Protein Science (1999), *8*:2580–2588. Cambridge University Press. Printed in the USA.

NMR assignments, secondary structure, and global fold of calerythrin, an EF-hand calcium-binding protein from *Saccharopolyspora erythraea*

HELENA AITIO,[1] ARTO ANNILA,[2] SAMI HEIKKINEN,[1] EVA THULIN,[3] TORBJÖRN DRAKENBERG,[2,3] AND ILKKA KILPELÄINEN[1]

[1]Institute of Biotechnology/NMR Laboratory, P.O. Box 56, FIN-00014 University of Helsinki, Finland
[2]VTT/Chemical Technology, P.O. Box 1401, FIN-02044 VTT, Finland
[3]Department of Physical Chemistry 2, Chemical Centre, Lund University, P.O. Box 124, S-22100 Lund, Sweden

(RECEIVED May 3, 1999; ACCEPTED September 6, 1999)

Abstract

Calerythrin is a 20 kDa calcium-binding protein isolated from gram-positive bacterium *Saccharopolyspora erythraea*. Based on amino acid sequence homology, it has been suggested that calerythrin belongs to the family of invertebrate sarcoplasmic EF-hand calcium-binding proteins (SCPs), and therefore it is expected to function as a calcium buffer. NMR spectroscopy was used to obtain structural information on the protein in solution. Backbone and side chain ^{1}H, ^{13}C, and ^{15}N assignments were obtained from triple resonance experiments HNCACB, HN(CO)CACB, HNCO, CC(CO)NH, and [^{15}N]-edited TOCSY, and HCCH-TOCSY. Secondary structure was determined by using secondary chemical shifts and characteristic NOEs. In addition, backbone N-H residual dipolar couplings were measured from a spin-state selective [^{1}H, ^{15}N] correlation spectrum acquired from a sample dissolved in a dilute liquid crystal. Four EF-hand motifs with characteristic helix-loop-helix patterns were observed. Three of these are typical calcium-binding EF-hands, whereas site 2 is an atypical nonbinding site. The global fold of calerythrin was assessed by dipolar couplings. Measured dipolar couplings were compared with values calculated from four crystal structures of proteins with sequence homology to calerythrin. These data allowed us to recognize an overall similarity between the folds of calerythrin and sarcoplasmic calcium-binding proteins from the sandworm *Nereis diversicolor* and the amphioxus *Branchiostoma lanceolatum*.

Keywords: calcium-binding protein; calerythrin; EF-hand; NMR; residual dipolar couplings; secondary structure

An EF-hand is defined as a helix-loop-helix structure, in which the 12-residue loop binds one calcium ion with a seven-coordinate geometry (Kretsinger & Nockolds, 1973). EF-hands usually occur in pairs even though one of the hands may not bind calcium, and proteins containing both single and multiple copies of pairs of EF-hands have been observed (Celio, 1996). Several proteins with two pairs of EF-hands have been well characterized and found to have different folds. Calmodulin, which functions as a mediator in cell signaling, has a dumbbell like shape, with two lobes connected by a long helix (Babu et al., 1988). The central region of the connecting helix is flexible in solution, as has been shown by NMR spectroscopy (Ikura et al., 1991; Barbato et al., 1992). Upon complex formation with target peptides, the flexible linker region allows the peptide to be sequestered into a hydrophobic channel formed by the two domains, resulting in a more globular structure (Ikura et al., 1992; Meador et al., 1992). Sarcoplasmic calcium-binding proteins (SCPs) (Hermann & Cox, 1995) fold into a globular structure, with the two halves oriented relative to each other to present the calcium-binding sites on opposite sides of the molecule. SCPs are four-EF-hand proteins found in the muscle of various invertebrates and probably function as intracellular calcium buffers. Not all of the SCPs' EF-hands are functional. Ca^{2+}-binding has been preserved in the first and third EF-hands of all known SCPs, and most SCPs also have a third site in either the second or fourth EF-hand. The nonfunctional site has noncoordinating amino acids that dramatically decrease calcium ion affinity. A still different fold of four-EF-hand proteins exists in recoverin (Flaherty et al., 1993), the calcium sensor in vision. Recoverin is one member of an emerging branch of the EF-hand superfamily,

Reprint requests to: Torbjörn Drakenberg, Department of Physical Chemistry 2, Chemical Centre, Lund University, P.O. Box 124, S-22100 Lund, Sweden; e-mail: trbjrn@afkem2.lth.se.

Abbreviations: B-SCP, sarcoplasmic calcium-binding protein from *Branchiostoma lanceolatum*; CSI, chemical shift index; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; N-SCP, sarcoplasmic calcium-binding protein from *Nereis diversicolor*; SCP, sarcoplasmic calcium-binding protein; TOCSY, total correlation spectroscopy.

which includes, for example, S-modulin (Kawamura et al., 1993) and visinin (Yamagata et al., 1990). It has a globular structure like the SCPs, but the calcium-binding sites are situated on the same side of the molecule.

Calerythrin is a 20 kDa (176 residues) calcium-binding protein isolated from the gram-positive bacterium *Saccharopolyspora erythraea* (Leadlay et al., 1984). Four potential EF-hands were predicted from its primary structure and it showed, on the whole, a marked homology to calmodulin (Swan et al., 1987). The loop of the second EF-hand motif does not contain, however, the usual calcium-binding residues. Positions 1 and 3 of the 12-residue loop are commonly occupied by aspartates and asparagines, which provide their oxygen-containing side chains for calcium ion coordination. In calerythrin these are replaced by glycines. Therefore, it was assumed that this site might not bind calcium. This was later confirmed with a ^{113}Cd NMR spectroscopic study (Bylsma et al., 1992). A sequence similarity to SCPs was also noticed (Cox & Bairoch, 1988). The atypical EF-hand of calerythrin has also been noted to bear a close similarity (almost 50% sequence identity) to the highly conserved consensus sequence of the annexin family of proteins (Moss & Crumpton, 1990).

To elucidate the structural relationship of calerythrin with other EF-hand calcium-binding proteins, we have undertaken to characterize the three-dimensional (3D) structure of calcium-saturated calerythrin in solution. We present here the ^{1}H, ^{13}C, and ^{15}N NMR assignments and the secondary structure based on secondary chemical shifts, characteristic NOEs, and residual dipolar couplings. Dipolar couplings were also used to gain insight into the global fold of calerythrin.

Results

Assignment

Main chain NH, N, Cα, and Cβ were assigned using HNCACB and HN(CO)CACB spectra (Yamazaki et al., 1994). A total of 163 residues out of 176 were assigned using the two spectra. Four additional backbone amide proton and nitrogen chemical shifts were identified based on the NH(i)–NH($i + 1$) correlations in the 3D [^{15}N]- and [^{15}N, ^{15}N]-edited NOESY spectra. Side-chain amide groups of asparagines and glutamines were assigned from the [^{15}N]-edited NOESY since these are not detected by the sensitivity enhanced gradient version of the HN(CO)CACB experiment optimized for AX-spin systems.

During the assignment it became clear that the amino acid sequence differed from that reported by Swan et al. (1987). The first methionine is missing, threonine 108 is replaced by an isoleucine and tryptophan 109 by a valine. The sequence obtained from the assignment was verified by determination of the sequence of a peptide obtained from an endoproteinase GluC digest of calerythrin, containing amino acids 95–124 (ASFNRVLGPVVKGTWGM CDKNADGQINADE).

The ^{15}N–^{1}H HSQC spectrum of calerythrin (Fig. 1) shows, in all, 200 peaks. Despite apparent overlap, especially in the central region of the heteronuclear single quantum coherence (HSQC) spectrum, 167 out of the 176 backbone amide shifts were assigned, leaving only the amides of the two N-terminal residues (T1, T2), and two stretches of five (R163-F167), and two (L174, L175) residues near the C-terminus unassigned. These amides remained unobservable probably due to line broadening caused by a hitherto unidentified exchange process. Three glycine residues (G22, G117, and G151) have amide proton and nitrogen chemical shifts ($\delta_H \approx$ 10.4 ppm, $\delta_N \approx$ 115 ppm) typical for glycines in position six of EF-hand calcium-binding loops. Thirty (15 pairs) of the remaining cross peaks in the HSQC spectrum belong to side-chain amides of nine asparagines and six glutamines. One resonance originates from the HN$^{\epsilon 1}$ of either W16 or W127 and one from the HN$^{\epsilon}$ of an arginine, aliased from lower ^{15}N frequencies. The latter two peaks remain so far unassigned due to lack of NOEs in the [^{15}N]-edited NOESY spectrum.

The carbonyl shifts, 163 out of 176, were obtained from HNCO. Side-chain proton and carbon chemical shifts were assigned using a combination of three spectra. Partial proton and carbon side-chain assignments obtained from [^{15}N]-edited TOCSY and CC(CO)NH were subsequently complemented with those from HCCH-TOCSY. Aromatic side-chains were not assigned at this stage. Chemical shifts of all assigned ^{1}H, ^{13}C, and ^{15}N nuclei are listed in the table in Supplementary material in the Electronic Appendix and have also been deposited in the BioMagResBank (accession number 4335).

Secondary structure and global fold

Hα, Cα, and C$'$ chemical shift indices (Wishart et al., 1992; Wishart & Sykes, 1994) indicate that calerythrin is composed of eight α-helices (helix A, A3–R15; B, R26–G41; C, A47–G68; D, E77–F89; E, V103–G109; F, A121–A130; G, K135–Q144; H, L155–A161) (Fig. 2). Furthermore, the indices suggest the presence of an additional helical region between helices D and E (residues E93–V99). Helices B–H are also revealed by the presence of many characteristic short (medium to strong NN(i, $i + 1$))- and medium-range (medium αN(i, $i + 1$), αN(i, $i + 3$)) NOEs (Wüthrich, 1986) observed in [^{15}N]-edited NOESY and confirmed with [^{15}N, ^{15}N]- and [^{15}N, ^{13}C]-edited NOESY in case of overlapping resonances. Altogether, the NOEs are generally in a good agreement with the chemical shift data, although some discrepancies regarding the exact beginnings and ends of helices can be observed. According to NOEs helix A begins only at I4 and helix C at E48, helix E has one additional residue at the N-terminal end and two at the C-terminal end and helix F continues at least to L131. The scarcity of characteristic NOEs in helix A might result from a lower stability of this N-terminal helix compared to the other helices. The backbone N–H dipolar couplings measured from a protein sample dissolved in a dilute liquid crystalline phase of lipid bicelles confirm the presence of eight helices (Fig. 2). In an α-helix the N–H bonds are, on loose criteria, uniformly oriented with respect to the axis of the helix. Since backbone N–H dipolar couplings are dependent on the orientation of the N–H bond vector relative to the principal axes of the alignment tensor, this regularity in the direction of the internuclear vectors results in similar values of dipolar couplings (Tolman et al., 1995; Tjandra et al., 1997; Tjandra & Bax, 1997). This is clearly observed in the dipolar couplings of calerythrin.

The chemical shift and NOE data do not allow us to unambiguously conclude how long the interhelical segment between helices D and E is, specifically whether the seven-residue segment E93–V99 is a part of either helix, due to the scarcity of αN(i, $i + 3$) and NN(i, $i + 1$) NOEs. The dipolar couplings imply that the polypeptide backbone is gradually changing its orientation over a segment of about 10 residues between helices D and E, which are roughly in a perpendicular orientation with respect to each other.

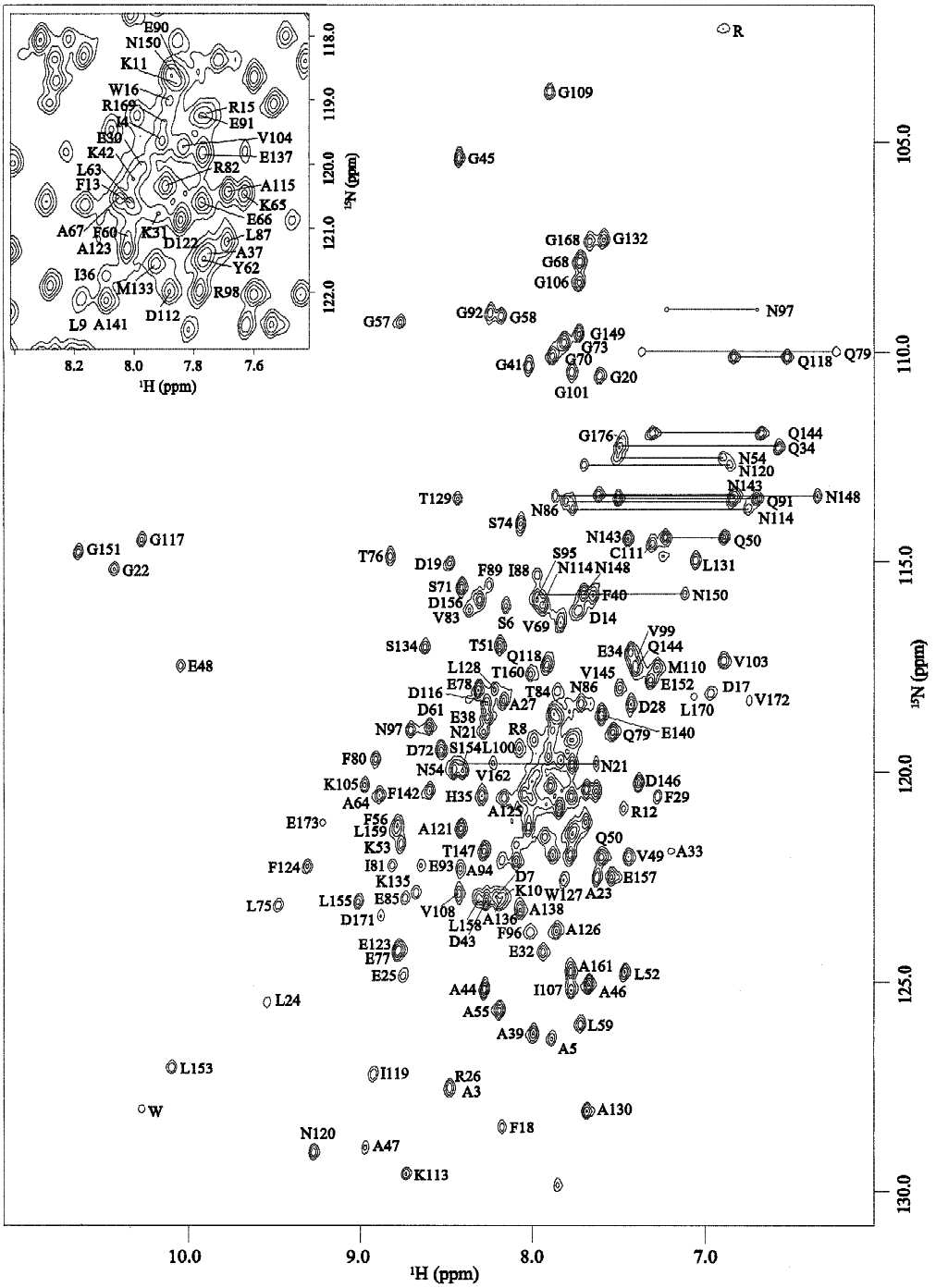

Fig. 1. ^{1}H-^{15}N HSQC spectrum of ^{15}N, ^{13}C double labeled calerythrin recorded at 600 MHz ^{1}H frequency. Assignments are indicated by residue type and number. The horizontal lines connect the side-chain NH_2 proton frequencies of asparagine and glutamine residues. Peaks originating from a tryptophan $NH^{\epsilon 1}$ and an arginine NH^{ϵ} are marked with W and R, respectively. An expansion of the most crowded region of the spectrum (δ_H 7.5–8.3, δ_N 117.5–123.0 ppm) is shown at the upper left hand corner of the spectrum.

Helix D is perpendicular to the long axis of the alignment tensor and helix E is, on the contrary, parallel to this axis.

Four short β-strands (A23–E25, G73–T76, Q118–N120, and E152–S154) were revealed by chemical shift indices, strong $\alpha N(i, i + 1)$, and the absence of $NN(i, i + 1)$ NOEs. (Interpretation concerning residue G73 was based on positive CSIs only.) Two of the strands form a small antiparallel β-sheet (Q118–N120 with E152–S154), as deduced from several long-range NOEs (N120 Hα–E152 Hα, N120 Hα–L153 HN, I119 HN–L153 HN). The two other extended structures (A23–E25, G73–T76) seem also to be spatially quite close to each other since a weak NOE was observed between the Hαs of E25 and S74 in the partially assigned [^{13}C]-

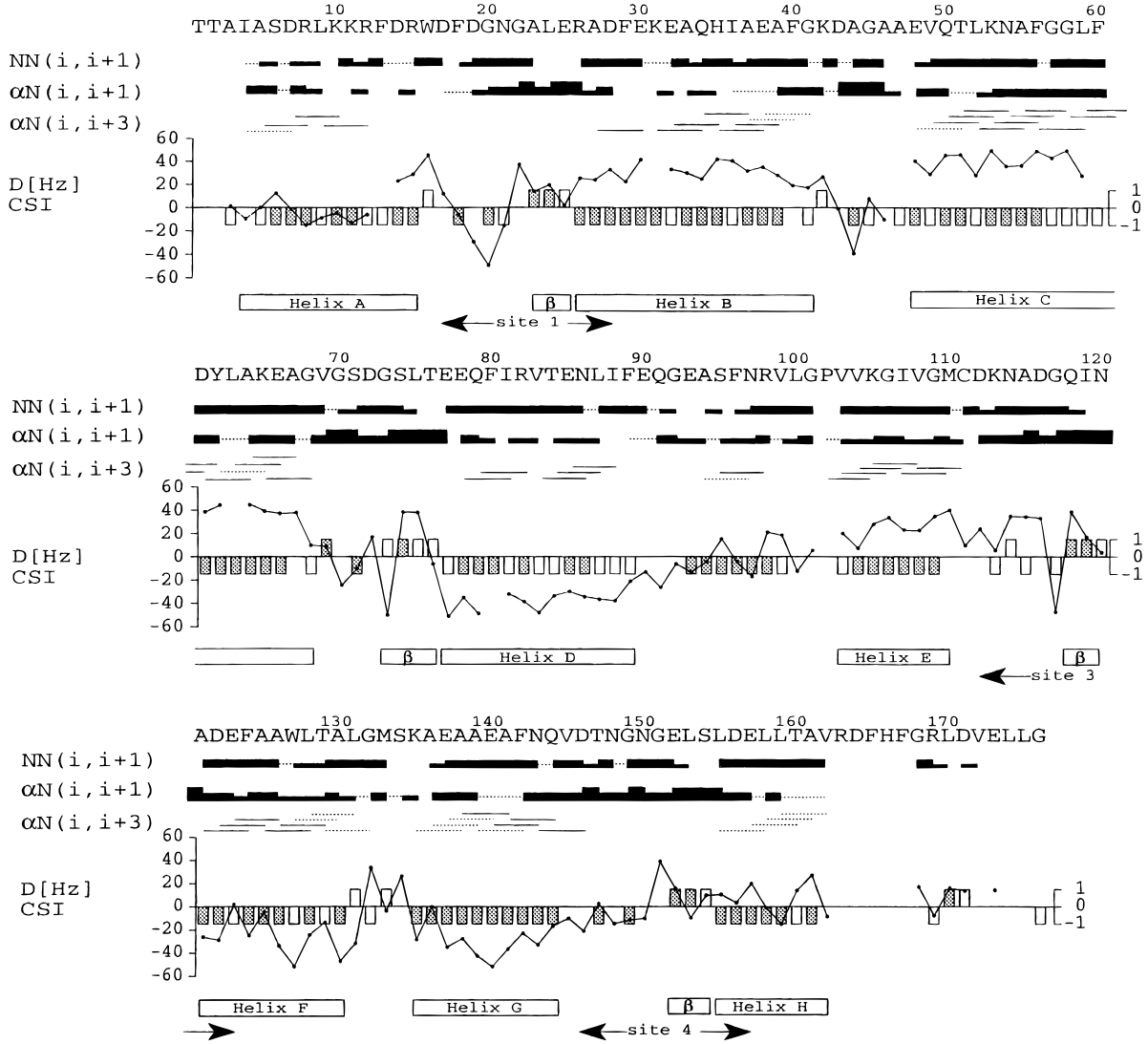

Fig. 2. Secondary structure prediction of calerythrin based on the chemical shift index, characteristic NOEs and residual dipolar couplings. For the CSI, filled boxes pointing down (up) imply that all three secondary chemical shift values (Hα, Cα, C′) represent α-helix (β-sheet). Open boxes mark that two out of three values represent a particular secondary structure. For the NOE data, the height of the bar corresponds to the intensity of the correlation (medium or weak for the NN(i, i + 1) peaks, strong, medium or weak for the αN(i, i + 1) peaks). The dashed lines indicate a tentative assignment, due to spectral overlap. The dipolar coupling values are represented with filled spheres connected with lines, gaps indicate undefined values. The secondary structure is represented with boxes.

edited NOESY spectrum. Presently, however, it cannot be defined as a β-sheet.

Characterization of the fold of calerythrin was based on comparison of measured dipolar couplings of calerythrin and corresponding values computed from X-ray crystal structures (Annila et al., 1999) of sarcoplasmic calcium-binding protein from sandworm *Nereis diversicolor* (N-SCP, Vijay-Kumar & Cook, 1992), SCP from amphioxus *Branchiostoma lanceolatum* (B-SCP, Cook et al., 1993), calmodulin (Babu et al., 1988), and recoverin (Flaherty et al., 1993), all in their Ca^{2+}-saturated forms (Fig. 3). Instead of relying merely on homology, sequences were aligned by matching corresponding secondary structures, in particular, the short β-strands in the EF-hand loops. Deletions were placed in turns, if possible. A good agreement was found between calerythrin and the N-SCP (Fig. 3A). All helices show similar mutual spatial orientations. A small deviation in the values of helix B is observed. For helices F–H there is substantial variation within a given helix in both experimental and calculated values. For most residues in the calcium-binding loops and turns, good agreement between measured and calculated values is found as well. An obvious misalignment by one residue occurs in the second atypical EF-loop. Calerythrin and B-SCP are also similar (Fig. 3B). The only pronounced difference is in mutual orientation of helices A and H. Although the second EF-hand is nonbinding in the N-SCP and the fourth in B-SCP, calerythrin's atypical second EF-loop matches much better with that of B-SCP than that of N-SCP. Neither calmodulin nor recoverin structures result in as good agreement as the SCPs. For calmodulin there is only sporadic similarities (Fig. 3C). The central helix of calmodulin has been shown to be flexible (Ikura et al., 1991; Barbato et al., 1992), thereby a single

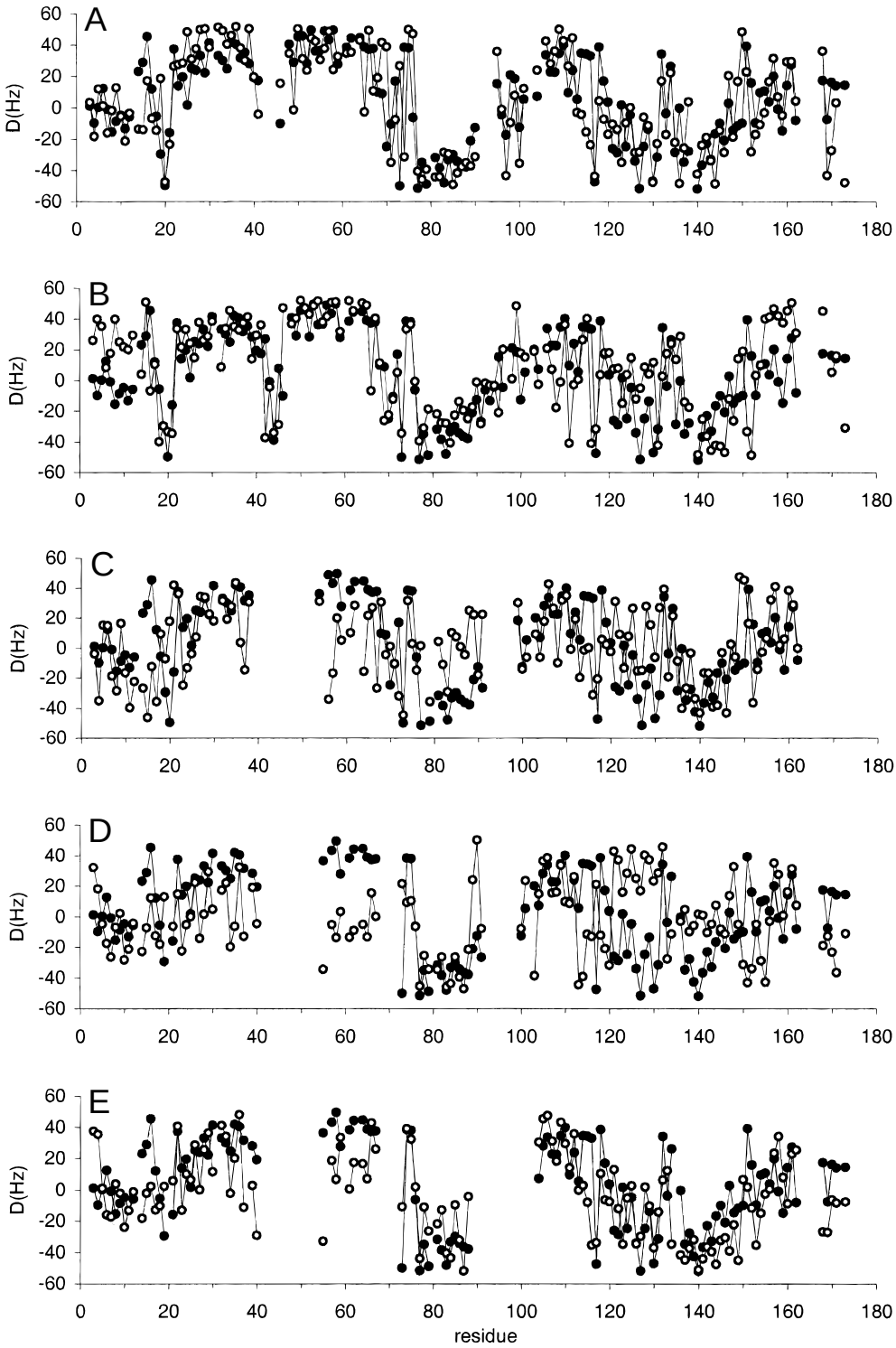

Fig. 3. Comparison of measured dipolar couplings of calerythrin (solid circles) with dipolar couplings computed (open circles) from the X-ray structures of **(A)** N-SCP, (**B**) B-SCP, (**C**) calmodulin, (**D**) recoverin, and (**E**) recoverin when the two halves (first half, residues 23–94; second half, residues 102–186) are calculated separately. Sequences were aligned according to secondary structures. Blanks correspond to residues without correspondence to the calculated protein (insertions) or for which dipolar couplings could not be determined due to spectral overlap, exchange phenomena, or inexistent coupling (prolines).

set of coordinates might not accurately represent the structure in solution. The two domains were therefore matched separately with the two halves of calerythrin, but no better agreement was found between measured and simulated data, although the small number of data points makes an accurate comparison difficult. Recoverin has helices A, D, and E in similar mutual orientations with caleryth-

rin (Fig. 3D). For recoverin, a better agreement between calculated and measured dipolar couplings was obtained for the C-terminal half of recoverin when the two halves were matched separately (Fig. 3E). This observation is consistent with the crystal structures of recoverin and N-SCP, which show that the C-terminal domains in the two proteins are more similar than the N-terminal domains.

Discussion

Long-range structural information is at this stage limited to the very few correlations between the short β-strands, since interhelical correlations are sparse in the [^{15}N]-edited NOESY. Thus, the comparison of the experimentally determined backbone N–H residual dipolar couplings with calculated values from crystal structures (Annila et al., 1999) of structurally potentially similar proteins, N-SCP (Vijay-Kumar & Cook, 1992), B-SCP (Cook et al., 1993), calmodulin (Babu et al., 1988), and recoverin (Flaherty et al., 1993), was most advantageous in view of the global fold determination (Fig. 3). This comparison allowed us to recognize an overall similarity between the folds of calerythrin and the sarcoplasmic proteins and to discard any potential similarity to the dumbbell form characteristic to calmodulin. Also, similarity to recoverin could be ruled out due to low match between measured and calculated values. The agreement between measured and calculated values was very good for the whole polypeptide length between calerythrin and N-SCP. This is not surprising in light of the high sequence identity (27%) between calerythrin and N-SCP. It is worth noting that B-SCP with a sequence identity to calerythrin of only 15% also shows a good overall similarity to calerythrin, whereas calmodulin and recoverin with sequence identities of 17 and 12%, respectively, do not. In Figure 4 the locations of dissimilarities are represented in a ribbon structure of N-SCP. Most residues with significant differences are located in loops and turns or in the first few residues immediately preceding or following the loops and turns. The slight dissimilarities observed between calerythrin and B-SCP could well be explained by the small differences in helix orientations observed between the two SCPs. Generally, some discrepancies can also be caused by different mobilities neglected in these calculations. Direct comparison of differences between different segments is although somewhat misleading, since the error between calculated and measured values is nonlinear, as discussed elsewhere (Annila et al., 1999).

The primary sequence of calerythrin (Swan et al., 1987) and a ^{113}Cd NMR study (Bylsma et al., 1992) indicate that only three of

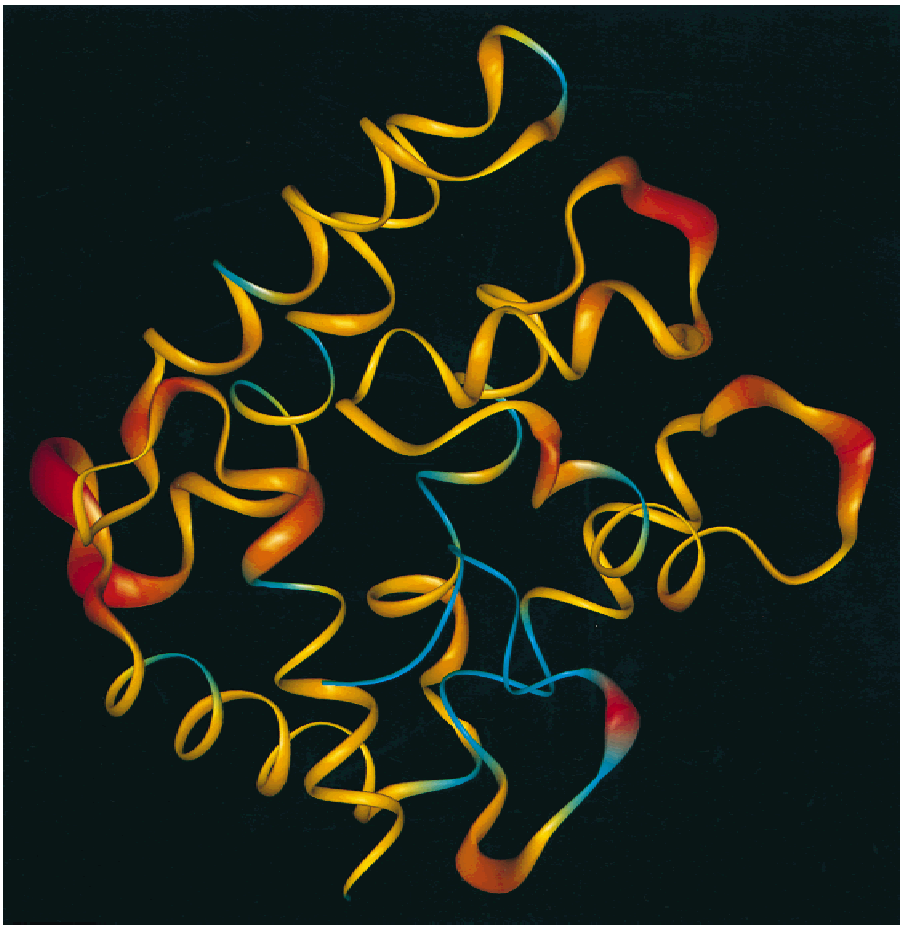

Fig. 4. Crystal structure of N-SCP, coded by differences observed between calculated (N-SCP) and measured (calerythrin) dipolar couplings. Darker color and wider ribbon indicate greater difference between calculated and measured values. Residues without correspondence to caleryhrin due to insertions and residues for which dipolar coupling could not be determined due to spectral overlap, exchange phenomena, or inexistent coupling (prolines) are colored in blue.

calerythrin's four potential EF-hands actually bind calcium and that site 2 (G68–Q79) is an atypical, nonbinding site. This conclusion is strengthened by the obtained chemical shifts of glycines in the sixth position of the EF-loops (G22, G73, G117, G151). Glycines situated in active loops have pronounced downfield shifts (G22 $\delta_H = 10.43$, $\delta_N = 115.0$; G117 $\delta_H = 10.27$, $\delta_N = 114.3$; G151 $\delta_H = 10.64$, $\delta_N = 114.6$ ppm compared to G73 $\delta_H = 7.82$, $\delta_N = 109.6$ ppm), likely caused by a hydrogen bond between backbone NH of the Gly residue and the side-chain carboxylate of Asp in loop position 1. Similar downfield chemical shifts of the glycine in the sixth position have been observed for example in calcium-saturated forms of calbindin D_{9k} (Kördel et al., 1989), calmodulin (Ikura et al., 1990), and troponin C (Slupsky et al., 1995). The main cause for inactivity of EF-hand 2 presumably is a replacement of residues with oxygen-containing side chains at loop positions 1 and 3 with aliphatic amino acids, glycines in calerythrin and alanines in N-SCP. The length of the nonhelical portion of the nonbinding loop corresponds to those of the binding loops, which was also the case with N- and B-SCPs.

Corresponding secondary structure elements are similar in calerythrin (Fig. 3) and N- and B-SCPs. The lengths of helices B–D and F–G are almost identical. The first and last helices are roughly one turn shorter in calerythrin than in the two sarcoplasmic proteins. The present view of the region between helices D and E suggests that helix E is also about one turn shorter in calerythrin. In these three proteins, three or four residues are in extended conformation in each of the EF-loops. As deduced from several long-range NOEs between the two strands, a small antiparallel β-sheet is formed between the functional sites 3 and 4 in calerythrin. A second antiparallel β-sheet might be formed between EF-loops 1 and 2, however presently less well defined. According to the crystal structures of the SCPs both these short β-sheets are present in N-SCP (Vijay-Kumar & Cook, 1992), whereas for B-SCP (Cook et al., 1993) a well-defined sheet is formed only between sites 1 and 2.

The region between helices D and E, which connects the two lobes, is determinative for the overall shape of a calcium-binding protein with two pairs of EF-hands. In SCPs this region contains a tight turn and the overall structure is globular (Vijay-Kumar & Cook, 1992; Cook et al., 1993). In calmodulin, two independent domains are connected by a long helix, with a flexible region of a few residues close to the middle (Ikura et al., 1991; Barbato et al., 1992). The present data point to two separate helices instead of a single long helix. A linker region of about 12 residues (E90–G101) is discerned from the dipolar couplings.

In both N- and B-SCPs, the polypeptide backbone past helix H forms a large loop on the surface. This loop ends with hydrophobic residues that are tucked into a hydrophobic pocket formed by several helices and the very last residues in the C-terminus are again on the surface. A similar structure in calerythrin could explain the disappearance of the backbone NH signals of residues R163–F167 and the reappearance for residues G168–E173. A loop structure on the surface of a protein is prone to movement and an exchange process on the order of 10^3 s^{-1} would cause severe line broadening. This explanation is strengthened by the observation of broad resonances for the residues close to this segment (V162, R169–E173, G176). Three residues remained unassigned in this irregular region also in N-SCP according to the recently published NMR assignments (Craescu et al., 1998).

EF-hand proteins have been divided in two groups according to the degree of conformational change they undergo upon Ca^{2+} binding (Ikura, 1996). The "sensors," including calmodulin, are involved in transducing calcium signals. In these proteins ion binding induces a large conformational change via reorientation of helices. The "buffers," e.g., calbindin D_{9k}, are involved in Ca^{2+} uptake, transport and buffering and only minor changes occur upon ion binding. It has been shown that the N-SCP is unstructured in the apo form at 27 °C and then to presumably switch to a nearly native state upon the binding of the first Ca^{2+} ion (Prêcheur et al., 1996). The chemical shift dispersion observed in the ^1H, ^{15}N HSQC spectrum of calcium-free calerythrin at 27 °C indicates that calerythrin is at least partially structured in the apo form. Interestingly, the chemical shifts of presumed EF-loop glycines in position 6 of calcium-free calerythrin, situated far left in the HSQC spectrum ($\delta_H = 11.1$, $\delta_N = 114.6$ and $\delta_H = 11.0$, $\delta_N = 115.3$ ppm), move toward even lower ^1H field upon Ca^{2+} removal, which is opposite to what is observed in, for example, calmodulin (Ikura et al., 1990) and calbindin D_{9k} (Kördel et al., 1989). In the future, more detailed comparison of the apo, partially and fully Ca^{2+}-saturated forms will form the basis for a classification of calerythrin in one of the aforementioned groups and as such give further insight into possible biological function of calerythrin, if any other than a Ca^{2+} buffer.

In summary, the presented structural data show that calerythrin belongs to the family of sarcoplasmic Ca^{2+}-binding proteins, although only a limited sequence homology exists between the primary sequences. It is thus clear that calerythrin has a compact rather than a dumbbell shape.

Materials and methods

Sample preparation

Calerythrin was produced uniformly ^{13}C, ^{15}N labeled in *Escherichia coli*. *E. coli* K38 containing a plasmid (pCBM1, pG-P1-2) overexpressing the gene for calerythrin (Swan et al., 1989) was grown in M9 minimal medium with 2*TY as an over night culture. In M9 glucose and NH_4Cl were substituted with 1.5 g/L of 99% ^{13}C-glucose and 1.0 g/L of 99% $^{15}NH_4Cl$. The growth took twice as long time as in rich medium. The cells were harvested by centrifugation (10 min, 6,000 *g*) at 4 °C. The pellet was suspended in 10 volumes of buffer 1 (10 mM PIPES-KOH, 1 mM $CaCl_2$, 2 mM beta-mercaptoethanol, pH 6.2), and sonicated 5 min on ice (Branson sonifier, 30 min, 80% duty cycle, output control 10) in 50 mL portions, followed by centrifugation 10 min at 27,000 *g* at 4 °C. The supernatant was poured into an equal volume of boiling buffer 1 and heated to 70 °C and then immediately put on ice. The precipitated proteins were removed by centrifugation 10 min 27,000 *g* at 4 °C. The supernatant was pumped onto a DEAE-cellulose column (10 mL DEAE-cellulose/mL pellet), equilibrated in buffer 1 and eluted with a linear salt gradient in buffer 1 ending at 0.5 M NaCl (total gradient volume was five times DEAE-cellulose volume). The fractions containing calerythrin were pooled and concentrated to 10 mL. The protein was then applied to a Sephadex G50 superfine column 3.4 × 180 cm in 50 mM ammonium acetate pH 6.0. The pure calerythrin fractions were pooled and lyophilized. The lyophilized protein was dissolved in 3 mL 0.1 M $CaCl_2$ at pH 7 and desalted on a 3.4 × 22 cm Sephadex G25 Superfine column equilibrated in H_2O. The calerythrin fractions were lyophilized. The purity of the protein was checked as described by Bylsma et al. (1992).

Calerythrin was digested with endoproteinase GluC and the obtained peptides were purified with HPLC and analyzed with MALDI-TOF mass spectrometry. As the molecular weight of one of the obtained peptides (aa. 95–124, SWISS-PROT sequence: ASFNRVLGPVVKGTWGMCDKNADGQINADE), differed from expected value, this peptide was sequenced with a protein sequenator. The peptide contained two replacements: T108 → I108 and W109 → V109, which matches well with the NMR assignments.

NMR spectra were measured from 1 mM protein samples dissolved in 250 μL of 4 mM $CaCl_2$, 10 mM DTT, 97% H_2O/3% D_2O at pH 6.0. The concentration of the calcium-free sample was ~0.6 mM, in 3 mM EDTA and 5 mM DTT in 95% H_2O/5% D_2O at pH 6.0. The liquid crystalline sample was prepared by dissolving ^{15}N-labeled calerythrin to 0.3 mM concentration in 5% w/v solution of diheptanoyl phosphatidylcholine (D7PC) and dimyristoyl phosphatidylcholine (DMPC) in approximately 1:3 molar ratio.

NMR spectroscopy

The ^{1}H-^{15}N HSQC, HNCACB (Muhandiram & Kay, 1994), HN(CO)CACB (Yamazaki et al., 1994), CC(CO)NH (Grzesiek et al., 1993), HNCO (Muhandiram & Kay, 1994), HCCH-TOCSY (Kay et al., 1993), [^{15}N]-edited NOESY (Zhang et al., 1994), and 3D [^{15}N, ^{13}C]-edited NOESY (derived from Kay et al., 1990; Jerala & Rule, 1995) spectra were acquired on a Varian Unity 600 spectrometer. The HSQC experiment was performed using 1,024(^{1}H)*256(^{15}N) points and 32 transients per increment. The triple resonance experiments were collected with 512 or 256 (^{1}H)*32(^{15}N)*64(^{13}C) real points and 16 or 32 transients per increment. The dimensions in the HCCH-TOCSY, [^{15}N]- and [^{15}N, ^{13}C]-edited NOESY spectra were 512(^{1}H)*64(^{13}C)*64(^{1}H), 512(^{1}H)*32(^{15}N)*128(^{1}H), and 512(^{1}H)*32(^{15}N)*56(^{13}C), respectively, acquired with 16 or 32 transients. Mixing times were 18 ms (CC(CO)NH), 22 ms (HCCH-TOCSY), 180 ms ([^{15}N]-edited NOESY), and 140 ms ([^{15}N,^{13}C]-edited NOESY). The spectral widths were 7,994 Hz for ^{1}H, with the carrier on water, referenced to 4.55 ppm, 1,800 Hz for ^{15}N, centered at 116 ppm, 9,000 Hz for $^{13}C\alpha/\beta$ (10,500 Hz in the [^{15}N, ^{13}C]-edited NOESY), centered at 43 ppm and 4,500 Hz for C′, centered at 179 ppm. The [^{15}N]-edited TOCSY (Zhang et al., 1994) and [^{15}N, ^{15}N]-edited NOESY (derived from Kay et al., 1990; Jerala & Rule, 1995) spectra were collected on a Varian Unity 500 spectrometer. The dimensions in the TOCSY experiment were 1,024(^{1}H)*32(^{15}N)*60(^{1}H) data points and in the NOESY 512(^{1}H)*55(^{15}N)*32(^{15}N), with 32 scans per increment in both spectra. The mixing times were 58 and 140 ms. The spectral widths were 8,000 Hz in ^{1}H dimension (carrier set at the frequency of water, 4.55 ppm), 1,700 Hz in ^{15}N (carrier set at 117 ppm). All spectra were collected at 45 °C. The ^{1}H chemical shifts were referenced to the water signal (4.55 ppm). The reference frequencies for ^{13}C and ^{15}N chemical shifts were calculated from the absolute proton frequency at 0 ppm using the value 0.25144954 (^{13}C) and 0.101329 (^{15}N) for the ratio of the gyromagnetic ratios.

Generally, the processing was performed by applying a squared cosine bell function in all dimensions. Linear prediction and zero-filling was used to improve the digital resolution of the spectra. A postacquisition water signal suppression by convolution of the time domain data was applied prior to Fourier transformation in all but the HCCH-TOCSY spectrum. All data were processed and analyzed with Felix 95.0 and 97.0 softwares (Biosym/Molecular Simulations, San Diego, California).

Dipolar couplings from calerythrin and corresponding computed values from X-ray structures were obtained as described in Annila et al. (1999). The orientation of the alignment tensor was obtained by a least-squares optimization of the measured and calculated values of dipolar couplings. Estimates of the axial component of the alignment tensor ($D_a = 27 \pm 3$) and the rhombicity ($R = 0.5 \pm 0.1$) used in the optimization were obtained from the distribution of measured dipolar couplings (Clore et al., 1998a, 1998b). Uncertainty in the parameters was noticed to be insignificant to the ranking of the fit between the homologous proteins. The following X-ray structures were used (Protein Data Bank entry code): N-SCP (2scp), B-SCP (2sas), calmodulin (1cll), and recoverin (1rec).

Supplementary material in the Electronic Appendix

Table 1 lists the ^{1}H, ^{13}C, and ^{15}N resonance assignments of calerythrin. Values marked in italics are tentative assignments.

Acknowledgments

We thank Dr. Jari Helin and Dr. Nisse Kalkkinen for the mass spectrometric analyses of calerythrin. We are grateful to Dr. Lewis E. Kay for providing the pulse sequences. This work was supported by the Academy of Finland and the Swedish National Sciences Research Council.

References

Annila A, Aitio H, Thulin E, Drakenberg T. 1999. Recognition of protein folds via dipolar couplings. *J Biomol NMR 14*:223–230.

Babu YS, Bugg CE, Cook WJ. 1988. Structure of calmodulin refined at 2.2 Å resolution. *J Mol Biol 204*:191–204.

Barbato G, Ikura M, Kay LE, Pastor RW, Bax A. 1992. Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy: The central helix is flexible. *Biochemistry 31*:5269–5278.

Bylsma N, Drakenberg T, Andersson I, Leadlay PF, Forsén S. 1992. Prokaryotic calcium-binding protein of the calmodulin superfamily. Calcium binding to a *Saccharopolyspora erythraea* 20 kDa protein. *FEBS Lett 299*:44–47.

Celio MR. 1996. *Guidebook to the calcium-binding proteins*. Oxford, UK: Oxford University Press, Inc.

Clore GM, Gronenborn AM, Bax A. 1998a. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J Magn Reson 133*:216–221.

Clore GM, Gronenborn AM, Tjandra N. 1998b. Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J Magn Reson 131*:159–162.

Cook WJ, Jeffrey LC, Cox JA, Vijay-Kumar S. 1993. Structure of a sarcoplasmic calcium-binding protein from amphioxus refined at 2.4 Å resolution. *J Mol Biol 229*:461–471.

Cox JA, Bairoch A. 1988. Sequence similarities in calcium-binding proteins. *Nature 331*:491.

Craescu CT, Prêcheur B, van Riel A, Sakamoto H, Cox JA, Engelborghs Y. 1998. ^{1}H and ^{15}N resonance assignment of the calcium-bound form of the *Nereis diversicolor* sarcoplasmic Ca^{2+}-binding protein. *J Biomol NMR 12*:565–566.

Flaherty KM, Zozulya S, Stryer L, McKay DB. 1993. Three-dimensional structure of recoverin, a calcium-sensor in vision. *Cell 75*:709–716.

Grzesiek S, Anglister J, Bax A. 1993. Correlation of backbone amide and aliphatic side-chain resonances in ^{13}C/^{15}N-enriched proteins by isotropic mixing of ^{13}C magnetization. *J Magn Reson B 101*:114–119.

Hermann A, Cox JA. 1995. Sarcoplasmic calcium-binding protein. *Comp Biochem Physiol 111B*:337–345.

Ikura M. 1996. Calcium binding and conformational response in EF-hand proteins. *Trends Biochem Sci 21*:14–17.

Ikura M, Clore GM, Gronenborn AM, Zhu G, Klee CB, Bax A. 1992. Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science 256*:632–638.

Ikura M, Kay LE, Bax A. 1990. A novel approach for sequential assignment of ^{1}H, ^{13}C, and ^{15}N spectra of larger proteins: Heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry 29*:4659–4667.

Ikura M, Spera S, Barbato G, Kay LE, Krinks M, Bax A. 1991. Secondary structure and side chain ^{1}H and ^{13}C resonance assignments of calmodulin in solution by heteronuclear multidimensional NMR spectroscopy. *Biochemistry 30*:9216–9228.

Jerala R, Rule GS. 1995. A 3D ^{1}H, ^{15}N and ^{13}C NOESY correlating experiment. *J Magn Reson 108*:294–298.

Kawamura S, Hisatomi O, Kayada S, Tokunaga F, Kuo CH. 1993. Recoverin has S-modulin activity in frog rods. *J Biol Chem 268*:14579–14582.

Kay LE, Clore M, Bax A, Gronenborn AM. 1990. Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1β in solution. *Science 249*:411–414.

Kay LE, Xu GY, Singer AU, Muhandiram DR, Forman-Kay JD. 1993. A gradient-enhanced HCCH-TOCSY experiment for recording side-chain ^{1}H and ^{13}C correlations in H_2O samples of proteins. *J Magn Reson B 101*:333–337.

Kördel J, Forsén S, Chazin WJ. 1989. ^{1}H NMR sequential resonance assignments, secondary structure, and global fold in solution of the major (*trans*-pro43) form of bovine calbindin D_{9k}. *Biochemistry 28*:7065–7074.

Kretsinger RH, Nockolds CE. 1973. Carp muscle calcium-binding protein. II. Structure determination and general description. *J Biol Chem 248*:3313–3326.

Leadlay PF, Roberts G, Walker JE. 1984. Isolation of a novel calcium-binding protein from *Streptomyces erythreus*. *FEBS Lett 178*:157–160.

Meador WE, Means AR, Quiocho FA. 1992. Target enzyme recognition by calmodulin: 2.4 Å structure of calmodulin-peptide complex. *Science 257*:1251–1255.

Moss SE, Crumpton MJ. 1990. The lipocortins and the EF hand proteins: Ca^{2+}-binding sites and evolution. *Trends Biochem Sci 15*:11–12.

Muhandiram DR, Kay LE. 1994. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J Magn Reson B 103*:203–216.

Prêcheur B, Cox JA, Petrova T, Mispelter J, Craescu T. 1996. *Nereis* sarcoplasmic Ca^{2+}-binding protein has a highly unstructured apo state which is switched to the native state upon binding of the first Ca^{2+} ion. *FEBS Lett 395*:89–94.

Slupsky CM, Reinach FC, Smillie LB, Sykes BD. 1995. Solution secondary structure of calcium-saturated troponin C monomer determined by multidimensional heteronuclear NMR spectroscopy. *Protein Sci 4*:1279–1290.

Swan DG, Cortes J, Hale RS, Leadlay PF. 1989. Cloning, characterization, and heterologous expression of the *Saccharopolyspora erythraea* (*Streptomyces erythraeus*) gene encoding an EF-hand calcium-binding protein. *J Bacteriol 171*:5614–5619.

Swan DG, Hale RS, Dhillon N, Leadlay PF. 1987. A bacterial calcium-binding protein homologous to calmodulin. *Nature 329*:84–85.

Tjandra N, Bax A. 1997. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science 278*:1111–1114.

Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A. 1997. Use of dipolar ^{1}H-^{15}N and ^{1}H-^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nature Struct Biol 4*:732–738.

Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH. 1995. Nuclear magnetic dipole interactions in field-oriented proteins: Information for structure determination in solution. *Proc Natl Acad Sci USA 92*:9279–9283.

Vijay-Kumar S, Cook WJ. 1992. Structure of a sarcoplasmic calcium-binding protein from *Nereis diversicolor* refined at 2.0 Å resolution. *J Mol Biol 224*:413–426.

Wishart DS, Sykes BD. 1994. The ^{13}C chemical-shift index: A simple and fast method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J Biomol NMR 4*:171–180.

Wishart DS, Sykes BD, Richards FM. 1992. The chemical shift index: A fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry 31*:1647–1651.

Wüthrich K. 1986. *NMR of proteins and nucleic acids*. New York: John Wiley & Sons.

Yamagata K, Goto K, Kuo CH, Kondo H, Miki N. 1990. Visinin: A novel calcium binding protein expressed in retinal cone cells. *Neuron 2*:469–476.

Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR, Kay LE. 1994. A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N, ^{13}C, ^{2}H labeled proteins with high sensitivity. *J Am Chem Soc 116*:11655–11666.

Zhang O, Kay LE, Olivier JP, Forman-Kay JD. 1994. Backbone ^{1}H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J Biomol NMR 4*:845–858.