Solution Structure of the Paramagnetic Complex of the N-Terminal Domain of Calmodulin with Two Ce^{3+} Ions by ^1H NMR†,‡

Detlef Bentrop,§ Ivano Bertini,*,§ Mauro A. Cremonini,|| Sture Forsén,⊥ Claudio Luchinat,# and Anders Malmendal⊥

Department of Chemistry, University of Florence, Via Gino Capponi 7, 50121 Florence, Italy, Food Science and Technology Laboratory, University of Bologna, Via Ravennate 1020, 47023 Cesena, Italy, Physical Chemistry 2, University of Lund, P.O. Box 124, 22100 Lund, Sweden, and Department of Soil Sciences and Plant Nutrition, University of Florence, P.le delle Cascine 18, 50144 Florence, Italy

Received May 1, 1997; Revised Manuscript Received July 1, 1997⊗

ABSTRACT: The solution structure of the dicerium(III) complex of the N-terminal domain of calmodulin (Ce_2-TR_1C hereafter) has been solved employing paramagnetic T_1 relaxation enhancements and pseudocontact shifts introduced by the Ce^{3+} ions, together with conventional NOE constraints. The use of pseudocontact shift constraints constitutes the first attempt to locate metal ions within a protein structure by NMR. Like calcium(II), paramagnetic cerium(III) has been found to bind to the two metal binding sites of the TR_1C fragment of calmodulin in a cooperative manner. Due to the presence of pseudocontact interactions between the Ce^{3+} ions and protons of the 76-residue protein, the ^1H NMR spectra of the complex show resonances shifted between +22 and −9 ppm. Eighty percent of its proton resonances could be assigned through a standard approach using TOCSY/COSY and NOESY spectra and through 1D NOE difference spectra for the broad resonances of protons close to the paramagnetic ions. A family of structures was calculated by means of the torsion angle dynamics program DYANA [Güntert, P., Mumenthaler, C., & Wüthrich, K. (1996) *XVIIth International Conference on Magnetic Resonance in Biological Systems* (Abstract)] using 1012 NOEs. Longitudinal proton relaxation times helped to roughly define the position of the metal ions within the protein. A total of 381 pseudocontact shift constraints, whose evaluation and use are critically discussed, have then been added to further refine the metal coordinates within the protein frame and to improve the structure resolution. A dramatic resolution improvement of the metal coordinates together with a sizable resolution improvement in the regions close to the paramagnetic centers, where the number of NOEs is low, is observed. The good quality of the solution structure permitted a meaningful comparison with the solid-state structure of calcium-loaded calmodulin at 1.7 Å resolution [Chattopadhyaya, R., Meador, W. E., Means, A. R., & Quiocho, F. A. (1992) *J. Mol. Biol. 228*, 1177]. The Ce_2-TR_1C complex is overall more compact than the Ca form.

Calmodulin (CaM)[1] is a ubiquitous protein that acts as a regulator of various metabolic processes by binding to specific target proteins (Forsén & Kördel, 1994; Alemà, 1984; Cox et al., 1984). The latter are involved in signaling, second messenger production, and motility (Klee, 1988; Crivici & Ikura, 1995). In turn, the regulatory properties of calmodulin are Ca^{2+}-dependent. The protein binds four Ca^{2+} ions via four helix−loop−helix motifs which are commonly called EF-hands (Kretsinger, 1980). These EF-hands are arranged in pairs; each of the two pairs in calmodulin forms a distinct domain comprising the N- and C-terminal half of the protein, respectively (Babu et al., 1988). The two domains are quite independent, as indicated by solution NMR studies which reveal high flexibility of the linker region as opposed to the well-ordered globular structures of the single domains (Barbato et al., 1992; Tjandra et al., 1995). The solution structure of the C-terminal fragment of calmodulin (Finn et al., 1995) confirmed that it maintains the same tertiary structure as in the holoprotein. The N-terminal domain, consisting of 76 amino acids (TR_1C fragment hereafter; Figure 1), has been expressed in *Escherichia coli* and shown to bind calcium essentially in the same manner (Linse et al., 1991) as in the holoprotein.

The present work aims at the solution structure of a dimetallic TR_1C fragment when the metal ions are paramagnetic and provide pseudocontact shifts (see below). Special emphasis is given to the information about the metal ions that can be obtained through NMR. The general challenge is an adequate treatment of the metal ions and their framing within the protein since solution structure information about metal ions is either missing or scarce. Indeed, the nuclei of the calcium isotopes, which are pertinent to the present case, have either zero or small magnetic moment associated with

† This work was supported by "Progetto Finalizzato Biotecnologie", "Comitato Biotecnologie e Biologia Molecolare", "Comitato Scienze Agrarie", and "Comitato Scienze e Tecnologie dell'Ambiente e dell'Habitat" of CNR, Italy, by the European Union via the award of a "Large-Scale Facility" grant to the Florence laboratory, and by Grant K-AA/KU 02545-326 from the Swedish Natural Science Research Council to S.F.

‡ The coordinates have been deposited in the Brookhaven Protein Data Bank (file name 1AK8).

* To whom correspondence should be addressed. Telephone: ++39 55 2757549. Fax: ++39 55 2757555. E-mail: bertini@risc1.lrm.fi.cnr.it.

§ Department of Chemistry, University of Florence.

|| University of Bologna.

⊥ University of Lund.

Department of Soil Sciences and Plant Nutrition, University of Florence.

⊗ Abstract published in *Advance ACS Abstracts,* September 1, 1997.

[1] Abbreviations: CaM, calmodulin; Ca_4-CaM, calcium-loaded CaM; TR_1C, N-terminal domain of CaM; Ce_2-TR_1C, cerium-loaded TR_1C; δ^{pc}, pseudocontact shift; RMSD, root mean square deviation.

S0006-2960(97)01022-2 CCC: $14.00

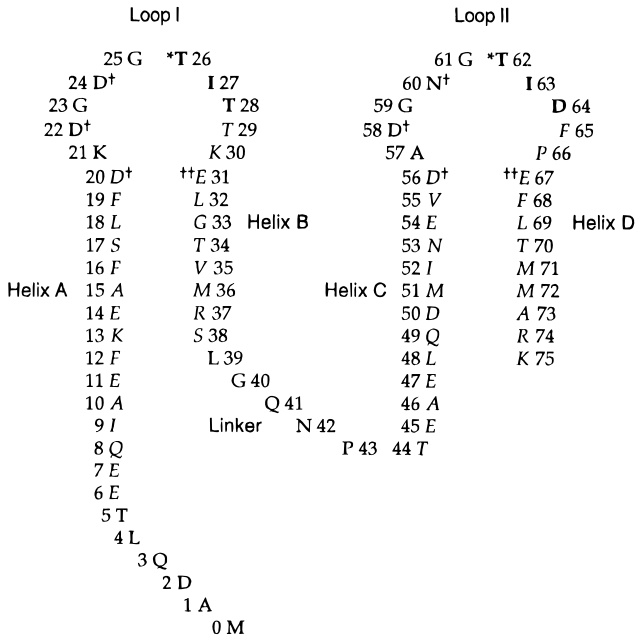

FIGURE 1: Amino acid sequence and secondary structure of TR_1C. The one-letter code is used for amino acid residues. The TR_1C secondary structure is indicated with α-helical residues in italics and β-sheet residues in boldface. Residues coordinating calcium with main-chain carbonyl oxygens are marked with an asterisk (*) and those using side-chain carboxylates with a dagger (†).

a nuclear quadrupolar interaction and cannot be of any help as far as NMR and solution structures are concerned. Successful attempts to gain information on the binding of metal ions to protein ligands have been based on heteronuclear correlation spectroscopy by using nuclei of large magnetic moments like ^{113}Cd (Coleman, 1993) and ^{199}Hg (Utschig et al., 1995; Blake et al., 1992) as metal probes. However, this information is limited to the metal−donor−carbon−proton dihedral angles and does not provide the full coordinates of the metal ions. Moreover, carbonyl and carboxylate ligands, as found in calcium binding proteins, are particularly unfavorable in terms of yielding measurable *J*-couplings to ligand nuclei.

Paramagnetic systems are characterized by broad NMR signals which render the determination of the conventional NOE or ^3J constraints for protein structure determination in solution experimentally difficult or impossible. Nevertheless, under favorable conditions the difficulties can be substantially overcome with the aid of suitably designed experiments, and the first NMR structure of a paramagnetic metalloprotein was published in 1994 (Banci et al., 1994). Since then, a considerable number of solution structures of paramagnetic proteins, or NMR-derived models for their structures, have been determined [for a recent review see Bertini et al. (1996b)]. Moreover, continuous attempts have been made to take advantage of relaxation enhancements due to the coupling between the paramagnetic metal ions and the resonating nuclei (Bertini et al., 1996a, 1997; Huber et al., 1996; Wang et al., 1996) in order to (1) obtain further structural constraints and (2) anchor the metal ion to the protein frame. The same goal has been pursued by the use of pseudocontact shifts (Banci et al., 1996; Gochin & Roder, 1995a; Gochin & Roder, 1995b; Guiles et al., 1996). The pseudocontact shifts (δ^{pc}) arise from the dipolar coupling between the resonating nuclei and the unpaired electron(s) when there is magnetic anisotropy. The latter occurs when there are low-lying excited states which also provide fast electron relaxation rates. The broadening of NMR lines is limited under fast electron relaxation conditions, and therefore the occurrence of pseudocontact shifts is often related to a relatively high quality of the NMR spectra of paramagnetic compounds. Lanthanides(III), except Gd(III), are appropriate metal ions. The magnitude of δ^{pc} depends on the extent of anisotropy of the paramagnetic susceptibility tensor:

$$\delta_i^{pc} = \frac{1}{12\pi r_i^3}\left[\Delta\chi_{ax}(3n_i^2 - 1) + \frac{3}{2}\Delta\chi_{rh}(l_i^2 - m_i^2)\right] \quad (1)$$

where $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and rhombic anisotropy parameters of the magnetic susceptibility tensor of the metal, r_i is the distance between an atom i and the metal ion, and l_i, m_i, and n_i are the direction cosines of the position vector of atom i ($\mathbf{r}_i$) with respect to the principal axes of the magnetic susceptibility tensor centered on the metal.

In this paper, we report the solution structure of the Ce_2-TR_1C complex. It is known that trivalent lanthanides have an ionic radius similar to that of calcium(II) and that they are good spectroscopic probes for the latter ion (Horrocks & Sudnick, 1981; Wu et al., 1996). Cerium(III) has $S = {}^1/_2$, one of the smallest magnetic moments among lanthanides (about 2.5 μ_B) and a relatively large magnetic anisotropy, yielding a pseudocontact shift/Curie line broadening ratio more favorable than other lanthanides for macromolecules at high field. Therefore, it represents an ideal probe to tackle the challenge of determining the coordinates of the metal ions in a metalloprotein and at the same time obtaining enough NOE structural constraints for the amino acids close to the paramagnetic center.

MATERIALS AND METHODS

Protein Preparation and Chemicals. The TR_1C used in this study is the N-terminal domain of bovine testes calmodulin. It was expressed in *E. coli* and purified as described previously for the TR_2C fragment of CaM (Finn et al., 1995). It is a 76-residue protein containing an N-terminal methionine that is not present in the sequence of bovine calmodulin. Ca^{2+}-free (apo)TR_1C was obtained according to a previously described procedure (Finn et al., 1995). The residual Ca^{2+} content was determined by atomic absorption spectroscopy. ApoTR_1C solutions of known concentration were obtained from the weight of the lyophilized protein and checked through total amino acid analysis. Analytical grade cerium(III) chloride was from Fluka and was used as obtained. Stock solutions for the titration of apoTR_1C were prepared by weight; their Ce^{3+} concentration was checked by titration with EDTA in 5 mM MES buffer at pH 6 using xylenol orange as indicator (Birnbaum & Sykes, 1978).

NMR Spectroscopy. ^{1}H NMR spectra were recorded at 298 and 290 K on a Bruker DRX 500 and a Bruker AC-P 300 spectrometer in 0.1 M KCl (pH = 6; 90% H_2O/10% D_2O or 100% D_2O) with 0.05% NaN_3 added. The sample for the titration with Ce^{3+} shown in Figure 2 was 4 mM in TR_1C; all other NMR experiments were carried out on 2.2 mM cerium-loaded TR_1C. The sample temperature was within ±0.1 K or better, ensuring that no extra line broadening was originated by temperature instability. The

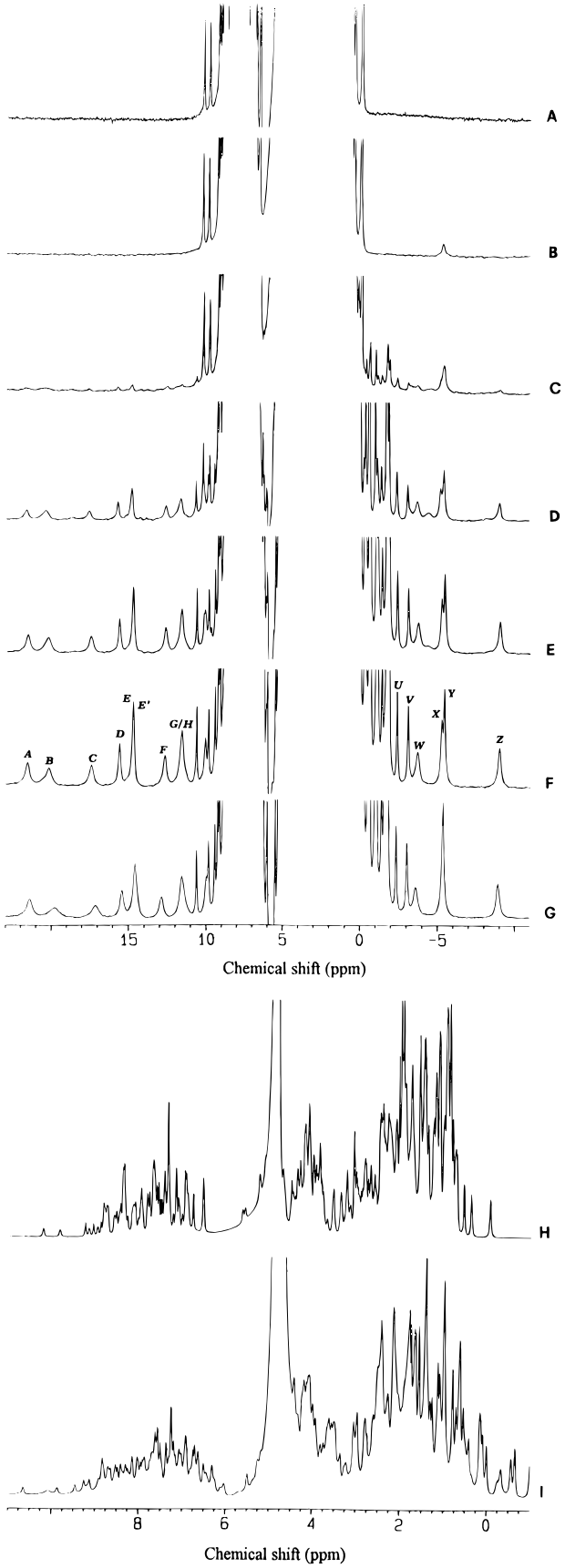

FIGURE 2: 600 MHz 298 K ^{1}H NMR spectrum of 4 mM apoTR_1C (A) and titration with increasing amounts of Ce^{3+}: molar ratio 0.05:1 (B), 0.3:1 (C), 0.8:1 (D), 1.4:1 (E), 1.7:1 (F), and 2.0:1 (G). The diamagnetic regions of spectra A and F are compared in panels H and I.

line shape of the broadest signals did not show evidence of important temperature gradients across the sample. NOESY (Wider et al., 1984) (mixing time 35 and 100 ms), CLEAN-TOCSY (Griesinger et al., 1988) (spin-lock mixing time 30 and 80 ms), and DQF-COSY (Rance et al., 1983) spectra were acquired in the phase-sensitive mode with TPPI (Marion & Wüthrich, 1983) for quadrature detection in the F_1 dimension over a spectral window of 15.3 ppm with a relaxation delay of 1.4 s. 4K data points in F_2 and 512 increments in F_1 were collected for these experiments. The carrier was always set on the water resonance. When necessary, water suppression was achieved either by continuous coherent irradiation prior to the first excitation pulse (and during the mixing time in the case of NOESY spectra) or by the WATERGATE technique (Sklenar et al., 1993). The data matrices were processed to a final size of 2K $\times$ 1K data points using shifted squared sine window functions prior to Fourier transformation. Fast repetition 2D spectra for the detection of connectivities between paramagnetically shifted resonances used a spectral window of 37 or 40 ppm with acquisition times of the order of 100 ms and short relaxation delays to allow for accumulation of a high number of scans. A total of 256 or 280 increments in F_1 were collected. A 10 ms CLEAN-TOCSY and a 30 ms NOESY were acquired in the phase-sensitive mode whereas the magnitude mode was used for a COSY spectrum. These spectra were Fourier-transformed to a final size of 1K $\times$ 1K data points after application of a squared cosine window function.

1D NOE difference spectra on pseudocontact-shifted resonances were recorded using previously reported methodology (Johnson et al., 1983; Banci et al., 1989) in H_2O (see above) and 100% D_2O solution. Solvent exchange was accomplished by lyophilization followed by dissolution in D_2O. Nonselective T_1 relaxation times of shifted signals were determined by the inversion recovery method (Vold et al., 1968) and an exponential three-parameter fit of the peak intensities versus the delay time. The spectra were calibrated assuming a chemical shift of 4.81 ppm for the water signal at 298 K with respect to DSS. All NMR data were processed with the Bruker UXNMR and XWINNMR software packages, respectively. The program XEASY (ETH Zürich) was used for the analysis of the 2D spectra.

Structure Calculations. The structure calculations were performed according to a recently introduced torsion angle dynamics (TAD) approach using the DYANA program package (Güntert et al., 1996). Pseudocontact shifts were introduced as constraints using the newly developed PSEUDYANA module (Banci et al., manuscript in preparation). The overall strategy of structure computation, developed as described in the Results section, comprised the following, subsequent steps:

(i) Pure TAD Calculations. Pure TAD calculations were carried out by the program DYANA (Güntert et al., 1996). DYANA generates a family of structures from a set of pairwise distance constraints, computed from the volumes of assigned NOESY cross peaks by the auxiliary program CALIBA (Güntert et al., 1991), and uses a simulated annealing algorithm in the torsional angle space. Additional proton−proton upper distance limits were determined from 1D NOE difference spectra recorded upon saturation of the pseudocontact-shifted signals, and proton−metal upper distance limits were determined from nonselective T_1 values measured on the pseudocontact-shifted signals. The 1D NOE connectivities were properly scaled into proton−proton distances using the steady-state approximation (Bertini &

Luchinat, 1996) with a rotational correlation time of 2.8 ns estimated from the Stokes−Einstein equation (Stokes, 1956; Einstein, 1956) assuming a molecular mass of 8.7 kDa for Ce_2-TR_1C. Proton−metal distances were estimated using the Solomon equation in the appropriate form for lanthanides (Bertini & Luchinat, 1996):

$$T_{1M}^{-1} = \frac{2}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_N^2 g_J^2 \mu_B^2 J(J+1)}{r_M^6}\left(\frac{7\tau_s}{1+\omega_S^2\tau_S^2} + \frac{3\tau_s}{1+\omega_I^2\tau_S^2}\right) \quad (2)$$

where μ_0 is the magnetic permeability of vacuum, γ_N is the nuclear magnetogyric ratio, J is the total angular momentum quantum number, g_J is the corresponding g-factor, μ_B is the Bohr magneton, r_M is the cerium(III)−proton distance, τ_s is the electronic relaxation time of cerium(III), and ω_S and ω_I are the electron and nuclear Larmor frequencies, respectively. An electronic relaxation time of 0.1 ps (Alsaadi et al., 1980) was used, and an uncertainty of ±0.3 Å was assigned to the derived distances. The diamagnetic contribution to the measured T_1^{-1} values is smaller than 2 s^{-1} for proteins of this size (Ishima et al., 1991) and therefore was considered negligible for the fastest relaxing signals used in the structure calculations (with T_1 values smaller than 240 ms; cf. Table 1). Furthermore, a relatively large tolerance was assigned to the derived distances which renders the error on the diamagnetic T_1 irrelevant. The reliability of proton−proton distance values estimated from NOEs on paramagnetically relaxed signals, and of metal−proton distance values estimated from nonselective T_1 measurements, has been demonstrated elsewhere (Bertini et al., 1997; Bertini & Luchinat, 1996).

Within DYANA each Ce^{3+} ion was represented by a special linker (pseudoresidue), called LTNS. A linker is a rigid sequence of six dummy atoms spaced 1 Å one from the other, making 90° angles, and allowing free rotation about the central bond between atoms 3 and 4. Linkers can be chained as many times as needed by overlapping the last three atoms of one with the first three atoms of the next, in such a way as to cover 1 Å each. The chain of linkers is invisible to the minimization algorithm except for a certain function, which typically is that of joining two different macromolecules together in the calculation. In the LTNS linkers, the fifth dummy atom represents the metal ion (ME). The unit vectors identifying the principal axes of the χ-tensor used in the subsequent PSEUDYANA calculation are represented by three additional dummy atoms (AX, AY, AZ) (Banci et al., 1996). To allow the two metal ions to span the whole space occupied by the macromolecule, the protein was extended with a chain of 29 linkers, covering a distance of 30 Å, followed by one special LTNS linker, 29 more linkers, and another LTNS linker, again covering a distance of 30 Å.

(ii) TAD Calculations with Pseudocontact Constraints. The family of structures obtained from the pure TAD runs was the starting point for an iterative structure refinement by the program PSEUDYANA (Banci et al., 1997), which can take into account and use during structure minimization both distance (NOE and T_1) and pseudocontact shift constraints.

PSEUDYANA is able to calculate pseudocontact contributions to both the DYANA target function and its gradient from a virtually unlimited number of paramagnetic centers, provided a suitable set of tensor parameters is available. The tensors are determined from a starting structure, or family of structures, with a new version of the program FANTASIA (Banci et al., 1996), which minimizes the anisotropy parameters *together with* the positions of the metals. Thus, for each coordination site, eight parameters are searched for, namely, the three direction cosines defining the χ-tensor with respect to the protein frame, the axial and rhombic anisotropy parameters ($\Delta\chi_{ax}$ and $\Delta\chi_{rh}$), and the three coordinates of the metal.

A set of pseudocontact shifts was calculated by subtracting the chemical shift values of the diamagnetic $(Ca^{2+})_4$ calmodulin (Ca_4-CaM) (Ikura et al., 1990; Ikura et al., 1991) from the ones measured for Ce_2-TR_1C. The choice of Ca_4-CaM as diamagnetic reference assumes that no major structural modifications are induced either by substitution of Ce^{3+} for Ca^{2+} or by the detachment of the TR_1C fragment from the other macromolecular moiety. The validity of this choice has been verified *a posteriori*, as will be discussed later. A fixed tolerance of 0.3 and 0.1 ppm (for HN and for nonexchangeable protons, respectively) or a proportional tolerance of 30% computed on the absolute value of the pseudocontact shift (whichever tolerance was larger) was used throughout the calculations. Under these conditions the pseudocontact contribution in PSEUDYANA was given the same weight as the NOE target function. Again, this choice was validated *a posteriori* by the relative contributions of NOE and pseudocontact contributions to the total target function (see below).

All RMSD values reported in this paper are RMSDs from the mean structure.

RESULTS

Titration of the ApoTR₁C Fragment with Cerium. The addition of microliter amounts of a cerium(III) chloride stock solution to apoTR_1C (both solutions in 0.1 M KCl, pH = 6) was monitored by 1D ^1H NMR. After the first addition of Ce^{3+}, corresponding to a Ce^{3+}/TR_1C molar ratio of 0.05, the spectrum of the apoprotein (Figure 2A) remained unaltered except for a new signal appearing upfield (Figure 2B). Above a Ce^{3+}/TR_1C molar ratio of 0.1 a new species is seen with signals spread from 22 to −9 ppm. Their intensity grows during the titration with a simultaneous decrease of the apoprotein signals (Figure 2C−G). A continuation of the titration beyond a Ce^{3+}/TR_1C molar ratio of 2.0 was not attempted. Above a molar ratio of ca. 1.8, a broadening and slight shifts of some signals are observed (Figure 2G), presumably due to exchange phenomena; the spectrum at a molar ratio of 1.7 (Figure 2F) is therefore taken as representative of the final Ce_2-TR_1C species. The above mentioned upfield-shifted signal is overlaid by two signals of the final Ce_2-TR_1C species at a molar ratio above 0.3 and reappears at the end of the titration. No effort was made to characterize the species represented by this signal. In the 0.3−1.7 molar ratio range the binding of Ce^{3+} appears to be cooperative like that of calcium (Linse et al., 1991). The sample on which the solution structure has been carried out did not show signals of the apoprotein, indicating a population of the latter of less than 10%, and showed a well-shaped spectrum of the Ce_2-TR_1C species.

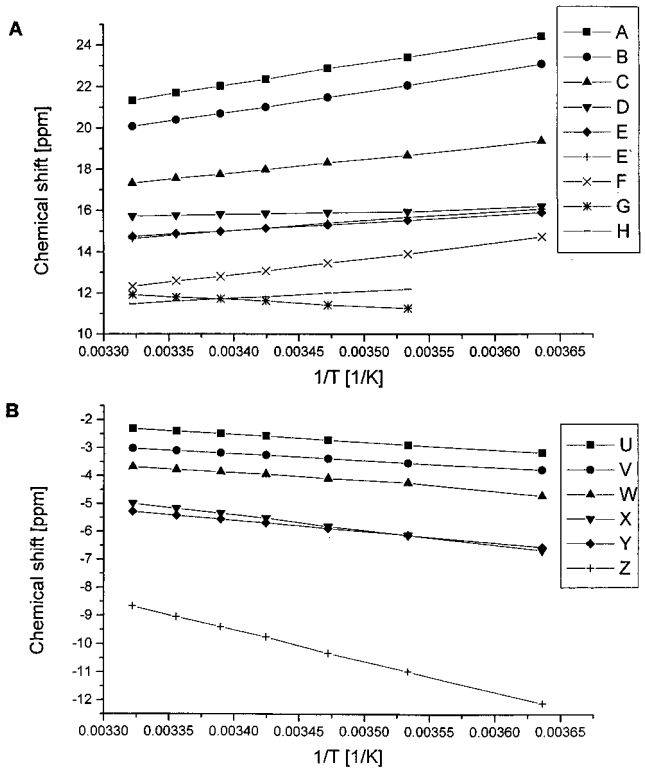

FIGURE 3: Temperature dependence of the pseudocontact-shifted resonances in the ^{1}H NMR spectrum of Ce_2-TR_1C. With the exception of signal G, the pseudocontact shifts decrease with increasing temperature. The labeling corresponds to Figure 2F and Table 1. Panels: (A) downfield-shifted resonances; (B) upfield-shifted resonances.

NMR Spectroscopy and Assignment. The large chemical shifts introduced by paramagnetic lanthanide ions are generally assumed to be pseudocontact shifts. The 1D ^{1}H NMR spectrum of Ce_2-TR_1C displays numerous resonances (labeled A−Z in Figure 2F) that are shifted out of the diamagnetic region (0−10 ppm); moreover, the diamagnetic region of the spectrum is completely different from the spectrum of apoTR_1C (Figure 2H,I). Table 1 provides a summary of the properties of the strongly pseudocontact-shifted signals observed in Figure 2 and their assignment (see below). It should be noted that all of the downfield-shifted resonances (A−H) are solvent exchangeable and are thereby identified as HN protons. None of the upfield-shifted signals (U−Z) exchanges with the solvent.

The pseudocontact shifts are strongly temperature dependent, as shown in Figure 3 by a Curie plot for the temperature range 275−301 K. With the exception of signal G (Asp-24 HN) all downfield-shifted resonances (Figure 3A) move upfield with increasing temperature, whereas all the upfield signals (U−Z; Figure 3B) move downfield with increasing temperature. Thus, the pseudocontact shifts uniformly decrease with increasing temperature, i.e., obey Curie behavior, as expected on theoretical grounds (Bertini & Luchinat, 1986, 1996). The unexpected anti-Curie temperature dependence of signal G will be discussed later.

TOCSY and COSY spectra in the diamagnetic region allowed the identification of 58 spin systems which were sequence-specifically assigned using dipolar connectivities in a 100 ms NOESY spectrum (sequential Hα−HN, Hβ−HN, Hγ−HN, and HN−HN NOEs and in the case of prolines Hα−Hδ and HN−Hδ NOEs). By this standard

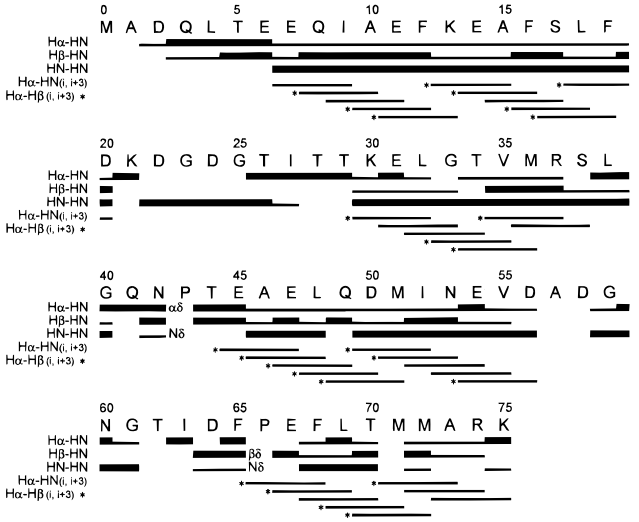

FIGURE 4: Pattern of sequential and (i, i+3) connectivities involving HN, Hα, and Hβ protons in the Ce_2-TR_1C complex. The relative NOE intensity is indicated by the width of the bars in the diagram. The asterisk (*) indicates a Hα−Hβ (i, i+3) NOE between the amino acids connected by the bar to the right of the asterisk.

methodology (Wüthrich, 1986), the sequential stretches Asp-2−Asp-20, Thr-29−Asp-56, and Asp-64−Lys-75 could be assigned unambiguously (Figure 4); Asp-56 was assigned on the basis of sequential dipolar connectivities with Val-55 only. No scalar connectivities within this residue could be detected. The pronounced temperature dependence of the chemical shifts of numerous residues in the paramagnetic Ce_2-TR_1C complex facilitated the assignment of spin patterns with overlapped backbone proton resonances considerably.

The presence of pseudocontact shifts arising from the bound Ce^{3+} ions precludes the use of the Hα chemical shift index (Wishart et al., 1992, 1991) for the recognition of secondary structural elements in Ce_2-TR_1C. However, the four α-helices of TR_1C could easily be delineated from their typical NOE pattern of strong sequential HN−HN NOEs and characteristic Hα−HN (i, i+3) and Hα−Hβ (i, i+3) NOEs (Figure 4). Helix A runs from Glu-6 to Asp-20 and forms together with helix B (Thr-29−Ser-38) part of the first EF-hand. EF-hand II starts with helix C (Thr-44−Asp-56) and ends with helix D extending from Phe-65 to the C-terminus (Lys-75).

The N-terminal residue Met-0 and the amide proton of Ala-1 could not be assigned, probably due to the high flexibility of this part of the protein. The Hα and the methyl group of Ala-1 were assigned on the basis of a strong COSY cross peak at the position of the corresponding resonances of Ala-1 in Ca_4-CaM and a weak NOE between Ala-1 Hα and Asp-2 HN at 290 K. The residues not assigned at this point were Lys-21−Thr-28 and Ala-57−Ile-63, all of which are located in the two metal binding loops of TR_1C (Figure 1). Extension of the assignment was based on NOESY spectra with fast repetition rates and 1D NOE difference experiments (for signals A−H and W−Z). The resulting network of connectivities permitted the assignment of protons that were unambiguously matching the available structural data. The X-ray structure of Ca_4-CaM (Chattopadhyaya et al., 1992), which in most cases provided a clear match with the NOE-estimated distances, can be used. Alternatively, a family of structures can be calculated by DYANA from the set of NOEs provided by the assignment available at this

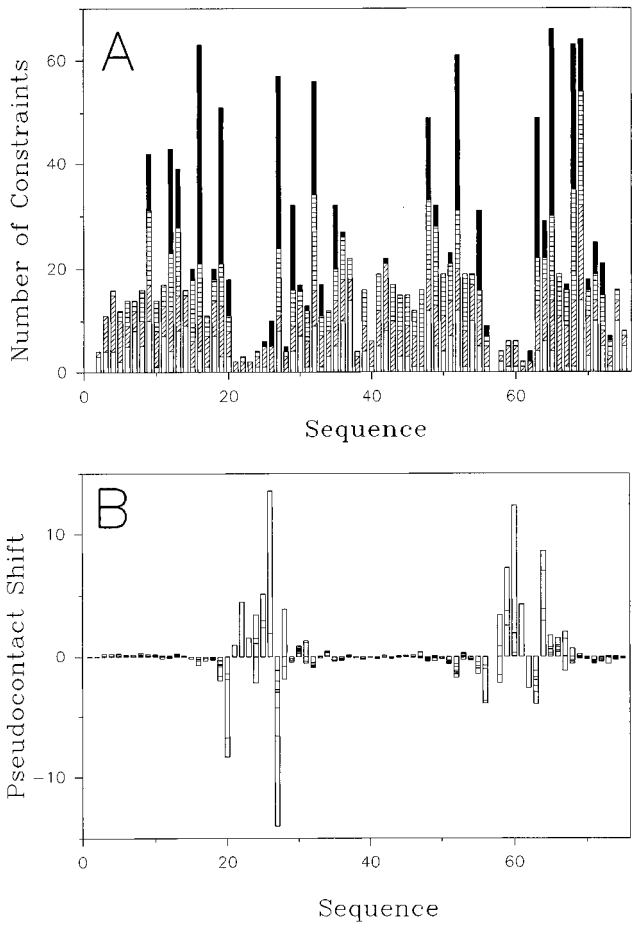

FIGURE 5: The NOE and pseudocontact shift information is not equally distributed over the sequence of Ce_2-TR_1C. (A) The number of meaningful intraresidual (white), sequential (cross-hatched □), medium-range (□), and long-range NOEs (black) per residue plotted versus the amino acid sequence. (B) Plot of the individual pseudocontact shifts $\delta_{obs} - \delta_{dia}$ of all assigned protons of Ce_2-TR_1C against residue number.

stage. If the chemical information given by the presence of two EF sites is exploited by imposing a 2−3 Å distance interval between the two cerium ions and the conserved EF-hand calcium ligands, i.e., the carbonyl oxygens of Thr-26 (site I) and Thr-62 (site II) and either of the two carboxylate oxygens of Asp-20, Asp-22, Asp-24, and Glu-31 (for site I) and Asp-56, Asp-58, Asn-60, and Glu-67 (for site II), a family of structures with a backbone RMSD of less than 2 Å is readily obtained. This family of structures can then be used as an aid to the assignment of the missing residues in an iterative process. The two processes, i.e., with or without the use of X-ray information, eventually lead to the same assignment, which is reported as Supporting Information. As examples, the following assignments are described.

The side chains of Ile-27 and Ile-63 could be identified from their spin patterns in COSY and TOCSY spectra. The sequence-specific assignment of these more (Ile-27) or less (Ile-63) upfield-shifted spin systems was established by analysis of numerous NOEs (Figure 5A) with resonances in the diamagnetic region and their comparison with the dipolar connectivities expected for the side chains of these two amino acids on the basis of either the CaM crystal structure or the DYANA family with the coordination constraints described above. This analysis included also folded cross peaks from the upfield-shifted resonances V and Y (Ile-27 Hγ_{12} and Hγ_{11}) that were folded in the F_1 dimension of the COSY and NOESY spectra recorded under slow repetition conditions.

The 1D NOE difference experiment on resonance Y confirms the assignment of Ile-27. The experiment on signal Z shows connectivities to all resonances of the Ile-27 side chain and to the βCH_2 protons of Glu-31 and identified this proton as Ile-27 Hα. Moreover, signal Z shows two NOEs to protons at 6.4 and 8.5 ppm, which are HN protons as they are not visible in 1D NOE difference spectra in D_2O solution and a COSY cross peak to the HN resonance at 8.5 ppm. These two HN protons were assigned to the amide protons of Ile-27 (8.5 ppm) and Thr-28 (6.4 ppm). Dipolar connectivities between side-chain resonances of Ile-27 and the resonance at 8.5 ppm were also detected in NOESY spectra and confirm this assignment.

The upfield-shifted resonances W and X show dipolar connectivities to each other. Moreover, both of these show 1D NOEs to resonances at 6.3 and 0.5 ppm that were previously assigned as part of Asp-20 (HN Asp-20: 6.29 ppm). Both signal W and signal X exhibit a scalar connectivity with the resonance at 0.5 ppm. 1D NOEs of signal X with the side chain of Ile-27 and the aromatic ring of Phe-16 identify it as a β-proton of Asp-20. A COSY cross peak between signal W and the HN of Asp-20 (6.3 ppm) and NOEs with the Ile-27 side chain allowed the assignment of W to the Hα of Asp-20. A strong NOE of resonance W with a HN proton at 8.76 ppm is tentatively assigned to the backbone amide proton of Lys-21. No further assignments for Lys-21 could be established.

Among the nine downfield-shifted amide proton resonances of Ce_2-TR_1C (signals A−H), strong HN−HN NOEs indicate a sequence G−D−A or A−D−G. Signal D can be safely assigned as HN of one of the missing glycine residues due to scalar and dipolar connectivties to resonances at 7.4 and 6.1 ppm, assumed to be a pair of geminal α-protons because of strong NOESY, COSY, and TOCSY cross peaks to each other. Since signal A, but not G, gives NOEs to the two Hα resonances, the sequence G−D−A can be deduced. The remaining 1D NOEs of signal A allow its tentative assignment to Thr-26 HN: connectivities with the aromatic ring of Phe-16 and a weak NOE to Ile-27 HN are consistent with the structural features of CaM. A resonance at 7.26 ppm is assigned to Hα of Thr-26, due to a strong NOE to signal A and a strong NOE to Ile-27 HN in the 100 ms NOESY. Moreover, this resonance gives NOEs with Asp-64 Hα and Phe-16 Hζ as expected for Thr-26 Hα on the basis of the CaM structure. Besides the sequential HN−HN NOEs and the NOEs to its presumed α-protons, signal D does not show any other connectivities. Signal G (Asp-24 HN) displays only a few 1D NOEs in addition to the HN−HN connectivity to Gly-25. There is a weak NOE with signal F and NOEs with resonances at 9.21, 8.45, 6.06, and 0.35 ppm. The last two resonances together with a third resonance at 4.18 ppm belong to a spin system clearly experiencing pseudocontact shifts and paramagnetic relaxation that could not previously be assigned due to the absence of interresidual NOE cross peaks. It was assigned to the AMX spin system of Asp-24 (Hα, 6.06 ppm. Hβ, 4.18 and 0.35 ppm). Signal F gives 1D NOEs to the presumed Lys-21 HN at 8.7 ppm, to signal G, and to resonances at 9.21, 3.7, 3.0, 2.55, and 0.5 ppm and is assigned to Asp-22 HN. This allows for the assignment of Gly-23 HN to the

resonance at 9.21 ppm, which is the only common neighbor of signals G (Asp-24 HN) and F (Asp-22 HN). No further assignments can be deduced for the residues Asp-22 and Gly-23.

All of the above given assignments for the backbone amide protons in the sequential stretch Asp-22−Thr-26 are consistent with their mutual HN−HN distances in the structure of CaM. It is important to note that they are also consistent with the T_1 relaxation times of the corresponding resonances (F, G, D, and A) which reflect semiquantitatively the distances of the corresponding amide protons to the Ca^{2+} ion in the CaM structure.

At this point, a full sequential assignment of the first metal binding site of TR_1C has been derived (Figure 4). The amide protons that give rise to the pseudocontact-shifted resonances B, C, E, E′, and H must therefore belong to the still unassigned residues of the second site. The NOEs between these signals allow to establish a sequence H−E′−B−E or vice versa. Analogous to signal D for the first site, signal E′ can be identified as HN of a glycine with the Hα protons at 7.7 and 6.4 ppm. The sequential direction H−E′−B−E follows from the NOEs of B with the α-protons of the preceding glycine (E′). As expected from the CaM structure, most of the backbone amide protons of the second metal binding site lack dipolar connectivities with the rest of the protein. Therefore, their assignment has to rely on the consistency of the observed sequential NOEs and T_1 relaxation times with the CaM X-ray structure, only. On this basis, signals H, E′, B, and E are assigned to Asp-58 HN, Gly-59 HN, Asn-60 HN, and Gly-61 HN, respectively. A new DYANA family of structures has been generated with the NOE constraints available at this point. It is much better defined than the first one (backbone RMSD of about 1 Å). The structural information about the backbone of the metal binding loops in these structures is consistent with the assignment procedure described above.

A COSY/TOCSY/NOESY spin pattern with resonances at 5.49, 1.47, and 0.49 was assigned to the AMX spin system of Asp-58 (Hα, 5.49 ppm; Hβ, 1.47 and 0.49 ppm) on the basis of 1D NOEs to Asp-58 HN (signal H) and an NOE between signal E′ (Gly-59 HN) and a resonance at 5.5 ppm. Similarly, resonances at 6.62, 3.39, and 3.00 ppm which also form an up to now unassigned spin pattern were identified as the side-chain protons of Asn-60 by 1D NOEs to Asn-60 HN (signal B) and Gly-61 HN (signal E). The only remaining NOE of signal E with a resonance at 4.0 ppm is assigned to one of the α-protons of Gly-61.

Signal C is the only pseudocontact-shifted resonance that does not exhibit any HN−HN NOEs. However, its irradiation gives rise to numerous NOEs with resonances in the diamagnetic shift range: dipolar connectivities with the side chain of Ile-63, with Hα and the βCH_2 protons of Asp-64, with HN and the βCH_2 protons of Glu-67, and with the HN's of Phe-65 and Phe-68 allow the unambiguous assignment of signal C to the backbone amide proton of Asp-64. A very strong NOE of signal C to a resonance at 0.4 ppm is tentatively assigned to Ile-63 Hα. Further inspection of the diamagnetic NOESY spectra allowed the assignment of Thr-28 (Hα and Hγ) and of Thr-62 Hα in the metal binding loops.

A summary of all assigned resonances of the Ce_2-TR_1C complex is provided as Supporting Information (Table S1). Besides Met-0, there is only Ala-57 as a completely unassigned residue. A total of 357 out of 449 theoretically expected proton resonances (about 80%) have been assigned.

Calculation of the Solution Structure of Ce_2-TR_1C. The protein studied in the present work contains two paramagnetic metal ions. Thus, NOEs, T_1 relaxation times, and chemical shifts sense the combined effect of two centers. In all the subsequent analyses, this is taken into account rigorously by considering the two contributions together. The calculations are trivial for NOEs and T_1 values as the relaxation effects are additive and scale with the sum of the r^{-6} values from the two metals. For the pseudocontact shifts, two χ-tensors need to be defined, and the two contributions depend on both r^{-3} and the tensor orientation for each center.

(i) Use of NOE and T_1 Data. For the DYANA calculations more than 1236 NOESY cross peaks were integrated and converted into upper distance limit constraints by means of the program CALIBA (Güntert et al., 1991). A total of 83 additional distance constraints were derived from 1D NOE difference spectra on the pseudocontact-shifted resonances. Finally, the 12 shortest T_1 values from Table 1 were calibrated and used as further distance constraints to the two metals. T_1 relaxation depends on the dipolar coupling between the unpaired electron and the nucleus according to eq 2. The T_1 relaxation constraints, although few, add precious structural information on the two metal binding sites. Similar information may be obtained from the line widths of the corresponding signals (Table 1). A total of 39 stereospecific assignments were provided by the auxiliary program GLOMSA (Güntert et al., 1991).

After DYANA runs with a total of 993 meaningful NOE and T_1 constraints, but without the above mentioned coordination constraints for the cerium ions and the oxygen atoms that usually ligate Ca^{2+} in an EF-hand, a family of 23 structures without consistent NOE violations and a target function below 0.70 Å^2 was selected. Figure 5A shows a plot of the number of NOEs per residue versus the residue number. The paucity of NOEs for the amino acids in the two metal binding loops, especially for residues 21−26 and 56−62, is obvious and presumably due both to the lack of some crucial assignments and to the enhanced relaxation introduced by the paramagnetic Ce^{3+} ion. The RMSD from the mean structure within this family was 0.79 ± 0.13 Å for the backbone atoms (1.21 ± 0.12 Å for all heavy atoms). The two metal ions were found to be reliably positioned in the metal binding loops, as expected, with a RMSD of 1.43 and 0.82 Å for sites I and II, respectively. Thus, NOE and T_1 constraints by themselves turned out to be able to generate a family of structures of acceptable quality, including the position of the metal ions. However, several side chains of the coordinating residues in the metal binding loops showed relatively high RMSD values (up to 4.17 Å for Asp-22).

(ii) Use of Pseudocontact Shifts. A total of 381 pseudocontact shifts have been obtained for Ce_2-TR_1C (see Figure 5B and Table 2). The pseudocontact shifts, as discussed before, connect the metal ions to the protein and may thus help in defining the structure around the metal ions. Achieving this goal, however, is not an obvious task, because it requires absolute consistency of two kinds of constraints, which, in addition, have different weights in different regions of the protein. We pursued also the ambitious goal of locating the metal ions inside the protein by using a combination of pseudocontact shifts and nuclear relaxation data. A procedure which was found effective was that of

Table 1: Properties of the Pseudocontact-Shifted Resonances of the Ce_2-TR_1C Complex

signal[a]	assignment	chemical shift at 298 K (ppm)	T_1 (ms)	line width[b] (Hz)	solvent exchangeable	chemical shift in Ca_4-CaM[c] (ppm)
A	HN Thr-26	21.7	42	172	yes	8.14
B	HN Asn-60	20.4	33	231	yes	8.07
C	HN Asp-64	17.5	46	172	yes	8.87
D	HN Gly-25	15.7	152	95	yes	10.57
E	HN Gly-61	14.8	∼120[d]	87[e]	yes	10.51
E′	HN Gly-59	14.8	∼120[d]	87[e]	yes	7.55
F	HN Asp-22	12.6	80	128	yes	8.10
G	HN Asp-24	11.8	∼70[d]	144[e]	yes	8.37
H	HN Asp-58	11.6	∼70[d]	144[e]	yes	8.18
U	Hγ_{11} Ile-63	−2.4	424	49	no	1.58
V	Hγ_{12} Ile-27	−3.1	429	54	no	1.22
W	Hα Asp-20	−3.8	81	149	no	4.53
X	Hβ_2 Asp-20	−5.2	239		no	1.57
Y	Hγ_{11} Ile-27	−5.4	388		no	1.22
Z	Hα Ile-27	−9.0	94	118	no	4.99

[a] Labeling as in Figure 2F. [b] Width at half-height at 500 MHz from the spectrum in Figure 2F; uncertainty ±10%. [c] Data for amide protons from Ikura et al. (1990) and for Hα and side-chain protons from Ikura et al. (1991). [d] Resonances E/E′ and G/H are overlapped, and only an apparent T_1 relaxation time can be measured. Gross deviations from exponentiality were not observed. [e] Apparent line width of overlapped peaks.

Table 2: Summary of the Experimental Constraints Used in the Present Work

Constraints for the Final DYANA Family (NOEs and T_1 Values Only)	
NOESY cross peaks	1236
1D NOEs	83
upper limits from T_1's	12
lower limits from T_1's	12
total no. of meaningful distance constraints	1012
intraresidual	311
short range	227
medium range	217
long range	257
stereospecific assignments[a]	39
Constraints for the PSEUDYANA Family	
total no. of meaningful distance constraints	989
intraresidual	304
short range	221
medium range	210
long range	254
stereospecific assignments[b]	51
pseudocontact shifts (in absolute values)	
>3 ppm	21
1−3 ppm	43
0.3−1 ppm	78
0.1−0.3 ppm	100
<0.1 ppm	139
total	381

[a] Obtained from the program GLOMSA (Güntert et al., 1991). [b] There are 12 stereospecific assignments more than for the DYANA family: 4 from GLOMSA and 8 from comparison of experimental and calculated pseudocontact shifts of diastereotopic proton pairs.

using the pseudocontact shifts as constraints through the program FANTASIA (Banci et al., 1996) to determine the two tensors and their origins (i.e., the locations of the metals) for each family with a 16-parameter fit (8 parameters for each tensor; see Materials and Methods). A single average tensor for each site is obtained for the whole family. PSEUDYANA (Banci et al., manuscript in preparation) was then used by retaining the two pairs of tensor anisotropy parameters and allowing the torsion angle dynamics algorithm to readjust the position of the origin and the tensor orientation for each site during the structure calculations.

Initially, the 24 largest pseudocontact shifts observed for protons belonging to the ill-determined side chains in the metal binding loops were left out from the calculations. The rationale was that modest movements of all other protons upon further structural refinement would not grossly alter the tensor parameters when recalculated, and therefore the starting tensor would already be a reasonable approximation of the final one. From a FANTASIA run on the DYANA family with a backbone RMSD of 0.79 Å, a first set of the five tensor parameters and the three Ce^{3+} coordinates in each site was obtained. PSEUDYANA calculations provided then a first simultaneous minimization of NOE, T_1, and pseudocontact target function and generated a new family of 23 structures. This family had a remarkably low RMSD for the two metal ions (0.29 and 0.42 for sites I and II, respectively). Apparently, the pseudocontact shifts are immediately able to define the position of the metal ions within the protein to a high degree of accuracy. In successive PSEUDYANA runs including all 381 pseudocontact shifts the RMSD of the two metal ions remained low; however, no improvement was noted for the side-chain RMSDs of the residues in the binding loops. Therefore, the oxygen atoms that usually coordinate Ca^{2+} in an EF-hand (cf. Figure 1; OD2 Asp-20, OD2 Asp-22, OD2 Asp-24, O Thr-26, and OE2 Glu-31 for site 1 and OD2 Asp-56, OD2 Asp-58, OD2 Asn-60, O Thr-62, and OE2 Glu-67 for site 2) (Strynadka & James, 1989) were again restrained to be at a distance between 2 and 3 Å from the metals in all final calculations.

A new starting family of 23 structures was generated by DYANA from a total of 1012 meaningful constraints (Table 2), which had a RMSD of 0.68 ± 0.13 Å for the backbone atoms and 0.98 ± 0.13 Å for all heavy atoms, followed again by a first FANTASIA run without the 24 largest pseudocontact shifts observed for protons belonging to the side chains in the metal binding loops. PSEUDYANA/FANTASIA cycles were then carried out by using all 381 pseudocontact shifts with tolerances proportional to the shifts themselves (±30%; in the very last PSEUDYANA run, ±20%) in addition to a fixed tolerance of 0.1 ppm for nonexchangeable and 0.3 ppm for exchangeable protons. The larger tolerance on the larger shifts ensured that even gross displacements of the involved protons at later stages would not disturb convergence of the calculations.

New families of structures were generated until the tensor parameters did not change by more than 5%. This occurred after three cycles of calculations. During the PSEUDYANA/FANTASIA cycles the obtained tensors were used to predict the shifts of signals from diastereotopic proton pairs. In eight cases, the predictions were sufficiently unambiguous to allow

Table 3: Parameters Defining the Magnetic Anisotropy Tensors of Ce_2-TR_1C in Metal Binding Sites I and II

	site I	site II
$\Delta\chi_{ax} \pm 3\sigma$ ($m^3 \times 10^{32}$)	-1.52 ± 0.02	1.29 ± 0.02
$\Delta\chi_{rh} \pm 3\sigma$ ($m^3 \times 10^{32}$)	-0.62 ± 0.03	0.59 ± 0.04

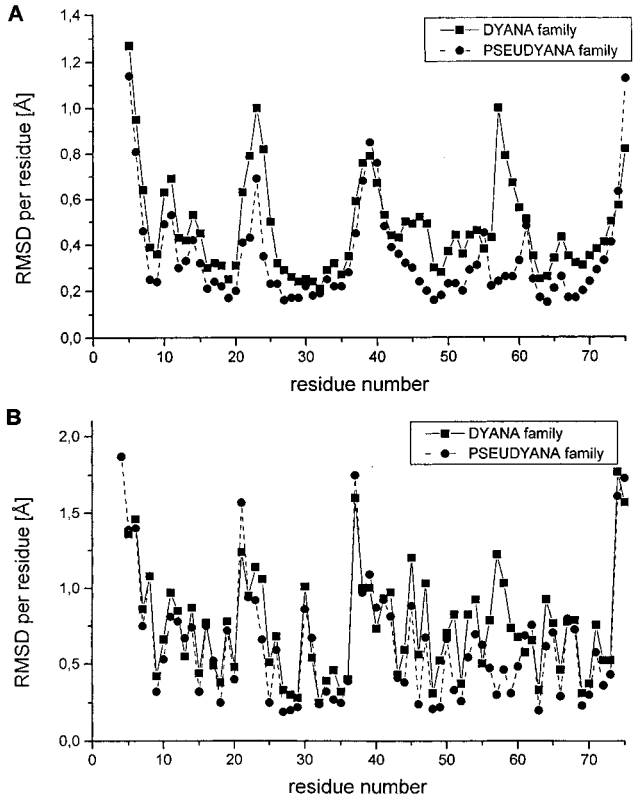

FIGURE 6: Diagram of the RMSD per residue for the backbone (A) and heavy atoms (B) of the final 23 Ce_2-TR_1C structures from DYANA (■ and solid line; calculated with NOE and T_1 constraints only) and of the final 23 Ce_2-TR_1C structures from PSEUDYANA (● and broken line; calculated with the inclusion of pseudocontact shifts).

for stereospecific assignments. These assignments turned out to be helpful for the convergence of the calculations.

The obtained tensor parameters and their uncertainty, calculated as 1 SD among the values calculated on each structure of the final family, are reported in Table 3. It can be noted that the values of the two anisotropy parameters for the two cerium ions are rather similar. The final family had a target function from 2.2 to 2.7 Å^2, backbone RMSD of 0.49 ± 0.15 Å, and heavy atom RMSD of 0.85 ± 0.13 Å. The contributions to the target function were about 75% from NOEs and 25% from pseudocontact shifts, a ratio that is deemed reasonable in view of the fact that the number of pseudocontact shift constraints is about one-fourth of the NOE constraints. This finding confirms that the relative weights used for the two classes of constraints were well balanced. Figure 6 shows the RMSD per residue for the final structure families with and without the use of pseudocontact shifts.

DISCUSSION

The Ce_2-TR_1C structure is schematically shown as a tube with a width proportional to the backbone RMSD in Figure 7A. The good quality of the structure and the improvement obtained by introducing pseudocontact shifts are demonstrated by the structural statistics in Figure 6 and Table 4. The improvement is not evenly distributed along the whole polypeptide chain. The most striking effect is observed on the resolution of the metal ions themselves. In a family of structures obtained with the NOE constraints alone, even when the side chains of the conserved coordinating EF-hand residues were restrained in the proximity of the metals, the RMSD for the metal ions was 0.65 Å in site I and 1.03 Å in site II. The RMSD already decreases sizably by introducing the T_1 constraints (0.40 and 0.49 Å, respectively) and finally drops to 0.18 and 0.16 Å, respectively, by introducing the pseudocontact shifts. It should be noted that such a sizable decrease is virtually independent of the detailed refinement pathway, i.e., with or without inclusion of X-ray-aided assignments in the early stages. It is also apparent from the very first inclusion of the set of pseudocontact shifts from distant residues and in the absence of the restraints imposed later on the coordinating atoms.

Although an overall decrease of the backbone RMSD by about 0.2 Å is introduced by pseudocontact shifts, the decrease is larger in the two metal binding sites and more marked in site II than in site I (Figure 6A). The peak of the backbone RMSD at Gly-23 is due to the incomplete assignment of this residue (one Hα is missing). The high heavy atom RMSDs of Lys-21 and Asp-22 are rationalized by the absence of side-chain assignments in these two cases. Virtually no improvement by the introduction of pseudocontact shifts is seen in the ill-defined connecting loop of residues 36−41. This loop has already been found to show marked flexibility in solution NMR studies of this class of proteins (Barbato et al., 1992; Tjandra et al., 1995; Akke et al., 1993). In the metal sites both T_1 and pseudocontact shifts play a dramatic role. However, the definition of the coordinating side chains in the loops is only moderately satisfactory. The absence of violations in the additional constraints imposed by chemical considerations as well as on the T_1 constraints indicates that the overall coordination geometries are in good agreement with expectations from analogous systems. However, details like a mono- versus a bidentate behavior of some of the coordinating carboxylate amino acids cannot be definitely assessed by the data (see Table 5 for a summary of the metal−oxygen distances in the two sites). Apparently, in the absence of NOEs, the pseudocontact shifts alone of the few side-chain protons are not sufficient to unambiguously orient the carboxylate heads.

The backbone arrangement in the NOE-only structures differs from that in the final structures in that helix A is closer to the C-terminal end of helix B and the linker and that loop I is farther away from helices B and C. It is worth noting that loop I is relatively ill-defined by NOEs only.

Particularly interesting is the behavior of Asp-64 (Table 5). This residue has been found noncoordinating in the X-ray structure of Ca_4-CaM (Chattopadhyaya et al., 1992) but hydrogen bonded to a metal-coordinated water molecule. No water molecule has been introduced in the present analysis, but the orientation of Asp-64, both at the NOE-only and at the pseudocontact shift level, is consistent with such geometric arrangement even in the absence of the water molecule. Indeed, the present structure clearly shows a relatively open coordination site on both metals which is very close to where water is bound in Ca_4-CaM (see below).

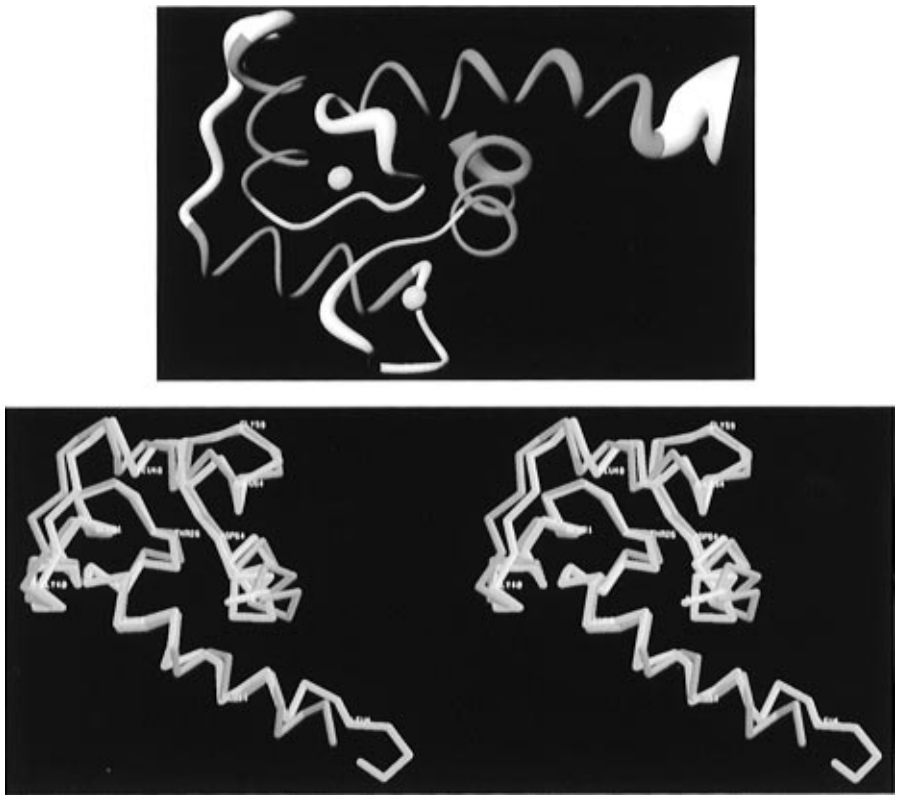

FIGURE 7: (A, top) Solution structure of Ce_2-TR_1C, represented as a tube with width proportional to the backbone RMSD (helices are in red, the β-sheet is in cyan, and the Ce^{3+} ions are in yellow). The backbone of residues 3−75 is shown. [Figure prepared by the program MOLMOL (Koradi et al., 1996).] (B, bottom) Stereoview of the "mean" solution structure of Ce_2-TR_1C (backbone only; yellow) superimposed on the backbone of the Ca_2-TR_1C domain from the X-ray structure of CaM (magenta) (Chattopadhyaya et al., 1992). The "mean" structure of Ce_2-TR_1C is the member of the PSEUDYANA family that has the smallest RMSD to the average structure calculated from the superposition of all 23 structures.

Table 4: Ramachandran Plot Statistics for Ce_2-TR_1C[a]

	"NOE-only" family		PSEUDYANA family	
	all[b]	6−74[b]	all[b]	6−74[b]
percentage of residues				
in most favored regions	77.9	82.6	76.0	81.0
in additionally allowed regions	18.4	14.8	20.4	16.7
in generally disallowed regions	3.0	2.2	3.0	2.2
in disallowed regions	0.7	0.4	0.6	0.1

[a] The percentages of residues in the four classes of Ramachandran regions are given for the final DYANA family (NOEs and T_1 values only) and for the PSEUDYANA family. [b] Values are reported for an analysis including all 76 residues and for an analysis without the disordered N-terminus (residues 6−74).

Farther from the metal binding sites the structural improvements are less dramatic but still locally significant. As an example, the detailed orientation of Ile-63, already clearly suggested by the NOE-only family, is confirmed beyond any doubt by the pseudocontact shifts. In the present structure, the Ile side-chain orientation is opposite to that found in the X-ray structure of CaM, i.e., the β-proton is in trans rather than in cis orientation with respect to the α-proton. Interestingly, the corresponding residue in EF site I, Ile-27, has a trans β-proton−α-proton orientation in both the X-ray and the present solution structure.

When compared to the crystal structure of calcium-loaded CaM at 1.7 Å resolution (Chattopadhyaya et al., 1992), the general fold is essentially the same (Figure 7B; backbone RMSD of 1.28 Å between the crystal structure and the "mean" solution structure which is the one member of the PSEUDYANA family that has the smallest RMSD to the calculated mean structure). As can be seen in Figure 8, the cerium-loaded protein has a generally more compact structure; however, the solvent-accessible surface area of the calcium- and cerium-loaded form is virtually the same (about 5000 Å^2). A more detailed analysis of the structural differences to Ca_4-CaM reveals that the C-terminal end of helix A and the N-terminal half of loop I are closer to all other secondary structure elements and helix D and the C-terminal half of loop II are closer to helices B and C and to the connecting loop, while the C-terminal end of helix B is more distant to the N-terminal end of helix C. The cerium−cerium distance is 0.8 ± 0.1 Å shorter than the calcium−calcium distance in the native protein. We deem this finding significant due to the very good definition of the metal coordinates. A shorter distance is also plausible from a chemical point of view, as the larger positive charge of cerium decreases the overall negative charge of loops I and II and therefore decreases the electrostatic repulsions between these two sides of the molecule. As far as the metal−Cα distances are concerned, it is found that the metal ion in site I is closer to the C-terminal EF-hand and the metal ion in site II is closer to the rest of the protein except for its own binding loop.

The detailed differences in the metal binding sites are shown in Figure 9A,B. The backbone arrangements in the two metal binding loops are very similar. However, the metal positions are offset by 0.95 Å (site I) and 0.60 Å (site II) with respect to the calcium positions in the Ca_4-CaM structure when only the backbones of the binding sites are

Table 5: Oxygen−Metal Distances in the Cation Binding Loops of the 23 Ce_2-TR_1C Structures of the PSEUDYANA Family and in the Crystal Structure of Ca_4-CaM[a]

	mean distance in the family	min distance	max distance	distance in Ca_4-CaM[b]
Asp-20 OD1	4.83	4.42	5.26	4.30
Asp-20 OD2	3.02	2.93	3.09	2.42
Asp-22 OD1	3.67	0.66[c]	4.96	4.12
Asp-22 OD2	2.80	2.13	3.04	2.50
Asp-24 OD1	3.01	2.57	4.31	4.13
Asp-24 OD2	2.63	2.04	3.01	2.39
Thr-26 O	2.40	2.23	2.78	2.35
Glu-31 OE1	2.57	1.84	4.90	2.43
Glu-31 OE2	2.38	2.00	3.01	2.38
Asp-56 OD1	4.44	2.15	5.20	4.0
Asp-56 OD2	2.72	2.13	3.04	2.13
Asp-58 OD1	4.42	1.99	5.01	3.64
Asp-58 OD2	2.53	2.00	3.01	2.30
Asn-60 OD1	2.65	2.39	2.87	2.38
Thr-62 O	2.74	2.24	3.02	2.29
Asp-64 OD1[d]	3.91	2.72	5.01	5.14
Asp-64 OD2[d]	3.96	2.69	5.37	5.11
Glu-67 OE1	4.05	2.11	5.13	2.55
Glu-67 OE2	2.82	2.03	3.05	2.40

[a] All distances are given in angstroms. The coordinating Asp and Glu residues were assumed to bind through OD2 and OE2, respectively. [b] PDB entry 1CLL (Chattopadhyaya et al., 1992); X-ray structure at 1.7 Å resolution. [c] This value is due to the fact that the unassigned side chain of Asp-22 is restrained only by a 2−3 Å distance interval between its OD2 atom and the metal ion in the first binding site and that no van der Waals radius was used for the metal ions throughout the structure calculations. [d] No restraints on the oxygen atoms were imposed in the calculations.

superimposed (when the backbones of the crystal structure and the mean solution structure are matched, as done for residues 4−75 in Figure 7B, the metals are offset by 0.61 and 0.68 Å, respectively). These offsets can be partially due to the invisibility of the metal-bound water molecule with the method used in this study. Despite the intrinsic uncertainties in the metal−oxygen donor distances (Table 5) discussed above, and despite the lack of the metal-bound water molecule, it can be speculated that the cerium ions have a tendency to increase their coordination number compared to calcium, as expected from their solution chemistry (Kowall et al., 1995). All differences in distances involving the metals are also summarized in Figure 8.

Figure 9C,D shows stereoviews of the metal sites with the orientation of the magnetic anisotropy tensors. It appears that the pseudosymmetry relating the two sites is not reflected in the orientations of the magnetic axes, which are considerably different in the two sites.

The peculiar anti-Curie temperature dependence of the Asp-24 HN resonance (signal G) can be analyzed in the light of the present structure. From Figure 9C it appears that Asp-24 HN is close to a zero pseudocontact shift surface (the equivalent of the magic angle cone surface in axial systems) and, due to its position, is expected to be extremely sensitive to even slight geometrical changes. Such changes are not unlikely even in a small temperature range, as Asp-24 HN is involved in H-bonding with at least two metal−ligand oxygens. Actually, even a small contribution from ligand-centered dipolar or contact shifts, also very sensitive to the H-bonding strength, cannot be ruled out in the present case. This finding underlines the potential information content of pseudocontact shifts when analyzed in detail.

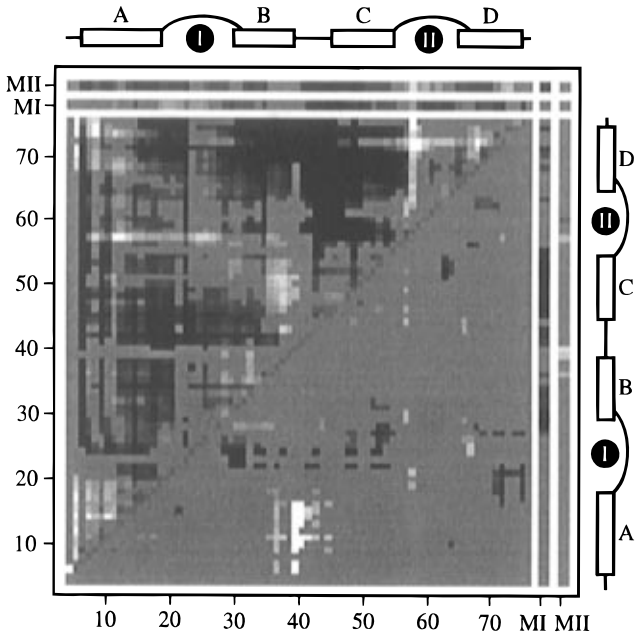

FIGURE 8: Difference distance matrix (Akke et al., 1995) describing the mean difference in Cα−Cα, Cα−metal, and metal−metal distances between different structures. The upper left shows the mean pairwise distance difference between all Cα atoms and metals of the final 23 PSEUDYANA structures of Ce_2-TR_1C and the X-ray structure of Ca_4-CaM at 1.7 Å resolution (Chattopadhyaya et al., 1992) as a function of amino acid numbers. The lower right shows the mean pairwise distance difference between all Cα atoms of the final 23 PSEUDYANA structures and the final 23 DYANA structures of Ce_2-TR_1C. The graph is shaded so that distances not significantly different are shown in gray, distances that are shorter in the final Ce_2-TR_1C structures tend toward black, and distances that are longer in the final Ce_2-TR_1C structures tend toward white. Full white or black indicates a distance difference of more than 1.5 Å. Distance differences larger than 2 SD are considered significant. MI and MII represent the metal ion in the first and second binding site, respectively.

About the Determination of Pseudocontact Shifts. A comment is due to the feasibility of a work of the present quality without the availability of reference diamagnetic chemical shifts. There is no perfect diamagnetic reference system, as this would imply no chemical difference whatsoever from the system under study except for the absence of unpaired electrons. The closest reference in this case would be the diamagnetic lanthanum analogue, which was not available. The chemical shifts of the calcium protein (Ikura et al., 1990, 1991) turned out to be more than adequate. The NMR data on the calcium protein were collected under experimental conditions similar to those of the present ones, as clearly shown by the very good match (typically within 0.03 ppm) of the shifts in the regions of the protein far from the metal sites. There is the possibility that some HN protons experience different hydrogen bonds and therefore different diamagnetic shift contributions in the calcium and cerium proteins. If this were the case to an extent larger than the tolerance for the pseudocontact shifts, it would lead to a violation of the corresponding pseudocontact shift constraint. *A posteriori*, it can be concluded that this is not the case.

When a diamagnetic reference protein is not available, diamagnetic reference values can be estimated from statistical data on classes of protein proton resonances, and deviations, due for instance to secondary structure and ring current effects, can be estimated from calculations (Ösapay & Case,

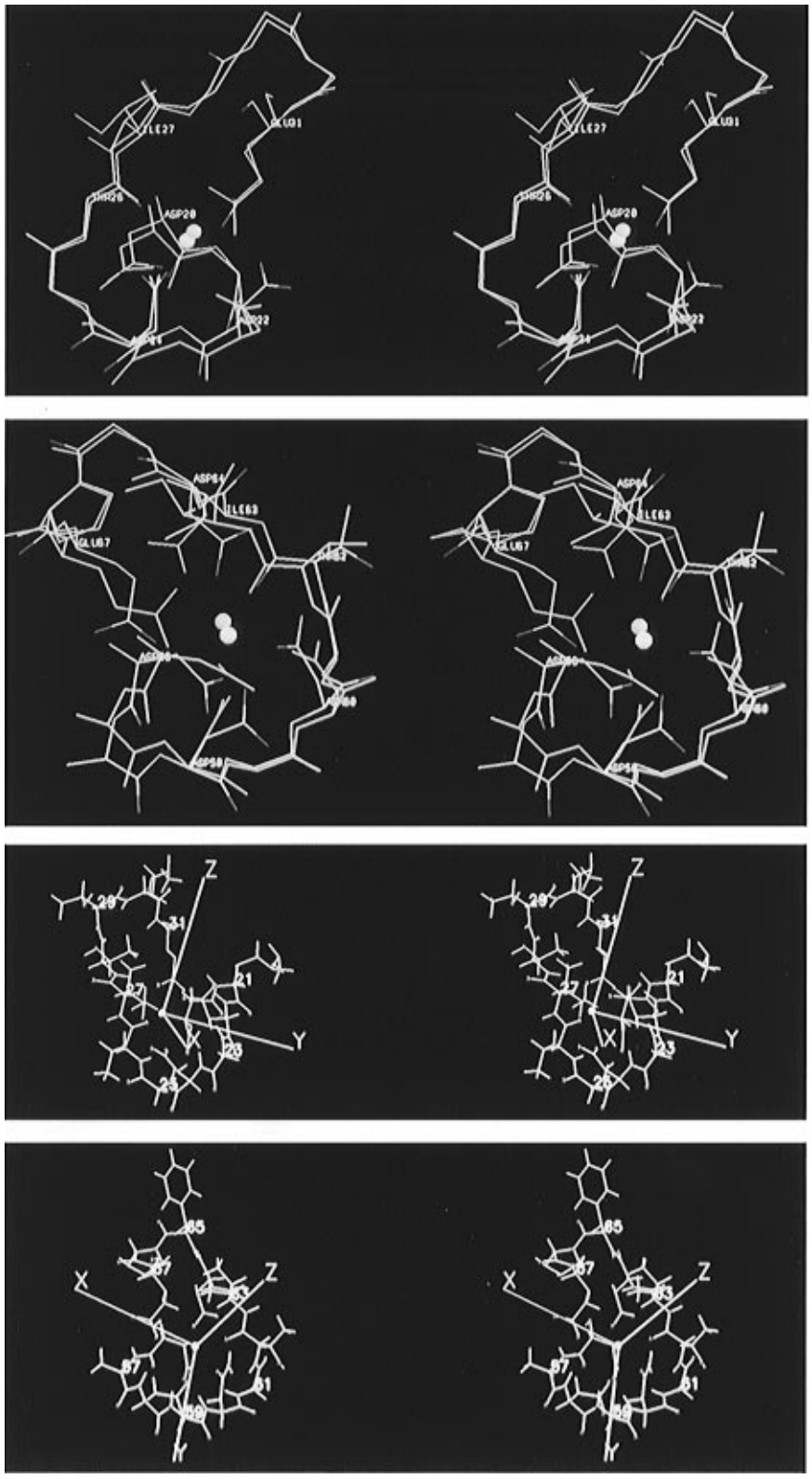

FIGURE 9: (A−D, from top to bottom) Comparison of metal binding sites I (A) and II (B) in the presence of Ce^{3+} and Ca^{2+}. All backbone atoms and the side chains of the metal-coordinating residues and of Ile-27 (site I, A) and Ile-63 and Asp-64 (site II, B) are shown. All atoms of the calcium form (Chattopadhyaya et al., 1992) are in cyan. The atoms of the cerium form (mean solution structure) are color-coded (red, oxygen; blue, nitrogen; magenta, cerium). (C) and (D) depict stereoviews of the metal binding sites I (C) and II (D) showing the metal coordination and the orientation of the principal axes of the magnetic anisotropy tensors (red, oxygen; blue, nitrogen; yellow, cerium).

1991; Kuszewski et al., 1995). The accuracy of the final result is, however, modest, the uncertainty being on the average of the order of 1 ppm, and occasionally even larger. As a matter of fact, it is common belief that pseudocontact shifts cannot be estimated with the precision needed for structural purposes in the absence of experimental data on diamagnetic references.

We have considered the possibility of exploiting a qualitative knowledge of the temperature dependence of the pseudocontact shifts in a semiquantitative way. For all

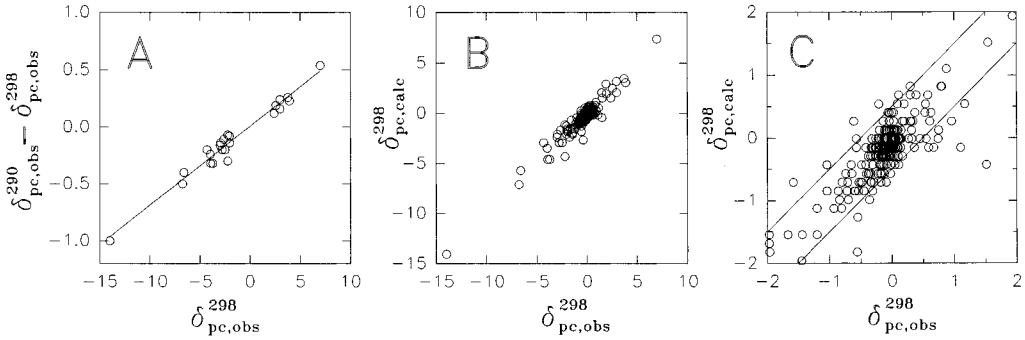

FIGURE 10: (A) Plot of δ_{obs}(290 K) − δ_{obs}(298 K) versus the observed pseudocontact shifts (δ^{pc}) at 298 K for all pseudocontact shifted resonances with an absolute pseudocontact shift value that is larger than 2 ppm. (B) Plot of the pseudocontact shifts calculated from a fit of the far-shifted signals in (A) versus the observed pseudocontact shifts. (C) Enlargement of the diamagnetic region of (B).

resonances with an absolute pseudocontact shift larger than 2 ppm, Figure 10A shows a plot of the difference in measured chemical shifts at two different temperatures (290 and 298 K) versus the pseudocontact shift derived from the available diamagnetic reference. A quite good correlation is apparent, as expected for the Curie-like behavior of almost all shifted signals. On the latter signals, relatively large errors in the estimate of the diamagnetic reference produce only minor effects, whereas the effects would be dramatic on the less shifted signals. Therefore, if the few strongly shifted signals are used to construct a calibration line (straight line in Figure 10A), the pseudocontact contributions to the less shifted signals can be estimated with reasonable accuracy from the calibration line. A plot of the pseudocontact shifts calculated in this way against the observed ones shows that the estimated uncertainty is on the average less than ±0.5 ppm (Figure 10B,C). This finding suggests that temperature-dependent data can constitute a valid alternative to the estimate of pseudocontact shifts in the absence of suitable diamagnetic references.

CONCLUDING REMARKS

In the present paper the solution structure of a two-site metalloprotein, including the positions of the two metal ions, has been solved by ^{1}H NMR methods. In order to determine independently the coordinates of the metal ions within the protein frame, we have used Ce^{3+}, a metal ion that gives rise to sizable pseudocontact shifts which in turn link the metal ion to the nuclei of the protein. With NOE constraints only, a properly folded protein structure could be calculated; with pseudocontact shifts as well as constraints derived from T_1 relaxation times, the metal ions have been located properly. In the present case, the metal ligands are mostly carboxylates, and therefore no direct information can be obtained from ^{1}H NMR on the donor atoms or their neighbors. Nevertheless, the metal coordination sites have been resolved to a satisfactory extent. The improvement in the overall structural quality extends fairly beyond the coordination sites. Pseudocontact shifts indeed provide long-range structural constraints as they depend on r^{-3} and may be useful for further studies on protein unfolding and protein-protein interactions. From the biochemical point of view, we have learned that cerium(III) binds to TR_1C in a cooperative manner as does Ca^{2+} and that the cerium-loaded protein (probably due to electrostatic effects) is more compact than the calcium form.

Generally speaking, the use of appropriate paramagnetic substituents for diamagnetic metals in proteins may be justified in order to define the position of the metal which would be invisible to NMR methods in the diamagnetic case. The loss of conventional NOE information due to fast relaxation of close nuclei can be more than compensated for by the gain of nonclassical constraints (nuclear relaxation rate enhancements and pseudocontact shifts). The latter constraints can be helpful in better defining the conformation of metal-coordinating residues, especially in the case of ligation via carboxylates and carbonyls as in calcium binding proteins. Heteronuclear NMR on uniformly or selectively ^{15}N- and ^{13}C-labeled proteins for further structural refinement of metal binding sites is expected to be very informative because paramagnetism provides structural constraints between nuclei of the coordinating residues and the paramagnetic center(s). Owing to the low magnetogyric ratio of the above heteronuclei, a modest line broadening is expected.

ACKNOWLEDGMENT

We thank Eva Thulin for protein expression and purification and Johan Evenäs for fruitful discussions.

SUPPORTING INFORMATION AVAILABLE

^{1}H NMR assignments of the Ce_2-TR_1C complex at 298 K (Table S1) (4 pages). Ordering information is given on any current masthead page.

REFERENCES

Akke, M., Skelton, N. J., Kördel, J., Palmer, A. G., III, & Chazin, W. J. (1993) *Biochemistry 32*, 9832.

Akke, M., Forsén, S., & Charm, W. S. (1995) *J. Mol. Biol. 252*, 102.

Alemà, S. (1984) in *Metal ions in biological systems: calcium and its role in biology* (Sigel, H., Ed.) pp 275−317, Marcel Dekker, New York.

Alsaadi, B. M., Rossotti, F. J. C., & Williams, R. J. P. (1980) *J. Chem. Soc., Dalton Trans.*, 2147.

Babu, Y. S., Bugg, C. E., & Cook, W. J. (1988) *J. Mol. Biol. 204*, 191.

Banci, L., Bertini, I., Luchinat, C., Piccioli, M., Scozzafava, A., & Turano, P. (1989) *Inorg. Chem. 28*, 4650.

Banci, L., Bertini, I., Eltis, L. D., Felli, I. C., Kastrau, D. H. W., Luchinat, C., Piccioli, M., Pierattelli, R., & Smith, M. (1994) *Eur. J. Biochem. 225*, 715.

Banci, L., Bertini, I., Bren, K. L., Cremonini, M. A., Gray, H. B., Luchinat, C., & Turano, P. (1996) *JBIC, J. Biol. Inorg. Chem. 1*, 117.

Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., & Bax, A. (1992) *Biochemistry 31*, 5269.

Bertini, I., & Luchinat, C. (1986) in *NMR of paramagnetic molecules in biological systems*, Benjamin/Cummings, Menlo Park, CA.

Bertini, I., & Luchinat, C. (1996) in NMR of paramagnetic substances, *Coord. Chem. Rev.*, 150.
Bertini, I., Couture, M. M. J., Donaire, A., Eltis, L. D., Felli, I. C., Luchinat, C., Piccioli, M., & Rosato, A. (1996a) *Eur. J. Biochem. 241*, 440.
Bertini, I., Luchinat, C., & Rosato, A. (1996b) *Prog. Biophys. Mol. Biol. 66*, 43.
Bertini, I., Donaire, A., Luchinat, C., & Rosato, A. (1997) *Proteins: Struct., Funct., Genet.* (in press).
Birnbaum, E. R., & Sykes, B. D. (1978) *Biochemistry 17*, 4965.
Blake, P. R., Lee, B., Summers, M. F., Adams, M. W. W., Park, J. B., Zhou, Z. H., & Bax, A. (1992) *J. Biomol. NMR 2*, 527.
Chattopadhyaya, R., Meador, W. E., Means, A. R., & Quiocho, F. A. (1992) *J. Mol. Biol. 228*, 1177.
Coleman, J. E. (1993) *Methods Enzymol. 227*, 16.
Cox, J. A., Comte, M., Malnoë, A., Burger, D., & Stein, E. A. (1984) in *Metal ions in biological systems: calcium and its role in biology* (Sigel, H., Ed.) pp 215−273, Marcel Dekker, New York.
Crivici, A., & Ikura, M. (1995) *Annu. Rev. Biophys. Biomol. Struct. 24*, 85.
Einstein, A. (1956) in *Investigations on the Theory of the Brownian Movement*, Dover, New York.
Finn, B. E., Evenäs, J., Drakenberg, T., Waltho, J., Thulin, E., & Forsén, S. (1995) *Nat. Struct. Biol. 2*, 777.
Forsén, S., & Kördel, J. (1994) in *Bioinorganic chemistry* (Bertini, I., Gray, H. B., Lippard, S. J., & Valentine, J. S., Eds.) pp 107−166, University Science Books, Mill Valley, CA.
Gochin, M., & Roder, H. (1995a) *Protein Sci. 4*, 296.
Gochin, M., & Roder, H. (1995b) *Bull. Magn. Reson. 17*, 1.
Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870.
Guiles, R. D., Sarma, S., DiGate, R. J., Banville, D., Basus, V. J., Kuntz, I. D., & Waskell, L. (1996) *Nat. Struct. Biol. 3*, 333.
Güntert, P., Braun, W., & Wüthrich, K. (1991) *J. Mol. Biol. 217*, 517.
Güntert, P., Mumenthaler, C., & Wüthrich, K. (1996) *XVIIth International Conference on Magnetic Resonance in Biological Systems* (Abstract).
Horrocks, W. D., Jr., & Sudnick, D. R. (1981) *Acc. Chem. Res. 14*, 384.
Huber, J. G., Moulis, J.-M., & Gaillard, J. (1996) *Biochemistry 35*, 12705.
Ikura, M., Kay, L. E., & Bax, A. (1990) *Biochemistry 29*, 4659.
Ikura, M., Spera, S., Barbato, G., Kay, L. E., Krinks, M., & Bax, A. (1991) *Biochemistry 30*, 9216.
Ishima, R., Shibata, S., & Akasaka, A. (1991) *J. Magn. Reson. 91*, 455.
Johnson, R. D., Ramaprasad, S., & La Mar, G. N. (1983) *J. Am. Chem. Soc. 105*, 7205.
Klee, C. B. (1988) in *Molecular Aspects of Cellular Regulation* (Cohen, P., & Klee, C. B., Eds.) pp 35−56, Elsevier, New York.
Koradi, R., Billeter, M., & Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51.
Kowall, T., Foglia, F., Helm, L., & Merbach, A. E. (1995) *J. Am. Chem. Soc. 117*, 3790.
Kretsinger, R. H. (1980) *CRC Crit. Rev. Biochem. 8*, 119.
Kuszewski, J., Gronenborn, A. M., & Clore, G. M. (1995) *J. Magn. Reson., Ser. B 107*, 293.
Linse, S., Helmersson, A., & Forsén, S. (1991) *J. Biol. Chem. 266*, 8050.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967.
Ösapay, K., & Case, D. A. (1991) *J. Am. Chem. Soc. 113*, 9436.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479.
Sklenar, V., Piotto, M., Leppik, R., & Saudek, V. (1993) *J. Magn. Reson., Ser. A 102*, 241.
Stokes, G. (1956) *Trans. Cambridge Philos. Soc. 9*, 5.
Strynadka, N. C. J., & James, M. N. G. (1989) *Annu. Rev. Biochem. 58*, 951.
Tjandra, N., Kuboniwa, H., Ren, H., & Bax, A. (1995) *Eur. J. Biochem. 230*, 1014.
Utschig, L. M., Bryson, J. W., & O'Halloran, T. V. (1995) *Science 268*, 380.
Vold, R. L., Waugh, J. S., Klein, M. P., & Phelps, D. E. (1968) *J. Chem. Phys. 48*, 3831.
Wang, P. L., Donaire, A., Zhou, Z. H., Adams, M. W. W., & La Mar, G. N. (1996) *Biochemistry 35*, 11319.
Wider, G., Macura, S., Kumar, A., Ernst, R. R., & Wüthrich, K. (1984) *J. Magn. Reson. 56*, 207.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1992) *Biochemistry 31*, 1647.
Wu, S. L., Franklin, S. J., Raymond, K. N., & Horrocks, W. D., Jr. (1996) *Inorg. Chem. 35*, 162.
Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, Wiley, New York.

BI971022+