Published in Biochemistry, May 15, 1990, pp. 4659–4667, by the American Chemical Society

A Novel Approach for Sequential Assignment of ^1H, ^{13}C, and ^{15}N Spectra of Larger Proteins: Heteronuclear Triple-Resonance Three-Dimensional NMR Spectroscopy. Application to Calmodulin†

Mitsuhiko Ikura, Lewis E. Kay, and Ad Bax*

Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, Maryland 20892

Received December 7, 1989; Revised Manuscript Received February 28, 1990

ABSTRACT: A novel approach is described for obtaining sequential assignment of the backbone ^1H, ^{13}C, and ^{15}N resonances of larger proteins. The approach is demonstrated for the protein calmodulin (16.7 kDa), uniformly (~95%) labeled with ^{15}N and ^{13}C. Sequential assignment of the backbone residues by standard methods was not possible because of the very narrow chemical shift distribution range of both NH and $C^\alpha H$ protons in this largely α-helical protein. We demonstrate that the combined use of four new types of heteronuclear 3D NMR spectra together with the previously described HOHAHA–HMQC 3D experiment [Marion, D., et al. (1989) *Biochemistry 28*, 6150–6156] can provide unambiguous sequential assignment of protein backbone resonances. Sequential connectivity is derived from one-bond J couplings and the procedure is therefore independent of the backbone conformation. All the new 3D NMR experiments use ^1H detection and rely on multiple-step magnetization transfers via well-resolved one-bond J couplings, offering high sensitivity and requiring a total of only 9 days for the recording of all five 3D spectra. Because the combination of 3D spectra offers at least two and often three independent pathways for determining sequential connectivity, the new assignment procedure is easily automated. Complete assignments are reported for the proton, carbon, and nitrogen backbone resonances of calmodulin, complexed with calcium.

Sequential resonance assignment of the backbone protons of a protein forms the basis for further solution structure studies by NMR.[1] This sequential assignment is most commonly accomplished by means of homonuclear ^1H 2D experiments that identify intraresidue through-bond $^3J(NH, H^\alpha)$ connectivity and sequential interresidue through-space (NOE) connectivity (Wüthrich, 1986; Kaptein et al., 1988; Clore & Gronenborn, 1989; Bax, 1989). Use of the interresidue NOE is essential in this approach because of the absence of a significant ^1H–^1H J coupling between protons of adjacent amino acids. This short-range interresidue NOE interaction depends strongly on the local conformation. In addition, many of these protons can also exhibit long-range NOE interactions, making unambiguous identification of sequential NOEs even more difficult.

In recent years, the sequential assignment procedure has been applied successfully to a large number of small proteins. For larger proteins (M_r > 10 kDa), or for proteins with a very narrow chemical shift distribution of the backbone proton resonances, the standard sequential assignment procedure may not yield unambiguous answers because of very extensive overlap in critical regions of the ^1H 2D NMR spectra. This overlap problem can be alleviated dramatically by the recording of isotope-edited 2D NMR experiments of proteins in which specific amino acids are isotopically labeled (McIntosh et al., 1987a,b; Senn et al., 1987; LeMaster & Richards, 1988; Fesik et al., 1988; Torchia et al. 1989). In a related double-labeling technique, one type of amino acid is labeled with ^{13}C in the carbonyl position and a second amino acid is labeled with ^{15}N. Any dipeptide fragment in which the amino acid labeled with ^{13}C precedes the ^{15}N-labeled residue can be easily identified on the basis of the $^1J_{CN}$ splitting

† This work was supported by the Intramural AIDS Antiviral Program of the Office of the Director of the National Institutes of Health. L.E.K. acknowledges financial support from the Medical Research Council of Canada and the Alberta Heritage Trust Foundation.

[1] Abbreviations: INEPT, insensitive nuclei enhanced by polarization transfer; HCACO, proton to α-carbon to carbonyl correlation; HCA(CO)N, proton to α-carbon (via carbonyl) to nitrogen correlation; HMQC, heteronuclear multiple quantum correlation; HNCO, amide proton to nitrogen to carbonyl correlation; HNCA, amide proton to nitrogen to α-carbon correlation; HOHAHA, homonuclear Hartmann–Hahn; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; 2D, two-dimensional; 3D, three-dimensional.

(~15 Hz) observed in the ^1H–^{15}N shift correlation spectrum (Kainosho & Tsuji, 1982; Griffey & Redfield, 1987; Torchia et al., 1989). More recently, nonselective three-dimensional experiments have been developed that separate the NOESY or J correlation spectra into a third dimension, the ^{15}N chemical shift (Fesik & Zuiderweg, 1988; Marion et al., 1989a; Zuiderweg & Fesik, 1989; Marion et al., 1989b). This approach is very useful for unraveling NOE and J correlations involving overlapping amide protons. However, when there is extensive overlap in the $C_\alpha H$ region, even this 3D approach may not be sufficient for making unambiguous sequential assignments for all residues.

For the protein calmodulin, which we have previously investigated with a combination of all experiments mentioned above (Ikura et al., 1990a), we were unable to assign more than ~80% of the backbone protons on the basis of these experiments alone. Calmodulin presents a particularly difficult case, despite its modest molecular weight (16.7 kDa), because it appears to be largely α-helical, with correspondingly narrow chemical shift distributions for the amide and $C^\alpha H$ protons; 76% of the backbone amide protons resonate in a spectral region of 1 ppm, and 50% of the $C^\alpha H$ protons resonate within a 0.52 ppm spectral region. Moreover, because calmodulin consists of four domains with substantial sequence homology, many connectivity patterns for these domains partially overlap, complicating their analysis.

It has long been recognized that sequential assignments can also be made based on backbone scalar connectivities (Llinas et al., 1977). Recently, experiments were proposed for making resonance assignments based on ^{13}C–^{13}C J couplings (Oh et al., 1988) and for obtaining sequential connectivity information from $^1J(C',N)$ couplings (Westler et al., 1988; Stockman et al., 1989; Niemczura et al., 1989). These experiments were performed in a two-dimensional fashion with direct observation of the ^{13}C resonances, resulting in relatively low sensitivity and substantial resonance overlap. Here we demonstrate that these limitations can be overcome by 3D experiments that integrate the J_{CC} and J_{CN} connectivity experiments with the high sensitivity of indirect detection. The new 3D methods yield high-sensitivity spectra almost free of spectral overlap that provide sufficient information for making complete and unequivocal sequential backbone ^1H, ^{13}C, and ^{15}N assignments of calmodulin, without relying on NOE information.

Assignment Approach. Our new assignment approach exploits the relatively large one-bond J couplings between the backbone ^{13}C and ^{15}N nuclei and between the backbone protons and the ^{15}N and C^α nuclei to which they are directly attached. The first 3D experiment (HNCO) correlates NH and ^{15}N chemical shifts of an amino acid with the carbonyl (C′) shift of the preceding residue. The second experiment (HNCA) correlates the NH shift with the ^{15}N shift and the intraresidue C^α shift. In addition, this experiment frequently also provides interresidue connectivity via the two-bond ^{15}N–$^{13}C^\alpha$ coupling. The third experiment (HCACO) correlates intraresidue H^α, C^α, and C′ shifts. The previously published HOHAHA–HMQC experiment (Marion et al., 1989b) correlates H^α with NH and ^{15}N shifts. Finally, a HCA(CO)N relay experiment correlates the H^α shift and the C^α shift via the carbonyl resonance with the ^{15}N shift of the next amino acid residue. These five types of connectivity are schematically represented in Figure 1.

The HNCA and HOHAHA–HMQC experiments firmly establish intraresidue correlations between pairs of ^{15}N–NH and C^α–H^α backbone resonances. Note that these pairs of resonances are linked through two separate and independent

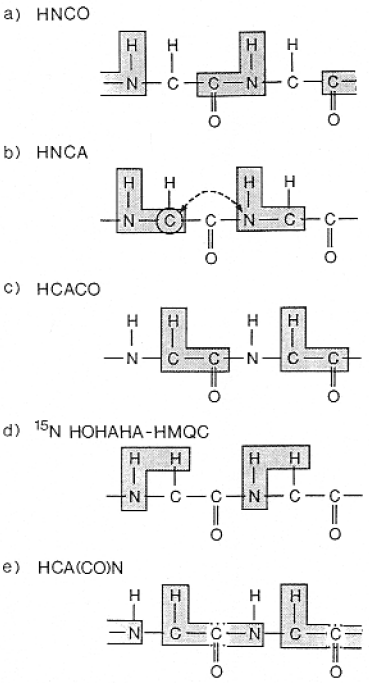

FIGURE 1: Connectivities observed in the five different types of 3D NMR experiments. (a) The HNCO experiment correlates the peptide bond C′ and ^{15}N resonances with the NH chemical shift. (b) The HNCA experiment correlates the NH and ^{15}N shifts with the intraresidue C^α shift. For ~60% of the residues in calmodulin, also a two-bond connectivity to the C^α resonance of the preceding residue is observed (broken line). (c) The HCACO experiment correlates intraresidue H^α, C^α, and C′ chemical shifts. (d) The ^{15}N HOHAHA–HMQC experiment correlates intraresidue ^{15}N, NH, and H^α shifts. (e) The HCA(CO)N experiment correlates the H^α and C^α shifts of one residue with the ^{15}N shift of the next residue.

pathways: via ^1H–^1H J coupling (HOHAHA–HMQC) and via $^1J(N, C^\alpha)$ (HNCA). For almost all residues this provides unambiguous correlation of the NH, H^α, ^{15}N, and C^α shifts, despite moderate overlap present in the ^{15}N–^1H and $^{13}C^\alpha$–^1H 2D correlation spectra.

The HCACO 3D spectrum links the C′ chemical shifts with the intraresidue C^α and H^α shifts. This experiment requires detection of the H^α protons during the t_3 period of the pulse scheme and is therefore most easily conducted in D_2O solution. The HNCO 3D experiment provides crucial sequential connectivity information, linking the ^{15}N–^1H pair of one residue to the C′ shift of the preceding residue. The HNCO experiment requires recording in H_2O. When the absolute shifts of the C′ resonances in H_2O and D_2O solutions are compared, a small isotope shift (~0.08 ppm) has to be taken into account. A more serious problem in comparing these shifts originates from small differences in pH values for the two samples, and from the limited digital resolution in the 3D spectra. In practice we find that, for the vast majority of nonoverlapping carbonyl resonances in the 3D spectra, C′ shifts reproduce to better than ±0.05 ppm. For resonances that show partial overlap, the reproducibility is ~±0.1 ppm. It should be noted, however, that the latter number may improve significantly if deconvolution of overlapping resonances is used instead of simple peak picking of local extrema.

The HCA(CO)N experiment presents a second independent source of sequential connectivity information, by correlating the H^α and C^α shifts of one amino acid with the ^{15}N shift of the next amino acid. This spectrum is recorded in D_2O solution, and in comparing the ^{15}N shifts measured in this experiment with values obtained from spetra recorded in H_2O, an upfield isotope shift of ~0.7 ppm must be taken into ac-

count. After this correction has been applied, ^{15}N chemical shift values measured for identical resonances in the different 3D spectra reproduce to better than ±0.2 ppm for nonoverlapping resonances and to better than ±0.3 ppm for resonances that exhibit partial overlap.

Finally, two-bond $^{13}C^{\alpha}$–^{15}N connectivity observed in the HNCA spectrum provides an additional and independent source of sequential connectivity information. Literature data suggest that this two-bond coupling (~7Hz) is somewhat smaller than the one-bond ^{15}N–$^{13}C^{\alpha}$ coupling (7–11 Hz) (Bystrov, 1976). It is clear from the variation in the intensities of the two-bond correlations in our 3D HNCA spectra that there is a substantial variation in the size of this two-bond coupling. So far, however, we have been unable to correlate the presence or absence of the two-bond $^{13}C^{\alpha}$–^{15}N J connectivity with secondary structure.

Figure 2 shows the pulse sequences of the 3D experiments used. These pulse schemes have been optimized with regard to removal of passive couplings and minimization of the number of required radiofrequency pulses. A detailed treatment of the mechanisms on which these pulse sequences rely and a discussion of their implentation on a commercial spectrometer are presented elsewhere (Kay et al., 1990a). The success of the methods used in the present work relies on the relatively narrow line widths of the backbone ^{15}N and carbonyl resonances, which permit detection of J coupling and connectivity via relatively small $^1J(C', N)$ (~15 Hz), $^1J(C^{\alpha}, N)$ (~10 Hz), and $^2J(C^{\alpha}, N)$ (≲7 Hz) couplings. For example, the HNCA scheme (Figure 1b) uses an INEPT-type transfer (Maudsley & Ernst, 1977; Morris & Freeman, 1979) to transfer ^1H polarization to the ^{15}N. The efficiency of a subsequent HMQC-type correlation (Bax et al., 1983; Bendall et al., 1983) between ^{15}N and C^{α} depends on the size of the $J(C^{\alpha}, N)$ coupling relative to the narrow ^{15}N line width.

The pulse schemes of Figure 2 all rely on multple-step magnetization transfers via well-resolved J couplings. In addition, the number of resonances observed in each 3D spectrum is relatively small, on the same order of magnitude as the total number of amino acids in the protein. As expected (Aue et al., 1978; Griesinger et al., 1989), the sensitivity of such experiments is therefore very high. Because of the moderate spectral dispersion of ^{15}N, C^{α}, and C' resonances (28, 22, and 10 ppm, respectively) the F_1 and F_2 spectral widths that have to be covered in the experiments of Figure 1 are narrow, despite the fact that these experiments are recorded in a nonselective manner. This makes it possible to use relatively small numbers of t_1 and t_2 increments, permitting extensive phase cycling for the removal of undesired coherence transfer pathways (Bodenhausen et al., 1984) while keeping the total measuring time within reasonable limits (1–2 days per 3D spectrum). Moreover, in experiments where the NH proton is detected during t_3 (Figure 2a,b), the t_3 spectral width can be restricted to the spectral width of the amide region of the spectrum. In Figure 2c and d, the spectral window only has to include the H^{α} resonances. As a consequence, the data matrix sizes used for such 3D spectra (before processing and zero filling) are quite small (~2–4 Mword), permitting rapid processing without introducing unusual demands on computer CPU time or disk storage.

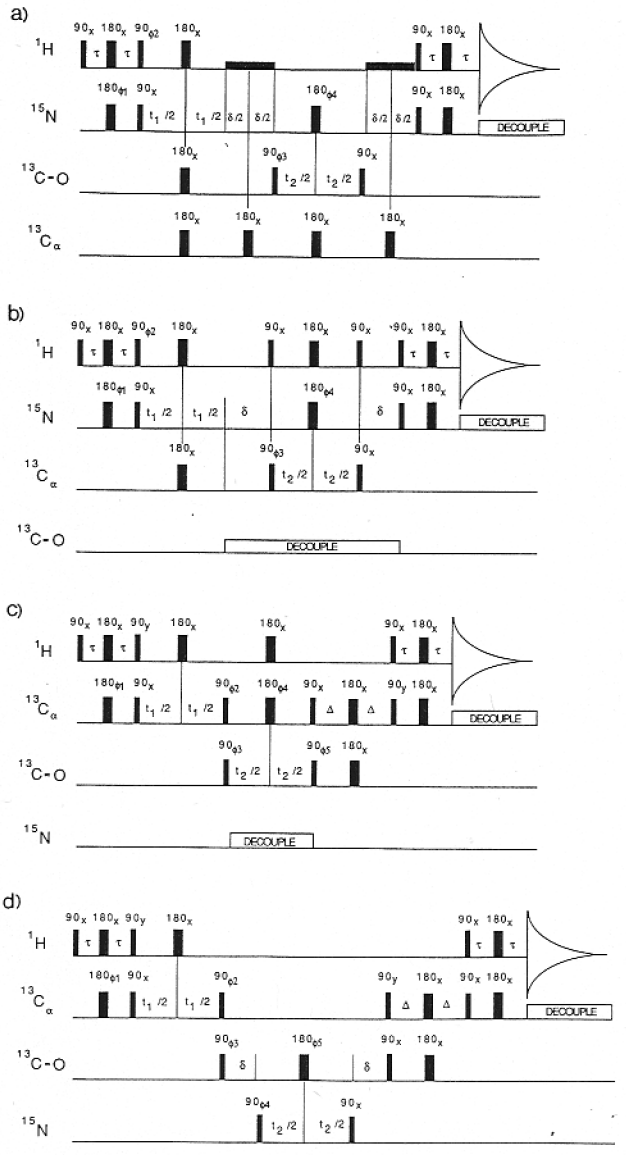

FIGURE 2: Pulse sequences of (a) the HNCO experiment, (b) the HNCA experiment, (c) the HCACO experiment, and (d) the HCA(CO)N experiment. For schemes a and b, H_2O presaturation is used and in addition, in (a) low-power water irradiation is used during the intervals δ (low solid bars). For the HNCO experiment the phase cycling is as follows: $\phi_1 = x, -x$; $\phi_2 = 2(y), 2(-y)$; $\phi_3 = 4(x), 4(-x)$; $\phi_4 = 8(x), 8(y), 8(-x), 8(-y)$; receiver = $2(x), 4(-x), 2(x), 2(-x), 4(x), 2(-x)$. For HNCA, the phase cycling is $\phi_1 = x, -x$; $\phi_2 = y, -y$; $\phi_3 = 2(x), 2(-x)$; $\phi_4 = 4(x), 4(y), 4(-x), 4(-y)$; receiver = $x, -x, -x, x, -x, x, x, -x$. Quadrature in the t_1 and t_2 domains is obtained by changing the phase of the first 90° ^{15}N pulse (HNCO) or the first 90° C^{α} pulse (HNCA) and by changing the phase ϕ_3 by 90°, respectively. For HCACO, the phase cycling used is as follows: $\phi_1 = x, -x$; $\phi_2 = y, -y$; $\phi_3 = 2(x), 2(-x)$; $\phi_4 = 4(x), 4(-x)$; $\phi_5 = 8(x), 8(-x)$; receiver = $2(x, -x, -x, x), 2(-x, x, x, -x)$. For HCA(CO)N, the phase cycling used is as follows: $\phi_1 = x, -x$; $\phi_2 = y, -y$; $\phi_3 = 2(x), 2(-x)$, $\phi_4 = 4(x), 4(-x)$; $\phi_5 = 8(x), 8(y), 8(-x), 8(-y)$; receiver = $x, 2(-x), x, -x, 2(x), 2(-x), 2(x), x, -x, 2(-x), x$. For both the HCACO and HCA(CO)N experiments, quadrature in the t_1 and t_2 domains is obtained by changing the phase of the first 90° C^{α} pulse and by changing the phase ϕ_3 (HCACO) or ϕ_4 (HCA(CO)N) by 90°, respectively.

Experimental Procedures

The *Drosophila* calmodulin gene (Smith et al., 1987) was overexpressed in *Escherichia coli* (strain AR58) using the pAS expression vector (Shatman et al., 1985). The amino acid sequence of *Drosophila* calmodulin is identical with vertebrate calmodulin except for three residues in the C-terminal domain: Y99 → F99, Q143 → T143, and A147 → S147. In addition, the *E. coli* expressed calmodulin lacks a trimethyl group at Lys-115, and an N-terminal acetyl group. Calmodulin labeled uniformly with ^{15}N and ^{13}C was obtained by growing the cells in M9 minimal medium with $^{15}NH_4Cl$ (1 g/L) and D-glucose-$^{13}C_6$ (3 g/L; Merck Isotopes) (Ikura et al., 1990a). No supplemental amino acids were required for growing the

cells. The cells were harvested by centrifugation and the pellets were lysed by hen egg white lysozyme (200 μg/mL). The protein was purified by employing two chromatography steps, phenyl-Sepharose and DEAE-Sephacel (Gopalakrishna & Anserson, 1982; Klee, 1977), combined with the TCA precipitation method (Yazawa et al., 1980). Purity of the protein was checked by SDS–polyacrylamide gel electrophoresis and UV and NMR spectroscopy. Typically, the yield was 15–20 mg from 2.4 L of bacterial culture, sufficient for two NMR samples (one D_2O and one H_2O sample). NMR experiments were performed on a sample of 1.5 mM calmodulin, complexed with Ca^{2+} (6.2 mM), in a 93:7 H_2O to D_2O mixture, pH 6.3, 100 mM KCl, and on a sample of 0.98 mM calmodulin (4.1 mM Ca^{2+}) in 99.9% D_2O, pH 6.3, 100 mM KCl. All experiments were performed on a Bruker AM-500 spectrometer, equipped with a triple-resonance probe optimized for ^1H detection.

All 3D spectra were recorded with sequential quadrature detection during the detection period. In all experiments 512 real t_3 data points were sampled. In the t_1 and t_2 dimensions, complex data were acquired in a hybrid states–TPPI manner (Marion et al., 1989c). Processing of the 3D data sets was accomplished in two steps. First, for processing in the t_1 dimension, routines previously developed in our laboratory (Kay et al., 1989a) were used. Subsequently, commercially available software (NMRi, Syracuse, NY) was used to process data in the t_2 and t_3 dimensions. Doubly phase shifted sine-bell digital filtering (Kay et al., 1989) and zero filling were used in all dimensions and for all data sets. The sine-bell window typically extended from 30 or 60° on the left side of the window, to ~165° on the right side.

For the HNCO experiment (Figure 2a), acquisition times in the t_1 (^{15}N), t_2 (C′), and t_3 (NH) dimensions were 30, 23, and 64 ms, respectively. The fixed delay period, δ, was set to 18 ms. For the HNCA experiment, acquisition times in the t_1 (^{15}N), t_2 (C^α), and t_3 (NH) dimensions were 30, 14.8, and 64 ms, respectively. The fixed delay period, δ, was set to 33 ms. For both the HNCA and the HNCO experiments the delay times, τ, were set to 2.25 ms. For the HNCO experiment, a 32-step phase cycle was repeated four times to obtain complex data in both the t_1 and t_2 dimensions, resulting in 128 scans per complex (t_1, t_2) pair. With a relaxation delay between scans of 1 s, the total measuring time was 42 h and the size of the acquired matrix was (32 complex) × (32 complex) × (512 real). For the HNCA experiment, a 16-step phase cycle was repeated four times to obtain complex data in both the t_1 and t_2 dimensions. The total measuring time was 43 h and the size of the acquired matrix was (32 complex) × (64 complex) × (512 real).

The HCACO 3D spectrum resulted from a (32 complex) × (64 complex) × (512 real) data matrix with acquisition times of 11.5, 46, and 64 ms for the t_1 (C^α), t_2 (C′), and t_3 (H^α) dimension, respectively. With a 16-step phase cycle (64 scans per complex (t_1, t_2) increment) and a 0.75-s relaxation delay, the total measuring time was 32 h. The HCA(CO)N spectrum resulted from a (32 complex) × (32 complex) × (512 real) data matrix with acquisition times of 11.5, 30, and 64 ms for the t_1 (C^α), t_2 (^{15}N), and t_3 (H^α) dimensions, respectively. With a 32-step phase cycle (128 scans per complex (t_1, t_2) increment) and a 0.75-s relaxation delay, the total measuring time was 32 h. The delay δ was set to 18 ms. For both the HCACO and the HCA(CO)N experiments the delay times τ and Δ were set to 1.5 and 3 ms, respectively.

The HOHAHA–HMQC 3D spectrum was recorded for a sample uniformly labeled with ^{15}N, without ^{13}C enrichment. Further experimental details can be found elsewhere (Marion et al., 1989b; Ikura et al., 1990a).

Results and Discussion

The 3D experiments sketched in Figure 2 can be recorded as 2D experiments, yielding the more familiar looking 2D correlation spectra. Indeed, we have recorded such spectra (supplementary material) but, unfortunately, for calmodulin resonance overlap in these 2D spectra is too severe to allow extensive resonance assignments. Nevertheless, these 2D spectra give a good impression of the chemical shift distribution of NH, H^α, C′, and C^α resonances. In addition, recording of a 2D spectrum just prior to recording the corresponding 3D spectrum provides a rapid method for verifying the correct experimental hardware setup.

Although recording and processing of 3D spectra do not present any particular difficulty once appropriate hard- and software have been developed, display of 3D spectra remains a practical problem. Here we follow the commonly used format of illustrating the use of the 3D methods by selecting 2D slices taken from the 3D spectra. Figure 3 shows small sections of the slices that are relevant for making the sequential connectivities involving residues Lys-21 and Asp-22. The solid and dotted lines in Figure 3 outline the pattern of correlations that has to be analyzed.

Figure 3A is a region of the (C′, NH) slice taken from the HNCO spectrum at an ^{15}N frequency of 117.4 ppm, showing a correlation between the NH of Lys-21 and the C′ frequency of Asp-20. Figure 3B is part of the corresponding slice of the HNCA spectrum, identifying the Lys-21 C^α chemical shift (58.5 ppm). The corresponding slice from the HOHAHA–HMQC spectrum (Figure 3C) identifies the H^α shift (4.01 ppm) of Lys-21. Note that this slice is not taken at exactly the same ^{15}N frequency because the digitization in the ^{15}N dimension of the HOHAHA–HMQC spectrum is slightly different from that in the HNCA and HNCO spectra.

Figure 3D is part of the (H^α, C′) slice of the HCACO spectrum, taken at a C^α shift nearest to 58.5 ppm, identifying the C′ chemical shift (178.29 ppm) of K21. A slice taken at the same C^α frequency from the HCA(CO)N spectrum (Figure 3E) identifies the ^{15}N shift (114.1 ppm) of the next amino acid, Asp-22. A slice from the HNCO spectrum (Figure 3F) taken at 114.1 ppm confirms the connectivity by identifying the previously assigned (Figure 3D) C′ shift of Lys-21 at 178.24 ppm. The assignment process then continues in a similar manner, identifying the C^α shift of Asp-22 (Figure 3G) and the H^α shift of Asp-22 (Figure 3H). Note that the slice of Figure 3G also shows a weak two-bond connectivity between the Asp-22 amide nitrogen and the C^α carbon of Lys-21, providing yet another confirmation of the connectivity between Lys-21 and Asp-22.

The broken lines in Figure 3 represent fragments of connectivity patterns observed for other amino acids. For example, parts A–C of Figure 3 also show the connectivity patterns for Leu-116, yielding the Lys-115 C′ shift (Figure 3A), the Leu-116 and Lys-115 C^α shifts (Figure 3B), and the Leu-116 H^α shift (Figure 3C).

In the conventional sequential assignment of homonuclear 2D NMR spectra, ambiguities occur if there is spectral overlap in the amide region or in the $C^\alpha H$ or $C^\beta H$ region of the 1D ^1H NMR spectrum. With the present procedure, analogous ambiguities are much rarer since they only occur for cases where ^{15}N–^1H or C^α–H^α correlations in the 2D spectra overlap with other such correlations. For amides, such overlap occurs when two amides have ^{15}N chemical shifts that differ by less than 0.2 ppm and NH shifts that differ by less than 0.02 ppm.

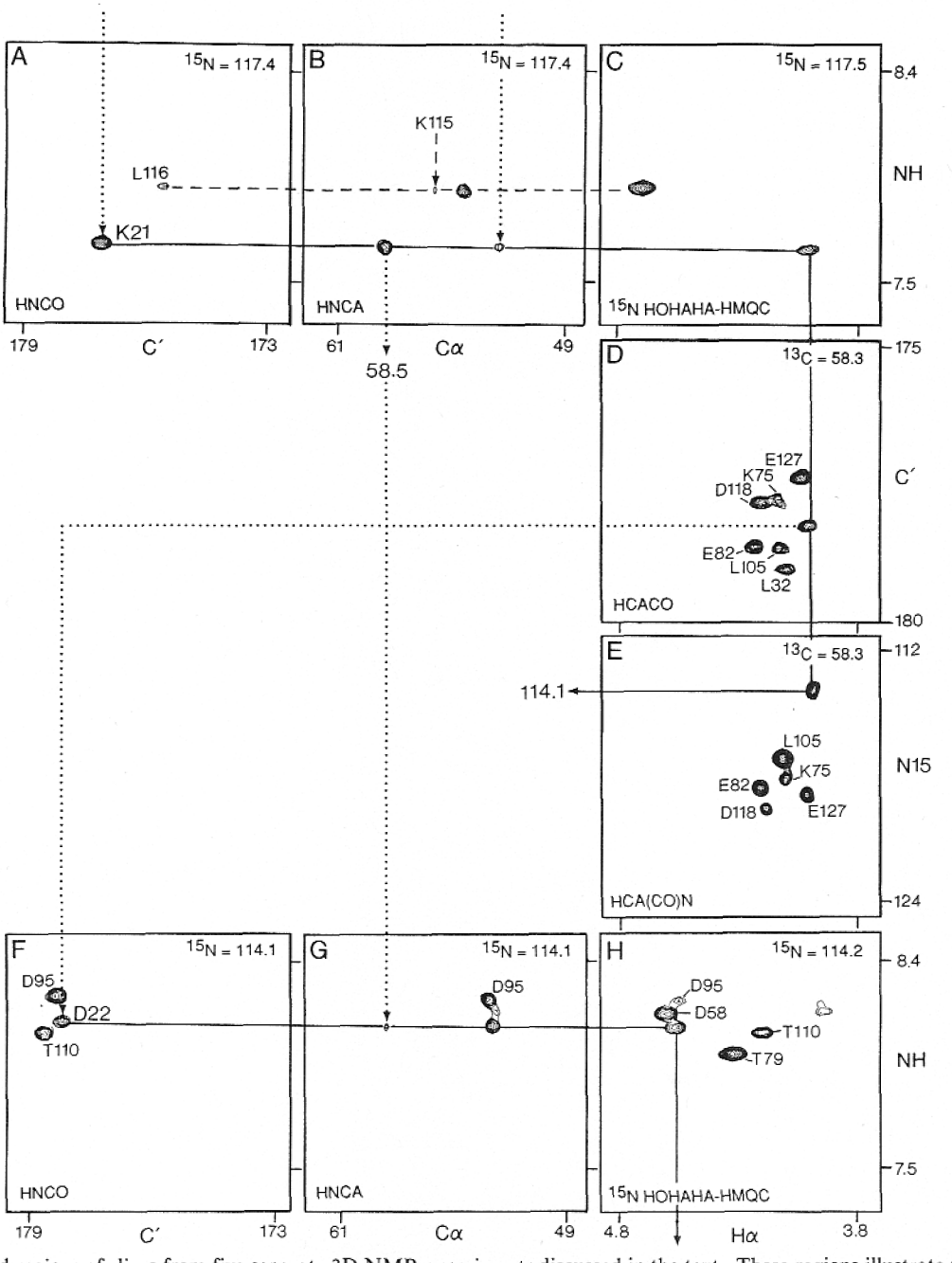

FIGURE 3: Selected regions of slices from five separate 3D NMR experiments discussed in the text. These regions illustrate the *J* correlation between Lys-21 and Asp-22. Solid and dotted lines trace the connectivity patterns for these two residues. Broken lines correspond to parts of the connectivity patterns observed for other residues. Slices A–C are taken at the Lys-21 ^{15}N chemical shift. Slices D and E are taken at the Lys-21 C$^{\alpha}$ shift, observed in B. Slices F–H are taken at the ^{15}N frequency of Asp-22, as measured in E. The analysis of the connectivity patterns is discussed in the text. No base-line correction or any other cosmetic procedures were used for any of the 3D spectra.

In calmodulin there are 19 residues that, according to this definition, do not have a unique ^{15}N–^{1}H correlation. For C$^{\alpha}$–H$^{\alpha}$ correlations, using a similar definition of overlap (H$^{\alpha}$ shifts differing by less than 0.02 ppm and C$^{\alpha}$ shifts by less than 0.2 ppm), there are 40 such nonunique H$^{\alpha}$–C$^{\alpha}$ correlations, and this type of 2-fold chemical shift degeneracy constitutes the most common source of ambiguity in the sequential assignment process.

Finally, ambiguities in the sequential linking process can also occur if both the C′ and ^{15}N chemical shifts of one particular peptide bond overlap with those of a second peptide bond. There are seven such cases in calmodulin. However, for all of these cases a two-bond ^{2}J(N, C$^{\alpha}$) connection was observed, eliminating any ambiguity in the sequential connectivity.

There are only two proline residues in calmodulin, Pro-43 and Pro-66. Connecting proline C$^{\alpha}$, H$^{\alpha}$, and C′ shifts to its succeeding residue does not constitute any unusual problems and is conducted in the same manner as for all other residues. The Pro ^{15}N shift can be identified from the correlation with the preceding residue in the HCA(CO)N spectrum. However, because no HNCA correlation can be obtained for proline, determination of connectivity to the preceding residue is not possible.

Semiautomated Assignment Strategy. The actual assignment process is largely done by several small computer programs, as schematically shown in Figure 4. In the first step of analysis, a peak-picking program determines resonance positions in all five 3D spectra and in strongly resolution enhanced 2D ^{1}H–^{15}N and ^{1}H$^{\alpha}$–^{13}C$^{\alpha}$ correlation spectra, pro-

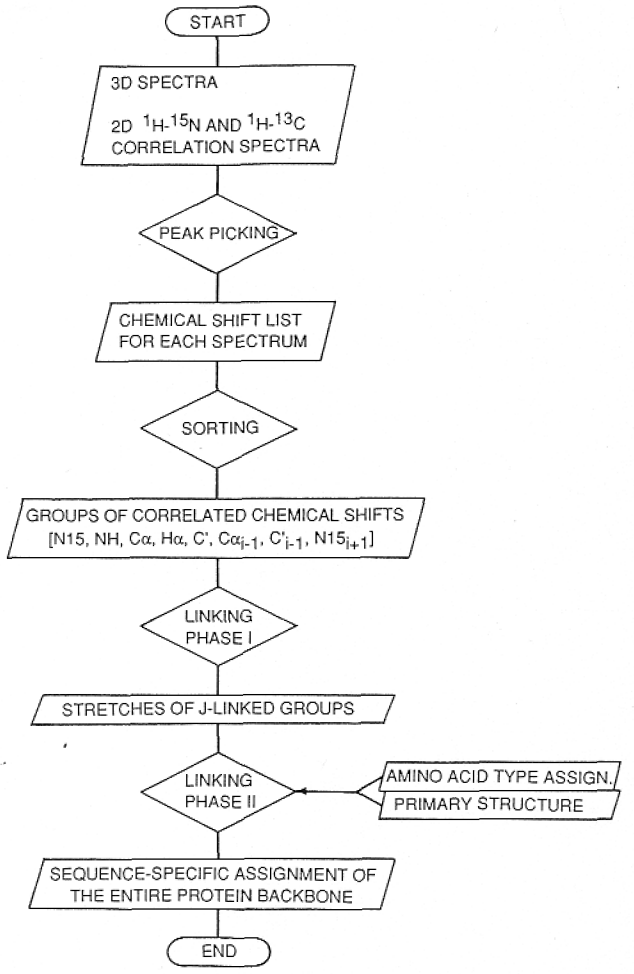

FIGURE 4: Flow diagram of the semiautomated assignment procedure.

ducing a list of chemical shifts for each data set. A sorting program then generates all possible groups of correlated chemical shifts, i.e., the ^{15}N, NH, C^{α}, H^{α}, and C' shifts of residue i, the C^{α}_{i-1} (if observed) and C'_{i-1} shifts of the preceding residue, and the $^{15}N_{i+1}$ shift of the next residue.

In the first stage of the linking phase, a third program searches for groups of chemical shifts (corresponding to individual amino acids) that are linked unequivocally. Thus, residue m and n are adjacent and unique if there is only one pair of residues for which $\delta(^{15}N_{i+1})^m = \delta(^{15}N)^n$, $\delta(C^{\alpha})^m = \delta(C^{\alpha}_{i-1})^n$, and $\delta(C')^m = \delta(C'_{i-1})^n$. In this way, stretches of J-linked amino acids are generated. For the first and last amino acid of each such backbone fragment there are multiple possibilities for connecting another amino acid.

At this stage, in addition to knowledge of the primary structure, knowledge of the residue types of some of the linked amino acids is required for putting the stretches in the correct order. This final step in the linking procedure proceeds in much the same manner as in linking the various NOE-connected fragments in regular homonuclear ^{1}H NMR studies (Wüthrich, 1986). However, because the average number of unequivocally J-linked amino acids is typically much larger than the fragment sizes obtained in homonuclear NOE studies, the number of amino acids for which the residue type must be known can be quite small. For example, if for any of the J-linked fragments two amino acids can be identified, for example, two Ala residues that are separated by five peptide bonds, this makes it possible to position this fragment along the polypeptide backbone, provided that the Ala-X-X-X-X-Ala sequence only occurs once in the protein. At either end of this

Table I: The Number of Sequential Connectivity Possibilities for Dipeptide Sequences in the Third Calcium Binding Domain of Calmodulin

dipeptide	C'^{a}	$C' + N^{b}$	$C' + N + C^{\alpha c}$	$d_{NN} + d_{\alpha N}{}^{d}$
86–87	10	1		2
87–88	6	1		3
88–89	8	1	1	2
89–90	4	1	1	2
90–91	5	1	1	1
91–92	6	1		1
92–93	4	1	1	4
93–94	7	1	1	6
94–95	5	2		1
95–96	12	3	2	1
96–97	3	1		2
97–98	4	1	1	1
98–99	1	1		2
99–100	2	1	1	4
100–101	5	1		3
101–102	6	1		2
102–103	5	1		2
103–104	2	1		2
104–105	2	1	1	
105–106	7	1	1	
106–107	10	1		3
107–108	8	1	1	7
108–109	9	2	1	1
109–110	7	1	1	
110–111	6	1		1
111–112	4	1		

[a] Number of sequential connectivities possible when comparing C' chemical shifts, assuming error margins of ±0.02 ppm for the NH proton and ±0.1 ppm for the C' ^{13}C chemical shift. [b] Number of sequential connectivities possible when comparing both C' and ^{15}N shifts. The ^{15}N shift error margin is assumed to be ±0.2 ppm. [c] Number of sequential connectivities possible when, in addition to the C' and ^{15}N shifts, the $^{2}J(N, C^{\alpha})$ connectity is observed. $^{13}C^{\alpha}$ shift error margin is assumed to be ±0.2 ppm. [d] Number of sequential connectivities possible based on NH–NH and H^{α}–NH NOE connectivities as observed in the ^{15}N 3D NOE–HMQC spectrum after the pertinent sequential NOE cross peaks were identified. Error margins for both NH and H^{α} are assumed to be ±0.02 ppm.

fragment, which now has been positioned along the polypeptide backbone, there are two or sometimes three possibilities for connecting a second fragment. However, even if only a single amino acid type is known in one of the connecting fragments, this usually is sufficient for connecting the two fragments; the unconnecting fragment would not have the identified amino acid in a position that would align with the known amino acid sequence of the protein. This process is repeated until all fragments have been connected and the polypeptide backbone assignment has been completed. For calmodulin, we previously have obtained a large number of residue-type assignments by selective isotope labeling (Ikura et al., 1990a), and positioning the fragments along the backbone was straightforward, even without using a computer program.

Table I shows the number of possible interresidue connections that are found during the first linking phase for amino acids of the third calcium binding region of calmodulin. Note that the position of these residues (86–112) is not known at this stage of the assignment process. The second column indicates that up to a dozen sequential connectivities may be possible if only the C' chemical shift value is used for linking two amino acids. However, if both the ^{15}N and the C' shifts are used, for most dipeptide fragments there is only one possibility for sequential assignment (third column of Table I). In addition, for ~50% of the residues, a two-bond connectivity is observed between ^{15}N and the $^{13}C^{\alpha}$ of the preceding residue. As can be seen in the fourth column of Table I, only two dipeptides remain with ambiguous linkage. Chemical shifts of the three residues involved (later assigned to Asp-93,

Table II: Polypeptide Backbone ^1H, ^{13}C, and ^{15}N Chemical Shifts for *Drosophila* Calmodulin (1.5 mM) at pH 6.3 and 47 °C, 6.1 mM $CaCl_2$ and 0.1 M KCl[a]

	^{15}N	C^α	C′	NH	H^α		^{15}N	C^α	C′	NH	H^α		^{15}N	C^α	C′	NH	H^α
A1		51.9	174.0		4.15	M51	119.3	59.5	178.8	7.84	4.09	I100	127.3	60.5	175.8	10.15	4.87
D2	120.5	54.7	175.8		4.67	I52	118.0	65.0	177.9	7.74	3.57	S101	123.8	56.0	175.5	9.01	4.90
Q3	119.6	55.7	175.8	8.24	4.42	N53	117.5	56.0	177.5	8.47	4.42	A102	123.0	56.1	179.4	9.16	3.97
L4	122.8	54.6	177.7	8.16	4.70	E54	116.4	59.0	177.8	7.58	4.09	A103	118.1	55.3	181.3	8.15	4.08
T5	112.9	60.7	175.8	8.56	4.52	V55	108.9	61.1	176.0	7.22	4.47	E104	119.4	59.4	179.9	7.79	4.09
E6	120.5	60.2	179.4	8.93	4.03	D56	121.6	54.0	176.4	7.72	4.67	L105	120.8	58.5	178.7	8.53	4.16
E7	119.3	60.0	179.1	8.63	4.12	A57	131.7	54.3	178.8	8.47	4.25	R106	117.3	60.0	178.8	8.51	3.85
Q8	119.8	58.7	178.3	7.67	3.93	D58	113.9	52.8	178.1	8.18	4.67	H107	118.6	59.5	177.4	7.91	4.38
I9	119.4	66.1	177.9	8.31	3.77	G59	108.2	47.4	175.2	7.55	3.95/3.85	V108	118.7	66.2	178.1	7.93	3.56
A10	121.3	55.6	180.9	7.97	4.15	N60	118.4	52.7	177.0	8.07	4.67	M109	116.5	57.8	178.7	8.18	4.37
E11	119.1	55.4	180.2	7.78	4.17	G61	113.3	45.8	173.6	10.51	4.22/3.55	T110	114.4	65.8	177.6	8.10	4.22
F12	119.6	59.5	178.8	8.52	5.01	T62	109.0	59.7	173.5	7.68	4.79	N111	121.9	55.7	176.5	7.85	4.54
K13	123.1	60.3	179.4	9.09	4.06	I63	124.0	60.0	175.6	8.90	5.19	L112	118.7	55.6	177.5	7.78	4.37
E14	120.2	59.5	179.5	7.82	4.20	D64	128.3	52.4	176.7	8.87	5.39	G113	106.3	45.6	174.6	7.79	4.27/3.78
A15	122.2	55.4	179.0	8.01	4.28	F65	119.0	63.7	173.8	8.90	4.04	E114	120.3	55.5	175.8	7.90	4.46
F16	119.0	62.1	177.4	8.73	3.31	P66	117.2	66.7	180.1		3.95	K115	123.5	55.8	175.7	8.41	4.41
S17	112.4	61.6	175.2	7.90	4.18	E67	117.4	58.8	178.8	7.80	4.15	L116	124.2	54.4	178.0	7.92	4.78
L18	120.8	57.3	178.0	7.45	4.04	F68	123.3	61.3	177.1	8.76	4.00	T117	114.0	60.8	175.6	8.91	4.52
F19	115.0	59.3	176.6	7.29	4.29	L69	118.8	58.1	179.1	8.31	3.39	D118	122.5	58.5	177.9	8.44	4.22
D20	117.5	52.5	177.3	7.74	4.52	T70	114.9	66.5	176.2	7.47	3.83	E119	119.8	59.5	178.6	8.27	4.11
K21	124.4	58.6	178.3	7.66	4.01	M71	121.4	59.0	178.1	7.75	3.84	E120	118.7	59.5	179.9	8.15	4.19
D22	114.0	53.0	177.9	8.10	4.62	M72	116.4	56.0	178.4	7.98	4.06	V121	120.8	67.0	177.4	8.02	3.68
G23	109.1	47.4	175.4	7.65	3.94/3.94	A73	121.8	54.5	179.6	8.08	4.13	D122	119.6	57.8	179.2	8.02	4.38
D24	120.8	53.8	177.7	8.37	4.54	R74	116.5	58.1	177.8	7.52	4.15	E123	119.2	59.2	178.1	7.89	4.05
G25	113.1	45.6	174.0	10.57	4.38/3.74	K75	118.7	57.1	177.6	7.71	4.28	M124	119.4	59.5	179.1	7.80	4.06
T26	112.9	60.1	173.2	8.14	5.31	M76	118.4	56.5	176.5	7.85	4.42	I125	118.2	64.2	177.3	7.92	3.55
I27	127.1	60.5	176.3	9.71	5.00	K77	120.3	60.2	176.5	7.81	4.37	R126	118.4	59.9	179.3	8.13	4.05
T28	116.6	59.8	176.8	8.53	4.86	D78	121.4	54.6	176.7	8.21	4.74	E127	115.8	58.5	177.4	7.87	4.05
T29	112.8	66.8	177.2	8.99	3.83	T79	114.0	62.3	174.7	8.01	4.37	A128	119.1	52.3	177.8	7.37	4.45
K30	120.4	59.3	179.9	7.65	4.22	D80	122.8	54.8	176.9	8.35	4.75	N129	117.3	54.2	176.4	7.83	4.57
E31	121.3	59.4	179.2	7.60	4.10	S81	116.7	59.9	175.7	8.30	4.50	I130	127.4	63.2	178.0	8.50	4.03
L32	120.4	58.2	179.1	8.64	4.13	E82	120.9	58.2	178.7	8.76	4.26	D131	117.1	53.8	178.5	8.27	4.61
G33	105.3	48.4	175.5	8.59	4.00/3.62	E83	118.8	60.0	179.2	8.63	4.13	G132	108.3	47.6	175.5	7.57	4.01/3.85
T34	117.8	67.0	177.2	7.85	3.99	E84	120.2	59.3	179.4	7.68	4.06	D133	120.6	53.8	177.8	8.34	4.51
V35	122.2	66.5	179.1	7.64	3.50	I85	121.3	64.5	178.0	7.93	4.07	G134	112.8	46.0	173.2	10.27	4.07/3.47
M36	117.9	59.1	179.2	8.38	4.16	R86	121.6	60.2	179.3	8.32	4.20	Q135	115.4	53.3	175.0	7.96	4.88
R37	118.9	59.2	181.1	8.38	4.81	E87	118.4	59.1	178.7	8.07	4.18	V136	125.2	61.8	176.1	9.12	5.24
S38	118.4	61.8	174.9	7.93	4.42	A88	121.8	55.2	179.2	7.93	4.23	N137	129.0	51.4	175.2	9.62	5.28
L39	120.1	54.7	177.6	7.38	4.52	F89	118.5	62.2	176.7	8.47	3.23	Y138	118.2	62.9	176.3	8.36	3.50
G40	106.4	45.7	174.7	7.82	4.24/3.81	R90	115.4	58.9	178.3	7.64	3.94	E139	118.2	60.5	180.5	8.00	3.72
Q41	118.4	54.6	174.4	7.82	4.53	V91	118.1	65.7	177.4	7.52	3.56	E140	120.1	58.7	179.2	8.73	4.08
N42	116.5	51.4	172.3	8.59	5.20	F92	116.4	60.0	177.0	7.47	4.25	F141	124.0	61.6	177.0	8.86	4.05
P43	115.6	62.7	177.9		4.80	D93	116.9	52.4	177.6	7.74	4.59	V142	119.4	67.2	179.6	8.48	3.21
T44	112.6	60.7	175.4	8.60	4.50	K94	125.8	59.0	178.4	7.66	3.97	T143	116.5	66.7	176.1	7.67	3.81
E45	120.7	60.3	179.0	8.74	4.02	D95	114.2	53.2	177.9	8.24	4.62	M144	121.8	58.7	178.0	7.93	4.13
A46	120.5	55.2	180.2	8.16	4.14	G96	109.1	47.3	175.4	7.76	3.86/3.82	M145	114.8	55.7	177.8	7.89	4.34
E47	118.6	59.2	180.1	7.63	4.08	N97	119.4	52.8	176.2	8.32	4.68	T146	110.0	62.5	175.1	7.64	4.45
L48	120.2	58.0	178.6	8.11	4.12	G98	112.8	45.3	172.8	10.59	4.04/3.44	S147	117.8	59.1	173.7	7.74	4.52
Q49	118.1	58.8	178.6	8.13	3.89	F99	115.6	56.0	174.9	7.64	5.19	K148	127.7	57.7	181.2	7.65	4.21
D50	120.0	57.7	178.8	8.01	4.46												

[a] ^1H Chemical shifts are expressed relative to (trimethylsilyl)propionic-d_4 acid, ^{15}N shifts relative to liquid NH_3 (Live et al., 1984), and ^{13}C shifts relative to hypothetical internal (trimethylsilyl)propionic acid (Bax & Subramanian, 1986).

Lys-94, and Asp-95) overlap with those of three other residues, later assigned to the homologous region of the first calcium binding loop (Asp-20, Lys-21, and Asp-22). Visual inspection of the pertinent regions of the 3D spectra often suggests the solution to such an ambiguity. For example, although the difference in C′ shifts of Asp-95 and Asp-22 is too small for the search program to make an unambiguous distinction, inspection of Figure 3F suggests that Asp-95 does not correlate with Lys-21. Moreover, Asp-22 has been unequivocally correlated with Lys-21 via the ^2J(N, C) correlation in Figure 3G, excluding the possibility of Lys-21 to Asp-95 connectivity. Thus, the numbers presented in Table I are upper estimates for the number of possible sequential connectivities, before any additional reasoning or visual inspection of the 3D spectra is used.

The last column of Table I shows the number of sequential assignment possibilities that remain open if the conventional NOE connectivity method is used, with data obtained from the NOESY–HMQC and HOHAHA–HMQC 3D spectra. In preparing the number of linking possibilities shown in this column, it is assumed that the correct sequential NOE cross peaks has already been identified, although in practice we were only able to do this after the 3D triple-resonance spectra revealed the correct assignments. Nevertheless, even in this case, only a relatively small number of dipeptide and tripeptide sequences can be identified unambiguously and for the majority of the peptide linkages no unique assignments can be made.

Positioning the stretch of 29 amino acids presented in Table I at the correct position along the backbone can be done on the basis of the multitude of Ala and Gly residues in this sequence for which the residue type is apparent from the HOHAHA–HMQC spectrum or from selective labeling experiments. Moreover, the resolved resonances of Gly-98 and Ile-100 were assigned previously (Ikura et al., 1985).

The H^α, NH, ^{15}N, C^α, and C′ chemical shifts of calmodulin

are summarized in Table II. These assignments are in general agreement with partial (~30%) assignments published for the C-terminal domain of calmodulin (Ikura et al., 1985) and the assignments of eight NH protons for the N-terminal domain (Ikura et al., 1987). Differences were found for Asp-97 H^{α} (misassignment) and NH protons of Leu-116, Thr-117, and Asp-118. The previous H^{α} assignments for Leu-116, Thr-117, and Asp-118 agree with our present results, however. Since the C-terminal domain used in the previous study is a tryptic fragment of natural scallop testis calmodulin, it has a trimethyllysine residue in position 115, whereas *E. coli* expressed calmodulin has a regular lysine in this position. The differences in NH chemical shift for residues 116–118 may therefore be attributed to the differences in the primary sequence of the protein.

For intact bovine calmodulin, a limited number (~18%) of backbone proton resonances were reported recently (Seeholzer & Wand, 1989). Since no data were reported for residues 104–132, we are unable to compare the chemical shifts we observed in the vicinity of Lys-115 with those of natural calmodulin. Comparison of chemical shifts for bovine calmodulin with our present results shows generally good agreement, with the exception of the H^{α} resonances of Thr-29, Gly-132, and Tyr-138, and the NH resonance of Lys-30. Since these residues are not close to any of the residues that differ between the two types of calmodulin, these differences most likely must be attributed to misassignments caused by the limitations of the conventional 2D NMR techniques for studying proteins with the spectral complexity of calmodulin.

Concluding Remarks. We have demonstrated that our new approach can yield complete backbone resonance assignments for proteins of considerably complexity, without recourse to the sometimes ambiguous NOE connectivity information. Comparison of the chemical shift values shown in Figure 1 with the secondary structure as determine by X-ray crystallography (Babu et al., 1985) suggests that carbonyl resonances that are part of α-helical segments are shifted toward higher frequency (downfield) relative to nonhelical residues. However, to obtain more rigorous information about the secondary structure of the protein in solution, interpretaton of 2D or 3D NOESY spectra is required. With complete and unambiguous assignments of the backbone resonances in hand, this is relatively straightforward. Moreover, recently it also has become feasible to make complete side-chain assignments for proteins of the complexity of calmodulin (Fesik et al., 1990; Kay et al., 1990b; Bax et al., 1990a,b; Ikura et al., 1990b; Zuiderweg et al., 1990). This latter methodology requires knowledge of the C^{α} chemical shift assignments, provided by the HNCA experiment discussed above. Using this information, we are currently in the process of determining a detailed calmodulin solution structure.

As discussed above, analysis of the 3D spectra can be automated more easily than the analysis of 2D NMR spectra since overlap is much less of a problem. We have analyzed our data only in a very coarse manner, using Fourier transformations of truncated signals that resulted in poorly digitized spectra. It is expected that better spectra can be obtained in even shorter measurement times if more sophisticated data-processing algorithms are used that are based on linear predictions or maximum entropy methods.

The sequential assignment approach outlined above provides a very powerful and direct method for analyzing the NMR spectra of proteins that can be isotopically enriched with ^{15}N and ^{13}C. The sensitivity of the new pulse schemes presented is sufficient to permit the recording of a 3D spectrum with a high signal-to-noise ratio in less than 2 days for 1 mM protein sample concentrations. However, it is expected that for the HNCA experiment sensitivity will become a limiting factor if the protein size exceeds 25 kDa. For these large proteins, the ^{15}N line width becomes greater than the $^{1}J(N, C^{\alpha})$ scalar coupling, decreasing the efficiency of the $N-C^{\alpha}$ J-correlation step of the HNCA experiment. The HCA(CO)N experiment, which includes an extra relay step, also suffers significantly in sensitivity when molecular weight is increased. However, on the basis of our experience with calmodulin, we expect that sequential assignment of proteins smaller than 200 amino acids will be quite straightforward with the methodology outlined in this paper.

Acknowledgments

We thank Claude Klee for stimulating discussions and continuous encouragement, Marie Krinks for preparing the calmodulin used in this study, and Kathy Beckingham and John F. Maune for providing us with the *Drosophila* calmodulin coding construct.

Supplementary Material Available

Five 2D correlation spectra showing $NH-C^{\alpha}$, $NH-C'(i-1)$, $H^{\alpha}-C'$, $H^{\alpha}-N(i+1)$, and $NH-H^{\alpha}$ correlations (6 pages). Ordering information is given on any current masthead page.

References

Aue, W. P., Bachmann, P., Wokaun, A., & Ernst, R. R. (1978) *J. Magn. Reson. 29*, 523–534.

Babu, Y. S., Sack, J. S., Greenbough, T. J. Bugg, C. E., Means, A. R., & Cook, W. J. (1985) *Nature (London) 315*, 37–40.

Bax, A. (1988) *Isr. J. Chem. 28*, 309–317.

Bax, A. (1989) *Annu. Rev. Biochem. 58*, 223–256.

Bax, A., & Subramanian, S. (1986) *J. Magn. Reson. 67*, 565–569.

Bax, A., Griffey, R. H., & Hawkins, B. L. (1983) *J. Magn. Reson. 55*, 301–315.

Bax, A., Clore, G. M., Driscoll, P. C., Gronenborn A. M., Ikura, M., & Kay, L. E. (1990a) *J. Magn. Reson.* (in press).

Bax, A., Clore, G. M., & Gronenborn, A. M. (1990b) *J. Magn. Reson.* (submitted).

Bendall, M. R., Pegg, D. T., & Doddrell, D. M. (1983) *J. Magn. Reson. 52*, 81–117.

Bodenhausen, G., Kogler, H., & Ernst, R. R. (1984) *J. Magn. Reson. 58*, 370–388.

Bystrov, V. F. (1976) *Prog. NMR Spectrosc. 10*, 41–81.

Clore, G. M., & Gronenborn, A. M. (1989) *CRC Crit. Rev. Biochem. Biol. 24*, 479–564.

Fesik, S. W., & Zuiderweg, E. R. P. (1988) *J. Magn. Reson. 78*, 588–593.

Fesik, S. W., Luly, J. R., Erickson, J. W., & Abad-Zapatero, C. (1988) *Biochemistry 27*, 8297–8301.

Fesik, S. W., Eaton, H. L., Olejniczak, E. T., Zuiderweg, E. R. P., McIntosh, L. P. & Dahlquist, F. W. (1990) *J. Am. Chem. Soc. 112*, 886–888.

Gopalakrishna, R., & Anderson, W. B. (1982) *Biochem. Biophys. Res. Commun. 104*, 830–836.

Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870–7872.

Griesinger, C., Sørenson, O. W., Ernst, R. R. (1989) *J. Magn. Reson. 84*, 14–63.

Ikura, M., Minowa, O., & Hikichi, K. (1985) *Biochemistry 24*, 4264–4269.

Ikura, M., Minowa, O., Yazawa, M., Yagi, K., & Hikichi, K. (1987) *FEBS Lett. 219*, 17–21.