Four-Dimensional Heteronuclear Triple-Resonance NMR of Isotopically Enriched Proteins for Sequential Assignment of Backbone Atoms

LEWIS E. KAY, MITSUHIKO IKURA, GUANG ZHU,* AND AD BAX

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

Any two-dimensional NMR experiment that involves a relay step to transfer magnetization between two nuclei via an intermediate nucleus can be converted into a three-dimensional experiment where the third axis provides the resonance frequency of the intermediate spin. Indeed, many of the 3D experiments proposed so far are constructed according to this principle and have popular 2D relay analogs. For example, the 3D heteronuclear NOESY–HMQC (*1–3*), HMQC–COSY (*4*), HOHAHA–HMQC (*5*), and ROESY–HMQC (*6*) experiments are very similar to their 2D counterparts (*6–11*). Similarly, the 3D homonuclear NOESY–HOHAHA/TOCSY experiments (*12, 13*) are directly related to analogous 2D experiments (*14, 15*). Here we show an example where a 3D relay pulse scheme is advantageously converted into a 4D experiment.

Recently, it has been demonstrated that four-dimensional NMR is a feasible and very practical approach for simultaneously reducing overlap and obtaining additional information regarding NOE interactions between amide and aliphatic protons (*16*). The 4D experiment described in this paper is of a different type and can be considered the analog of a previously published 3D relay experiment, known as HCA(CO)N (*17, 18*). The new 4D experiment yields the combined information that can be obtained from the 3D HCA(CO)N experiment, and a 3D experiment known as HCACO, removing ambiguities that may be present in each of the corresponding 3D spectra. The HCA(CO)N 3D spectrum correlates the Hα and Cα resonances of one amino acid residue in a protein with the ^{15}N resonance of the next residue, using the carbonyl resonance (CO) as an intermediate nucleus to transfer coherence from Cα to ^{15}N, and vice versa. The HCACO spectrum correlates intraresidue Hα, Cα, and CO chemical shifts. Both 3D and 4D experiments require a protein that is labeled uniformly at a high level (>90%) with both ^{15}N and ^{13}C.

In the 4D experiment, named HCACON, the four frequency coordinates of a resonance are determined by the intraresidue Hα (F_4), Cα (F_1), and CO (F_2) shifts and the ^{15}N shift (F_3) of the next residue. The pulse sequence of this experiment is sketched in Fig. 1. The magnetization transfer pathway that gives rise to the desired resonances

---

* Also Chemical Physics Program, University of Maryland at College Park.
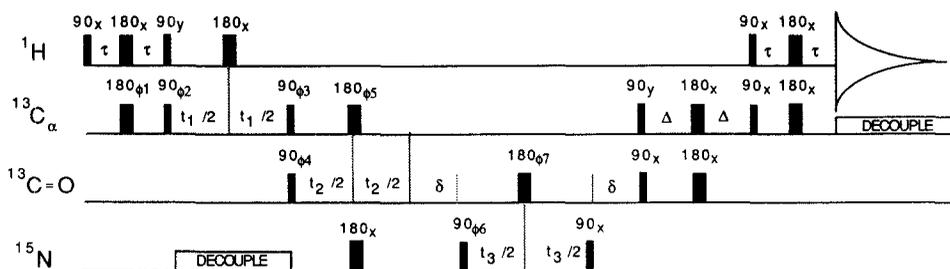
FIG. 1. Pulse sequence of the 4D HCACON sequence. Typical delay durations used are $\tau = 1.5$ ms, $\delta = 20$ ms, and $\Delta = 3.25$ ms. ^{13}C pulse widths are adjusted such that a 90° Cα pulse causes minimal excitation of the carbonyl resonances, and vice versa. Carbonyl pulses are generated in a DANTE-like fashion (18) using the Cα channel. The phase cycling employed is as follows: $\phi_1 = 16(x)$, $16(-x)$; $\phi_3 = 16(x)$, $16(-x)$; $\phi_4 = 2(x)$, $2(-x)$; $\phi_5 = 8(x)$, $8(-x)$; $\phi_6 = x$, $-x$; $\phi_7 = 4(x)$, $4(y)$, $4(-x)$, $4(-y)$; Acq. $= x$, $2(-x)$, x, $-x$, $2(x)$, $2(-x)$, $2(x)$, $-x$, x, $2(-x)$, x, $-x$, $2(x)$, $-x$, x, $2(-x)$, $2(x)$, $2(-x)$, x, $-x$, $2(x)$, $-x$. Complex data in the t_1, t_2, and t_3 dimensions are generated by incrementing the phases ϕ_2, ϕ_4, and ϕ_6 independently by 90°, using the States–TPPI method (26).

in the 4D spectrum will be briefly discussed below in terms of the product operator formalism. In this description, the effects of the pulse sequence between the end of the magnetization recovery period and the beginning of the evolution period are described by the operator A; the effect of the 90° (Cα) and 90° (CO) pulses between t_1 and t_2 is described by the operator B; the delay δ and the 90° (^{15}N) pulse are described by the operator C; and finally, the operator for the pulses and delays between the end of the t_3 period and the beginning of data acquisition is given by D. In the discussion below, pulse phases are assumed to correspond to the first step of the phase cycle, but for simplicity only terms that remain after the full 32-step phase cycle are considered. The Hα, Cα, CO, and ^{15}N spins are labeled I, A, S, and N, respectively.

First a regular INEPT type transfer converts Hα magnetization into antiphase Cα magnetization:

$$I_z \xrightarrow{\quad A \quad} -2I_z A_y \sin 2\pi J_{IA}\tau. \qquad [1]$$

Because in this 4D experiment Hα magnetization is transferred to Cα, CO, and ^{15}N spins, before being transferred back to Hα for observation, it is not necessary to use a refocused INEPT transfer, which would require additional delays and RF pulses. Thus, during t_1, magnetization remains antiphase with respect to Hα,

$$-2I_z A_y \xrightarrow{\quad t_1 \quad} -4I_z A_x S_z \cos \Omega_A t_1 \sin \pi J_{AS} t_1 \cos \pi J_{AB} t_1, \qquad [2]$$

where J_{AB} denotes the passive Cα–Cβ J coupling, and the sine term at the right-hand side of expression [1] has temporarily been omitted. The following polarization transfer to the carbonyl and t_2 evolution is described by

$$-4I_z A_x S_z \xrightarrow{\quad B, t_2 \quad} 4I_z A_z S_y \cos \Omega_S t_2. \qquad [3]$$

In the above expression and in what follows, small and unresolved long-range couplings have been neglected. Next, the S_y magnetization is correlated with ^{15}N using the principle of heteronuclear multiple-quantum coherence (HMQC) (19, 20), which then

evolves during the time t_3:

$$4I_zA_zS_y \xrightarrow{\quad C, t_3 \quad} 8I_zA_zS_xN_y\sin \pi J_{SN}\delta \cos \Omega_N t_3. \qquad [4]$$

When this coherence is transferred back into ^1H magnetization, following the reverse of the pathway described above, this results in

$$-I_x\cos \Omega_A t_1 \sin \pi J_{AS}t_1 \cos \pi J_{AB}t_1 \cos \Omega_S t_2 \sin^2\pi J_{SN}\delta \cos \Omega_N t_3$$

$$\times \sin 2\pi J_{AS}\Delta \cos 2\pi J_{AB}\Delta \sin^2 2\pi J_{IA}\tau, \quad [5]$$

where all previously omitted sine and cosine terms have been reintroduced. As can be seen from expression [5], the detected ^1H magnetization, I_x, is modulated in amplitude by $\cos \Omega_A t_1$, $\cos \Omega_S t_2$, and $\cos \Omega_N t_3$ in the t_1, t_2, and t_3 dimensions. In addition there is a modulation by $\sin \pi J_{AS}t_1$ in the t_1 dimension. As discussed previously (18), we prefer to phase the resulting signal after Fourier transformation to the antiphase dispersive mode, which gives rise to peaks with lineshapes that resemble absorptive resolution-enhanced singlets. In the other dimensions, the signal is purely absorptive.

The 4D HCACON experiment has been applied to a 0.95 mM sample of calmodulin complexed with Ca^{2+} and a 26-residue N-terminal peptide of myosin light-chain kinase. The total molecular weight of the complex is 19.7 kDa. Experiments were conducted at 35°C, on a Bruker AM-500 spectrometer, equipped with a Bruker triple-resonance probehead and with homebuilt hardware to generate the RF needed for the third channel in the triple-resonance experiments (18). In order to ensure the absence of artifacts in the final 4D spectrum, we found it necessary to execute the entire 32-step phase cycle presented in the legend to Fig. 1. In addition, 8 steps are needed to obtain complex data in the t_1, t_2, and t_3 dimensions, resulting in a total of 256 scans per (t_1, t_2, t_3) increment, severely limiting the number of increments that could be used in a reasonable amount of time. The size of the acquired data matrix was (32 complex) $\times$ (8 complex) $\times$ (8 complex) $\times$ (512 real) in the t_1, t_2, t_3, and t_4 dimensions respectively. In order to increase the digital resolution in the ^{15}N dimension, the spectral width in this dimension was set to 20 ppm, slightly less than the total range of ^{15}N chemical shifts. This causes aliasing of some of the resonances in this dimension, however without introducing any ambiguity (21). No aliasing was used in any of the other frequency dimensions. To further increase the resolution in the t_2 and t_3 dimensions, the length of the truncated data table is extended to 16 complex data points using linear prediction. Prior to Fourier transformation, zero filling in all dimensions was used, resulting in a manageable size (32 Mword) for the real part of the 4D spectrum. Note that if imaginary data were not discarded, 16 times this amount of data storage would be required. A flow diagram of the processing procedure is shown in Fig. 2. During data processing, the 256 (t_1, t_4) slices of the raw data matrix are treated as separate 2D data files. Processing of these 2D files was performed using commercial software (NMRi, Syracuse, New York). Fourier transformation, phasing, and digital filtering in the remaining two dimensions (t_2 and t_3) are performed with in-house written software, described previously (21). The requirement to increase the resolution in the t_2 and t_3 dimensions by linear prediction, however, adds some complication to the processing. A good solution would be to use a true 2D linear prediction algorithm (22) for each of the (t_2, t_3) slices. However, we have not yet successfully implemented such a program and therefore have decided to use one-dimensional
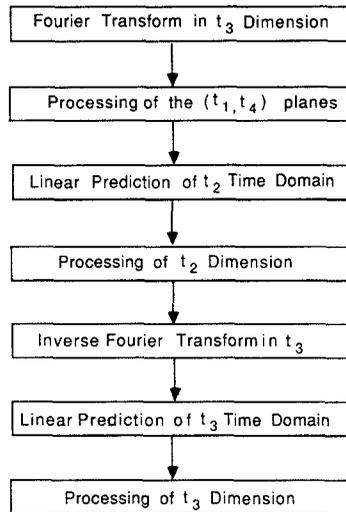
FIG. 2. Flowchart of the steps used for processing the 4D data set.

linear prediction routines for extending the time domain in both the t_2 and the t_3 dimension separately. Before such a one-dimensional linear prediction can be performed, it is essential to minimize the number of frequency components present in the time domain that is to be extended. To this extent, Fourier transformation in all other dimensions must be performed prior to the linear prediction. Thus, in the scheme of Fig. 2, we start with a Fourier transformation in the t_3 dimension, followed by digital filtering, zero filling, Fourier transformation, phasing, and discarding of imaginary data in the t_1 and t_4 dimension. Subsequently, in the t_2 dimension, linear prediction, digital filtering, zero filling, and Fourier transformation are performed and imaginary data are subsequently discarded. Finally, the last step which serves to increase the resolution in the F_3 dimension consists of an inverse Fourier transformation in this dimension, followed by linear prediction, digital filtering, zero filling, Fourier transformation, phasing, and discarding of the imaginary data. Because the time-domain data are to be extended by a factor of two, significantly longer than that used by Olejniczak and Eaton (23), extra care is needed in this process. Thus, it is essential to treat the data as complex numbers, rather than predicting the future of the real and imaginary parts of the signal separately. Also, as pointed out previously (24), when extending the time domain substantially it becomes necessary to use a process known as root reflection (25) which ensures that the predicted part of the time domain does not increase in amplitude. For the short time domain used (8 complex data points) this involves calculating the roots of a cubic polynomial, in the case considered here, which adds an insignificant fraction to the total data processing time. All data processing, including linear prediction, Fourier transformation, and phasing, was done in a fully automated fashion, requiring two days of processing time on a SUN Sparc-1 workstation.

The digital resolution in the final spectrum is 23.4 Hz (F_1, Cα), 43.4 Hz (F_2, CO), 31.3 Hz (F_3, ^{15}N), and 9.8 Hz (F_4, Hα). Although by conventional high-resolution standards this digitization is rather coarse, there is virtually no overlap in this spectrum

and peak positions can therefore be determined with reasonable accuracy. Comparison of 4D peak coordinates as determined by software using parabolic interpolation with those determined in a similar manner from 3D spectra recorded with much higher resolution indicates that the accuracy of the peak picking in the 4D spectrum is at least three times better than the spectral digital resolution, yielding sufficient precision for sequential assignment purposes.

Figure 3 illustrates regions of representative (F_1, F_4) slices taken from the 4D spectrum. These slices are perpendicular to both the F_2 and the F_3 axis, and show a correlation between directly attached $C\alpha$ and $H\alpha$ resonances for residues that have a C' frequency corresponding to the F_2 frequency indicated in each panel and are followed by residues with a ^{15}N shift corresponding to the indicated F_3 coordinate. Apart from the backbone connectivities, the 4D spectrum also shows a number of weak $H\beta$–$C\beta$–CO–N and $H\gamma$–$C\gamma$–CO–N connectivities for the side chains of Asn and Gln residues, respectively. Thus, the total number of resonances in the 4D spectrum is only slightly larger than the total number of residues in the protein. Consequently, the entire 4D HCACON spectrum for a relatively large protein contains at most several hundred resonances, and overlap is therefore rare.
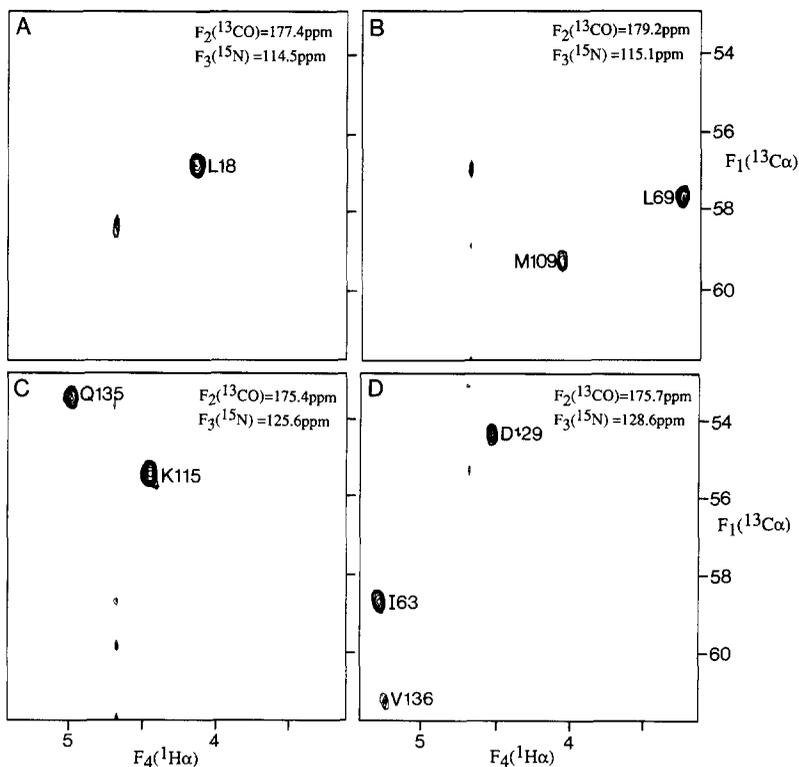
FIG. 3. Selected regions of (F_1, F_4) slices of the 4D HCACON spectrum of a complex of calmodulin, Ca^{2+}, and a 26-residue peptide. The total molecular weight of the complex is 19.7 kDa. The protein was uniformly enriched with ^{13}C and ^{15}N at a high level (>90%). The spectrum was recorded on a Bruker AM-500 spectrometer, using a sample concentration of 0.9 mM at 35°C, p^{2}H 6.8. Slices have been taken perpendicular to the F_2 and F_3 axes, at coordinates indicated in each panel.

Automated determination of resonance positions is straightforward for nonoverlapping peaks, as found in the 4D spectrum. However, peak picking for the HCACO and HCA(CO)N 3D spectra is not as straightforward because a number of cases of partial overlap occur in these spectra. Thus, the 4D spectrum makes automated analysis much easier. Moreover, in the substantial number of cases where pairs of $C\alpha$–$H\alpha$ resonances overlap, the combination of these two 3D spectra does not make it possible to decide which carbonyl resonance is attached to which ^{15}N resonance. However, such an ambiguity is immediately resolved by inspection of (F_2, F_3) slices taken from the 4D spectrum, as shown in Fig. 4. The panels shown are taken perpendicular to the F_1 and F_4 axes, at $H\alpha$ and $C\alpha$ frequencies that correspond to four of the resonances shown in Fig. 3. These slices thus represent ^{15}N–CO correlations for a given combination of $H\alpha$ and $C\alpha$ frequencies. Because in four-dimensional space there are six orthogonal planes, apart from the slices shown in Figs. 3 and 4, four more types of slices can be extracted from the 4D spectrum, i. e., (F_1, F_2), (F_1, F_3), (F_2, F_4), and (F_3, F_4) slices. In practice, however, we make a hard-copy printout of the spectrum for only one set of slices (F_1, F_4 for the present spectrum). This hard copy is used for visual inspection which sometimes is needed for distinguishing between real resonances and t_1 noise, such as observed in the panels shown in Fig. 3 at the HDO F_4 frequency (4.67 ppm).
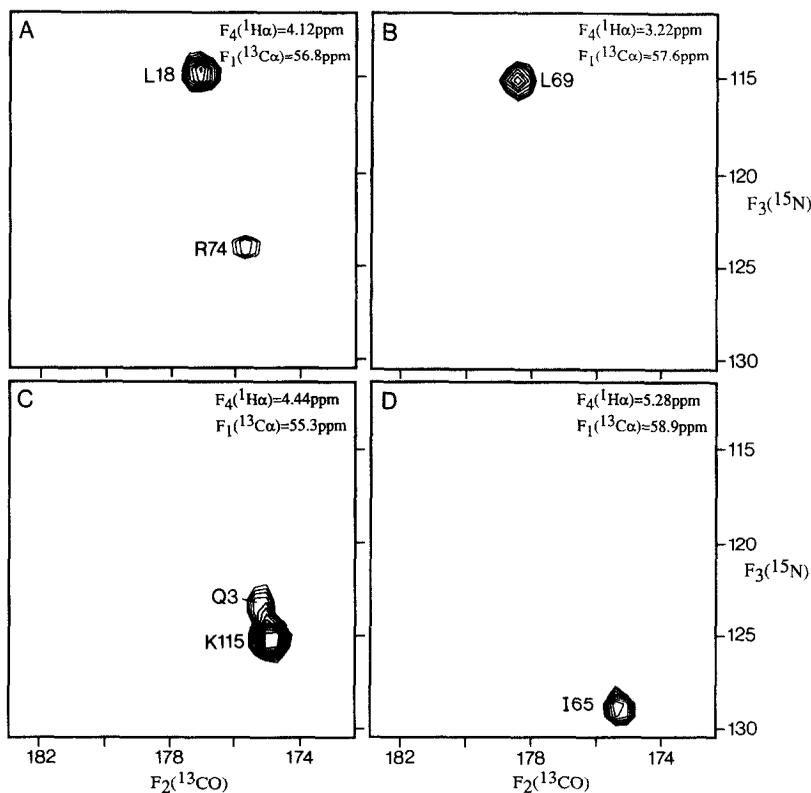
FIG. 4. Selected regions of (F_2, F_3) slices of the 4D HCACON spectrum. Slices have been taken perpendicular to the F_1 and F_4 axes, at F_1 and F_4 values indicated in each panel. The slices have been selected such that four of the resonances present in Fig. 2 are again visible in the orthogonal planes.

We have shown that 4D NMR is a feasible approach for the study of larger proteins. The experiment is performed in a nonselective manner. Because ^{15}N acts as a filter for resonances to be observed, the chemical-shift distribution in each of the four dimensions is relatively narrow, allowing data to be recorded in an acceptable amount of measuring time (5 days) with reasonable digital resolution despite the relatively small size of the acquired data matrix (8 Mword).

ACKNOWLEDGMENTS

REFERENCES

1. S. W. FESIK AND E. R. P. ZUIDERWEG, *J. Magn. Reson.* **78,** 588 (1988).
2. D. MARION, L. E. KAY, S. W. SPARKS, D. A. TORCHIA, AND A. BAX, *J. Am. Chem. Soc.* **111,** 1515 (1989).
3. E. R. P. ZUIDERWEG AND S. W. FESIK, *Biochemistry* **28,** 2387 (1989).
4. S. W. FESIK, R. T. GAMPE, AND E. R. P. ZUIDERWEG, *J. Am. Chem. Soc.* **111,** 770 (1989).
5. D. MARION, P. C. DRISCOLL, L. E. KAY, P. T. WINGFIELD, A. BAX, A. M. GRONENBORN, AND G. M. CLORE, *Biochemistry* **28,** 6150 (1989).
6. G. M. CLORE, A. BAX, P. T. WINGFIELD, AND A. M. GRONENBORN, *Biochemistry* **29,** 5671 (1990).
7. A. M. GRONENBORN, A. BAX, P. T. WINGFIELD, AND G. M. CLORE, *FEBS Lett.* **243,** 93 (1989).
8. K. SHON AND S. J. OPELLA, *J. Magn. Reson.* **82,** 193 (1989).
9. L. LERNER AND A. BAX, *J. Magn. Reson.* **69,** 375 (1986).
10. D. G. DAVIS, *J. Magn. Reson.* **84,** 417 (1989).
11. G. WAGNER AND D. BRÜWHILER, *Biochemistry* **25,** 5839 (1986).
12. H. OSCHKINAT, C. GRIESINGER, P. J. KRAULIS, O. W. SØRENSEN, R. R. ERNST, A. M. GRONENBORN, AND G. M. CLORE, *Nature (London)* **332,** 374 (1988).
13. G. W. VUISTER, R. BOELENS, AND R. KAPTEIN, *J. Magn. Reson.* **80,** 176 (1988).
14. G. WAGNER, *J. Magn. Reson.* **57,** 497 (1984).
15. H. KESSLER, G. GEMMECKER, AND S. STEUERNAGEL, *Angew. Chem. Int. Engl. Ed.* **27,** 564 (1988).
16. L. E. KAY, G. M. CLORE, A. BAX, AND A. M. GRONENBORN, *Science* **249,** 411 (1990).
17. M. IKURA, L. E. KAY, AND A. BAX, *Biochemistry* **29,** 4659 (1990).
18. L. E. KAY, M. IKURA, R. TSCHUDIN, AND A. BAX, *J. Magn. Reson.* **89,** 496 (1990).
19. L. MÜLLER, *J. Am. Chem. Soc.* **101,** 4481 (1979).
20. A. BAX, R. H. GRIFFEY, AND B. L. HAWKINS, *J. Magn. Reson.* **55,** 301 (1983).
21. L. E. KAY, D. MARION, AND A. BAX, *J. Magn. Reson.* **84,** 72 (1989).
22. H. GESMAR AND J. J. LED, *J. Magn. Reson.* **76,** 575 (1988).
23. E. T. OLEJNICZAK AND H. L. EATON, *J. Magn. Reson.* **87,** 628 (1990).
24. G. ZHU AND A. BAX, *J. Magn. Reson.,* in press.
25. W. H. PRESS, B. P. FLANNERY, S. A. TEUKOLSKY, AND W. T. VETTERLING, "Numerical Recipes in C: The Art of Scientific Computing," p. 463, Cambridge Univ. Press, Cambridge, 1989.
26. D. MARION, M. IKURA, R. TSCHUDIN, AND A. BAX, *J. Magn. Reson.* **85,** 393 (1989).