# $^1$H NMR Sequential Resonance Assignments, Secondary Structure, and Global Fold in Solution of the Major (*trans*-Pro43) Form of Bovine Calbindin $D_{9k}$†

Johan Kördel,‡,§ Sture Forsén,§ and Walter J. Chazin*,‡

*Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, California 92037, and Department of Physical Chemistry 2, Chemical Center, University of Lund, S-221 00 Lund, Sweden*

*Received February 22, 1989; Revised Manuscript Received May 4, 1989*

ABSTRACT: A wide range of two-dimensional $^1$H NMR experiments have been used to completely assign the 500-MHz $^1$H NMR spectrum of recombinant $Ca^{2+}$-saturated bovine calbindin $D_{9k}$ (76 amino acids, $M_r$ = 8500). In solution, calbindin $D_{9k}$ exists as an equilibrium mixture of isoforms with trans (75%) and cis (25%) isomers of the peptide bond at Pro43 [Chazin et al. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2195–2198], which results in two sets of $^1$H NMR signals from approximately half of the amino acids. The complete $^1$H NMR assignments for the major, *trans*-Pro43 isoform are presented here. By use of an integrated strategy for spin system identification, 62 of the 76 spin systems could be assigned to the appropriate residue type. Sequence-specific assignments were then obtained by the standard method. Secondary structure elements were identified on the basis of networks of sequential and medium-range nuclear Overhauser effects (NOEs), $^3J_{HN\alpha}$ spin coupling constants, and the location of slowly exchanging amide protons. Four helical segments and a short $\beta$-sheet between the two calcium binding loops are found. These elements of secondary structure and a few additional long-range NOEs provide the global fold. Good agreement is found between the solution and crystal structures of the minor A form of bovine calbindin $D_{9k}$ and between the solution structures of the minor A form of bovine calbindin $D_{9k}$ and intact porcine calbindin $D_{9k}$.

The $Ca^{2+}$ ion plays an important role in the regulation of a wide variety of cellular functions (Rasmussen, 1986a,b). Over the past decade the ion has come to be regarded as a new kind of second messenger—or perhaps better, a "third messenger" if the intermediary action of inositol tri- or tetraphosphates is taken into consideration (Berridge, 1987). In the resting cell the $Ca^{2+}$ concentration is generally found to be very low, on the order of $10^{-7}$ M, while as a consequence of an external stimulus the concentration may transiently reach 10–100-fold higher levels. At these higher levels a number of intracellular $Ca^{2+}$ binding proteins, or complexes involving such proteins, will have their binding sites fully or partially occupied by $Ca^{2+}$. In many cases the binding of $Ca^{2+}$ is accompanied by protein conformational changes essential for the regulation of other proteins and enzymes.

Most, if not all, regulatory $Ca^{2+}$ binding proteins show a high degree of homology, and together they constitute the major part of the calmodulin superfamily (Kretsinger, 1987). Detailed information regarding the structure–function relationships in this protein family is still very limited. X-ray crystallographic studies on four members of this class of proteins (parvalbumin, calbindin $D_{9k}$,[1] troponin C, and calmodulin) have shown that the $Ca^{2+}$ binding sites always consist of a helix–loop–helix arrangement that is known as an "EF hand" (Kretsinger & Nockolds, 1973) or alternatively as "the calmodulin fold" (Kretsinger, 1987). The loop in the EF-hand arrangement is typically 12 amino acids long and contains the residues that are directly coordinated to the $Ca^{2+}$ ion. There appears to be no straightforward correlation between $Ca^{2+}$ affinity (which ranges from $10^5$ to $10^9$ $M^{-1}$) and the nature and disposition of the ligands in the binding site (Herzberg & James, 1985; Szebenyi & Moffat, 1986), although some structure–affinity relationships have been inferred from studies of peptide analogues of EF hands (Marsden et al., 1988). It is now generally accepted that the functional unit in these $Ca^{2+}$ binding proteins is a pair of EF hands rather than the individual binding sites. Interactions between two such sites have been observed, and cooperative binding of $Ca^{2+}$ is often found (Seamon & Kretsinger, 1982; Forsén et al., 1986; Linse et al., 1987, 1988).

In an attempt to better understand the molecular mechanisms underlying the $Ca^{2+}$ binding process in the calmodulin family of proteins, we have concentrated our efforts on bovine calbindin $D_{9k}$. This protein has a size ($M_r$ = 8500) and tertiary structure very similar to that of the globular domains of calmodulin and troponin C. According to X-ray data on the bovine calbindin $D_{9k}$, refined to 2.3-Å resolution (Szebenyi & Moffat, 1986), the protein contains two EF-hand-type $Ca^{2+}$ binding sites. While site II (the C-terminal site) is an archetypal EF hand, site I contains a loop that is two residues longer than that normally found. A gene encoding the native bovine calbindin $D_{9k}$ has been synthesized and expressed in *Escherichia coli* (Brodin et al., 1986). Equilibrium and dynamic properties of recombinant calbindin corresponding to the wild type and of a considerable number of mutants with amino acid substitutions and/or deletions have been studied by using different biophysical methods (Forsén et al., 1988; Linse et al., 1987, 1988; Wendt et al., 1988).

The molecular details of the conformational changes accompanying $Ca^{2+}$ binding in the regulatory calcium binding proteins are not yet known, principally because crystals suitable for X-ray structure determination have not been obtained for the apo forms. Solution structure determination utilizing 2D $^1$H NMR[2] is uniquely suited to overcome this problem. The

† We gratefully acknowledge financial support from the National Institutes of Health (Grant GM 40120-01 to W.J.C.) and the Swedish Natural Science Research Council (graduate fellowship to J.K. and operating grants to S.F.).

* To whom correspondence should be addressed.

‡ Scripps Clinic.

§ University of Lund.

[1] Formerly the 9-kDa intestinal calcium binding protein (ICaBP).

0006-2960/89/0428-7065$01.50/0 

small size and high solubility of calbindin $D_{9k}$ make it very suitable for high-resolution 2D $^1H$ NMR studies. The work reported here and our ongoing study of the apoprotein are the first steps toward the determination of the conformational basis for calcium regulatory protein function.

During this work an equilibrium between two forms of the protein was detected and established by $^1H$ NMR to be caused by proline cis–trans isomerism of Pro43 (Chazin et al., 1989a). The global folds of the cis and trans forms appear identical, and local conformational differences are constrained to the polypeptide segment Ser38–Leu53. We report here the complete sequential resonance assignments of the major *trans*-Pro43 form of recombinant calbindin $D_{9k}$. On the basis of these data, secondary structure elements have been identified and a global fold is proposed. In a subsequent paper we will present a comparative analysis of the major and minor forms of r-calbindin and of the P43G mutant (Pro43 replaced with a glycine residue) lacking this conformational heterogeneity.

## Material and Methods

The r-calbindin samples were purified as described in Chazin et al. (1989b). Protein from bovine intestine was purified according to the method of Hitchman et al. (1973). The protein samples were dissolved in 0.42 mL of $^1H_2O$, containing 5% (v/v) $^2H_2O$ for the lock, or "100%" $^2H_2O$ (Merck Isotopes, Montreal, Canada) to a protein concentration of 4–5 mM. The pH was adjusted to 6.0 by addition of microliter amounts of 0.01 M NaOH, HCl, NaO$^2$H, or $^2$HCl. The $^2H_2O$ samples were prepared under argon, and the pH* was not corrected for isotopic effects.

All NMR experiments were performed by using Bruker AM-500 spectrometers equipped with Aspect 3000 computers and digital phase shifting hardware. Standard pulse sequences and phase cycling were utilized to obtain phase-sensitive COSY (Marion & Wüthrich, 1983), R-COSY and DR-COSY (Wagner, 1983) experiments in $^1H_2O$, and 2QF-COSY and 3QF-COSY (Rance et al., 1983; Rance & Wright, 1986) in $^2H_2O$. 2Q and 3Q experiments were performed by using the pulse sequence and phase cycling described by Braunschweiler et al. (1983), with a composite 180° pulse ($90°_{\phi}$–$180°_{\pi/2+\phi}$–$90°_{\phi}$) (Levitt & Freeman, 1979). TOCSY experiments were obtained by using the modification described by Rance (1987). Pure absorption Hahn–echo NOESY spectra were acquired by using the normal pulse sequence followed by a short Hahn–echo period to improve the quality of the base line (M. Rance, unpublished). Minimum phase cycling of the 90° pulses (Bodenhausen et al., 1984) was used along with phase alternation of the 180° echo pulse. All spectra were recorded in the phase-sensitive mode with quadrature detection in the $\omega_1$ dimension achieved by using time-proportional phase incrementation (Marion & Wüthrich, 1983). The carrier was set on the solvent resonance for all experiments except TOCSY acquired from $^2H_2O$ solution. A preacquisition delay of 1.2 s was used, and a total of 2048 quadrature pairs of data points were collected in the sequential acquisition mode for each $t_1$ value. Table S1 of the supplementary material lists the pertinent acquisition parameters for each experiment.

The raw data were processed on either an Aspect 3000 data station using Bruker software or on a SUN 3/160C computer with the FTNMR program provided by Dr. D. Hare, modified for use on the SUN by Dr. J. Sayre. Application of window functions and base-line corrections were performed as described elsewhere (Chazin et al., 1988). Data that were cosine modulated in the $\omega_1$ dimension were routinely treated for suppression of $t_1$ ridges by halving the first FID (Otting et al., 1985). Typically, the final data block was 1024 points in $\omega_1$ and 2048 points in $\omega_2$, giving a digital resolution of 4.6 and 3.1 Hz/point, respectively.

The slowly exchanging amide protons were identified on the basis of observation of at least one cross peak in a 17-h Hahn–echo NOESY spectrum ($\tau_m$ = 200 ms) started 45 min after dissolution of the protein in $^2H_2O$ at pH* 6.0. Backbone NH/C$^{\alpha}$H coupling constants were measured from the COSY spectrum at pH 6.0 and 300K, as described by Marion and Wüthrich (1983).

## Results

The characterization of r-calbindin by a number of biochemical and biophysical methods has clearly shown its identity with the native minor A form of the protein isolated from bovine intestines (Brodin et al., 1986). To further establish that the solution structures are identical, the backbone, methyl, and aromatic fingerprint regions of COSY spectra acquired under identical conditions from native and recombinant protein samples were carefully compared. Except for the extra Met0 signals of r-calbindin, all NMR parameters were found to be essentially identical. This included the relative distribution of the two isoforms, which result from cis–trans isomerism at Pro43.

*Sequential Assignment Strategy.* The procedure used for obtaining $^1H$ NMR resonance assignments was that of Wüthrich and co-workers [Wüthrich, 1983; reviewed in Wüthrich (1986)]. The $^1H$ spin systems were identified by using the strategy described in Chazin et al. (1988). The approach relies heavily on observation in $^1H_2O$ solution of relayed scalar connectivities from the side-chain protons to the backbone amide protons [discussed in Chazin and Wright (1987)]. To completely characterize the long side chains, additional assignments are obtained from experiments in $^2H_2O$ using a parallel strategy, but based at the side-chain terminus rather than the backbone amide proton. MQ and MQF-COSY experiments are analyzed to complete the side-chain assignments and to reliably discriminate between the various protons along the side chain. With complete or near-complete identification of spin systems, classification as unique, 3-spin, or 5-spin can be made on the basis of chemical shifts and scalar connectivity patterns. Characteristic through-space (NOE) connectivities between aromatic rings, side-chain amide groups, and methionine methyl groups and the corresponding backbone amide based spin systems are then used to identify many (if not all) of the aromatic, asparagine, glutamine, and methionine spin systems. Proline spin systems are assigned in the final stages of the analysis. This detailed analysis of spin systems provides an ample basis for the subsequent sequential assignment by the standard procedure (Billeter et al., 1982). The spin systems are placed in their sequential order on the basis of the known amino acid sequences and characteristic NOE connectivities between adjacent residues in the sequence.

[2] Abbreviations: r-calbindin, recombinantly expressed sequence of the minor A form of bovine calbindin $D_{9k}$ plus N-terminal methionine (Met0); NMR, nuclear magnetic resonance; 1D, one-dimensional; 2D, two-dimensional; COSY, correlated spectroscopy; R-COSY, relayed COSY; DR-COSY, double relayed COSY; TOCSY, total correlation spectroscopy; MQ (2Q, 3Q), multiple (two, three) quantum spectroscopy; MQF-COSY (2QF-COSY, 3QF-COSY), multiple (two, three) quantum-filtered COSY; NOE, nuclear Overhauser effect; NOESY, 2D NOE spectroscopy; FID, free induction decay; 3-spin, the C$^{\alpha}$H–C$^{\beta}$H$_2$ spin system of serine, cysteine, aspartic acid, asparagine, and the aromatic amino acid residues; 5-spin, the C$^{\alpha}$H–C$^{\beta}$H$_2$–C$^{\gamma}$H$_2$ spin subsystem of glutamic acid, glutamine, and methionine residues; N-cap, amino acid residue at the N-terminus of a helix [see Richardson & Richardson (1988)].

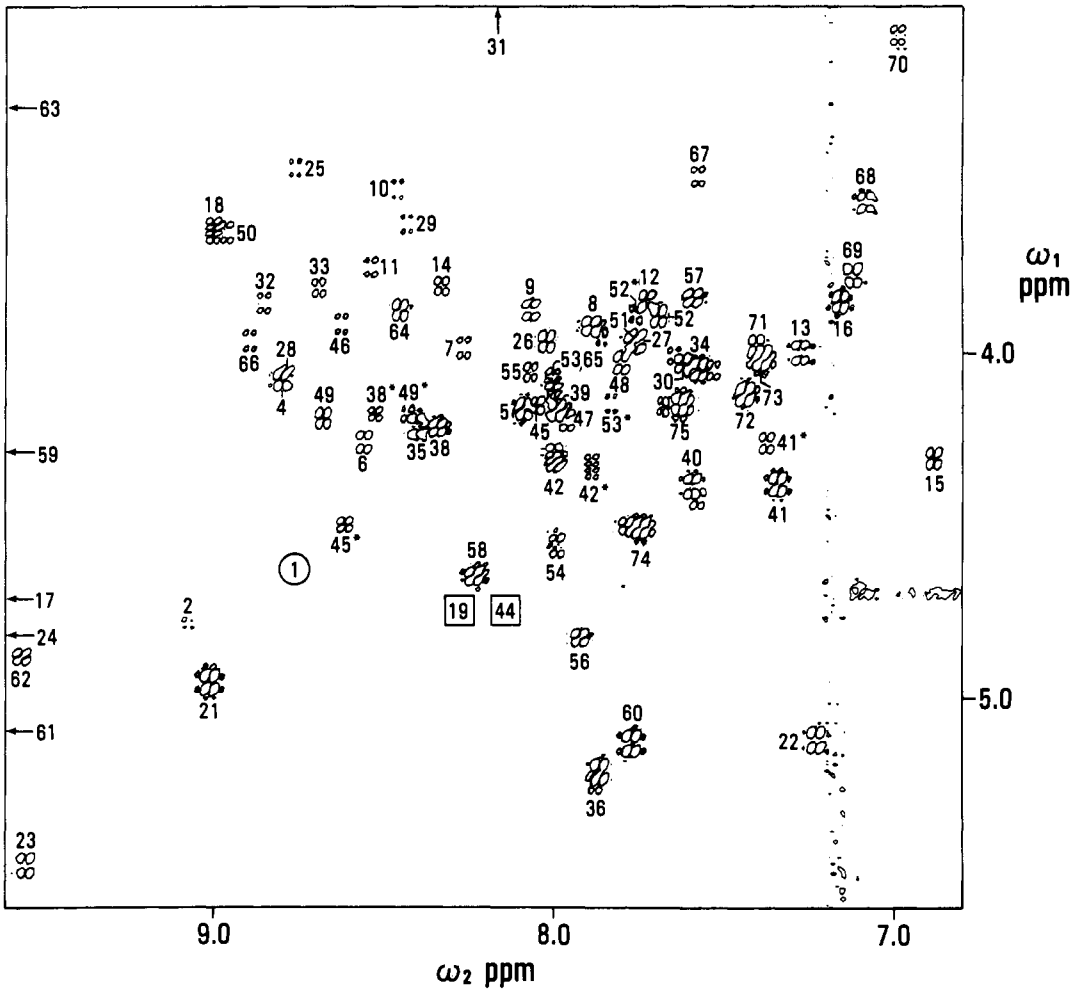


FIGURE 1: Backbone fingerprint region of a 500-MHz COSY spectrum of 5 mM r-calbindin at 300 K and pH 6.0. The $\omega_1$ = $C^{\alpha}H$, $\omega_2$ = NH cross peaks are labeled with the sequence-specific assignments. Arrows indicate cross peaks that appear outside the region shown. Boxes show the location of cross peaks not observed in this spectrum, but that could be identified in other 2D spectra. Circles are drawn around cross peaks that can only be observed at very low contour levels. Cross peaks from the minor isomer that are well separated from their major counterpart are marked with their sequence-specific assignment and an asterisk.

*Spin System Identification.* (*A*) *Preliminary Analysis.* Examination of the backbone fingerprint region ($\omega_1$ = $C^{\alpha}H$, $\omega_2$ = NH) of a COSY spectrum for r-calbindin should contain cross peaks for 71 residues (76 residues minus 4 prolines and the N-terminus). However, a much larger number of cross peaks were observed (Figure 1), due to the existence of an equilibrium mixture of *cis*- and *trans*-Pro43 isoforms (Chazin et al., 1989a). In the following, we present the detailed $^1H$ NMR analysis for the major *trans*-Pro43 isoform. Problems associated with the differentiation of major and minor cross peaks are discussed elsewhere (Chazin et al., 1989a; Drakenberg et al., 1989; J. Kördel, S. Forsén, and W. J. Chazin, unpublished results).

All but three of the backbone $C^{\alpha}H$/NH cross peaks expected for the trans form are observed at 300 K (Figure 1). The $C^{\alpha}H$ resonances of Asp19 and Ser44 are coincident with the $H_2O$ signal; thus, corresponding cross peaks are suppressed by saturation during the course of the experiment. The cross peak of Asp19 is not observed at 300, 315, or 325 K, apparently due to the temperature sensitivity of the $C^{\alpha}H$ resonance. The cross peak of Leu53 is overlapped by two other cross peaks (Leu39 and Glu65) and could not be identified in the preliminary analysis.

Most methyl group cross peaks are readily identified in the characteristic regions of the COSY spectrum. Degeneracy of the Ala15 and Thr45 methyl group cross peaks was revealed in R-COSY and DR-COSY experiments. Only 29 of the expected 32 methyl doublet resonances are identified for the 3 valine, 12 leucine, and 2 isoleucine residues (Figure 2). The absence of three cross peaks is caused by degeneracy of the pairs of $C^{\delta}$ methyl group resonances of Leu30, Leu39, and Leu46. Several cross peaks from the minor isoform can also be identified in the methyl regions. All of these minor peaks have chemical shift patterns that are very similar to their counterpart in the major form but clearly lower intensities.

In the aromatic region of the COSY spectrum, the complete ring proton spin systems of the tyrosine and three of the five phenylalanines are readily identified. For Phe10, the $C^{\delta}H$ resonance is significantly broadened and the $C^{\epsilon}H$ resonance slightly broadened at 300 K, and no COSY cross peak is observed. This effect is attributable to an intermediate (on the NMR time scale) 180° ring flip rate. The broadened resonances are coalesced at 315 K, and the corresponding cross peaks are observed. The aromatic ring proton resonances of Phe50 are almost degenerate and are also severely overlapped with the $C^{\delta}H$ and $C^{\zeta}H$ resonances of Phe36. Detailed analyses of 2Q and 3Q spectra [for an example see Chazin et al. (1988)] were essential for assigning the Phe50 resonances. COSY cross peaks arising from scalar couplings between the side-chain amide groups of the asparagine and glutamine residues were also observed in this region of the spectrum.

(*B*) *Identification of Backbone Amide Based Spin Systems.* On the basis of direct and relayed connectivities observed at the backbone amide proton in the series of COSY, 2Q, R-

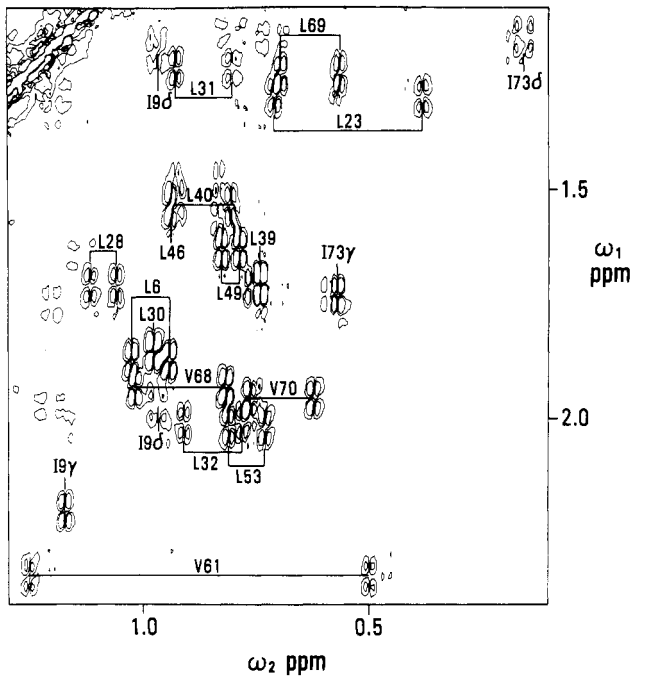


FIGURE 2: Valine/leucine/isoleucine fingerprint region of a 500-MHz DQF-COSY spectrum of 5 mM r-calbindin in $^2H_2O$ at 300 K and pH* 6.0. The valine ($\omega_1 = C^\beta H$, $\omega_2 = C^\gamma H_3$), leucine ($\omega_1 = C^\gamma H$, $\omega_2 = C^\delta H_3$), and isoleucine ($\omega_1 = C^\beta H$, $\omega_2 = C^\gamma H_3$ and $\omega_1 = C^\gamma H$, $\omega_2 = C^\delta H_3$, identified as $\gamma$ and $\delta$, respectively) cross peaks are labeled with their sequence-specific assignments. The most upfield shifted Ile73 $C^\gamma H/C^\delta H_3$ cross peak is not shown.

COSY, DR-COSY, and TOCSY spectra acquired from $^1H_2O$ solution, it was possible to identify a substantial number (in most cases all) of the resonances of the amino acid spin systems. Overall, 111 (92%) of the $C^\beta H$ and 61 (81%) of the $C^\gamma H$ resonances were identified in these experiments. All glycine, alanine, threonine, valine, isoleucine, and leucine spin systems, as well as 6 of 10 lysines, could be directly distinguished on the basis of the pattern of side-chain resonances and the corresponding chemical shifts. In addition, 17 of the remaining spin systems could be tentatively assigned as one of the 18 3-spin side-chain residues, and 21 others were assigned as one of the 17 5-spin or 4 remaining lysine residues. Finally, one spin system (subsequently identified as Ser62) showed no relayed connectivities in any of the $^1H_2O$ spectra.

(*C*) *Assignment of Residues with Unique Spin Systems.* The five glycine residues were identified from their characteristic multiplet structure in the COSY spectra and remote peaks at $\omega_1 = \omega_\alpha + \omega_{\alpha'}$, $\omega_2 = \omega_{NH}$ in 2Q spectra ($^1H_2O$ solution). R-COSY and DR-COSY spectra together provided the complete set of side-chain connectivities to the backbone NH for the alanine, threonine, and valine spin systems.

Spin system assignments could be made for both isoleucine residues from the combination of $C^\gamma H_3/C^\delta H_3$, $C^\gamma H_3$/NH, and $C^\beta H$/NH relayed connectivities. The leucine spin systems were completely identified by the observation in TOCSY spectra of relayed connectivities from each of the side-chain protons to the backbone NH or by the coincidence of relayed connectivities observed at the $C^\delta H_3$ and backbone NH resonances. The identification of Leu39 and Leu53 was greatly complicated by the degeneracy of both the backbone NH and $C^\alpha$ proton resonances (Table I). A careful examination of the relayed connectivities from the $C^\beta$ protons to backbone NH and of the $C^\beta H/C^\alpha H$ cross peaks in 3QF-COSY was instrumental in distinguishing these two spin systems. The assignment of all leucine $C^\beta H_2$ resonances and isoleucine $C^\gamma H_2$ resonances was confirmed in MQ spectra.

The 10 lysine spin systems were completely identified by the strategy described in Chazin et al. (1987). Multistep relayed connectivities observed at the $C^\epsilon H_2$ and backbone NH resonances provided the basis for assignments. MQ and MQF-COSY were essential for obtaining position-specific assignments along the side chain. Magnetic equivalence of $C^\delta H_2$ and $C^\epsilon H_2$ resonances was firmly established in 2Q experiments from remote peaks $\omega_1 = \omega_\delta + \omega_{\delta'}$, $\omega_2 = \omega_\epsilon$ and $\omega_1 = \omega_\epsilon + \omega_{\epsilon'}$, $\omega_2 = \omega_\delta$, respectively, and confirmed in 3Q spectra. The resonance frequency of all but one lysine proton could be unambiguously assigned. The lone exception is for the $C^\gamma H_2$ of Lys1, for which evidence of resonance degeneracy could not be found in MQ spectra.

(*D*) *Assignment of Residues with 3-Spin Side Chains.* Fourteen of 18 spin systems with 3-spin side chains could be completely identified from connectivities to the backbone amide proton. Asp19 had relayed connectivities from both $C^\beta$ protons, but the $C^\alpha$ proton was bleached out by the $H_2O$ saturation. Ser44 and Ser74 each had only one $C^\beta H$ relayed connectivity, and Ser62 had none. Relayed connectivities were all verified in 3QF-COSY spectra and by at least one of the following: (i) direct or remote peaks in 2Q spectra; (ii) direct peaks in 3Q spectra. These spectra also provided additional resonances for the Asp19 and Ser62 spin systems. Unambiguous evidence for the degeneracy of the $C^\beta H$ resonances of Ser44 and Ser74 was obtained from remote peaks at $\omega_1 = \omega_\beta + \omega_{\beta'}$, $\omega_2 = \omega_\alpha$ in the 2Q spectra.

The five phenylalanines, single tyrosine, and two asparagines were specifically identified by NOEs between ring or side-chain amide protons and the $C^\beta$, $C^\alpha$, or backbone amide protons. Six other spin systems were assigned to the six serine residues on the basis of the average of the $C^\beta H$ chemical shifts (>3.8 ppm). One of the serine residues exhibited a clearly resolved hydroxyl proton resonance at 5.98 ppm. Observation of this resonance is attributed to a reduced rate of solvent exchange presumably due to a strong hydrogen bond. NOESY connectivities identify this resonance as the $O^\gamma H$ of Ser62. NOEs to Glu65 $C^\gamma H_2$ indicate hydrogen bonding to the carboxylate of Glu65.

(*E*) *Identification of Residues with 5-Spin Side Chains.* For 9 of the 18 5-spin spin systems, all 4 side-chain resonances could be identified from relayed connectivities to the backbone amide proton. For Glu51, Glu52, Glu64, Glu65, and Gln75, only three side-chain relayed connectivities were observed and for Glu27, Glu35, and Glu60 only two. The identification of the Met0 spin system (with no amide resonance) is described below.

$C^\beta$ protons were distinguished from the $C^\gamma$ protons by the $C^\alpha H/C^\beta H$ connectivities in MQ and MQF-COSY. These spectra also provided additional connectivities for Glu51, Glu52, and Glu65. The complete assignments for Glu27, Glu35, and Glu60 were obtained from $C^\gamma H/C^\alpha H$ relayed connectivities in TOCSY spectra. For Glu64 and Gln75, only one strong resonance was observed in all spectra for $C^\gamma H_2$, and the chemical shift dispersion of the corresponding $C^\beta$ protons is small, so the $C^\gamma$ protons are assumed to be degenerate.

Three of the 5-spin spin systems could be identified as glutamine residues by characteristic NOE connectivities between the side-chain amide group and the corresponding $C^\alpha$, $C^\beta$, or backbone amide proton(s). A fourth pair of glutamine side-chain amide proton resonances is readily identified, but does not exhibit any of the characteristic NOEs. The latter were subsequently assigned by default to Q75, after all other glutamine residues were identified by sequential resonance assignment.

Table I: $^1$H NMR Chemical Shifts of the *trans*-Pro43 Isoform of Recombinant Minor A Bovine Calbindin $D_{9k}$ (pH 6.0, 300 K)[a]

| residue | NH | chemical shifts (ppm) | | | | |
|---|---|---|---|---|---|---|
| | | $C^\alpha$ | $C^\beta$ | $C^\gamma$ | $C^\delta$ | other |
| M 0 | | 4.27 | 1.88, 1.96 | 2.19 (2.19)[b] | | 2.13 ($C^\gamma H_3$) |
| K 1 | 8.76 | 4.60 | 1.78, 1.84 | 1.62 (1.62)[b] | 1.73, 1.73 | 2.95, 2.95 ($C^\epsilon H_2$) |
| S 2 | 9.07 | 4.78 | 4.06, 4.44 | | | |
| P 3 | | 4.32 | 2.00, 2.46 | 2.05, 2.24 | 3.98, 3.98 | |
| E 4 | 8.80 | 4.08 | 1.99, 2.10 | 2.30, 2.43 | | |
| E 5 | 8.08 | 4.16 | 2.08, 2.45 | 2.31, 2.44 | | |
| L 6 | 8.56 | 4.26 | 1.71, 2.20 | 1.88 | 0.95, 1.03 | |
| K 7 | 8.26 | 3.98 | 1.81, 1.92 | 0.81, 1.16 | 1.35, 1.35 | 2.59, 2.67 ($C^\epsilon H_2$) |
| G 8 | 7.89 | 3.92, 3.92 | | | | |
| I 9 | 8.06 | 3.87 | 2.21 | 1.20, 1.98 | 0.98 | 1.18 ($C^\gamma H_3$) |
| F10 | 8.46 | 3.52 | 2.68, 3.31 | | | 6.32, 7.11, 7.62 ($C^\delta H$, $C^\epsilon H$, $C^\zeta H$) |
| E11 | 8.53 | 3.75 | 1.96, 2.10 | 2.32, 2.68 | | |
| K12 | 7.72 | 3.84 | 1.74, 1.79 | 0.54, 1.08 | 1.43, 1.50 | 2.71, 2.71 ($C^\epsilon H_2$) |
| Y13 | 7.28 | 4.00 | 2.44, 2.81 | | | 7.43, 6.74 ($C^\delta H$, $C^\epsilon H$) |
| A14 | 8.33 | 3.80 | 0.43 | | | |
| A15 | 6.88 | 4.30 | 1.43 | | | |
| K16 | 7.15 | 3.85 | 1.98, 2.15 | 1.48, 1.56 | 1.64, 1.71 | 2.73, 2.94 ($C^\epsilon H_2$) |
| E17 | 9.70 | 4.72 | 1.90, 2.00 | 1.94, 2.22 | | |
| G18 | 9.00 | 3.64, 3.95 | | | | |
| D19 | 8.29 | 4.69 | 2.64, 2.87 | | | |
| P20 | | 4.80 | 2.01, 2.22 | 1.87, 2.06 | 4.00, 4.00 | |
| N21 | 9.00 | 4.95 | 2.70, 3.01 | | | 7.94, 6.96 ($N^\delta H_2$) |
| Q22 | 7.23 | 5.12 | 1.83, 2.12 | 2.02, 2.26 | | 6.58, 7.48 ($N^\epsilon H_2$) |
| L23 | 9.54 | 5.48 | 1.60, 2.03 | 1.30 | 0.38, 0.71 | |
| S24 | 10.08 | 4.83 | 4.17, 4.35 | | | |
| K25 | 8.76 | 3.46 | 0.40, 1.31 | 1.09, 0.64 | 1.37, 1.38 | 2.53, 2.58 ($C^\epsilon H_2$) |
| E26 | 8.02 | 3.96 | 1.89, 1.97 | 2.22, 2.32 | | |
| E27 | 7.75 | 3.97 | 1.88, 2.36 | 2.27, 2.45 | | |
| L28 | 8.78 | 4.07 | 1.56, 2.33 | 1.72 | 1.06, 1.12 | |
| K29 | 8.43 | 3.62 | 1.89, 2.10 | 1.18, 1.30 | 1.62, 1.62 | 2.81, 2.86 ($C^\epsilon H_2$) |
| L30 | 7.62 | 4.03 | 1.74, 1.86 | 1.87 | 0.99, 0.99 | |
| L31 | 8.14 | 2.33 | 1.20, 1.76 | 1.24 | 0.80, 0.93 | |
| L32 | 8.85 | 3.85 | 1.31, 1.97 | 2.02 | 0.79, 0.92 | |
| Q33 | 8.69 | 3.81 | 1.95, 2.17 | 2.36, 2.54 | | 7.34, 6.79 ($N^\epsilon H_2$) |
| T34 | 7.57 | 4.05 | 4.16 | 1.22 | | |
| E35 | 8.40 | 4.21 | 1.30, 1.46 | 2.16, 2.64 | | |
| F36 | 7.86 | 5.21 | 2.74, 3.33 | | | 7.14, 7.01, 7.12 ($C^\delta H$, $C^\epsilon H$, $C^\zeta H$) |
| P37 | | 4.16 | 2.42, 1.21 | 1.96 (1.96)[b] | 3.07, 3.51 | |
| S38 | 8.34 | 4.22 | 3.90, 3.97 | | | |
| L39 | 7.99 | 4.16 | 1.55, 1.86 | 1.72 | 0.74, 0.74 | |
| L40 | 7.59 | 4.38 | 1.61, 1.82 | 1.55 | 0.81, 0.94 | |
| K41 | 7.34 | 4.38 | 1.73, 1.96 | 1.41, 1.48 | 1.64, 1.70 | 2.99, 2.99 ($C^\epsilon H_2$) |
| G42 | 8.00 | 4.10, 4.30 | | | | |
| P43 | | 4.39 | 2.04, 2.32 | 2.06, 2.06 | 3.60, 3.76 | |
| S44 | 8.15 | 4.74 | 3.86, 3.89 | | | |
| T45 | 8.05 | 4.15 | 4.31 | 1.43 | | |
| L46 | 8.62 | 3.92 | 1.59, 1.92 | 1.54 | 0.94, 0.94 | |
| D47 | 7.96 | 4.19 | 2.62, 2.73 | | | |
| E48 | 7.80 | 4.02 | 2.02, 2.21 | 2.27, 2.39 | | |
| L49 | 8.68 | 4.19 | 1.51, 1.82 | 1.64 | 0.79, 0.83 | |
| F50 | 8.96 | 3.64 | 3.00, 3.22 | | | 7.13, 7.12, 7.15 ($C^\delta H$, $C^\epsilon H$, $C^\zeta H$) |
| E51 | 7.77 | 3.97 | 1.90, 2.10 | 2.15, 2.37 | | |
| E52 | 7.68 | 3.89 | 2.09, 2.15 | 1.98, 2.26 | | |
| L53 | 8.00 | 4.15 | 1.13, 1.62 | 2.03 | 0.74, 0.82 | |
| D54 | 7.99 | 4.56 | 1.59, 2.52 | | | |
| K55 | 8.06 | 4.05 | 1.88, 1.94 | 1.50, 1.60 | 1.74, 1.74 | 3.06, 3.12 ($C^\epsilon H_2$) |
| N56 | 7.92 | 4.82 | 2.86, 3.30 | | | 6.62, 8.03 ($N^\delta H_2$) |
| G57 | 7.58 | 3.84, 3.84 | | | | |
| D58 | 8.23 | 4.64 | 2.46, 3.14 | | | |
| G59 | 10.49 | 3.71, 4.27 | | | | |
| E60 | 7.77 | 5.13 | 1.44, 1.91 | 2.03, 2.20 | | |
| V61 | 10.35 | 5.10 | 2.36 | 0.50, 1.26 | | |
| S62 | 9.55 | 4.88 | 4.20, 4.51 | | | 5.98 ($O^\gamma H$) |
| F63 | 9.62 | 3.28 | 2.46, 2.66 | | | 6.49, 7.13, 7.35 ($C^\delta H$, $C^\epsilon H$, $C^\zeta H$) |
| E64 | 8.45 | 3.87 | 1.89, 2.05 | 2.24 (2.24)[b] | | |
| E65 | 8.00 | 4.07 | 1.65, 1.84 | 2.36, 2.56 | | |
| F66 | 8.89 | 3.96 | 3.10, 3.28 | | | 6.91, 7.16, 7.09 ($C^\delta H$, $C^\epsilon H$, $C^\zeta H$) |
| Q67 | 7.57 | 3.49 | 1.90, 1.98 | 2.15, 2.36 | | 5.74, 6.15 ($N^\epsilon H_2$) |
| V68 | 7.08 | 3.56 | 1.94 | 0.82, 1.03 | | |
| L69 | 7.12 | 3.78 | 1.20, 1.32 | 1.25 | 0.57, 0.70 | |
| V70 | 6.99 | 3.08 | 1.96 | 0.62, 0.77 | | |
| K71 | 7.40 | 3.98 | 1.75, 1.83 | 1.40, 1.48 | 1.60, 1.60 | 2.90, 2.90 ($C^\epsilon H_2$) |
| K72 | 7.43 | 4.12 | 1.84, 1.88 | 1.48, 1.54 | 1.42, 1.59 | 2.70, 2.82 ($C^\epsilon H_2$) |
| I73 | 7.39 | 4.01 | 1.74 | 0.71, 1.18 | 0.16 | 0.58 ($C^\gamma H_3$) |
| S74 | 7.74 | 4.50 | 3.86, 3.86 | | | |
| Q75 | 7.63 | 4.15 | 1.95, 2.11 | 2.31 (2.31)[b] | | 6.80, 7.50 ($N^\epsilon H_2$) |

[a] Chemical shifts are referenced to the $H_2O$ signal at 4.75 ppm and are generally accurate to ±0.01 ppm (±0.03 ppm for geminal protons separated by <0.1 ppm). [b] Degeneracy assumed.

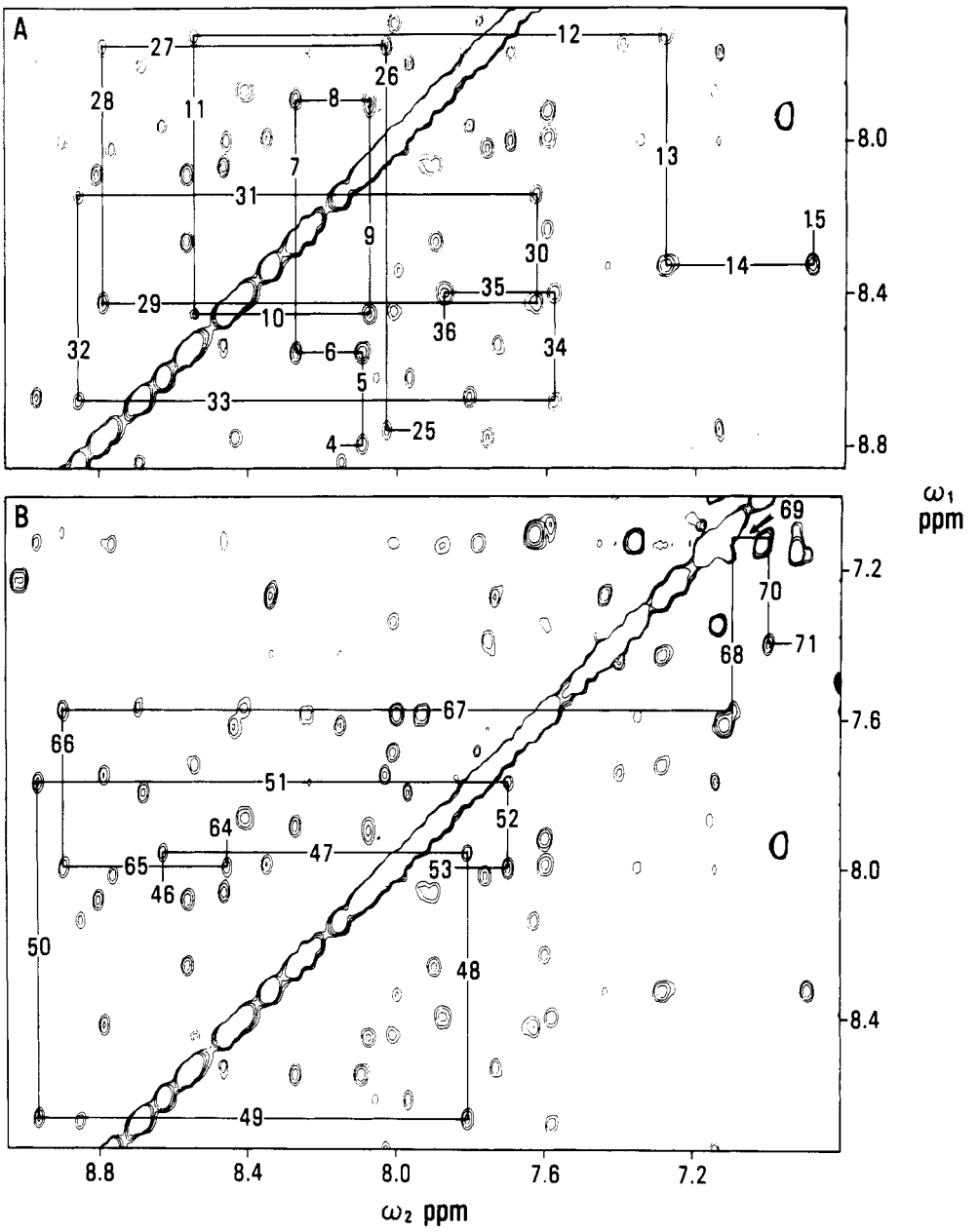


FIGURE 3: Sequential resonance assignments for polypeptide segments in helical conformation. Region of a 500-MHz Hahn-echo NOESY spectrum ($\tau_m$ = 200 ms) of a 5 mM solution of r-calbindin at 300 K and pH 6.0 showing $d_{NN}$ connectivities. Lines are drawn between pairs of $d_{NN}$ ($i, i-1$) and $d_{NN}$ ($i, i+1$) cross peaks and labeled with the sequence-specific assignments. The experimental data are shown for Glu4–Ala15 and Lys25–Phe36 (A) and for Leu46–Leu53 and Glu64–Lys71 (B).

(*F*) *Identification of Proline Spin Systems and Met0.* The identification of the four proline and the N-terminal Met0 spin systems was deferred until the $C^\alpha$ and $C^\beta$ of all other residues had been identified. This required that all resolved $C^\alpha H$–$C^\beta H_2$ spin subsystems of the minor form had also to be identified. A description of the analysis of the minor form will be given in a subsequent paper. With an otherwise complete set of assignments available, the $C^\alpha H$–$C^\beta H_2$ proline and Met0 spin subsystems were readily identified in the course of systematic analysis of MQF-COSY and MQ spectra. The proline assignments were completed by relayed connectivities observed at both the $C^\alpha$ and $C^\delta$ proton resonances. The one remaining spin system was assigned by default to Met0. The narrow methyl signal observed at 2.13 ppm in 1D experiments did not show any NOESY connectivities, but was assigned by default to Met0.

*Sequence-Specific Assignment.* The sequence-specific assignments of r-calbindin were obtained by using the standard sequential assignment procedure (Billeter et al., 1982) and are listed in Table I. The assignments were based on data from NOESY spectra recorded in $^1H_2O$ and $^2H_2O$ with 200-ms mixing times, which we estimate to give cross peaks for interproton distances of up to approximately 5 Å, either directly or through a limited amount of spin diffusion. The shorthand notation of Wüthrich et al. (1984) is adopted for specifying short proton–proton distances and corresponding NOE connectivities.

The sequence-specific assignments for residues in helical conformation were obtained from $d_{NN}$ connectivities. Two regions of the 200-ms NOESY spectrum containing sequential $d_{NN}$ connectivities for most of the four helices are shown in Figure 3. For residues in extended conformation, the sequence-specific assignments are based on strong $d_{\alpha N}(i, i+1)$ connectivities. A summary of the sequential connectivities is included in Figure 4. As can be seen in this figure, with the exception of M0/K1, S44/T45, and L53/D54, at least one $d_{NN}(i, i+1)$ or $d_{\alpha N}(i, i+1)$ connectivity is observed for each pair of consecutive residues at 300 K and pH 6.0. Under these


FIGURE 4: Summary of sequential NOE connectivities, slow amide proton exchange rates, and backbone coupling constants ($^3J_{HN\alpha}$) observed in r-calbindin and used to characterize secondary structure. The one-letter code for the amino acid sequence is given at the top. Characteristic NOE connectivities $d_{NN}(i, i + 1)$, $d_{\alpha N}(i, i + 1)$, and $d_{\beta N}(i, i + 1)$ are indicated by lines or bars between the two residues. The height of the bars gives a qualitative measure of the relative strength of the NOE in the 200-ms NOESY spectrum acquired at pH 6 and 300 K. NOEs for proline residues are to the $C^\delta$ protons rather than the backbone amide proton as discussed in Wüthrich et al. (1984). Connectivities that could not be identified in this spectrum due to degeneracy of resonances are labeled with an asterisk. Values of $^3J_{NH\alpha}$ are classified as small (<6 Hz) or large (>8 Hz) as indicated by an up arrow or a down arrow, respectively. Backbone amide protons that exchange slowly are indicated with a solid circle.

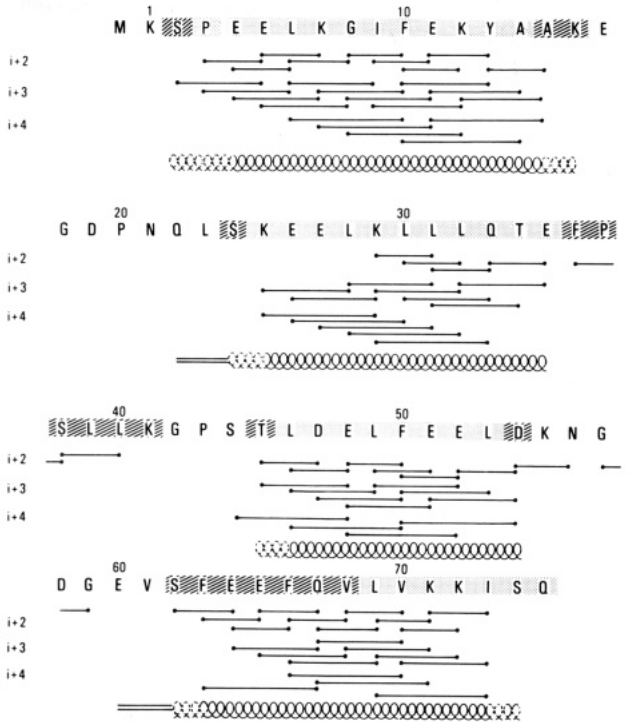


FIGURE 5: Summary of medium-range NOE connectivities used to identify the elements of secondary structure of r-calbindin in solution. The one-letter code for the amino acid sequence is given at the top. The locations of $\alpha$-helices in the X-ray structure (Szebenyi & Moffat, 1986) are indicated by shading. Characteristic NOE connectivities $d_{NN}(i, i + 2)$, $d_{\alpha N}(i, i + 2)$, $d_{\alpha N}(i, i + 3)$, $d_{\alpha N}(i, i + 4)$, and $d_{\alpha\beta}(i, i + 3)$ are indicated by lines between the two participating residues. The locations of secondary structure elements in solution are marked by coils for helices and by two parallel lines for the short antiparallel $\beta$-sheet.

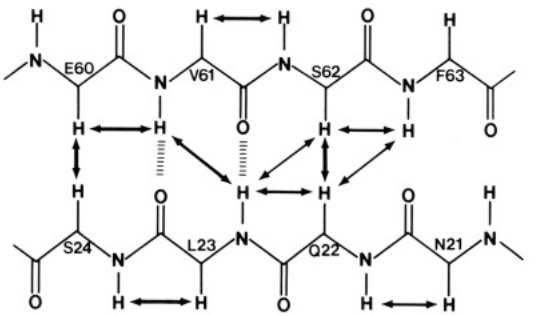


FIGURE 6: Diagram of the short antiparallel $\beta$-sheet identified in r-calbindin. The observed NOE connectivities are indicated with double-headed arrows, the width of the arrows giving a qualitative measure of the relative strength of the NOE connectivities observed in the 200-ms NOESY spectrum. The two hydrogen bonds that can be assigned on the basis of the NMR evidence are shown by cross-hatched vertical bars.

experimental conditions, the backbone amide proton of K1 has an elevated rate of solvent exchange, resulting in severe attenuation of the resonance by saturation transfer and greatly reduced intensities for corresponding NOESY cross peaks. For S44/T45, the $d_{\alpha N}$ connectivity is lost due to the coincidence at 300 K of the S44 $C^\alpha H$ resonance with the solvent signal, and the $d_{NN}$ connectivity is difficult to observe due to close proximity to the diagonal; however, a $d_{\beta N}(i, i + 1)$ connectivity is readily identified. Connectivities between L53 and D54 cannot be observed due to near degeneracy of the backbone amide proton resonances, separated only by 0.01 ppm (Table I). Medium-range connectivities [$(i, i + 2)$, $(i, i + 3)$; see Figure 5 below] do, however, bridge this gap and confirm the assignments.

*Secondary Structure.* Helical regions can be identified on the basis of networks of sequential $d_{NN}$ and medium-range [$(i, i + 3)$, $(i, i + 2)$, $(i, i + 4)$] NOE connectivities, slow amide proton exchange, and small values of $^3J_{HN\alpha}$ [reviewed in Wüthrich (1986)]. These data for r-calbindin are summarized in Figures 4 and 5. Helices are assigned to the segments Ser2–Lys16, Ser24–Glu35, Thr45–Asp54, and Ser62–Ser74 (indicated in Figure 5) and designated as helix I, II, III, and IV, respectively. Essentially no difference is found between the helical segments observed in solution and in the crystal structure (Szebenyi & Moffat, 1986).

Large spin coupling constants ($^3J_{NH\alpha} > 8$ Hz) combined with weak or not observable sequential $d_{NN}$ NOE connectivities indicate two short polypeptide segments with extended conformation, residues 21–24 and 60–62. Several NOEs are observed between the protons of these two segments as depicted in Figure 6, typical of an antiparallel $\beta$-sheet. The very low field shifts (≥9.0 ppm) of most of the backbone amide protons in these two segments are consistent with such a $\beta$-sheet-type interaction (Pardi et al., 1983).

*Global Fold.* By placing the defined elements of secondary structure in sequence and adding a few key tertiary NOEs, it is possible to sketch a schematic global fold of bovine calbindin $D_{9k}$ (Figure 7). The N-terminal helix I extends into the first $Ca^{2+}$ binding site, loop I. Contacts between Ala15 $C^\alpha H$ and Pro20 $C^\alpha H$ as well as Ala14 $C^\alpha H$ and both Leu23

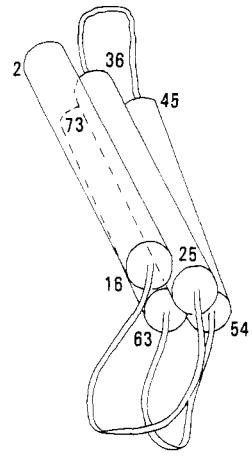


FIGURE 7: Schematic diagram of the chain folding of r-calbindin in solution. The four helical segments are depicted as cylinders and nonregular structures as ribbons.

$C^{\alpha}H$ and Ser24 NH establish that a loop is present in the polypeptide segment Ala14–Ser24. NOEs from Leu32 $C^{\delta}H_3$ to Glu4 NH and from Tyr13 $C^{\delta}H_2$ to Leu30 $C^{\delta}H_3$ establish that helices I and II are packed in an antiparallel fashion. Helices II and III are separated by a loop that has no obvious tertiary contacts to the globular core. The helix–($Ca^{2+}$ binding loop)–helix motif is then repeated in the C-terminal half of the molecule. As in the first $Ca^{2+}$ binding site, the first helix (helix III) extends into the $Ca^{2+}$ binding loop. The ends of this second $Ca^{2+}$ binding loop are close in space (established by a NOE between Val61 $C^{\gamma}H_3$ and Asp54 $N^{\delta}H_2$); thus, helices III and IV are packed in antiparallel fashion. At the present stage of analysis no other NOEs that would further clarify the nature of the packing of these two helices are evident. However, both helix III and helix IV have close contacts with helix II, as indicated by NOEs between Phe50 and Leu28 and between Ile73 and Phe36, respectively. Contacts between helix I and helix IV (NOEs between Lys7 and Phe63 and between Phe10 and Phe63) are also observed. The $\beta$-type contact between the two $Ca^{2+}$ binding sites acts as an anchor and serves to further align the two EF hands, resulting in a structure reminiscent of a four-helix bundle. To meet the criteria of appropriate pairwise helical contacts, the four-helix bundle has to be skewed, as sketched in Figure 7. The results of ongoing structure calculations based on all of the available NMR data should serve to fully clarify the details of the global fold in solution.

## Discussion

The global structure of bovine calbindin $D_{9k}$ in solution is clearly very similar to the published crystal structure (Szebenyi & Moffat, 1986). Of particular note is the excellent agreement of the solution and crystal structural analyses with respect to the identity and location of elements of secondary structure (Figures 5 and 6). The only significant difference is the absence in solution of the single, irregular turn of helix (F36–K41) in the linker segment connecting helices II and III. Insights into the cause of this discrepancy should be provided by ongoing calculations of the complete three-dimensional structure.

In the crystal structure, the N-termini of the four helices are stabilized by hydrogen bonds with the backbone amide groups of Glu5, Glu27, Glu48, and Glu65 as donors and the side-chain hydroxyl groups of Ser2, Ser24, Thr45, and Ser62, respectively, as acceptors. In solution, NOE connectivities can be observed between the backbone amide proton of Glu5 and one of the $C^{\beta}$ protons of Ser2, between the backbone amide proton of Glu27 and one of the $C^{\beta}$ protons of Ser24 in helix II, between the NH of Glu48 and the $C^{\gamma}$ methyl group of Thr45 in helix III, and between the Glu65 NH and the slowly exchanging hydroxyl proton of Ser62 in helix IV. These NOEs indicate a close spatial proximity of the corresponding protons and strongly suggest that the helix-stabilizing hydrogen bonds also exist in solution. We note that in the N-terminal portion of helix I there is a proline residue which is normally considered helix breaking (Chou & Fasman, 1978). However, this proline occurs at precisely the "high preference" N-cap +1 position found by Richardson and Richardson (1988).

Table II: Discrepancies in Minor A Bovine Calbindin $D_{9k}$ Assignments

| resonances (ppm) | previous assignments[a] | this work |
|---|---|---|
| 10.12 | G18 NH | S24 NH |
| 5.51 | G18 $C^{\alpha 1}H$ | L23 $C^{\alpha}H$ |
| 2.41 | G18 $C^{\alpha 2}H$ | V61 $C^{\beta}H$ |
| 0.58 | L23 $C^{\delta 1}H_3$ | L69 $C^{\delta}H_3$ |
| 5.23 | S24 $C^{\alpha}H$ | F36 $C^{\alpha}H$ |
| 7.17 | F50 ring | F36 $C^{\delta}H$ |

[a] From Dalgarno et al. (1983).

The hydrophobic core in solution is similar to that found in the X-ray crystal structure (Szebenyi & Moffat, 1986). For example, the $C^{\delta}$ protons of Phe10, Phe63, and Phe66 are all within 5 Å from each other as deduced from NOESY experiments and observed in the crystal structure. Even hydrogen-bonding patterns appear to be faithfully reproduced. For example, of the 18 amide protons defined as not hydrogen bonding (or possibly hydrogen bonded to water molecules) in the crystal structure, only 1, Gln22, is found to exchange slowly. This residue is part of the short $\beta$-sheet, but the backbone NH is facing out into the solution (see Figure 6). The slow exchange of Gln22 NH might be due to hydrogen bonding to the carboxyl group of Asp19, suggested by the observed NOE connectivities between this amide proton and one of the $C^{\beta}$ protons of Asp19.

In addition to comparison to the crystal structure, we are able to compare our solution data with the independent study of native porcine calbindin $D_{9k}$ (Drakenberg et al., 1989). As would be expected on the basis of the very high (>90%) sequence homology between the bovine and porcine proteins (Hoffmann et al., 1979; Fullmer & Wasserman, 1981), we find that the solution structures are essentially identical at this low level of resolution. The similarity of the minor A and intact structures is important, validating the large number of studies that have been carried out on the minor A form of the protein.

A comparison of previous $^1H$ chemical shift assignments based on the available crystal structure (Szebenyi et al., 1981) and 1D NOE experiments (Dalgarno et al., 1983) versus the complete assignments reported here is informative. In the earlier studies, no evidence of cis–trans isomerism at Pro43 could be found, and of the 19 resonances assigned to specific protons, there are six obvious discrepancies (Table II). In one instance, two resonances (at 10.12 and 2.41 ppm) were assigned to the wrong protons as a consequence of another misassigned resonance (5.51 ppm). These three missassignments were in part based on the fact that the NH resonance at 10.12 ppm appears as a multiplet and was therefore assumed to belong to a glycine spin system. In fact, the signal observed at 10.12 ppm consists of the doublet resonance of Ser24 of the major (*trans*-Pro43) form partially overlapped by the Ser24 doublet NH resonance from the minor (*cis*-Pro43) form, which is shifted downfield by 0.02 ppm at pH 6.0 and 300 K. Calbindin $D_{9k}$ represents a very clear example of the necessity

of obtaining a complete and self-consistent set of assignments for the entire $^1H$ spectrum to make reliable assignments.

Ikura et al. (1987) have proposed that a low-field shift for the amide proton resonance of the glycine residue in the sixth position of the $Ca^{2+}$ binding loop is characteristic of regular EF hands. The chemical shift observed for the Gly59 NH in loop 2 of calbindin $D_{9k}$ is consistent with this proposed diagnostic. The absence of such a characteristic NH shift for the first $Ca^{2+}$ binding site is expected, since this loop differs slightly from the EF-hand template by having two residues (Pro20 plus Asp21) at the sixth (consensus glycine) position. The chemical shift diagnostic for regular EF hands has been extended to include $C^{\alpha}H$ chemical shifts >4.75 ppm for residues 7 and 8 of the $Ca^{2+}$ binding loop (Drakenberg et al., 1989).

In summary, an integrated approach to spin system identification and the sequential resonance assignment procedure provide the complete sequence-specific assignment of the $^1H$ NMR spectrum of the major (*trans*-Pro43) form of recombinant minor A bovine calbindin $D_{9k}$. The 71 backbone amide protons and all 486 nonlabile side-chain protons have been assigned, along with the resonances of all 12 side-chain amide protons and the $O^{\gamma}$ proton of Ser62. All assignments are unambiguous, apart from unproven degeneracy of the $C^{\gamma}H_2$ resonances of Met0, Lys1, Pro37, Glu64, and Gln75. Sixty-two of the 76 spin systems could be assigned to the appropriate amino acid, while the remaining 14 were identified as either glutamine or glutamic acid residues. This greatly facilitated the sequential resonance assignment. The complete assignments provide the necessary background for in-depth analysis of solution structure and dynamics. The results described herein provide a starting point for ongoing studies to determine high-resolution three-dimensional structures in solution using distance geometry and molecular dynamics calculations with distance constraints from NMR. The determination of the conformational (and dynamical) changes that accompany binding of $Ca^{2+}$ are of particular importance, since this should provide information, directly and with high resolution, on the structural basis for the function of the calmodulin superfamily of calcium regulatory proteins.

## ACKNOWLEDGMENTS

We gratefully acknowledge Dr. Theo Hofmann for supplying the sample of calbindin $D_{9k}$ from bovine intestines, Peter Brodin and Dr. Thomas Grundström for expression and isolation of r-calbindin, Eva Thulin for purification of the calbindin samples, Linda Tenant for assistance in sample preparation, Dr. Mark Rance for continued support with experimental methods, and Dr. Torbjörn Drakenberg for helpful discussions. We thank Kathy Carpenter for assistance in preparing the manuscript and Yvonne Andréasson for preparing Figure 7.

## SUPPLEMENTARY MATERIAL AVAILABLE

Table of acquisition parameters for the 2D experiments (1 page). Ordering information is given on any current masthead page.

## REFERENCES

Berridge, M. J. (1987) *Annu. Rev. Biochem. 56*, 159–193.

Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.

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

Braunschweiler, L., Bodenhausen, G., & Ernst, R. R. (1983) *Mol. Phys. 48*, 535–560.

Brodin, P., Grundström, T., Hofmann, T., Drakenberg, T., Thulin, E., & Forsén, S. (1986) *Biochemistry 25*, 5371–5377.

Chazin, W. J., & Wright, P. E. (1987) *Biopolymers 26*, 973–977.

Chazin, W. J., Rance, M., & Wright, P. E. (1987) *FEBS Lett. 222*, 109–114.

Chazin, W. J., Rance, M., & Wright, P. E. (1988) *J. Mol. Biol. 202*, 603–622.

Chazin, W. J., Kördel, J., Drakenberg, T., Thulin, E., Brodin, P., Grundström, T., & Forsén, S. (1989a) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2195–2198.

Chazin, W. J., Kördel, J., Drakenberg, T., Thulin, E., Brodin, P., Grundström, T., Hofmann, T., & Forsén, S. (1989b) *Biochemistry* (in press).

Chou, P. Y., & Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251–276.

Dalgarno, D. C., Levine, B. A., Williams, R. J. P., Fullmer, C. S., & Wasserman, R. H. (1983) *Eur. J. Biochem. 137*, 523–529.

Drakenberg, T., Hofmann, T., & Chazin, W. J. (1989) *Biochemistry 28*, 5946–5954.

Forsén, S., Vogel, H., & Drakenberg, T. (1986) Biophysical Studies of Calmodulin, in *Calcium and Cell Function* (Cheung, W. Y., Ed.) Vol. 6, pp 113–157, Academic Press, New York.

Forsén, S., Linse, S., Thulin, E., Lindegård, B., Martin, S. R., Bayley, P. M., Brodin, P., & Grundström, T. (1988) *Eur. J. Biochem. 177*, 47–55.

Fullmer, C. S., & Wasserman, R. H. (1981) *J. Biol. Chem. 256*, 5669–5674.

Herzberg, O., & James, M. N. G. (1985) *Biochemistry 24*, 5298–5302.

Hitchman, A. J. W., Kerr, M. K., & Harrison, J. E. (1973) *Arch. Biochem. Biophys. 155*, 221–222.

Hofmann, T., Kawakami, M., Hitchman, A. J. W., Harrison, J. E., & Dorrington, K. J. (1979) *Can. J. Biochem. 57*, 737–748.

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

Kretsinger, R. H. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 499–510.

Kretsinger, R. H., & Nockolds, C. E. (1973) *J. Biol. Chem. 248*, 3313–3326.

Levitt, M., & Freeman, R. (1979) *J. Magn. Reson. 33*, 473–476.

Linse, S., Brodin, P., Drakenberg, T., Thulin, E., Sellers, P., Elmdén, K., Grundström, T., & Forsén, S. (1987) *Biochemistry 26*, 6723–6735.

Linse, S., Brodin, P., Johansson, C., Thulin, E., Grundström, T., & Forsén, S. (1988) *Nature 335*, 651–652.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.

Marsden, B. J., Hodges, R. S., & Sykes, B. D. (1988) *Biochemistry 27*, 4198–4206.

Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson. 66*, 187–193.

Pardi, A., Wagner, G., & Wüthrich, K. (1983) *Eur. J. Biochem. 137*, 445–454.

Rance, M. (1987) *J. Magn. Reson. 74*, 557–564.

Rance, M., & Wright, P. E. (1986) *J. Magn. Reson. 66*, 372–378.

Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 131*, 1094–1102.

Rasmussen, H. (1986a) *N. Engl. J. Med. 314*, 1094–1101.

Rasmussen, H. (1986b) *N. Engl. J. Med. 314*, 1164-1170.
Richardson, J. S., & Richardson, D. C. (1988) *Science 240*, 1648-1652.
Seamon, K. B., & Kretsinger, R. H. (1983) Calcium Binding Proteins, *Met. Ions Biol. 6*, 1-52.
Szebenyi, D. M. E., & Moffat, K. (1986) *J. Biol. Chem. 261*, 8761-8777.
Szebenyi, D. M. E., Obendorf, S. K., & Moffat, K. (1981) *Nature 294*, 327-332.
Wagner, G. (1983) *J. Magn. Reson. 55*, 151-156.
Wendt, B., Hofmann, T., Martin, S. T., Bayley, P., Brodin, P., Grundström, T., Thulin, E., Linse, S., & Forsén, S. (1988) *Eur. J. Biochem. 175*, 439-445.
Wüthrich, K. (1983) *Biopolymers 22*, 131-138.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715-740.

# Assignment of the Proton NMR Spectrum of Reduced and Oxidized Thioredoxin: Sequence-Specific Assignments, Secondary Structure, and Global Fold†

H. Jane Dyson,‡ Arne Holmgren,§ and Peter E. Wright*,‡

*Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, California 92037, and Department of Physiological Chemistry, Karolinska Institutet, S-10401 Stockholm, Sweden*

*Received February 6, 1989; Revised Manuscript Received May 5, 1989*

ABSTRACT: Complete proton assignments are reported for the $^1$H nuclear magnetic resonance (NMR) spectrum of *Escherichia coli* thioredoxin in the oxidized (with active-site disulfide bridge) and reduced (with two sulfhydryl groups) states. The assignments were obtained by using an integrated assignment strategy in which spin systems were identified from a combination of relayed and multiple quantum NMR techniques prior to sequential assignment. Elements of secondary structure were identified in each protein from characteristic nuclear Overhauser effects (NOE), coupling constants, and slowly exchanging amide protons. In both oxidized and reduced thioredoxin, approximately 33% of the 108 amino acid residues participate in a $\beta$-sheet containing four major strands (three antiparallel and one parallel). A further short $\beta$-strand is connected in a parallel fashion at the N-terminal end of the sheet. Two of the antiparallel $\beta$-strands are connected by a 7-residue $\beta$-bulge loop. Three helical segments, also containing approximately 33% of the amino acid residues, are well-defined in both oxidized and reduced thioredoxin. The remaining third of the molecule apparently consists of reverse turns and loops with little defined secondary structure. The global folds of oxidized and reduced thioredoxin are shown to be essentially identical. Both NOE connectivities and chemical shift values for the two proteins are very similar, except in the immediate vicinity of the active site where significant variations in the chemical shift indicate subtle conformational changes. While the overall fold of oxidized thioredoxin is the same in solution and in the crystalline state, some small differences in local conformation are apparent.

Thioredoxin is a ubiquitous protein ($M_r$ = 11 700) with many functions, particularly in thiol-dependent redox reactions [for a review, see Holmgren (1985)]. The active site of thioredoxin, Cys-Gly-Pro-Cys, contains a disulfide in the oxidized form. This is usually reduced by NADPH and the FAD-containing enzyme thioredoxin reductase to a dithiol in the reduced form of thioredoxin. Reduced thioredoxin is a powerful general protein disulfide reductase and also a hydrogen donor for ribonucleotide reductase, methionine sulfoxide reductases, and sulfate reductase. In addition, the reduced form of *Escherichia coli* thioredoxin is essential for phage T7 DNA replication as a subunit of T7 DNA polymerase and for assembly of filamentous phages f1 and M13.

The 108-residue amino acid sequence of *E. coli* thioredoxin (Holmgren, 1968) and the nucleotide sequence of the corresponding *trxA* gene (Höög et al., 1984; Lim et al., 1985) are known. The three-dimensional structure of oxidized thioredoxin has been solved by X-ray crystallography at 2.8-Å resolution (Holmgren et al., 1975). The active-site disulfide (between Cys 32 and Cys 35) is located in a protrusion of the molecule at the end of a $\beta$-strand and immediately followed by an $\alpha$-helix. The X-ray structure of oxidized thioredoxin shows about 75% of the molecule in well-defined secondary structure elements, including four helices and five strands of $\beta$-sheet. Thioredoxin has been isolated from many different organisms and, for example, the primary structures of about 10 prokaryotic thioredoxins show about 50% positional identity to *E. coli* thioredoxin [for a review, see Gleason and Holmgren (1988)], consistent with a protein ancient on the evolutionary time scale, with a conserved global fold.

The three-dimensional structure of reduced thioredoxin is not known, since it has not been possible to crystallize the protein in this state. Previous spectroscopic studies (Stryer et al., 1967; Holmgren, 1972; Holmgren & Roberts, 1976) suggest that no gross structural changes occur on reduction of thioredoxin. Functional data show that enzymes such as thioredoxin reductase and ribonucleotide reductase discriminate between oxidized and reduced thioredoxin. Furthermore, only the reduced conformation of thioredoxin is able to activate gene 5 protein of phage T7 DNA polymerase (Mark &

†Supported by Grant GM 36643 from the National Institutes of Health and Grant 13x - 3529 from the Swedish Medical Research Council and by the Karolinska Institutet.
‡Research Institute of Scripps Clinic.
§Karolinska Institutet.

0006-2960/89/0428-7074$01.50/0 