Volume 266, number 1,2, 155-158 FEBS 08521 June 1990

An efficient NMR approach for obtaining sequence-specific resonance assignments of larger proteins based on multiple isotopic labeling

Mitsuhiko Ikura[1], Marie Krinks[2], Dennis A. Torchia[3] and Ad Bax[1]

[1]*Laboratory of Chemical Physics, NIDDK,* [2]*Laboratory of Biochemistry, NCI* and [3]*Bone Research Branch, NIDR,*
National Institutes of Health, Bethesda, MD 20892, USA

Received 13 March 1990

By simultaneously incorporating in a protein ^{13}C-carbonyl- and ^{15}N-labeled amino acids with different levels of enrichment, characteristic asymmetric doublet-like patterns are observed for ^{15}N nuclei that are directly adjacent to the ${}^{13}C_1$-labeled residues, providing unambiguous identification of a large number of unique dipeptide fragments of the protein. Additional assignments and qualitative structural information can be obtained from such a selectively labeled protein by recording multiple bond correlation spectra. The procedure is demonstrated for the protein calmodulin, complexed with calcium.

Two-dimensional NMR; Isotopic labeling; Sequential assignment; Calmodulin; Central helix

1. INTRODUCTION

An unambiguous but labor intensive approach for making sequence specific resonance assignments in larger proteins relies on incorporation of isotope labeled amino acids. With this procedure, a unique dipeptide fragment in the protein sequence, say Leu-Val, can be identified by incorporating in the protein [^{15}N]Val and L-[^{13}C]Leu [1,2]. In a ^{1}H-^{15}N shift correlation spectrum, the amide ^{15}N of this peptide bond can be identified uniquely by the presence of a ${}^{1}J_{CN}$ splitting [9]. To obtain a sufficient number of such unambiguous amide identifications, preparation of a large number of protein samples with different amino acid labels is required [3]. Moreover, separate 2D NMR spectra must be recorded for each of these samples.

In this paper, we demonstrate a modified selective labeling procedure that yields a large number of sequence-specific assignments from a single protein preparation. This modified procedure relies on simultaneous labeling of the protein with a substantial number of amino acids at a variety of ^{15}N and ^{13}C enrichment levels. The intensity of a ^{15}N ^{1}H correlation identifies the different types of ^{15}N-labeled amino acids. As will be demonstrated, the type of amino acid preceding the ^{15}N-labeled amino acid can be determined from the J splitting pattern observed for ^{15}N. In addition, the multiple labeled protein sample can be used for ^{1}H-^{15}N and ^{1}H-^{13}C heteronuclear multiple-bond correlation (HMBC) experiments [4], providing NH, Hα, and C_1 assignments.

Correspondence address: M. Ikura or A. Bax, Building 2, Room 109, NIH, Bethesda, MD 20892, USA

2. MATERIALS AND METHODS

The *Drosophila* calmodulin gene [5], kindly provided to us by Dr Kathy Beckingham, was overexpressed in *Escherichia coli* strain AR58, using the pAS expression system. Details regarding the expression and purification of calmodulin are presented elsewhere [6]. A mixture of the following labeled and unlabeled amino acids was used in the growth medium: L-Lys (99% ^{15}N) 60 mg/l, L-Ile (95% ^{15}N) 25 mg/l, L-Ile (99% ${}^{13}C_1$) 30 mg/l, L-Ile (unlabeled) 5 mg/l, L-Arg (99% ^{15}N) 10 mg/l, L-Arg (99% ${}^{13}C_1$) 40 mg/l, L-Arg (unlabeled) 10 mg/l, L-Phe (99% ${}^{13}C_1$) 60 mg/l, L-Met (99% ${}^{13}C_1$) 40 mg/l, L-Met (unlabelled) 14 mg/l, L-Leu (99% ${}^{13}C_1$) 40 mg/l, L-Leu (unlabeled) 20 mg/l, L-Thr (99% ${}^{13}C_1$) 30 mg/l, L-Thr (unlabeled) 30 mg/l. Starting from 2.4 l of the final cell cuture, 16 mg of purified protein was obtained.

One-bond ^{1}H-^{15}N shift correlation spectra were recorded at 600 MHz on a Bruker AM-600 spectrometer, using the decoupled-Overbodenhausen [7,8] experiment. Acquisition times were 102 ms (t_2) and 205 ms (t_1). The total measuring time was 16 h. Heteronuclear multiple bond correlation (HMBC) spectra for correlation between amide protons and carbonyl carbon resonances were also recorded on the Bruker AM-600 spectrometer, using a slight modification of the standard HMBC [4] pulse sequence. By replacing the 180° refocusing pulse in this experiment by a jump-and-return pulse, modulation caused by the NH-Hα scalar coupling is suppressed [8], enhancing the sensitivity of NH-C′ correlations. For the NH-C′ HMBC spectrum, the acquisition times were 102 ms (t_2) and 50 ms (t_1); the delay for excitation of multiple quantum coherence was set to 30 ms. The total measuring time was 20 h. Spectra were recorded and processed according to the mixed mode (F_1 absorption; F_2 absolute value) recipe [9]. A Hα-C′ HMBC spectrum was recorded in D_2O solution on a Nicolet NT-500 spectrometer using the standard HMBC experiment with a 33 ms delay period for excitation of multiple quantum coherence. Acquisition times were 102 ms (t_2) and 51 ms (t_1). Because of the low inherent sensitivity of the NT-500 spectrometer, it was necessary to acquire the data for 60 h.

3. RESULTS AND DISCUSSION

Our procedure for identifying different types of ^{15}N-labeled amino acids relies on a quantitative interpreta-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

Table I

Position of ^{15}N-labeled amino acids in calmodulin that have characteristic double-labeling multiplet patterns

	[^{15}N]Lys (99%)	[^{15}N]Ile (40%)	[^{15}N]Arg (16%)
[^{13}C]Phe (99%)	13	100	90
[^{13}C]Met (77%)	77	52,125	37
[^{13}C]Arg (66%)	75	-	-
[^{13}C]Leu (66%)	-	-	106
[^{13}C]Thr (50%)	30	27,63	-
[^{13}C]Ile (50%)	-	-	86,126
[^{13}C]others (0%)	21,94,115,148	9,85,130	74

tion of the ^{15}N-^{1}H correlation intensities in the 2D correlation spectra. Because differences in ^{1}H and ^{15}N line width also result in differences in intensities, it is safest to compare the intensities with those observed for a sample that has been uniformly enriched with ^{15}N, and that is needed at a later stage for completing the sequential assignment [6]. The ratio of the resonance intensities observed in the selectively labeled sample and in the uniformly labeled sample gives a reliable indication of the level of ^{15}N enrichment obtained for the selectively labeled sample. Transamination and other metabolic processes often cause the ^{15}N isotopic enrichment in the selectively labeled sample to be lower than expected. The degree of transamination depends strongly on the

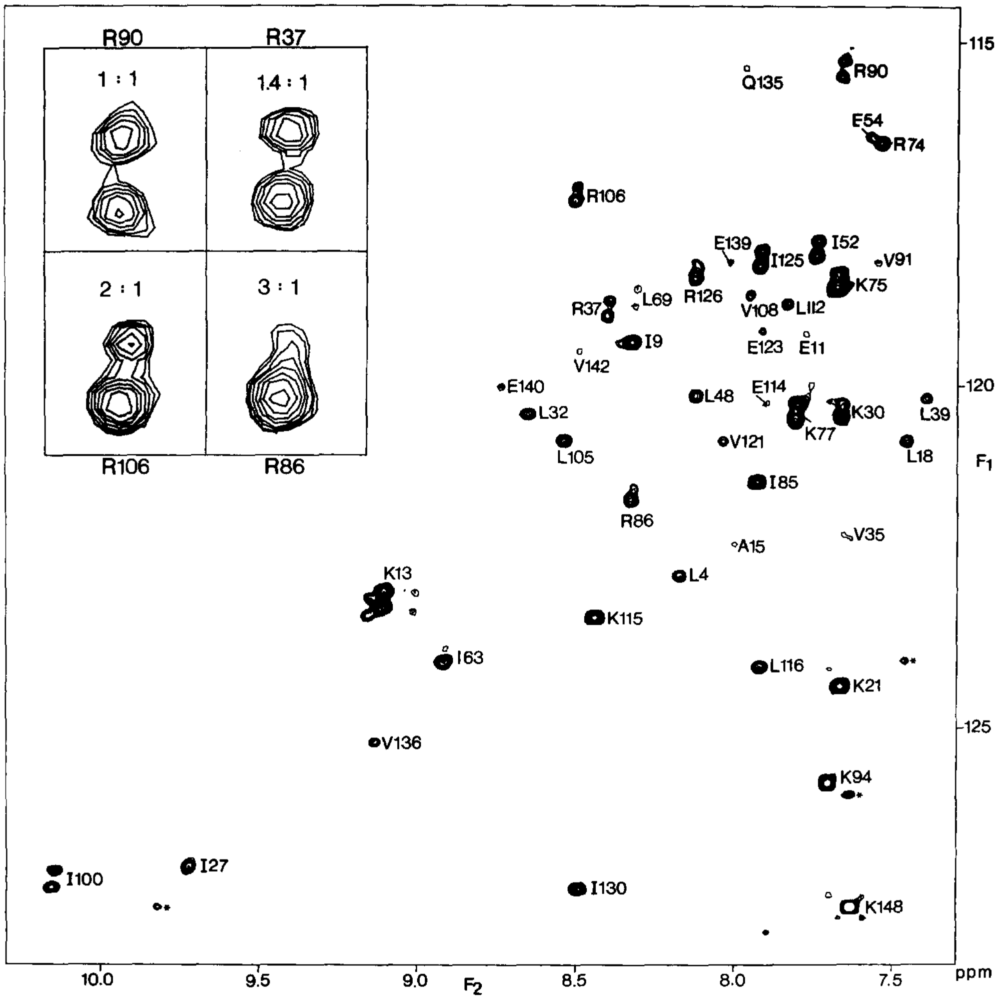

Fig. 1. Decoupled Overbodenhausen ^{1}H-^{15}N shift correlation spectrum of calmodulin of multiple labeled calmodulin, complexed with calcium. The spectrum was recorded at 600 MHz ^{1}H frequency, 47°C, pH 6.3. Amide correlations that are preceded by a ^{13}C carbonyl appear as asymmetric doublets; the degree of asymmetry is indicative of the level of ^{13}C enrichment. Assignment of resonances not adjacent to ^{13}C$_1$ was obtained at a later stage using triple resonance experiments (unpublished).

particular strain of *E. coli* used for expressing the protein, and also on the growth conditions. In practice, however, we have observed that many of the more complicated amino acids, i.e. all but Asx, Glx, Gly, Ala and Ser usually label in a semi-quantitative manner, i.e. using 99% ^{15}N-labeled amino acid in the growth medium, the incorporation level in the protein is at least 50%, sufficient for our purpose. Similarly, we have observed that for these 'complicated' amino acids, the labeling of ^{13}C in the C_1 position is close to 100%. Thus, it is possible to use isotopic labeling with different amino acids in a single protein preparation, greatly enhancing the effectiveness of the selective isotopic labeling approach.

The selection of amino acids to use for isotopic labeling depends on the primary structure of the protein. This selection should maximize the number of dipeptide fragments that can be identified on the basis of the selected amino acids. For the case of calmodulin, we used three ^{15}N-enriched amino acids, Lys, Ile, and Arg at isotopic enrichment levels of 99%, 40%, and 16%, respectively, and six 1-^{13}C-enriched amino acids, Phe, Met, Arg, Leu, Thr, and Ile at enrichments of 99%, 77%, 66%, 66%, 50%, and 50%, respectively. The positions of dipeptide sequences in calmodulin that are doubly labeled, i.e. with both ^{13}C and ^{15}N enrichment in the pertinent peptide bond, are shown in Table I.

Fig. 1 shows the ^{15}N-^{1}H correlation spectrum. In principle, three different levels of ^{15}N-^{1}H correlation are excepted to be present with relative intensities of 1:0.4:0.16 for Lys, Ile and Arg, respectively. Indeed, as excepted, the spectrum shows 8 very intense correlations for the lysine residues. Four of them, K13, K77, K75 and K30 show the expected splitting due to coupling to ^{13}C. For lysines that have only partial ^{13}C enrichment in the preceding residue, the observed signal is a superposition of a singlet and a doublet (caused by the ~15 Hz ^{15}N-^{13}C J coupling). Because of a ^{13}C isotope shift effect, the ^{15}N doublet is shifted to lower frequency (upfield) by ~0.08 ppm relative to the singlet. On high field spectrometers (>400 MHz) this causes one of the doublet components to overlap with the singlet, resulting in the appearance of an asymmetric doublet. The level of asymmetry, 1:1, 1.4:1, 2:1 or 3:1 corresponds to the level of ^{13}C enrichment of the preceding residue (99%, 77%, 66% or 50%). The inset in Fig. 1 shows an enlargement of correlations observed for four of the arginine residues (16% ^{15}N-enrichment), clearly showing these asymmetric doublet patterns, and thus identifying their positions in the protein backbone.

At the intermediate level of correlation intensity, 5 doublets and three singlets are observed corresponding to Ile residues. The 1:1 doublet can be assigned to I100, the two 1.4:1 doublets to I52 and I125, and the two 3:1 doublets to I27 and I63. Note that in the case of nonunique dipeptides, such as M51/I52 and M124/I125 additional NMR experiments are required to uniquely assign the amide resonances.

The isotopic ^{15}N enrichment level as compared to the uniformly ^{15}N-labeled sample was 87 ± 3, 20 ± 3 and 15 ± 2% for Lys, Ile and Arg, respectively. As can be

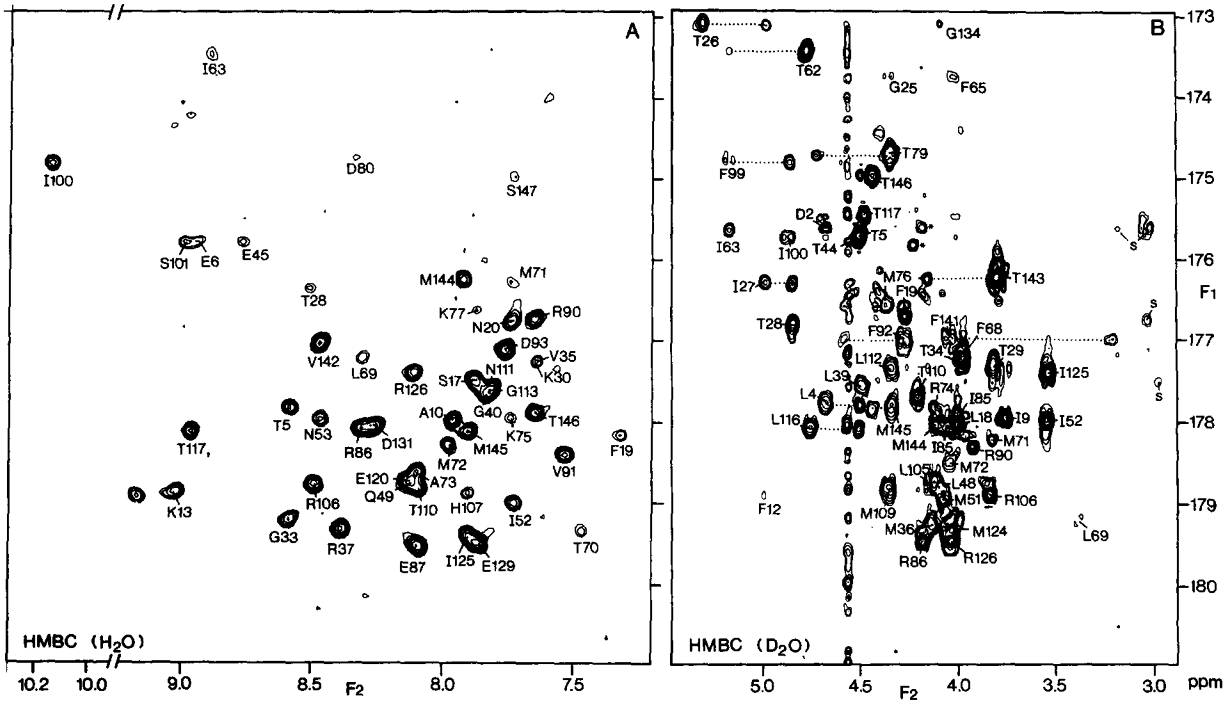

Fig. 2. Multiple bond ^{1}H-^{13}C$_1$ shift correlation spectra of calmodulin. (A) Recorded at 600 MHz, in H_2O solution, and showing NH-C_1 correlations. (B) Recorded at 500 MHz, in D_2O solution, and showing Hα-C_1 correlations. Correlations to side chain Hβ protons are marked 's', unidentified peaks are marked '*'. Only intraresidue correlations are labeled; dotted lines mark the correlation to the Hα resonance of the succeeding residue.

seen from these numbers, especially for Ile the level of ^{15}N enrichment was much lower than the expected level of 40%. This is most likely caused by transamination. Indeed, 9 additional residues are observed at enrichment levels of 10 ± 2%, and an additional number at much lower levels of ^{15}N enrichment, corresponding to the 9 Leu residues in calmodulin. Weaker correlations are observed for valine residues (enrichment level of 4 ± 2%) and at contour levels lower than shown in Fig. 1, a number of very weak Ala, Glu, Gln, Asp and Asn residues were observed.

In addition to the ^{15}N-^{1}H one-bond correlation experiment, the multiple labeled protein sample can be used for a number of additional experiments that provide assignment and structural information. Here we demonstrate for the first time the combined use of multiple bond ^{13}C$_1$-NH and ^{13}C$_1$-Hα correlation. The modified HMBC experiment mentioned in the experimental section provides a high quality NH-C$_1$ correlation spectrum, shown in Fig. 2a. A Hα-C$_1$ correlation spectrum, recorded with the standard HMBC experiment, is shown in Fig. 2b. Comparison of the two spectra reveals a substantial number of sequential connectivities. For example, at the top of Fig. 2A, I63-NH shows a correlation to T62-C$_1$. In the right panel, correlations to the same C$_1$ shift are observed for T62-Hα and I63-Hα. As pointed out previously [4], sequential three-bond Hα-Cα correlations are only observed for ϕ backbone angles in the vicinity of -120°. Therefore, it is interesting to note the correlation observed in the right panel of Fig. 2 between T79-C$_1$ and D80-Hα, suggesting a significant deviation from ideal α-helical geometry ($\phi \approx -50°$) in the 'central helix' part of calmodulin. This result is compatible with the X-ray crystal structure ($\phi = -99°$) [10], but does not exclude the possibility of rapid averaging in a 'random coil' type fashion. Additional NMR experiments are in progress to accurately determine the solution structure of this part of the protein.

Acknowledgement: This work was supported by the Intramural AIDS Antiviral Program of the Office of the Director of the National Institutes of Health.

REFERENCES

[1] Kainosho, M. and Tsuji, T. (1982) Biochemistry 21, 6273–6279.

[2] Griffey, R.H., Redfield, A.G., McIntosh, L.P., Oas, T.G. and Dahlquist, F.W. (1986) J. Am. Chem. Soc. 108, 6816–6817.

[3] Torchia, D.A., Sparks, S.W. and Bax, A. (1989) Biochemistry 28, 5509–5524.

[4] Bax, A., Sparks, S.W. and Torchia, D.A. (1988) J. Am. Chem. Soc. 110, 7926–7927.

[6] Ikura, M., Marion, D., Kay, L.E., Shih, H., Krinks, M., Klee, C.B. and Bax, A. (1990) Biochem. Pharm. in press.

[5] Smith, V.L., Doyle, K.E., Maune, J.F., Munjaal, R.P. and Beckingham, K. (1987) J. Mol. Biol. 196, 471–485.

[7] Bodenhausen, G. and Ruben, D.J. (1980) Chem. Phys. Lett. 69, 185–189.

[8] Bax, A., Ikura, M., Kay, L.E., Torchia, D.A. and Tschudin, R. (1990) J. Magn. Reson. 86, 304–318.

[9] Bax, A. and Marion, D. (1988) J. Magn. Reson. 78, 186–191.

[10] Babu, Y.S., Bugg, C.E. and Cook, W.J. (1988) J. Mol. Biol. 204, 191–204.