Journal of Biomolecular NMR, 5 (1995) 202–206
ESCOM

J-Bio NMR 264

High-resolution 3D HNCOCA experiment applied to a 28 kDa paramagnetic protein

Bernhard Brutscher, Florence Cordier, Jean-Pierre Simorre, Michael Caffrey and Dominique Marion*

Institut de Biologie Structurale-Jean Pierre Ebel, C.E.A.-C.N.R.S., 41 Avenue des Martyrs, F-38027 Grenoble Cedex, France

Received 9 December 1994
Accepted 23 January 1995

Keywords: Cytochrome *c'*; Resonance assignment; HNCOCA; Multiple quantum coherence

Summary

A new triple-resonance 3D HNCOCA pulse scheme is presented, designed to identify the backbone nuclei (H^N, N, CO, C^α) of doubly labelled proteins. The two carbon frequencies are labelled along the same indirect dimension and the corresponding dwell times can be independently scaled in order to account for the relaxation properties and chemical shift ranges of the CO and C^α. If one takes advantage of the symmetry properties of the spectra in the course of the peak picking, this 3D scheme has the same sensitivity as the 4D experiment, but with an improved resolution. The sequence is illustrated on a 0.5 mM sample of *Rhodobacter capsulatus* cytochrome *c'*, a homodimeric paramagnetic protein of 2 × 14 kDa. A resonance assignment strategy, based on a low-concentration $^{13}C/^{15}N$-labelled sample and a more concentrated ^{15}N-labelled sample, is proposed for proteins where the expression system shows a limited efficiency.

Multidimensional triple-resonance correlation (TRC) experiments, first proposed by the laboratories of Bax and Wagner (Ikura et al., 1990; Kay et al., 1990; Montelione and Wagner, 1990), have proven their ability for the backbone assignment of $^{15}N/^{13}C$-labelled proteins up to molecular weights of 30 kDa (Grzesiek and Bax, 1992). Backbone nuclei are sequentially correlated via one-bond (and occasionally two-bond) J couplings. Each nucleus is usually labelled along a separate dimension. In practice, however, the number of residues that can be correlated in a single TRC experiment is limited for several reasons: (i) the short T_2 relaxation times of larger molecules compared to the small J_{NC} coupling constants prompt one to minimize the number of transfer delays and labelling periods; (ii) each additional dimension causes a $\sqrt{2}$ loss in signal-to-noise (S/N) ratio; (iii) in a reasonable experimental time, only coarse digital resolution in the indirect dimensions can be obtained; and (iv) relatively long processing times, high data storage requirements and multidimensional space (>2D) make data analysis less convenient. Consequently, for larger proteins, a large collection of 2D and 3D data sets is generally required to resolve overlaps of the individual nuclei.

Recently, we introduced the notation of a 'pseudo-residue' (formed by four nuclei: $H^N(i)$, N(i), CO(i–1) and $C^\alpha(i-1)$) and presented a computer program for the automatic assignment of backbone resonances of $^{15}N/^{13}C$-labelled proteins. These building blocks (pseudoresidues) are automatically assigned to positions in the primary sequence by an appropriate optimization procedure (Morelle et al., 1995). The important points of the proposed strategy can be summarized as follows: (i) *unambiguous identification of pseudoresidues*: this unusual combination of nuclei has been chosen because they can be correlated in one very sensitive pulse scheme. The relatively large coupling constants of J_{NCO} (≈ 15 Hz) and J_{COC^α} (≈ 55 Hz) and the slow T_2 relaxation of N and CO make the 'out-and-back' version of HNCOCA the *most sensitive* TRC experiment. In comparison, the 'net transfer' (H)CACO-NH experiment, which might also be used for this purpose, is less sensitive due to the short T_2 relaxation of C^α and the $J_{C^\alpha C^\beta}$ coupling during the $C^\alpha \rightarrow CO$ transfer step. Therefore, the HNCOCA experiment yields good results, even for very large proteins (i.e., proteins with a molecular mass of 30 kDa or greater). High sensitivity and good spectral resolution of the four involved nuclei reduce the

*To whom correspondence should be addressed.

0925-2738/$ 6.00 + 1.00 © 1995 ESCOM Science Publishers B.V.

occurrence of errors in the identification step. In addition, the identified frequencies in this sensitive experiment may be used for improving the peak-picking procedure of any other (less sensitive) TRC experiment. (ii) *High flexibility in the sequential assignment step*: depending on the protein samples available, many types of information can be added to connect pseudoresidues or to limit the possible amino acid types for each of them.

Here we present a 3D version of the HNCOCA correlation experiment with the detection of all four pseudoresidue nuclei (Fig. 1). High sensitivity and high resolution in all dimensions can be obtained, even for very large proteins. The interesting features of this pulse sequence will be discussed. Finally, an application is shown for a 0.5 mM sample (heme concentration) of *Rhodobacter capsulatus* ferricytochrome c', a 28 kDa paramagnetic homodimeric protein.

The coherence transfer pathway is essentially the same as previously described (Grzesiek and Bax, 1992), except for the N → CO transfer which uses an HMQC instead of an INEPT. The first feature of interest in this pulse sequence is the reduction of the dimensionality of the experiment from 4D to 3D, using the method described previously for different applications (Szyperski et al., 1993a, b; Brutscher et al., 1994; Simorre et al., 1994). In the present case, the two ^{13}C frequencies (C^{α}, CO) are labelled in the same frequency dimension, resulting in two correlation peaks per pseudoresidue. The CO and C^{α} chemical shifts are then obtained from the separation (C^{α}) and the midpoint (CO) of the two peaks. As sensitivity is a crucial issue for larger molecules, this will be analyzed in detail. On the one hand, the splitting of a correlation peak into two peaks in the ^{13}C plane leads to a twofold decrease in sensitivity with respect to a 3D version with three detected nuclei. On the other hand, a fourth dimension would also reduce the sensitivity by a factor of $\sqrt{2}$, due to the magnetization splitting into two orthogonal components (for hypercomplex FT). Thus, the apparent experimental S/N ratio of the 4D version is $\sqrt{2}$ higher than that of a 3D version. However, if the central frequency of the two peaks is available (for instance, from an HNCO experiment which is highly sensitive), a better performance of the peak-picking routine is achieved by using the existing symmetry of the two correlation peaks. This sensitivity gain is illustrated in Fig. 2 for a ^{13}C strip extracted at a given $\{H^N,N\}$ frequency pair. In the lower part of Fig. 2B, the mirror image (reflected at the CO frequency) has been added to the original strip. This results in a completely symmetric spectrum with a $\sqrt{2}$ increase in S/N at the two peak positions. Although the experimental S/N ratio is $\sqrt{2}$ lower, the final sensitivity of the 3D experiment is the same as that of the 4D counterpart, provided that the symmetry information is astutely employed during the peak picking. Therefore, in view of the higher achievable resolution without sensitivity loss, the 3D HNCOCA experiment becomes very attractive.

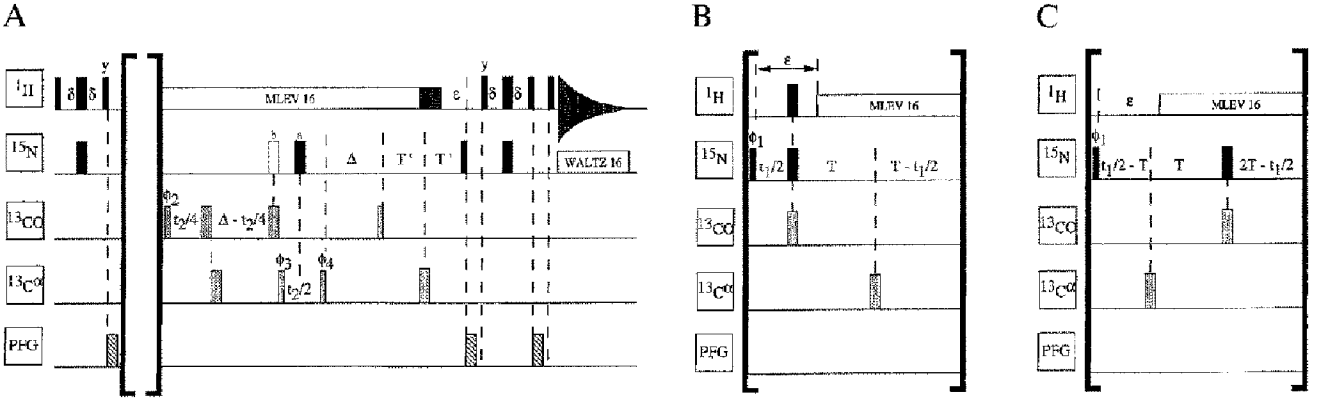

Fig. 1. Pulse sequence for the 3D HNCOCA experiment. The sequence is slightly different for times $t_1 < 2T$ (insert B) and $t_1 > 2T$ (insert C), as the 180° decoupling pulse on the C^{α} has to be moved at time 2T. The unchanged part of the sequence is shown in part (A). In order to get minimal excitation at the CO and C^{α} frequencies, respectively, the C^{α} and CO RF pulses were applied as square shaped pulses using RF fields ($\gamma B_1/2\pi$) of 4.4 kHz for 90° pulses and 8.8 kHz for 180° pulses. All frequency shifts on the ^{13}C channel were achieved by linear phase incrementation (as implemented in the Bruker software for shaped pulses) to avoid any phase shift. Gradient pulses of 200–400 μs (gradient strength ≈ 20 G/cm) were used for water suppression and coherence pathway selection. The carrier frequencies were set to $^1H = 4.73$ ppm, $^{15}N = 121.8$ ppm and $^{13}C = 174.8$ ppm (phase shifted to 58.2 ppm for C^{α} excitation). For the final excitation pulse, the ^1H carrier frequency was shifted to the center of the $^1H^N$ (8.38 ppm) resonances. For unambiguous determination of the C^{α} chemical shifts the demodulation frequency was shifted to 68 ppm. This was achieved by incrementing the phase ϕ_x after each complex point in t_2 ($\Delta\phi = 2\pi \times \Delta\nu \times \Delta t$). The spectral widths are: 3623 Hz ($^1H^N$), 1824 Hz (^{15}N) and 6000 Hz (^{13}C). On account of the relaxation properties of the cytochrome c' sample, the transfer delays δ ($1/4J_{HN}$), ε ($1/2J_{NH}$), T ($1/4J_{NCO}$) and Δ ($1/2J_{COC^{\alpha}}$) were optimized to δ = 2.2 ms, ε = 5.4 ms, T = 11 ms, T' = 11.0285 ms (b), T' = 11.0855 ms (a) and Δ = 6.5 ms. The pulse lengths of the ^{13}C pulses were taken into account (T') to avoid an important first-order phase correction in t_1. For the same reason (in t_2) the 180° refocusing pulse on ^{15}N was applied at position (b) for the first point in t_2 and then switched to position (a). Field strengths for composite decoupling are 2.8 kHz (for ^1H) and 2.1 kHz (for ^{15}N). At the end of the ^1H decoupling, a spin-lock pulse of 2 ms further removes the H_2O resonance. All pulses are aligned along the x-axis, unless indicated otherwise. A four-step phase cycle was applied as follows: $\phi_1 = x,-x$; $\phi_2 = 2(x),2(-x)$; $\phi_3 = x,-x$; and $\phi_{rec} = 2(x),2(-x)$. Quadrature detection was accomplished by incrementing the phases ϕ_1 and ϕ_2, respectively, according to the States-TPPI method (Marion et al., 1989).

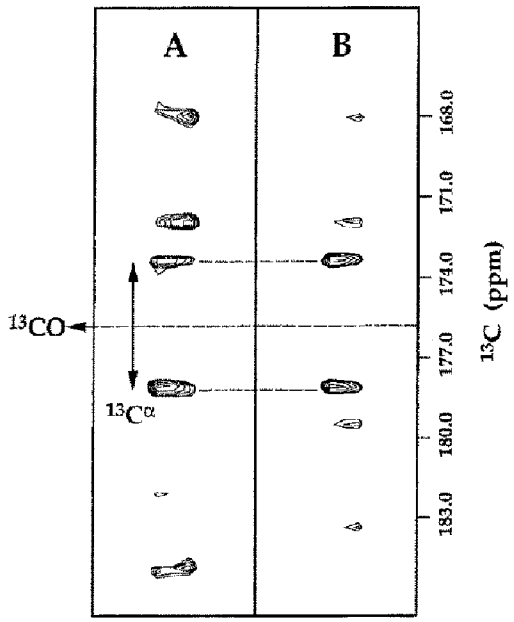

Fig. 2. ^{13}C strip of residue V6, drawn at the noise level (A). The intensity of the two correlation peaks is close to the noise level. Nevertheless, the symmetry of the two peaks with respect to a known center point allows one to distinguish between signal and noise. This is shown in part (B), where spectrum (A) has been added to its mirror image (reflected on the CO frequency) resulting in an increased S/N ratio for the two correlation peaks. All other peaks, which do not display the symmetry with regard to the same CO frequency (e.g., the signal at 172 ppm), are more or less cancelled, together with the experimental artefacts.

The second point deals with spectral resolution. In efforts to reduce the transverse relaxation periods, constant-time editing has been proposed, where the labelling of a nucleus is overlayed on an already existing INEPT transfer period (Bax et al., 1979). As long as constant-time evolution is performed, high resolution can be achieved without loss of sensitivity. More recently, for the so-called 'out-and-back' TRC experiments, different methods have been proposed to double the resolution in the constant-time dimensions for HMQC- (Madsen and Sørensen, 1992) and HSMQC-type transfer (Van Doren and Zuiderweg, 1994). In these so-called 'full sweep constant-time' versions, both the defocusing and the refocusing period of a transfer are used for spin labelling. This method is particularly appealing for transfer via relatively large coupling constants (and, consequently, short transfer delays). The HMQC version was implemented here for both $N \rightarrow CO$ and $CO \rightarrow C^{\alpha}$ transfers. In these two cases, no significant sensitivity difference is observed between the HSMQC and the HMQC version: both the MQ and SQ relaxation rates are essentially determined by the ^{1}H-$^{13}C^{\alpha}$ dipolar relaxation, because the CO does not bear a proton and the ^{15}N has a low gyromagnetic ratio. On the other hand, HMQC transfer entails fewer pulses and no additional phase cycling and thus results in less signal loss due to pulse imperfections. A basic phase cycle of only four scans is proposed to allow the recording of a high-resolution spectrum in less than three days. The resolution can be further improved by applying mirror-image linear prediction along the CT dimension (Zhu and Bax, 1990).

The labelling of the carbon frequencies can be optimized in the 3D HNCOCA on the basis of the chemical shift ranges and relaxation properties of the CO and C^{α}. The chemical shift range of the CO in proteins is about 10 ppm, as compared to 30 ppm for the C^{α}, and the transverse relaxation of the CO is slower than that of the C^{α}. The slower relaxation rate of the CO prompts us to use a slower sampling rate and a longer acquisition time for the CO. To achieve this relative scaling of the two dwell times, the slowly relaxing CO nucleus is labelled twice, as an MQ coherence together with the C^{α} and during the $CO \rightarrow C^{\alpha}$ CT transfer. For a complete discussion, the influence of the $J_{C^{\alpha}C^{\beta}}$ and $J_{NC^{\alpha}}$ coupling constants during the t_2 labelling period should be considered. The $J_{C^{\alpha}C^{\beta}}$ coupling is active during the C^{α} labelling. This coupling cannot be removed easily (by applying a 180° decoupling pulse), because of the overlapping chemical shift ranges of C^{α} and C^{β}. As the ^{15}N nucleus remains in the transverse plane, the $J_{NC^{\alpha}}$ coupling is active during the CT labelling of CO (the C^{α} coherence is not part of the {^{15}N-CO} MQ coherence) but not during the second part of the labelling ({^{15}N-CO-C^{α}} MQ coherence). Keeping in mind the simplicity of the pulse sequence, this coupling has not been removed as its value (≈ 10 Hz) can be neglected with respect to the natural ^{13}C linewidth ($\Delta\nu \approx 40$ Hz) and the $J_{C^{\alpha}C^{\beta}}$ coupling (≈ 35 Hz). The signal modulation (in t_2) is described as follows:

$$\exp\left(\frac{-2\Delta}{T_2(CO)}\right) \times \exp(i\Delta\Omega_{CO}t_2)$$
$$\times \cos\left(\frac{\Delta\Omega_{C^{\alpha}}}{\lambda}t_2\right) \times \cos\left(\pi\frac{J_{C^{\alpha}C^{\beta}}}{\lambda}t_2\right) \qquad (1)$$
$$\times \cos\left(\pi\frac{(\lambda-1)\times J_{NC^{\alpha}}}{\lambda}t_2\right) \times \exp\left(\frac{-t_2}{\lambda * T_2(COC^{\alpha})}\right)$$

where λ is the scaling factor between the CO and C^{α} dwell times ($DW(CO) = \lambda \times DW(C^{\alpha})$) and Δ the $CO \rightarrow C^{\alpha}$ transfer delay (Fig. 1). If λ is set to 1 (no additional CO labelling), the peak difference in Hz equals twice the frequency difference between the C^{α} chemical shift and the C^{α} demodulation frequency ($\Delta\Omega_{C^{\alpha}}$).

The scaling factor λ has two outcomes on the experimental spectrum: the location of the peaks and their linewidth. As the effective C^{α} modulation frequency is given by $\Delta\Omega_{C^{\alpha}}/\lambda$, the separation of the two peaks is scaled according to $1/\lambda$, as previously described (Brutscher et al., 1994). As a consequence, the scaling factor λ reduces the spectral width in the ^{13}C dimension. A total ^{13}C spectral width of $SW(CO) + 2 \times SW(C^{\alpha})/\lambda$ is sufficient to avoid any

signal folding. On the other hand, the apparent linewidth of the correlation peaks is altered by this scaling procedure. As the additional CO editing is achieved during CT evolution, the associated relaxation reduces the peak intensity (first exponential term of Eq. 1) but does not further broaden the peaks. Consequently, for the same overall acquisition time in the ^{13}C dimension (determined by the decay rate of the COC^{α} coherence during t_2) the effective transverse relaxation time is $\lambda * T_2(COC^{\alpha})$. As the effective $J_{C^{\alpha}C^{\beta}}$ coupling is reduced in the same way, the resulting signal linewidth in the ^{13}C dimension is decreased by a factor of λ. In the pulse sequence of Fig. 1, a scaling factor $\lambda = 2.0$ has been chosen, which is a good compromise taking into account the resulting ^{13}C spectral width and the values of Δ and $T_2(COC^{\alpha})$ for larger proteins. Thus, the spectral width was set to 40 ppm (CO scale) without folding of signals. Further experimental details are given in the legend of Fig. 1.

To test the performance of the sequence, a 3D data set was recorded on a $^{15}N/^{13}C$-labelled cytochrome *c'*. This protein forms a 28 kDa complex of two identical subunits. Each of the monomers consists of 129 residues and a paramagnetic heme group (spin state of the oxidized Fe(III) form: $S = 5/2$). The *Rb. capsulatus* cytochrome *c'* gene has not been cloned and expressed in an appropriate vector; thus, only a very small amount of $^{15}N/^{13}C$-labelled cytochrome *c'*, which was purified from *Rb. capsulatus* grown in minimal media, was available. For the NMR experiments the protein was dissolved in 200 µl, resulting in a concentration of approximately 0.5 mM (heme concentration). A Shigemi NMR tube (Shigemi, Allison Park, PA) was used to avoid homogeneity problems due to the small sample volume. The experiment was performed on a Bruker AMX-600 spectrometer equipped with a triple-resonance gradient probe. Due to the finite stability of the sample, the overall experimental time was limited to 35 h and 8 transients were recorded per $\{t_1, t_2\}$ increment to benefit from the larger NMR signal at the early stages of the experiment. As a result, the potential resolution was not attained in the ^{15}N dimension. The acquisition matrix contained $256(^1H) \times 41(^{15}N) \times 128(^{13}C)$ complex points (acquisition times were: 71 ms (^1H), 22.5 ms (^{15}N) and 21.4 ms (^{13}C)). The data were processed using the FELIX program, version 2.35 (Biosym Technologies, San

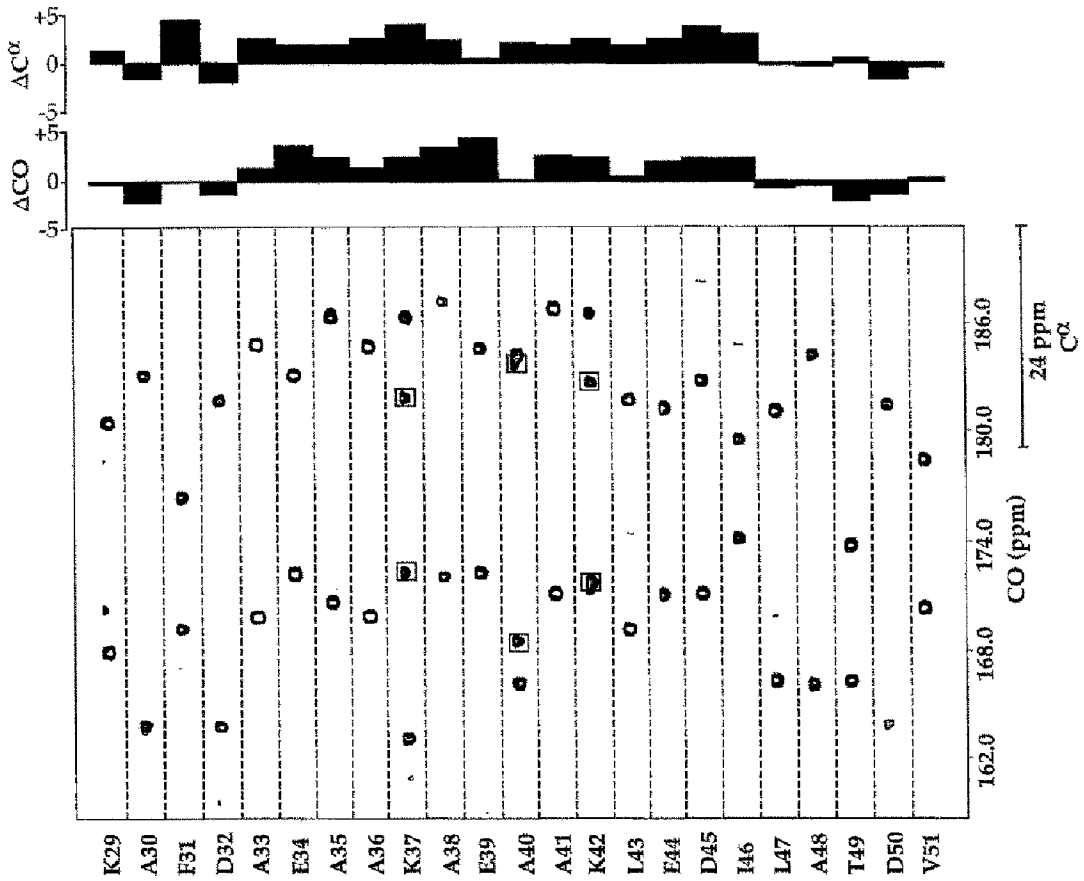

Fig. 3. 2D representation of the 3D HNCOCA experiment for the CO and C^{α} assignment of residues K29–V51. Each residue is taken at the $\{^1H^N, ^{15}N\}$ frequencies of the following residue. In cases of more than two correlation peaks (e.g. K37, A40, K42) the correct combination is indicated by open squares. A separate scale for C^{α} is given, taking into account the scaling factor $\lambda = 2.0$. The corresponding CO and C^{α} frequencies have been extracted in the following way: for each doublet, the center frequency is the CO shift and the distance to the center is the C^{α} shift (using the C^{α} scale and versus the demodulation frequency: 68 ppm). These chemical shifts, shown as deviations from random coil values $\delta_{obs} - \delta_{rc}$ (Richarz and Wüthrich, 1978), indicate a helical region from residue A33 to I46 (Wishart and Sykes, 1994).

Diego, CA). Prior to Fourier transformation the data were multiplied by Lorentzian-to-Gaussian transformation (^{1}H and ^{13}C) and skewed sine-bell functions (^{15}N) and zero-filled to a final matrix of $512 \times 128 \times 256$ real points. As stated above, the cytochrome *c′* sample was degrading during the course of the experiment, yielding a signal decay in the ^{15}N dimension recorded in the slowest loop (about one complex point per hour). For this reason, no mirror-image linear prediction was applied in this CT dimension. For convenient analysis of the transformed data, a ^{1}H-^{15}N correlation spectrum was recorded in order to extract ^{13}C strips from the 3D matrix. A FELIX macro extracts these strips and stores them into a 2D matrix, as previously described (Caffrey et al., 1994).

The good S/N ratio of the data is illustrated in Fig. 3 for the ^{13}C strips assigned to residues K29–V51. In most cases the CO(i–1) and C^{α}(i–1) frequencies could be determined directly by calculation of the midpoint (CO) and the distance (C^{α}) of the two correlation peaks seen in the strip. In cases of overlapping or nonresolved {^{1}H,^{15}N} HSQC peaks, the peak picking of a 2D H(N)CO experiment (recorded in 1 h) was used to differentiate among possible combinations. In this way 103 out of the 124 possible pseudoresidues (129 residues minus three prolines and the amino- and carboxy-terminal residues) could be unambiguously identified. For five {^{1}H,^{15}N} pairs detected in the HSQC experiment, no pseudoresidue could be built. Note that, for a few residues close to the paramagnetic center, no HSQC correlation peak could be detected, probably due to line broadening. This example shows that the 3D HNCOCA pulse sequence described here yields good results, even under very poor NMR conditions (low concentration, high molecular weight, presence of a paramagnetic center and unstable sample conditions). With the aid of an additional HNCA experiment (Grzesiek and Bax, 1992) the assignment previously obtained by ^{1}H-^{15}N NMR on an 8 mM sample labelled with ^{15}N has been confirmed (Caffrey et al., 1995). In addition, the ^{13}C chemical shifts have been used to identify the secondary structural elements of the protein (Wishart and Sykes, 1994). This is shown in Fig. 3 for one out of the four helices (A33–I46) which have been identified in the *Rb. capsulatus* cytochrome *c′*.

In conclusion, this 3D version of a high-resolution HNCOCA correlation experiment enables unambiguous identification of pseudoresidues in a doubly labelled protein using only one data set. In spite of difficulties usually encountered for TRC experiments of higher dimensionality, this pulse sequence still yields good results for proteins with high molecular weights (e.g. 28 kDa) and low sample concentration. Therefore, the HNCOCA experiment provides a sensitivity which is sufficient for a low-concentration ^{15}N/^{13}C sample when the protein cannot be expressed (or overexpressed) at a high rate. The derived information, combined with that of an HNCA experiment, yields a valuable basis for the assignment of larger proteins. In many cases, a second sample labelled with ^{15}N is available at a much higher concentration; a reliable assignment can be obtained when the information of the two sensitive TRC experiments (3D HNCOCA and 3D HNCA) is used in concert with additional and independent information from {^{1}H,^{15}N} NMR experiments. We believe that the approach of using two samples, a low-concentration ^{15}N/^{13}C-labelled sample together with a high-concentration ^{15}N-labelled sample, can be a very attractive and inexpensive alternative in cases where suitable expression systems are either missing or inefficient.

Acknowledgements

The authors are pleased to thank N. Morelle (IBS, Grenoble) for assistance in the automatic data analysis and Dr. S. van Doren and Prof. E.R.P. Zuiderweg (University of Michigan) for helpful discussions. The contribution of Prof. M. Cusanovich (University of Arizona) in the preparation of the labelled sample is acknowledged. This work was supported by the Centre National de la Recherche Scientifique, the Commissariat à l'Énergie Atomique and the Bruker company (Wissembourg). B.B. is the recipient of a CEA-Bruker Ph.D. fellowship. This is publication no. 259 of the Institut de Biologie Structurale-Jean Pierre Ebel.

References

Bax, A., Mehlkopf, A.F. and Smidt, J. (1979) *J. Magn. Reson.*, **35**, 167–169.

Brutscher, B., Simorre, J.-P., Caffrey, M.S. and Marion, D. (1994) *J. Magn. Reson. Ser. B*, **105**, 77–82.

Caffrey, M.S., Brutscher, B., Simorre, J.-P., Fitch, J., Cusanovich, M.A. and Marion, D. (1994) *Eur. J. Biochem.*, **221**, 63–75.

Caffrey, M.S., Simorre, J.-P., Brutscher, B., Cusanovich, M.A. and Marion, D. (1995) *Biochemistry*, in press.

Grzesiek, S. and Bax, A. (1992) *J. Magn. Reson.*, **96**, 432 440.

Ikura, M., Kay, L.E. and Bax, A. (1990) *Biochemistry*, **29**, 4659–4667.

Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) *J. Magn. Reson.*, **89**, 496 514.

Madsen, J.C. and Sørensen, O.W. (1992) *J. Magn. Reson.*, **100**, 431–436.

Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989) *J. Magn. Reson.*, **85**, 393–399.

Montelione, G.T. and Wagner, G. (1990) *J. Magn. Reson.*, **87**, 183–188.

Morelle, N., Brutscher, B., Simorre, J.-P. and Marion, D. (1995) *J. Biomol. NMR*, **5**, 154–160.

Richarz, R. and Wüthrich, K. (1978) *Biopolymers*, **17**, 2133–2141.

Simorre, J.-P., Brutscher, B., Caffrey, M.S. and Marion, D. (1994) *J. Biomol. NMR*, **4**, 325–333.

Szypersky, T., Wider, G., Bushweller, J.H. and Wüthrich, K. (1993a) *J. Am. Chem. Soc.*, **115**, 9307–9308.

Szypersky, T., Wider, G., Bushweller, J.H. and Wüthrich, K. (1993b) *J. Biomol. NMR*, **3**, 127–132.

Van Doren, S. and Zuiderweg, E.R.P. (1994) *J. Magn. Reson. Ser. B*, **104**, 171–180.

Wishart, D. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 127–333.

Zhu, G. and Bax, A. (1990) *J. Magn. Reson.*, **90**, 405–410.