before Fourier transformation the time domain data were multiplied in $t_1$ and $t_2$ with sine-bell windows (De Marco and Wüthrich, 1976) shifted by $\pi/2$. The digital resolution after zero-filling was 1.3 Hz along $\omega_1$ and 4.9 Hz along $\omega_2$. The assignment of the $^{1}H$ and $^{13}C$ resonances of the free amino acids in acidic aqueous solution was started using chemical shift data taken from Wüthrich (1976). In addition, a homonuclear 2D TOCSY spectrum (Braunschweiler and Ernst, 1983; Bax and Davies, 1985) with 150 ms mixing time and a homonuclear 2D ROESY spectrum (Bothner-By et al., 1984) with 400 ms mixing time were recorded to unambiguously assign the $^{1}H$ resonances. The $^{13}C$ resonance assignments were obtained using the [$^{13}C$,$^{1}H$]-COSY spectrum and a $^{1}H$ TOCSY-relayed [$^{13}C$,$^{1}H$]-COSY spectrum (Otting and Wüthrich, 1988) with 80 ms mixing time. For the intact protein P22 c2(1–76) the [$^{13}C$,$^{1}H$]-COSY spectrum was recorded with a data size of 2048 points in both $t_1$ and $t_2$, $t_{1max}$ = 90 ms, $t_{2max}$ = 188 ms, the total recording time was about 27 h. Before Fourier transformation the time domain data were multiplied with sine-bell windows (De Marco and Wüthrich, 1976) shifted by $\pi/5$ in $t_1$ and $\pi/9$ in $t_2$, respectively. The digital resolution after zero-filling was 2.8 Hz along $\omega_1$ and 1.3 Hz along $\omega_2$. A reference [$^{13}C$,$^{1}H$]-COSY spectrum of P22 c2(1–76) with natural $^{13}C$-abundance was recorded on a Bruker AM500 spectrometer. This spectrum had the same digital resolution and was processed with the same parameters as the 600 MHz spectrum of the fractionally $^{13}C$-labeled protein. For both experiments

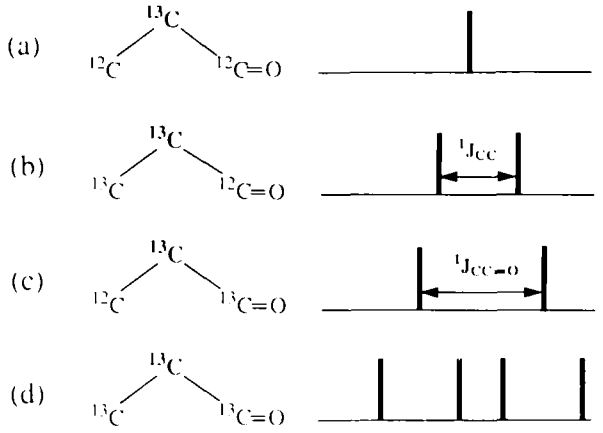


Fig. 2a–d. The 4 different types of labeling patterns (left) and the corresponding multiplets observed in $^{1}$H-decoupled $^{13}$C NMR spectra (right) encountered for aliphatic carbon positions in the amino acid residues of proteins with biosynthetically directed fractional $^{13}$C-labeling to the extent of ca. 10%. The multiplet components in the stick diagrams on the right are separated by the coupling constants $^{1}J(^{13}C,^{13}C)$, which are of the order of 35 Hz between two aliphatic carbons and 55 Hz between an aliphatic and a carbonyl or carboxyl carbon (Krivdin and Kalabin, 1989). In each fragment the bold letter identifies the observed carbon position.

the sample temperature was 28°C. The $^{1}$H NMR spectrum of P22 c2(1–76) had previously been completely assigned (Otting, 1987).

With the NMR experiments used, only proton-bound carbons are observed, and only the $^{1}J(^{13}C,^{13}C)$ coupling constants are sufficiently large to be resolved in the $^{13}$C dimension. The $^{13}$C fine structure is thus solely determined by the $^{13}$C-labeling pattern of the directly attached carbon atoms. From the known biosynthetic pathways (Neidhardt, 1987; Szyperski et al., 1992) one expects to observe exclusively singlet, doublet, triplet, and doublet of doublets fine-structure patterns for the aliphatic carbons in all 20 common amino acids (Fig. 2). The different values of $^{1}J(^{13}C,^{13}C)$ between chemically different carbon positions (caption to Fig. 2) may also be useful for obtaining resonance identifications. The spectral analysis is based on the fact that $2^n$ isotopomers exist for a molecule with $n$ carbon atoms. The scalar coupling fine structure of a given $^{13}$C resonance, which is observed in the $[^{13}C,^{1}H]$-COSY spectrum along $\omega_1$, is therefore a superposition of the fine structures of different isotopomers weighted by their relative abundance, which is in turn determined by the relative importance of the different possible pathways for the biosynthesis of the amino acid type considered (Szyperski et al., 1992; Wüthrich et al., 1992).

## RESULTS AND DISCUSSION

The $[^{13}C,^{1}H]$-COSY spectrum of the protein P22 c2(1–76) with biosynthetic fractional labeling to the extent of 8% (Fig. 3) is representative also for the spectra recorded with the protein hydrolysates, except that these contained narrower and hence better separated peaks. The spectral regions 1–9 are identified, which contain different classes of cross peaks as described in Fig. 4.

The 20 proteinogenic amino acids contain 59 proton-bearing aliphatic carbons. In the mixture of amino acids obtained after hydrolysis of P22 c2(1–76) or P22 c2 repressor in 6 N HCl at 160°C

for 2 h, cysteine and tryptophan were lost due to oxidation, and glutamine and asparagine were deamidated so that their NMR lines coincided with those of glutamate and aspartate. Furthermore, it turned out that the $^{13}C$-$^{1}H$ cross peaks of the α-carbons of lysine and arginine were overlapped in the [$^{13}C$,$^{1}H$]-COSY spectrum of the free amino acids. Therefore, 48 of the total of 59 aliphatic $^{13}C$-$^{1}H$ cross peaks expected in a mixture of the 20 common amino acids could be analysed in the hydrolysate. The $^{13}C$-$^{13}C$ fine structures of these 48 cross peaks are collected in Fig. 4, which is the basis for using biosynthetically directed fractional $^{13}C$-labeling to support $^{1}H$ spin-system identifications in proteins. Figure 4 also includes spectral data collected with the intact P22 c2(1–76) (Fig. 3), which relate to the $^{13}C$-$^{13}C$ fine structures of those 11 carbon positions that could not

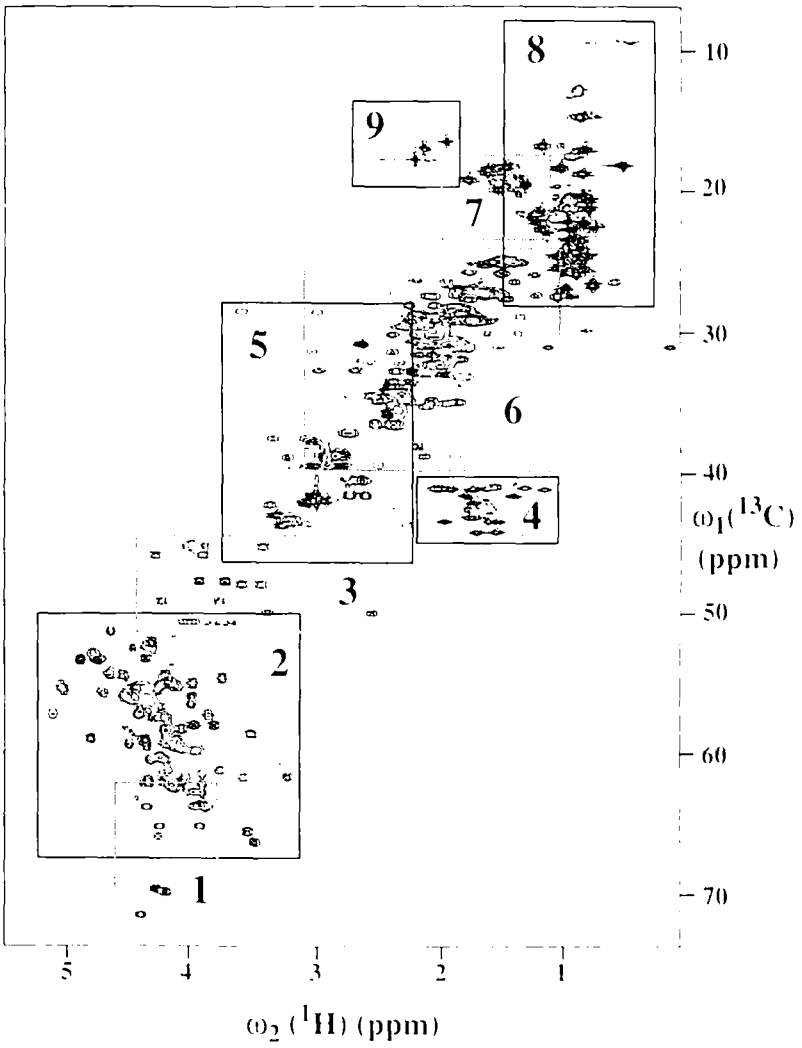


Fig. 3. [$^{13}C$,$^{1}H$]-COSY spectrum at 600 MHz of P22 c2(1–76) with biosynthetic fractional $^{13}C$-labeling to the extent of 8%. The protein was obtained from *E. coli* grown in a fermenter with vigorous aeration. The spectrum was recorded at 28 °C in $^{2}H_2O$ at p$^{2}H$ = 4.8. The rectangles define 9 spectral regions containing groups of cross peaks as listed in Fig. 4.

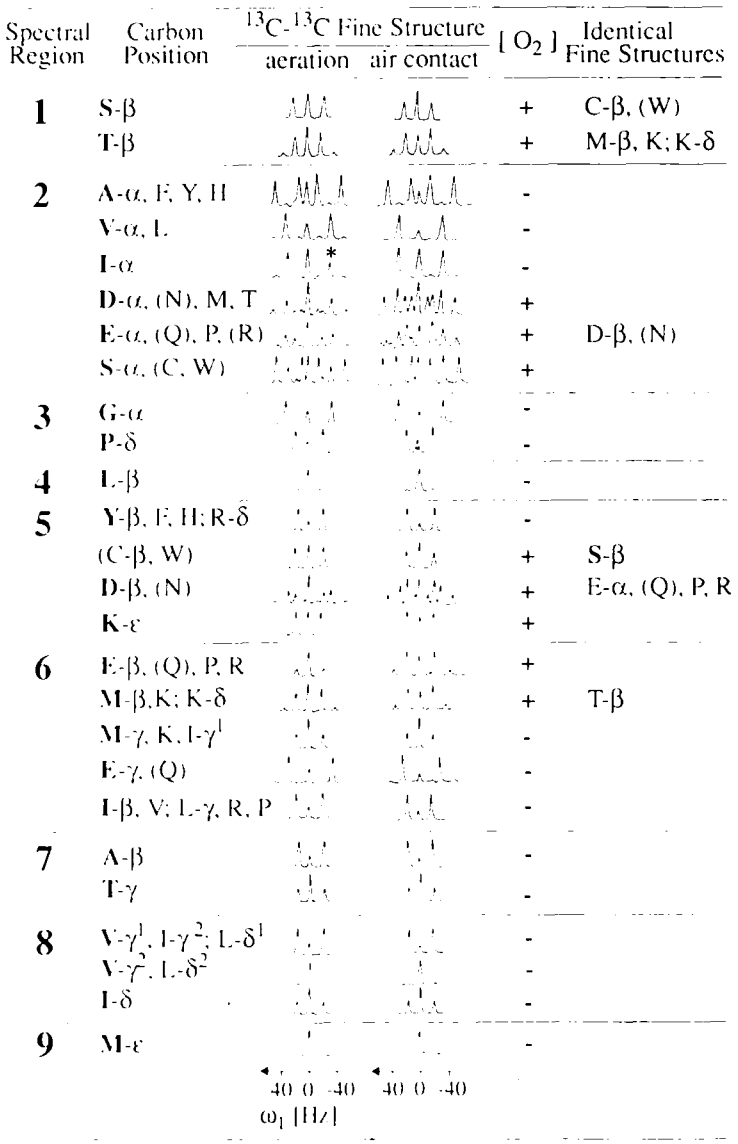


Fig. 4. Survey of the $^{13}C$-$^{13}C$ fine structures observed in the hydrolysates of P22 c2(1–76) from the aerated fermenter preparation and p22 c2 repressor from the culture grown with air contact in a flask. The spectral regions indicated in the left-most column by the numbers 1–9 are defined in Fig. 3. The second column lists the carbon positions represented by the spectra in the third and fourth columns. The bold letter identifies the amino acid for which the spectra are shown; the others are amino acids with identical multiplets in the *same* spectral region, and carbon positions in parentheses are those for which the fine structures could only be examined in the intact P22 c2(1–76) protein (see text). (In the row of the β-resonances of cysteine and tryptophan in the spectral region 5 the spectra of the β-resonance of serine from region 1 are shown. The asterisk indicates that the low intensity of the right fine-structure component of the α-carbon of isoleucine is an artifact.) In the fifth column the entries which are insensitive to changes in oxygen supply are marked '-' (this includes 6 carbon positions that are not strictly invariant; see text) while those which show differences between columns 3 and 4 are marked '+'. The right-most column lists carbon positions in *different* spectral regions that show the same multiplet fine structures.

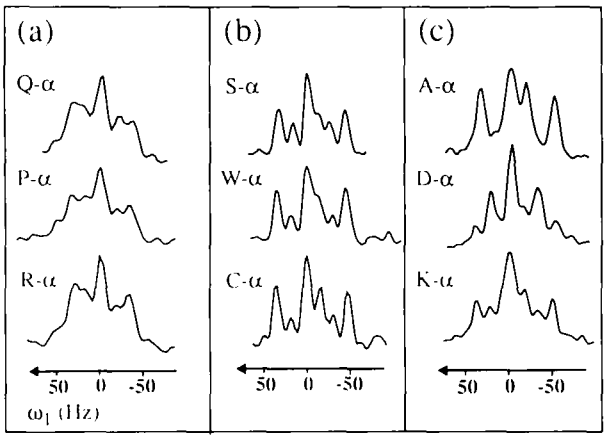


Fig. 5. Cross sections along $\omega_1$ from the spectral region 2 (Fig. 3) showing $^1J(^{13}C,^{13}C)$ fine structures observed in the $[^{13}C,^1H]$-COSY spectrum of P22 c2(1–76). (a) α-carbon resonances of Gln, Pro and Arg. (b) α-carbon resonances of Ser, Trp and Cys. (c) α-carbon resonances of Ala, Asp and Lys. The fine structure observed for Lys corresponds to an equally weighted superposition of the fine structures of Ala and Asp.

be investigated in the hydrolysates: (i) the fine structures of glutamine and asparagine, which turned out to be identical to those of glutamate and aspartate; (ii) the α-carbon fine structure of arginine, which is identical to that of the α-carbons of glutamine and proline (Fig. 5a); (iii) the α-carbon fine structures of cysteine and tryptophan, which were indistinguishable from those observed in serine (Fig. 5b). In addition to the data in Fig. 4 there is the α-carbon fine structure of lysine (Fig. 5c), which, in agreement with the expected biosynthetic pathways (Neidhardt, 1987), corresponds to a 1:1 superposition of the fine structures observed for the α-carbons of alanine and aspartate. Note also that P22 c2(1–76) does not contain phenylalanine and histidine and that these amino acids were therefore not investigated in the fermenter preparation.

Figures 5 and 6 show expanded fine structures from the $[^{13}C,^1H]$-COSY spectrum of P22 c2(1–76) (Fig. 3), which are quite representative for the data obtained with proteins in the approximate molecular-weight range 5000–12 000. The salient features of the fine structures seen in the hydrolysates (Fig. 4) can readily be identified either in cross sections (Fig. 5) or in contour plots (Fig. 6) of the spectrum recorded with the protein. In practice, a spectrum of the unlabeled protein (Fig. 6a) is also helpful as a reference for the spectral analysis.

Overall, Fig. 4 shows that for 33 of the total of 59 aliphatic carbon positions the $^{13}C$-$^{13}C$ fine structures do not vary with the amount of oxygen supplied during cell growth under aerobic conditions. This includes 6 carbon positions (Ile-α, Ile-$\gamma^1$, Ile-δ, Thr-γ, Met-γ and Lys-γ) which are not strictly invariant to the oxygen supply as demonstrated by samples obtained from anaerobically grown cell cultures (Szyperski et al., 1992; Wüthrich et al., 1992), but can for all practical purposes be considered invariant in the context of the present investigations. The $^{13}C$-$^{13}C$ multiplet fine structures of these 33 invariant carbon positions are obviously a reliable source of information in support of spin-system identification. The fine structures of the remaining 26 carbon positions are sensitive to variations in oxygen supply. Nonetheless, as we shall see below, many of these $^{13}C$-$^1H$ cross peaks can also provide useful information. Practical applications of the data in Fig. 4 should

also take account of knowledge on NMR chemical shifts (Wüthrich, 1976,1986) and on biosynthetic pathways (Umbarger, 1978; Szyperski et al., 1992; Wüthrich et al., 1992), in addition to the spin-spin coupling fine structures. Thus, limitations imposed by the fact that the 59 aliphatic carbon positions in the common amino acids give rise to only 16 different $^{13}C$-$^{13}C$ fine structures (Figs. 4 and 5) can often be overcome because different carbon positions with identical fine structures occur in different spectral regions (Fig. 3). Alternatively, for situations where certain cross

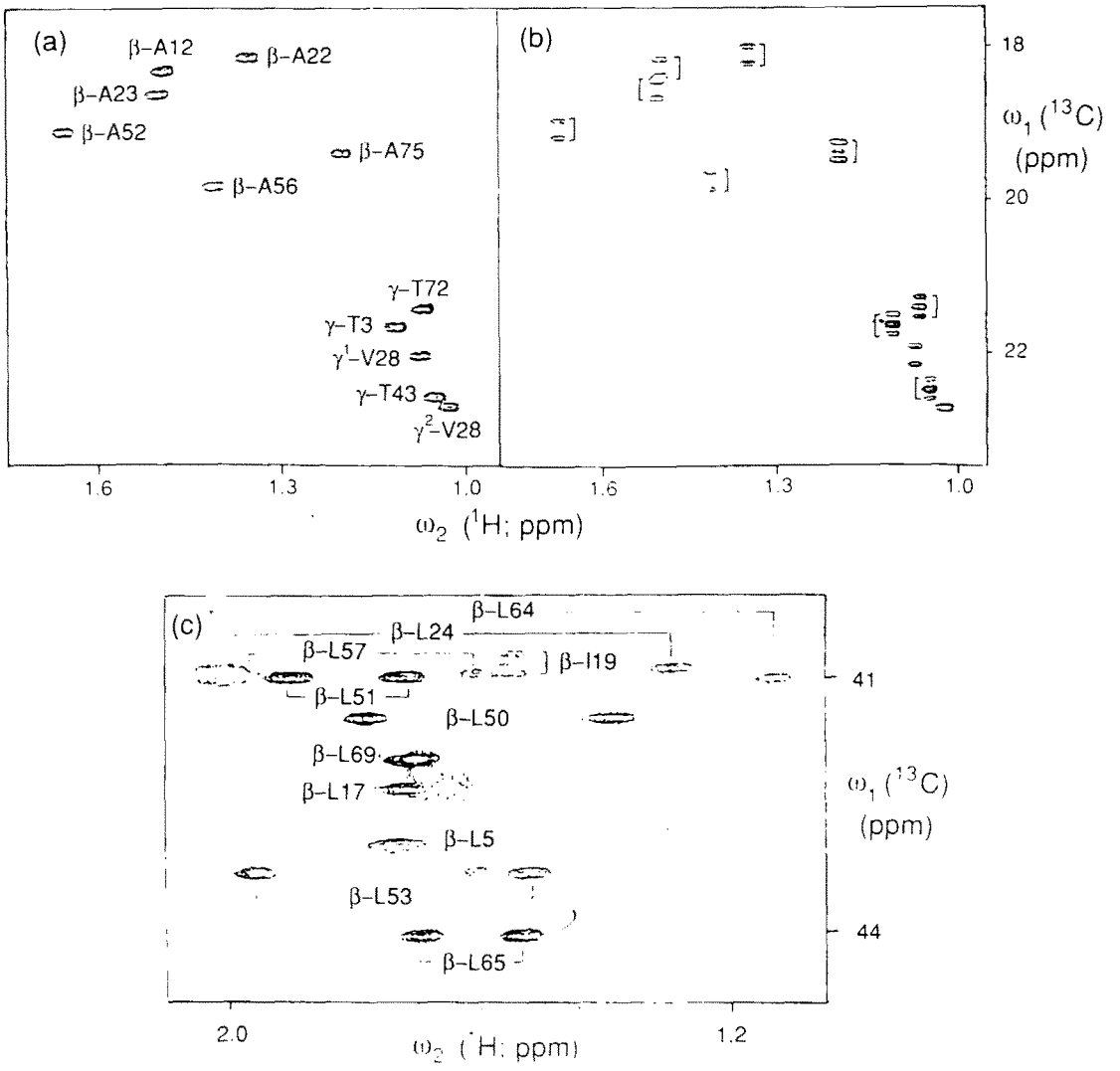


Fig. 6. Illustrative examples for the observation of $^{1}J(^{13}C,^{13}C)$ fine structures listed in Fig. 4 in expanded contour plots of the [$^{13}C$,$^{1}H$]-COSY spectrum of the protein P22 c2(1-76) of Fig. 3, which was recorded with a digital resolution of 2.8 Hz in $\omega_1$ (see *Materials and Methods* for other parameters used). (a) and (b): Methyl cross peaks of Ala and Thr in the spectral region 7, without and with 8% $^{13}C$-labeling, respectively. In (b) the doublet splittings of the cross peaks of Ala and Thr are indicated with brackets. (c) β-$CH_2$ cross peaks of Leu in the spectral region 4 of the labeled protein. The individual cross peaks are identified with the carbon position, the one-letter amino acid symbol and the sequence position.

peaks are poorly resolved in the [$^{13}$C,$^{1}$H]-COSY spectra, detailed biosynthetic studies (Szyperski et al., 1992) showed that information on cross-peak fine structures from different spectral regions can often be combined in order to define or confirm the fine-structure type for these carbon positions (column 6 in Fig. 4). In the following we take account of these special features when discussing the information content of the individual spectral regions outlined in Fig. 3.

The *spectral region 1* contains the β-carbon resonances of serine and threonine (Figs. 3 and 4). Both are dominated by a superposition of a doublet split by approximately 35 Hz arising from the scalar coupling constant $^{1}J_{C\alpha C\beta}$ with a singlet of similar line intensity. Depending on the oxygen supply the $^{13}$C-$^{13}$C fine structures may be somewhat different, but there are two other criteria that usually enable an unambiguous distinction between the two residues. One is that the β-methylene group of serine has two cross peaks at the same $^{13}$C chemical shift with identical $^{13}$C-$^{13}$C fine structures. Furthermore, from the biosynthetic pathways it is known (Szyperski et al., 1992) that the $^{13}$C-$^{13}$C fine structure of Ser-β must be the same as that of Cys-β and Trp-β (Fig. 4).

*Region 2* contains all α-carbon resonances except those of glycine. According to their $^{13}$C-$^{13}$C fine structures they can be grouped into 7 classes. Alanine, phenylalanine, tyrosine and histidine are insensitive to oxygen supply and exhibit a doublet of doublets arising from the scalar coupling constants $^{1}J_{C\alpha C'}$ of ~ 50 Hz and $^{1}J_{C\alpha C\beta}$ of ~ 35 Hz; this multiplet is superimposed on a singlet originating from glucose with natural isotope abundance (Fig. 4). Here, as well as in other multiplets with fine-structure components near the central position, the influence of $^{13}$C isotope effects on the $^{13}$C chemical shifts (Hansen, 1988) must be properly allowed for: the doublet of doublets is shifted by ~ 2 Hz to higher field relative to the singlet line, so that with the line broadening in protein NMR spectra the singlet merges with one of the components of the doublet of doublets (Fig. 5c). Valine and leucine are also insensitive to oxygen supply and are represented by a doublet split by $^{1}J_{C\alpha C'}$ of ~ 50 Hz which is superimposed on a singlet of somewhat smaller line intensity arising from unlabeled glucose. The α-carbon fine structure of isoleucine is unique in that the relative line intensities of the singlet and the doublet are nearly identical. The $^{13}$C-$^{13}$C fine structures of the α-carbons of glutamate, glutamine, proline and arginine are identical to those of the β-carbon resonances of aspartate and asparagine. Similar to the α-carbon resonances of aspartate, asparagine, methionine and threonine they show 9 lines arising from a doublet of doublets as described above, a doublet split by $^{1}J_{C\alpha C'}$ of ~ 50 Hz, a doublet split by $^{1}J_{C\alpha C\beta}$ of ~ 35 Hz and a singlet. Serine, cysteine and tryptophan have the same fine structure, except that the intensity of the doublet split by $^{1}J_{C\alpha C\beta}$ is negligibly small. The $^{13}$C-$^{13}$C fine structure of the α-carbon of lysine (not represented in Fig. 4) is again unique. Figure 5c shows that it is identical to a 1:1 superposition of the α-carbon resonances of alanine and aspartate. Due to the increased line widths in proteins, several lines in the more complex $^{13}$C-$^{13}$C fine structures may merge (see, for example, Fig. 5b).

*Region 3* contains the resonances of the α-carbons of glycine and the δ-carbons of proline, which are both insensitive to oxygen supply. In spite of the fact that both are represented by a doublet superimposed on a singlet, they can be distinguished by the larger doublet splitting of ~ 50 Hz for glycine as compared to the splitting of ~ 35 Hz for proline $C^{\delta}$. In the area where regions 2 and 3 overlap (Fig. 3), proline $C^{\delta}$ and glycine $C^{\alpha}$ can be unambiguously distinguished from the other α-carbon resonances because they are manifested by two cross peaks. If the two methylene protons have identical chemical shifts, an unambiguous distinction between the α-carbon resonances of glycine, valine and leucine may be difficult (Fig. 4), but the proline $C^{\delta}$ resonance will still be unique because of the small doublet splitting (Fig. 4).

*Region 4* contains exclusively the β-carbon resonances of leucine (Fig. 4). These are singlets which are insensitive to oxygen supply. Since the only other spectral regions which contain singlets are 8 and 9, there is no danger of mistakes (Fig. 3), and non-leucine residues which are in region 4 due to unusual chemical shifts can readily be identified. For example, in Fig. 6c the resonance of the β-carbon of $Ile^{19}$ can be distinguished from the resonances of leucine because of its doublet fine structure.

*Region 5* contains the resonances of the β-carbons of all residues with AMX proton spin systems except serine, the δ-resonances of arginine and the ε-resonances of lysine (Fig. 4). As for the α-carbon resonances, the β-carbons of the AMX spin systems fall into 3 distinct groups. The β-resonances of phenylalanine, tyrosine, histidine and the δ-resonances of arginine, each of which is represented by a doublet split by approximately 35 Hz superimposed on a weaker singlet, are insensitive to oxygen supply. Depending on the oxygen supply, the $^{13}C$-$^{13}C$ fine structures observed for the β-resonances of cysteine and tryptophan, which must be identical to that observed for the β-carbon of serine (Fig. 4), may be identical to the fine structure of the ε-carbon of lysine. The resonances of the β-carbons of aspartate and asparagine show a complex multiplet fine structure, which is dominated by the doublet split by the large $^1J_{C\beta C\gamma}$ coupling constant of ~ 50 Hz, similar to that for the α-carbon resonances of glutamate, glutamine, arginine and proline in the spectral region 2 (Fig. 3). In the overlap area with region 3, care must be taken because the δ-resonances of proline show the same fine structure as the β-resonances of phenylalanine, tyrosine and histidine, and the δ-resonances of arginine.

*Region 6* contains the remaining 17 resonances of nonmethyl carbons in the long side chains. They can be grouped into 5 classes, of which 3 are insensitive to oxygen supply (Fig. 4). The γ-resonances of glutamine and glutamate are represented by a doublet split by approximately 50 Hz superimposed on a singlet of small intensity, and can therefore readily be distinguished from all other resonances in this region. The resonances of the β-carbons of valine and isoleucine and the γ-carbons of leucine, arginine and proline exhibit the same multiplet components but with a doublet splitting of only approximately 35 Hz. The γ-methylene resonances of methionine, lysine and isoleucine show a characteristic superposition of a strong singlet and a doublet split by approximately 35 Hz. The remaining $^{13}C$-$^{13}C$ fine structures in this region are sensitive to oxygen supply and may in some situations become virtually identical to those of the γ-carbons of methionine, lysine and isoleucine. In the overlap area with region 5, the β-resonances of cysteine and tryptophan show again $^{13}C$-$^{13}C$ fine structures very similar to those of the β-carbons of methionine and lysine and the δ-carbon of lysine, but here one may arrive at an unambiguous distinction from the fact that the latter fine structures must be the same as that of the β-carbon resonance of threonine in region 1 (Fig. 4).

*Region 7* contains the methyl groups of alanine and threonine, which can be unambiguously distinguished due to their pronounced differences in $^{13}C$-$^{13}C$ fine structures. This is nicely illustrated by the spectra of intact P22 c2(1–76) in Figs. 6a and b.

*Region 8* contains the methyl resonances of valine, leucine and isoleucine. The stereospecific assignment of the diastereotopic methyl groups of valine and leucine with biosynthetically directed fractional labeling (Neri et al., 1989; Senn et al., 1989) is robust with respect to variations in oxygen supply. Since *all* methyl resonances are insensitive to oxygen supply, it is further possible to unambiguously distinguish between γ- and δ-methyl groups of isoleucine, and also between δ-methyl groups of isoleucine and methyl groups of valine and leucine. Special care may have to be

exercised in the overlap area with region 7, where the γ-resonances of threonine exhibit virtually identical $^{13}C$-$^{13}C$ fine structures as the δ-methyl groups of isoleucine, and in the overlap area with region 6, where similar $^{13}C$-$^{13}C$ fine structures to that for Ile-δ prevail for the γ-carbon resonances of methionine and lysine and the γ-methylene group of isoleucine.

*Region 9* contains the resonances of the ε-methyl group in methionines, which is the only aliphatic carbon position in the 20 proteinogenic amino acids that has no neighbouring carbon atom. Therefore, the resonances are represented by a clean singlet (Fig. 4). In Fig. 3 this region shows the singlet methyl resonances of the 3 methionyl residues in P22 c2(1–76), which are broadened along $\omega_1$ by tails arising from truncation effects.

## CONCLUSIONS

The spectral analyses in the preceding section show that fractional $^{13}C$-labeling of recombinant proteins can greatly facilitate the assignment of *all* methyl resonances in a protein, which are otherwise often poorly separated in homonuclear $^1H$ NMR spectra. Due to the high inherent sensitivity of [$^{13}C$,$^1H$]-COSY for observation of methyl groups, only a small amount of protein enriched to about 10% in $^{13}C$ is needed for these assignments, corresponding to 0.5 ml of an approximately 0.1 mM protein solution. Similarly, because of the large values of the $^{13}C$-$^{13}C$ coupling constant with the carbonyl or carboxyl carbon, the $C^\gamma$ resonances of glutamine and glutamic acid can usually be unambiguously identified. All these assignments are robust in the sense that they are not affected by variation of the oxygen supply. The additional experimental observations listed in Fig. 4 do not lend themselves for an independent assignment strategy, but they can present the evidence needed to resolve ambiguities in $^1H$ NMR assignments. For example, among the α- and β-carbon resonances of the 8 $C^\alpha H$-$C^\beta H_2$ fragments giving rise to AMX proton systems in amino acids (Wüthrich, 1986), the 3 groups (i) phenylalanine, tyrosine, histidine, (ii) serine, cysteine, tryptophan, (iii) asparagine, aspartate, can be distinguished. Similarly, among the residues with 5 proton spin systems the α-, β- and γ-carbon resonances of glutamine and glutamate are clearly different from those of methionine (Fig. 4). Quite generally, the data of Fig. 4 can support the $^1H$ spin-system identification for any of the 20 common amino acid residues in situations where spectral overlap prevents a complete analysis of the $^1H$ NMR spectra.

For recombinant polypeptides the biosynthetic fractional $^{13}C$-labeling technique is the most direct and reliable way for obtaining individual assignments of the diastereotopic methyl groups of Val and Leu (Neri et al., 1989,1990; Senn et al., 1989). In globular proteins it is often the only method that can provide stereospecific assignments for *all* Val and Leu residues, and for nonglobular polypeptides there is quite generally no other approach that can provide stereospecific assignments for *any* of the Val and Leu residues. As a consequence, biosynthetic fractional $^{13}C$-labeling will foreseeably become a standard experiment in many structural studies of polypeptides and proteins. Whenever a fractionally labeled protein sample is prepared and used to record [$^{13}C$,$^1H$]-COSY spectra, Fig. 4 of the present paper will be a useful guide for obtaining additional $^1H$ and $^{13}C$ NMR assignments from the same data that are needed for obtaining stereospecific assignments of the methyl groups of Val and Leu. A special advantage of the method lies also in the fact that only small quantities of $^{13}C$-labeled precursor compounds are needed, so that even experiments with relatively low-yield expression systems are usually affordable.

Strictly speaking, the data in Fig. 4 are valid only with proteins obtained with the high-level ex-

pression system *E. coli* W3110 *lac* $I^Q$/pTP125, which is under the control of the *lac* promoter, and other organisms which grow on a minimal medium and use the same biosynthetic pathways for the amino acid synthesis. When in doubt, reference data corresponding to those in Fig. 4 can in principle be collected for any new system. However, there are indications that Fig. 4 may be a quite reliable guideline for a rather wide spectrum of organisms, expression systems and growth conditions. Examples are that stereospecific assignments of the methyl groups of valine and leucine with the presently described approach were obtained for cyclosporin A expressed in *Tolypocladium inflatum* (Senn et al., 1989), and that no significant increase of $^{13}C$-scrambling was observed with the presently used expression system under anaerobic growth conditions, where the cell growth rate and the yield of P22 c2 repressor were reduced 5-fold when compared to aerobic growth (Szyperski et al., 1992).

## ACKNOWLEDGEMENTS

We acknowledge financial support from the Schweizerischer Nationalfonds (Project 31.25174.88). T. Szyperski and B. Leiting are indebted to the Stipendien-Fonds im Verband der Chemischen Industrie and the Deutsche Forschungsgemeinschaft, respectively, for a scholarship. We thank Mr. R. Marani for the careful processing of the manuscript.

## REFERENCES

Bax, A. and Davies, D.G. (1985) *J. Magn. Reson.*, **65**, 355-360.
Bodenhausen, G. and Ruben, D. (1980) *Chem. Phys. Lett.*, **69**, 185-188.
Bothner-By, A., Stephens, R.L., Lee, J., Warren, C.D. and Jeanloz, R.W. (1984) *J. Am. Chem. Soc.*, **106**, 811-813.
Braunschweiler, L. and Ernst, R.R. (1983) *J. Magn. Reson.*, **53**, 521-528.
De Anda, J., Poteete, A.R. and Sauer, R.T. (1983) *J. Biol. Chem.*, **258**, 10536-10542.
DeMarco, A. and Wüthrich, K. (1976) *J. Magn. Reson.*, **24**, 201-204.
Driscoll, P.C., Gronenborn, A.M. and Clore, G.M. (1989) *FEBS Lett.*, **243**, 223-233.
Güntert, P., Braun, W., Billeter, M. and Wüthrich, K. (1989) *J. Am. Chem. Soc.*, **111**, 3997-4004.
Hansen, P.E. (1988) *Prog. NMR Spectrosc.*, **20**, 207-255.
Krivdin, L.B. and Kalabin, G.A. (1989) *Prog. NMR Spectrosc.*, **21**, 293-448.
Marion, D. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967-974.
Neidhardt, F.C. (1987) *Escherichia coli and Salmonella typhimurium*, American Society for Microbiology, Washington, DC.
Neri, D., Szyperski, T., Otting, G., Senn, H. and Wüthrich, K. (1989) *Biochemistry*, **28**, 7510-7516.
Neri, D., Otting, G. and Wüthrich, K. (1990) *Tetrahedron*, **46**, 3287-3296.
Otting, G. (1987) *Strukturermittlung an kleinen Proteinen mit NMR: Neue Methoden und Anwendungen am Beispiel der Cardiotoxine von Naja Mossambica Mossambica und des P22λ2 Repressors*. Ph. D. thesis No. 8314, ETH Zürich.
Otting, G. and Wüthrich, K. (1988) *J. Magn. Reson.*, **76**, 569-574.
Senn, H., Werner, B., Messerle, B.A., Weber, C., Traber, R. and Wüthrich, K. (1989) *FEBS Lett.*, **249**, 113-118.
Shaka, A.J., Keeler, J. and Freeman, R. (1983) *J. Magn. Reson.*, **53**, 313-340.
Szyperski, T., Leiting, B., Otting, G. and Wüthrich, K. (1992) *Biochemistry*, submitted.
Umbarger, H.E. (1978) *Annu. Rev. Biochem.*, **47**, 1127-1162.
Wüthrich, K. (1976) *NMR in Biological Research: Peptides and Proteins*, North-Holland, Amsterdam.
Wüthrich, K. (1983) *Biopolymers*, **22**, 131-138.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wüthrich, K., Szyperski, T., Leiting, B. and Otting, G. (1992) *Proc. 1st Biennial International Conference on Amino Acid Research*, Kyoto 1991, in press.