Journal of Biomolecular NMR, **21:** 85–98, 2001.
KLUWER/ESCOM

Paramagnetism-based versus classical constraints: An analysis of the solution structure of Ca Ln calbindin D_{9k}

Ivano Bertini[a,*], Antonio Donaire[b], Beatriz Jiménez[b], Claudio Luchinat[c], Giacomo Parigi[c], Mario Piccioli[a] & Luisa Poggi[a]

[a]*Magnetic Resonance Center (CERM) and Department of Chemistry, University of Florence Via L. Sacconi, 6, I-50019 Sesto Fiorentino, Italy;* [b]*Department of Chemistry, Universidad Cardenal Herrera-CEU, Avda. Seminario s/n, 46113-Montcada, Valencia, Spain;* [c]*Magnetic Resonance Center (CERM) and Department of Agricultural Biotechnology, University of Florence, Via L. Sacconi, 6, I-50019 Sesto Fiorentino, Italy*

Received 17 April 2001; Accepted 16 July 2001

Key words: calcium binding proteins, paramagnetic NMR, paramagnetism based constraints, solution structure, sparse NOEs, structure refinement

Abstract

The relative importance of paramagnetism-based constraints (i.e. pseudocontact shifts, residual dipolar couplings and nuclear relaxation enhancements) with respect to classical constraints in solution structure determinations of paramagnetic metalloproteins has been addressed. The protein selected for the study is a calcium binding protein, calbindin D_{9k}, in which one of the two calcium ions is substituted with cerium(III). From 1823 NOEs, 191 dihedral angles, 15 hydrogen bonds, 769 pseudocontact shifts, 64 orientational constraints, 26 longitudinal relaxation rates, plus 969 pseudocontact shifts from other lanthanides, a final family with backbone r.m.s.d. from the average of 0.25 Å was obtained. Then, several families of structures were generated either by removing subsets of paramagnetism-based constraints or by removing increasing numbers of NOEs. The results show the relative importance of the various paramagnetism-based constraints and their good complementarity with the diamagnetic ones. Although a resolved structure cannot be obtained with paramagnetism-based constraints only, it is shown that a reasonably well resolved backbone fold can be safely obtained by retaining as few as 29 randomly chosen long-range NOEs using the standard version of the program PSEUDYANA.

Introduction

In the early 1980s the basis were laid down for the solution structure determination of proteins (Wüthrich, 1996). In the beginning the structural constraints were essentially NOEs and ^3J coupling values and the software was essentially based on optimisation of the dihedral angles of polypeptides in order to match the NOE based distances and the ranges of dihedral angles determined by the ^3J values (Güntert et al., 1991; Güntert and Wüthrich, 1991). The NOEs do not provide precise distances as they are affected by mobility and spin diffusion. Even the angle constraints were used with large tolerance. As a consequence, a good structure was, and still is, the result of as many as possible constraints.

During the years new constraints have been used into structure calculations: angles obtained from cross-correlation effects (Reif et al., 1997), residual dipolar couplings (rdc) when the protein is partially oriented in high magnetic fields (Tolman et al., 1995; Tjandra et al., 1996, 1997; Vold and Prosser, 1996; Bax and Tjandra, 1997) and hydrogen bonds (Cordier and Grzesiek, 1999), together with constraints derived from the Chemical Shift Index (CSI) (Wishart et al., 1991; Gagne' et al., 1994). The strong interest for new constraints arises from the need of solving structures without NOEs or with relatively few of them. Indeed,

*To whom correspondence should be addressed. E-mail: bertini@cerm.unifi.it

the assignment of protein signals is less limited by the protein size than the obtainment of NOE cross-peaks (Yamazaki et al., 1997) (Salzmann et al., 1999; Hus et al., 2000). In proteins containing paramagnetic metal ions paramagnetism-based constraints can be exploited (Bertini et al., 2001c). They are the contact shifts, the pseudocontact shifts, the hyperfine shifts as sum of the two, and the relaxation enhancements. The contact shifts, whenever available, contain structural information in principle, although only in a few cases a relationship was proposed (Bertini et al., 1994). The use of hyperfine shifts as such was suggested in the case of low spin heme proteins (Bertini et al., 1999). The pseudocontact shifts (pcs) (Banci et al., 1996, 1998a; Kechuan and Gochin, 1999), which are operative in the presence of magnetic anisotropy, are given by the following equation (Kurland and McGarvey, 1970):

$$\delta_i^{pcs} = \frac{1}{12\pi r_i^3}[\Delta\chi_{ax}(3\cos^2\theta_i - 1) + \tfrac{3}{2}\Delta\chi_{rh}\ \sin^2\theta_i\ \cos 2\phi_i], \quad (1)$$

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 the atom i and the metal ion, and θ_i and ϕ_i are the spherical polar angles of atom i with respect to the principal axes of the magnetic susceptibility tensor centred on the metal ion.

The occurrence of magnetic anisotropy causes partial orientation in high magnetic field, which generates rdc (Tolman et al., 1995; Tjandra et al., 1996; Vold and Prosser, 1996; Bax and Tjandra, 1997). The equation describing this effect is similar to that of pcs. However, the metal nucleus distance is not present as a parameter, and the magnetic anisotropy is that of the whole protein instead of that of the metal ion (Banci et al., 1998a). In the system studied here the two anisotropies are very close.

R_{1p} and R_{2p} (where p stays for paramagnetic) are proportional to r^{-6}, where r is the distance between the unpaired electron and the resonating nucleus. As such, they are constraints similar to NOEs (Bertini et al., 1996). R_{1p} involves many more nuclei than R_{2p}, as diamagnetic $R_2 > R_1$. For R_{1p}, cross-relaxation tends to equalize the values (Granot, 1982; La Mar and de Ropp, 1993), and for both R_{1p} and R_{2p} the assumption that the electron is localized on the metal ion(s) may be a problem (Wilkens et al., 1998). These issues have been analysed and assessed in previous works, so that now safe procedures are available (Bertini et al., 1996). Finally, NMR approaches have been developed to identify metal ion ligands (Bertini et al., 2001a). The latter can be used as structural constraints.

The solution structure of an artificial paramagnetic molecule was solved with a wealth of constraints both classical/diamagnetic and paramagnetism-based. The protein is the mono Ce^{3+} substituted Calbindin D_{9k}, which has 76 amino acids and bears two Ca^{2+} ions in its native state (Linse et al., 1987). The structure is well resolved, especially in the backbone part, on which the efforts were concentrated. Such a well resolved structure of a paramagnetic molecule is already a meaningful result, which we would like to pose to the attention of the scientific community. In fact, paramagnetism decreases the number and quantitative meaning of classical NOEs, especially around the metal ion, which is then compensated by the new constraints. The use of paramagnetism-based constraints on this system had already been explored by us (Allegrozzi et al., 2000). In this research, a number of diamagnetic constraints was determined by extending the whole assignment on a ^{13}C, ^{15}N-labeled sample (i) in order to obtain a very well defined structure as a starting point, and (ii) to make sure that the diamagnetic constraints were by themselves sufficient for a high resolution structure. Then, one class at a time of paramagnetism-based constraints has been neglected, to assess its contribution in the definition of the structure. Therefore, this study shows the relevance of each paramagnetism-based constraint. The ultimate goal is to contribute to the discussion on the possibility of solving solution structure of proteins without, or with a limited number, of NOEs. This is essential in the framework of high throughput structure determination aiming at the obtainment of quick, reliable information on protein scaffold.

Material and methods

Sample preparation

Protein expression (Brodin et al., 1986) and purification (Johansson et al., 1990) of both the Ca^{2+} and the apo form of the bovine Pro43→Met43 (P43M) mutant (Chazin et al., 1989; Malmendal et al., 1998) of calbindin D_{9k} was performed as reported. The expression system was a generous gift of Prof. S. Forsén. The labelling and lanthanide substitution procedure is described elsewhere (Allegrozzi et al., 2000; Bertini et al., 2001a). The pH was adjusted to 6.0 by

means of 0.1 M NaOH or 0.1 M HCl. The samples were kept at 4 °C in between measurements. Sample concentrations range from 1.5 to 2.0 mM.

NMR spectroscopy

Experiments were performed at 300 K and on Bruker Avance spectrometers operating at 600, 700 or 800 MHz. Three-dimensional HNCA and HNCO experiments (Kay et al., 1990) were performed to assign backbone resonances. For the above experiments spectral windows of 12.8 ppm for ^{1}H, 28 ppm for ^{15}N, and 14 ppm (HNCO) or 27 ppm (HNCA) for ^{13}C dimensions were typically used. The number of real data points acquired were 128 or 256 in the t_1 dimension (^{13}C), from 40 to 64 in the t_2 dimension (^{15}N), and 1024 or 2048 in acquisition (t_3 dimension). Three dimensional HNHA (Vuister and Bax, 1993), NOESY-HSQC (Kay et al., 1989), and HNHB (Archer et al., 1991) experiments were carried out to determine ϕ, ψ and χ dihedral angles restraints. Spectral widths of 12 ppm for both ^{1}H dimensions and 30 ppm for ^{15}N dimension were used. The number of real data points acquired were from 64 to 96 points in the ^{15}N dimension, from 128 to 256 in the indirect ^{1}H dimension, and 1024 or 2048 in acquisition (t_3 dimension). The 3D NOESY-HSQC experiment was repeated at 278 K to observe the exchangeable amide protons of Gly42 and Met43. For all the above experiments, from 8 to 16 scans were collected, using a relaxation delay of 800 ms. The total time for each scan was, in all cases, *ca*. 2 seconds. Decoupling of ^{15}N nuclei during acquisition was performed using a GARP pulse sequence (Shaka et al., 1985).

A 2D HNCO experiment with a ^{15}N-^{13}C inept transfer delay of 66.6 ms was also acquired. This delay optimizes the coherence transfer between an amide proton and the carbonyl of a group that is making hydrogen bond with it to provide spectroscopic evidence for H-bond patterns (Cordier and Grzesiek, 1999).

Two dimensional non-selective inversion-recovery ^{13}C and ^{15}N HSQC experiments were performed to measure non-selective longitudinal relaxation rates of protons (Bertini et al., 1996). In order to measure T_1 values of very fast relaxing protons, all (inept transfer, acquisition and relaxation) delays were shortened. The inept transfer delays were 1.4 and 0.8 ms for ^{15}N and ^{13}C, respectively, relaxation delays were typically 500 ms. In order to measure T_1 values of moderately fast relaxing protons, ^{13}C and ^{15}N non-selective IR-HSQC were collected with usual delays for inept transfer and relaxation ($1/4^{1}J$ and 1 s, respectively).

All the data were zero-filled in the indirect dimensions and apodized using cosine squared functions. All NMR data were processed with the Bruker XWIN-NMR software packages. The program XEASY (ETH Zürich) (Eccles et al., 1991) was used for the analysis of the NMR spectra.

NMR restraints

NOESY restraints

The dipole-dipole restraints used in the present work are those obtained from NOESY and 1D NOE experiments for calbindin D_{9k} in which the native calcium in site II is replaced by cerium (CaCeCb hereafter), as previously reported (Allegrozzi et al., 2000). However, the 1D NOE intensities were re-calibrated as the refinement process advanced, according to the CALIBA protocol (Güntert et al., 1991).

Dihedral angle constraints

The $^{3}J_{\mathrm{HN}\alpha}$ coupling constants were obtained from the ratio between the intensity of the diagonal peak ($\nu_{1\mathrm{HN}}$, $\nu_{2\mathrm{N}}$, $\nu_{3\mathrm{HN}}$) and that of the cross peak ($\nu_{1\mathrm{H}\alpha}$, $\nu_{2\mathrm{N}}$, $\nu_{3\mathrm{HN}}$) integrated on 3 to 5 ^{1}H-^{1}H planes in the HNHA experiment (Vuister and Bax, 1993). They were converted into constraints for the backbone torsion angle ϕ by means of the appropriate Karplus curve (Bax and Wang, 1995). An uncertainty of $\pm 10°$ was given to all angles. Among possible solutions of the Karplus equations, dihedral angle values were selected to be consistent with those found in structures obtained without the use of these constraints. For 8 $^{3}J_{\mathrm{HN}\alpha}$ values out of 59 a range of accepted dihedral angles larger than 20° was considered.

The ψ dihedral angles were obtained in the ^{15}N NOESY-HSQC experiment from the ratio (Gagne' et al., 1994):

$$I_{\mathrm{HN}\alpha}/I_{\mathrm{HN}\alpha-1},$$

where $I_{\mathrm{HN}\alpha}$ is the intensity of the NOESY cross-peak between an amide proton and the α proton of the same residue, and, analogously, $I_{\mathrm{HN}\alpha-1}$ is the intensity of the same amide proton and the α proton of the precedent residue. When the value of this ratio was larger than 2.0, a ψ angle ranging from -65 to $+5$ was assumed for all residues but Ile, Thr and Ala, for which values between -65 and -15 were assumed. When the ratio was lower than 1.0, the ψ angle was assumed to be included in the 100 to 170 range for all residues.

Stereospecific assignments of methylene proton pairs were determined from $^3J_{HNH\beta}$ values. The latter were obtained from the relative intensity of the two H_N-H_β cross-peaks in the 3D HNHB experiment (Archer et al., 1991).

Relaxation rates

Longitudinal relaxation rates of protons were measured from non selective inversion-recovery ^{13}C and ^{15}N HSQC experiments (Bertini et al., 1996). The intensities of the signals were plotted versus the recovery delays and fit to a mono-exponential three parameters function. The diamagnetic contribution (R_{1dia}) was subtracted from the experimental relaxation rates (R_1^{obs}) according to the following expression:

$$R_1^{obs} = R_{1dia} + R_{1p},$$

where R_{1dia} was estimated from the average of the R_1^{obs} values for protons far from the metal ion (for which $R_{1p} = 0$). The R_{1dia} values used in the present work were 2.0 s^{-1} for amide protons and 1.2 s^{-1} for protons bound to carbon nuclei. In turn, the obtained R_{1p} values were plotted versus r^{-6} (r being the proton-metal distance) in different steps of the calculations and then converted in to lower and upper distance limits according to the following equation:

$$R_{1p} = K \cdot r^{-6}.$$

The value of K varies as the degree of resolution is increased. The final K value used was 2.6×10^{-55} s^{-1} m^6 for both amide and aliphatic protons.

Additionally, the longitudinal relaxation rates of four ^{13}C (Asp54 Cγ, Asn56 Cγ, Asp 65 Cδ, and Val 61 Cα) nuclei in the CaCeCb derivative were determined and translated into upper and lower distance limits by applying the same methodology explained above. Due to the weaker effect of the hyperfine interaction on the relaxation rates of low γ nuclei (Bertini et al., 2001a), only the above resonances were found to be unambiguously affected. The diamagnetic contribution was estimated to be 0.7 s^{-1}.

Pseudocontact shifts

Pseudocontact shifts are introduced as constraints in the target function of PSEUDYANA calculations as a sum of differences between the experimental and the calculated pcs for the observed nuclei in each family of structures according to the previously reported procedure (Banci et al., 1998b). The experimental pseudocontact shifts for ^{15}N and backbone ^1H of CaLnCb derivatives (Ln = Ce, Pr, Nd, Sm, Eu, Tb, Dy, Ho, Er, Tm, Yb) were obtained from the observed chemical shifts according to the expression:

$$\delta^{exp} = \delta^{dia} + \delta^{pcs},$$

where δ^{dia} was the chemical shift of the corresponding nucleus in the CaLaCb native protein. Pseudocontact values for ^{13}C atoms and side chains protons were obtained using the CaCaCb derivative as a diamagnetic analogue. The calculated δ^{pcs} values were obtained by assuming the metal-center point dipole approximation, i.e., by applying Equation (1) to the corresponding nuclei. In turn, the magnetic anisotropy components as well as the orientation of the tensor was recalculated as the refinement of the structures increased, until convergence. The recalculated tensor after each refinement step was obtained using the program FANTASIA(Banci et al., 1997) by omitting the pcs values from the metal ligands as well as those from the nearby residues (57, 62–64).

Hydrogen bonds

When the existence of a hydrogen bond was observed in the specific HNCO experiment (Cordier and Grzesiek, 1999) (see above) the following restrictions were given:

$$r_{H...O} < 2.2 \text{ Å} \quad \text{and} \quad 2.6 \text{ Å} < r_{N...O} < 3.3 \text{ Å}$$

where $r_{H...O}$ and $r_{N...O}$ are the distance of the oxygen of the hydrogen bond with the amide proton and its nitrogen nucleus, respectively. The lower distance limits were imposed to keep the linearity of the hydrogen bond.

Residual dipolar coupling

The rdc constraints used in the present work are those ones previously reported for the CaCeCb derivative. The procedure followed to obtain them and for their introduction as restraints in PSEUDYANA calculations have been previously reported in detail (Bertini et al., 2001d).

Structure calculation

The modified version of the program DYANA (Güntert et al., 1997), called PSEUDYANA (Banci et al., 1998b), which permits the simultaneous use of pcs and of rdc as constraints, was used for the structure calculations. Torsion Angle Dynamics (TAD) combined with a simulated annealing algorithm was used to calculate a family of 200 structures starting from randomly generated conformers in 10 000 annealing steps. The quality of structures calculated by

Table 1. Type and number of constraints used in PSEUDYANA calculations

Type of constraints	Number of constraints
Diamagnetic constraints	
NOESY	2146 (1793 meaningful)
1D NOE	30 (13 meaningful)
Hydrogen bonds	15 (30 upper distance limits, 15 lower distance limits)
ϕ angle constraints	59
ψ angle constraints	46
χ angle constraints	86 (used as stereo specific assignment constraints)
Total diamagnetic constraints	2042 meaningful
Paramagnetism-based constraints	
T_1	26
rdc	64
pcs (from 11 lanthanides)	1738 (562 N, 573 H_N, 156 C of Ce^{3+}, 447 HC of Ce^{3+})
Metal-ligand constraints	10
Total Paramagnetism-based constraints	1838
Total constraints	3880

PSEUDYANA can be assessed by a properly defined function (target function) proportional to the squared deviations of the calculated constraints from the experimental ones, plus the squared van der Waals contact violations. In DYANA (and in PSEUDYANA as well), the Target Function TF has the role of a pesudo potential energy and is defined such that TF = 0 if and only if all experimental distance constraints and torsion angle constraints are fulfilled and all non-bonded atom pairs satisfy a check for the absence of steric overlap (Güntert et al., 1997). TF values are expressed in $Å^2$. Mean structures obtained were minimized using the program PSEUDYANA with a 1000 steps conjugate gradient minimization.

The programs MOLMOL and PROCHECK were subsequently used to analyse the calculated structures.

Results

NMR assignment

The assignment of the backbone ^{13}C nuclei in the CaCeCb derivative was extended, as well as that in the native CaCaCb protein, by the analysis of the HNCO and HNCA experiments. These new assignments are included in the overall assignment, which is given in the Supplementary Material section. Furthermore, an HNCO experiment performed with a ^{15}N-^{13}CO inept transfer delay of 66.6 ms (Cordier and Grzesiek, 1999) allowed us to identify 15 amide protons and their corresponding carbonyl groups to which they are linked through hydrogen bonds.

NMR constraints

The number of constraints used in PSEUDYANA calculations are given in Table 1. They are grouped according to their origin in diamagnetic and paramagnetism-based constraints. The list of all the new constraints used in the present work is given in the Supplementary Material section.

The diamagnetic constraints contain both short range constraints derived from NOE and ^3J coupling constants and long range constraints derived from the observation of hydrogen bonds. The NOESY constraints used in the present work (1793) are those previously reported (Allegrozzi et al., 2000), without any change in the calibration.

Constraints relative to the ϕ, ψ and χ dihedral angles (Table 1) were obtained from the measurements of the $^3J_{HNH\alpha}$, the $I_{HNH\alpha}/I_{HNH\alpha-1}$ ratio, and the $^3J_{HNH\beta}$ values, respectively. Thirteen 1D NOE meaningful constraints (Table 1) are those used in the previous work (Allegrozzi et al., 2000), but they have been re-calibrated as the resolution of the structure was improving during the calculations.

As in the case of diamagnetic constraints, the paramagnetism-based constraints contain both short range and long range constraints. The former arise

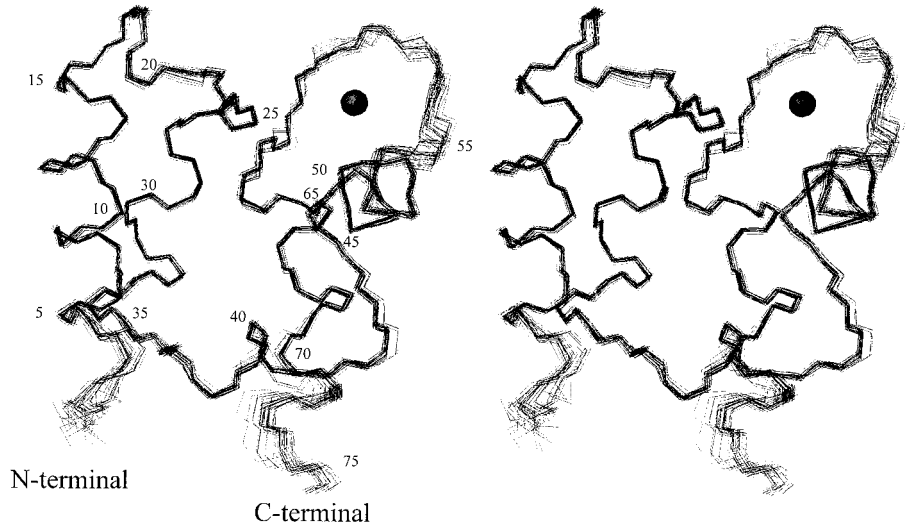

Figure 1. Stereo view of family of accepted conformers (30 structures).

from relaxation rates from signals in the vicinity of the paramagnetic center. Longitudinal relaxation rates corresponding to 22 protons obtained from inversion recovery HSQC experiments and 4 longitudinal relaxation times of ^{13}C nuclei were introduced as restraints in the structure calculations (Table 1).

As far as long range paramagnetism-based constraints are concerned, 166 pcs from peptide N and NH nuclei and 156 carbon nuclei assigned for the CaCeCb derivative were also introduced as constraints, together with pcs from 969 peptide N and NH obtained from ten different CaLnCb derivatives. Finally, the rdc obtained for 64 NH pairs in the CaCeCb derivative were also included as constraints in the structure calculations.

In total, as it is described in Table 1, 3880 meaningful constraints were used in the structure calculations, the diamagnetic constraints being 52.3%, versus 47.7% arising from the presence of the paramagnetic center.

Table 2. Contribution to the total target function of the final structure of single groups of constraints

Type of constraint	Contribution (%)
Diamagnetic constraints	
Van der Waals	7.4
Hydrogen Bonds	2.3
Tensors	1.0
Angles	3.4
1D NOE	0.0
NOESY	16.3
Total diamagnetic constraints	30.4
Paramagnetism-based constraints	
T_1	1.5
pcs*	46.7
rdc	21.4
Total paramagnetism-based constraints	69.6

*Single contributions of each metal are summarized in Table 3.

Solution structure calculations

Structure calculations were performed with the program PSEUDYANA using the constraints shown in Table 1. The structure consists of four helices and three loops, of which the first and the third constitute the lanthanide-binding site. A stereo view of backbone atoms for the 30 conformers with lowest target function is shown in Figure 1. Such an ensemble of conformers has a root mean square deviation (r.m.s.d.) for the backbone atoms of 0.25 ± 0.07 Å (the r.m.s.d. values presented here are all given with regard to the average value). The mean global heavy atom r.m.s.d. for this ensemble of conformers is 1.09 ± 0.10 Å. Figure 2 (open dots) depicts the r.m.s.d. per residue. Besides C- and N-terminal sites, a slightly larger r.m.s.d. is observed for the three loops with a maximum for the cerium-binding loop. The target function obtained for this family of structures ranges from 1.74 to 1.90 $Å^2$. The contributions of each kind of constraint to the global value are given

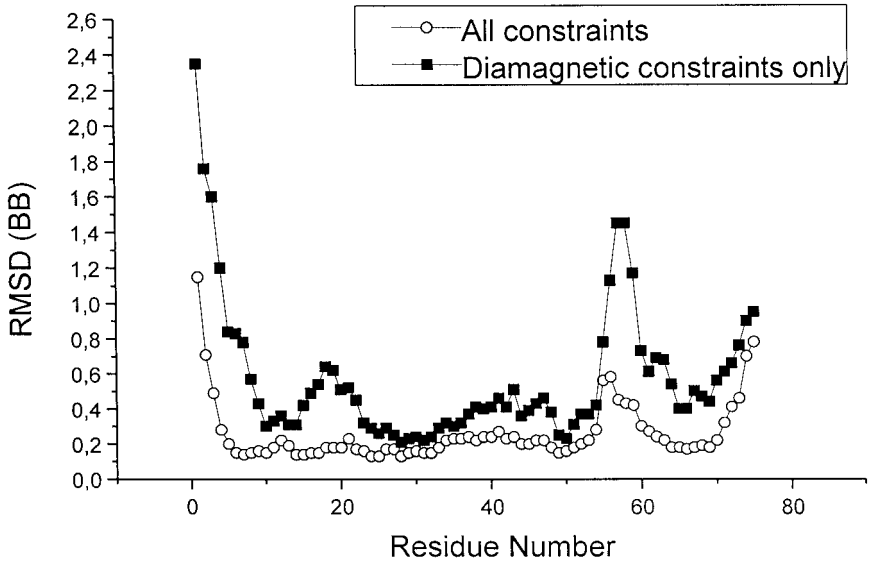

Figure 2. Backbone r.m.s.d. plot of backbone atoms for the structure obtained using only diamagnetic constraints and from that obtained using all available constraints.

in Table 2. Most of the contributions arise from long range paramagnetism-based constraints (68%). This is not surprising since PSEUDYANA (as all structure calculations programs written up to now) is thought to build the protein tertiary structures by modulating local changes, i.e., from short range constraints.

The orientation of the magnetic axes and the magnitude of the main components of the magnetic susceptibility tensor for the CaCeCb derivative are similar to those already reported. In fact, the values obtained in the present work are $\Delta\chi_{ax} = 2.10 \times 10^{-32}$ m^3 and $\Delta\chi_{rh} = 0.80 \times 10^{-32}$ m^3, in good agreement with those previously reported ($\Delta\chi_{ax} = 2.08 \times 10^{-32}$ m^3 and $\Delta\chi_{rh} = 0.71 \times 10^{-32}$ m^3) (Bertini et al., 2001b).

Table 3 shows the contribution to the target function of the pcs of each lanthanide ion. It is clear that the high contribution of such set of constraints to the target function is due to the CaCeCb derivative, in agreement with the fact that this derivative is the one with the highest number of constraints, including those arising from non exchangeable protons and from ^{13}C backbone resonances.

We have analysed the effect that paramagnetism-based constraints (pcs, rdc, and relaxation rates) have both on the final resolution of the family of structures and on the overall target function. In order to do this, we have performed several PSEUDYANA runs eliminating, in each calculation, a class or part of a class of constraints. We have performed in total thirteen of these runs. In Table 4, the r.m.s.d. and target functions

Table 3. Contribution of each metal to the total pcs target function of the final structure

Metal	Contribution (%)
Ce (N, HN)	8.2
Ce (all others)	37.9
Dy	5.6
Er	0.0
Eu	1.7
Ho	10.2
Nd	3.3
Pr	10.1
Sm	9.7
Tb	5.8
Tm	1.0
Yb	6.5

(TF) of the obtained families (all of them of 30 conformers) are given. For all calculations with subsets of constraints, the r.m.s.d. of the corresponding average structure from the final average structure is given.

Discussion

The effect of introducing different classes of structural constraints in a structure calculation should be considered with respect to two points: (i) the effect of the new

Table 4. Effect of each kind of paramagnetism-based constraint in the final resolution of the structures

Restraints used	r.m.s.d. from the final average structure	Structural statistics for the family of structures		Ramachandran plot statistics for the family of structures			
	Backbone r.m.s.d. (Å)	Backbone r.m.s.d.	Target functon (Å^2)	A (%)	B (%)	C (%)	D (%)
All	–	0.25 ± 0.07	1.74–1.90	92.3	7.7	0	0
Without pcs from ^{1}HC and ^{13}C	0.32	0.30 ± 0.04	1.32–1.50	92.3	7.7	0	0
Without pcs	0.59	0.43 ± 0.09	0.66–0.74	84.6	15.4	0	0
Without rdc	0.40	0.32 ± 0.07	1.22–1.45	89.2	10.8	0	0
Without T_1	0.35	0.25 ± 0.07	1.19–1.31	89.2	10.8	0	0
Without short-range pcs	0.41	0.40 ± 0.08	0.92–1.00	90.8	9.2	0	0
Without medium- and long-range pcs	0.44	0.39 ± 0.08	1.12–1.35	91.2	8.8	0	0
Without T_1 and rdc	0.46	0.33 ± 0.05	0.74–0.88	86.2	13.8	0	0
Without T_1 and pcs	0.67	0.54 ± 0.13	0.61–0.69	87.7	12.3	0	0
Without pcs and rdc	0.65	0.55 ± 0.11	0.23–0.33	83.1	16.9	0	0
Without T_1, rdc, and medium- and long-range pcs	0.61	0.54 ± 0.08	0.43–0.70	90.8	9.2	0	0
Without T_1, rdc, and short-range pcs	0.45	0.48 ± 0.11	0.43–0.58	87.7	12.3	0	0
Without T_1, pcs and rdc	0.90	0.61 ± 0.13	0.22–0.30	84.6	15.4	0	0
Without T_1, pcs, rdc and ligands	1.01	0.69 ± 0.18	0.22–0.32	83.1	15.4	1.5	0

A = most favoured regions.
B = additionally allowed regions.
C = generously allowed regions.
D = disallowed regions.

constraints on the accuracy and precision of the structure and (ii) the consistency with the old constraints. Comparing the structure obtained with the available diamagnetic constraints only with the final structure of Calbindin D_{9k}, a significantly lower r.m.s.d. in the latter is observed (from 0.69 to 0.25 Å) , which is a measure of the precision of the structure. The overall consistency of the constraints can be appreciated from the fact that the r.m.s.d. decreases in the presence of paramagnetism-based constraints (empty circles in Figure 2) without increasing the target function of the diamagnetic constraints and without any increase in the number of consistent violations (only 2 consistent violations being observed with a 0.1 Å threshold). This proves that the two sets of diamagnetic and paramagnetism-based constraints are fully consistent.

In the following, a series of calculations is presented where several paramagnetism-based constraints are, in turn, removed. These calculations are instructive on the intrinsic value and complementarily of these constraints. Even more importantly, calculations excluding increasing numbers of NOE constraints are also shown. The results are relevant for the general issue of how much non-NOE constraints can substitute for NOEs in actual solution structure determinations.

Inspection of Table 4 indicates that pcs are important for the final quality of the solution structure. In fact, without these constraints the global r.m.s.d. increases from 0.25 Å to 0.43 Å. Rdc constraints are also important although to a lower degree, the r.m.s.d. increasing from 0.25 Å to 0.32 Å without these constraints. As only 64 rdc values are used vs. about 1700 pcs constraints, their contribution is still remarkable. Furthermore, they improve the quality of the Ramachandran plot of the structure more than the 1738 pcs constraints, as observed in Table 4. This confirms that rdc constraints in general, and those arising from the orientation induced by a paramagnetic center in particular, are precious constraints.

Figure 3 allows us to analyse in a finer detail how pcs and rdc affect the local resolution of the structure. As already discussed in previous works (Bertini et al., 2001b), the combined use of different lanthanides removes the intrinsic limitation provided by a single metal ion to act only within a spherical shell from the metal center whose inner and outer radii are metal ion dependent. The availability of different active shells for different lanthanides allows pcs to be operative all over the protein scaffold, from the region from residue

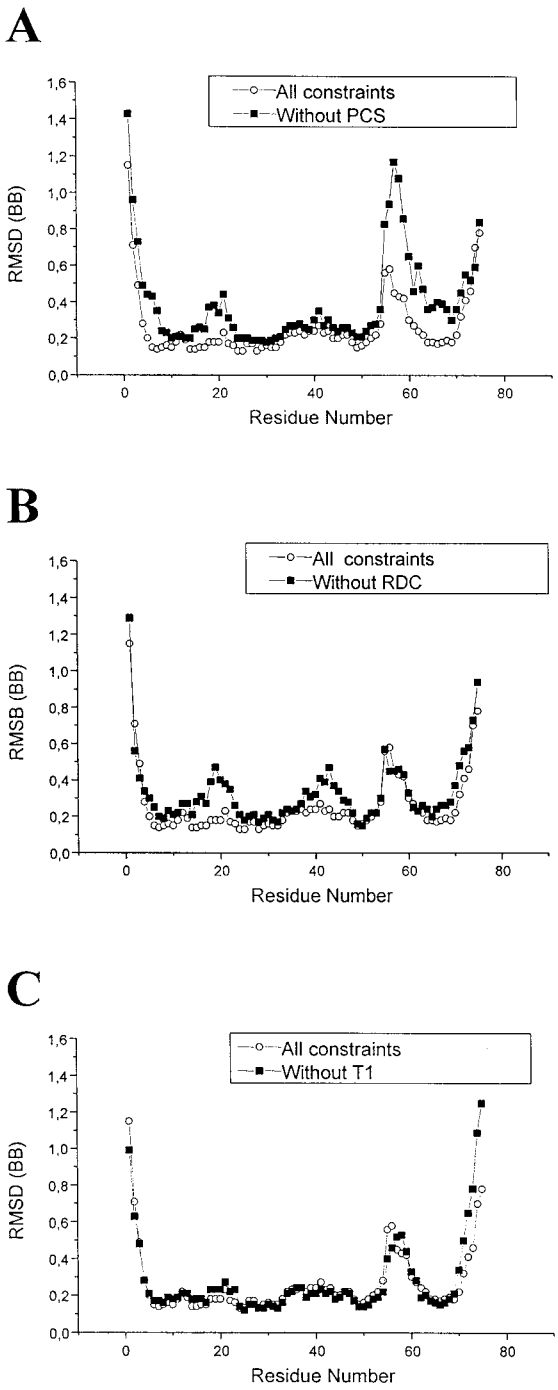

Figure 3. Backbone r.m.s.d. for (A) para vs. para without pcs; (B) para vs. para without rdc; (C) para vs. para without T_1.

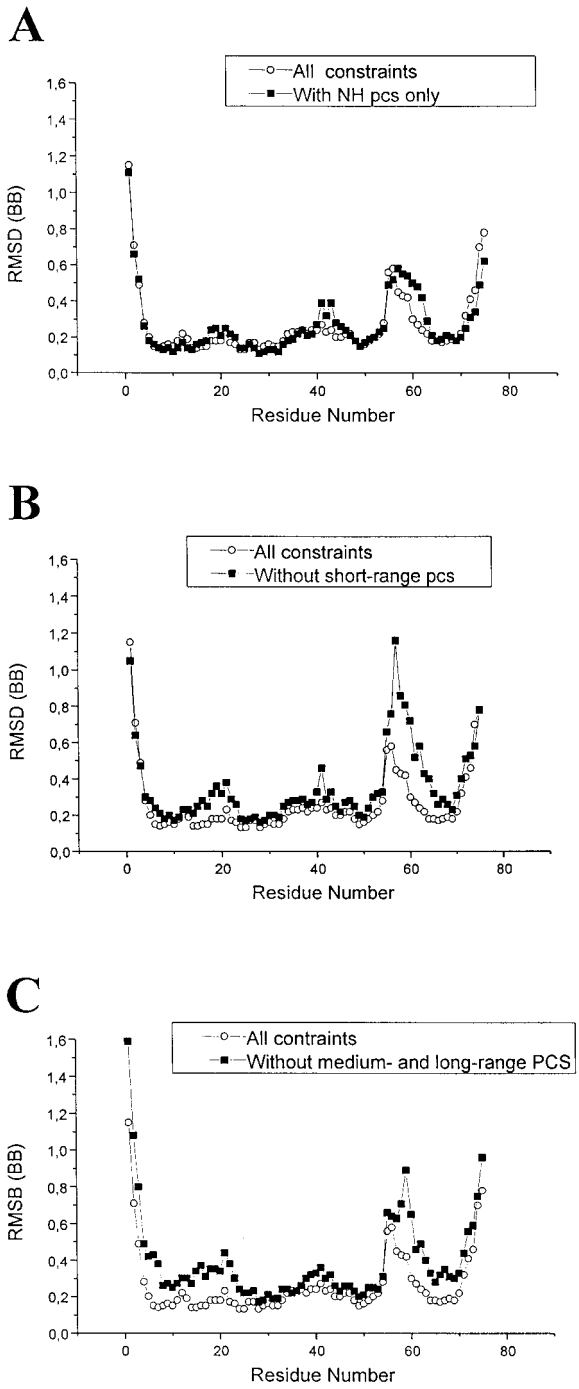

Figure 4. Backbone r.m.s.d. for (A) para vs. para with NH pcs only; (B) para vs. para without short-range pcs; (C) para vs. para without medium- and long-range pcs.

53 to residue 67, encompassing the Ln^{3+} binding loop, to the farthest residues in the linker between helices II and III. Figure 3A also shows that pcs constraints are more localized that rdc. Indeed, when rdc are excluded from calculations (Figure 3B), a higher r.m.s.d. is observed all over the molecule. Again, in regions like helix II (residues from 24 to 35) in which the structure is better defined due to dihedral angle constraints, the contribution of rdc to the overall resolution is smaller than in the loop regions. Furthermore, it seems that rdc do not simply affect the residues for which rdc are available. Indeed, even if there are no rdc constraints available for residues 55–62 (but Val 61), the exclusion of the 64 rdc constraints from calculations causes a remarkable increase of local r.m.s.d. in that region.

The 26 constraints arising from T_1 measurements do not, apparently, contribute to decrease the global r.m.s.d. value (Figure 3C). However, T_1 constraints significantly contribute to the quality of the Ramachandran plot, providing an increase of the number of residues falling in the core region. Therefore, although the structure without T_1 has the same r.m.s.d. and a slightly lower target function with respect to the structure with T_1 constraints, the structure obtained including T_1 constraints should be considered as the 'best' structure arising from our calculations. On the other hand, T_1 constraints are helpful in the absence of other paramagnetism-based constraints. As shown in Table 4, they provide a decrease of r.m.s.d. from 0.61 Å to 0.55 Å in the absence of other paramagnetism-based constraints. This is almost the same effect provided by rdc.

An interesting result is observed when running PSEUDYANA with the inclusion of only certain categories of pcs constraints. Figure 4A shows the r.m.s.d. diagram obtained when only pcs arising from amide nitrogens and protons are used. It is evident that, in the regions in which a consistent set of amide pcs is available (all but residues 53–63), no substantial improvement in the quality of the structure is provided by the inclusion of ^{13}C and non-exchangeable protons. The latter, on the other hand, are crucial to refine the region close to the metal site, for which proton relaxation prevents the observation of ^{15}N-HSQC peaks.

The exclusion from the calculation of one group of amide pcs constraints leads to another interesting conclusion: the structure obtained without short-range pcs (Figure 4B) is globally worse then the structure obtained without long- and medium-range pcs (Figure 4C). This is also apparent from Table 4, in terms of both global r.m.s.d. and Ramachandran plot. Consistently, the structures obtained with the exclusion of only one group of amide pcs show, as expected, worsening in different regions of the protein. The most striking worsening by the exclusion of short-range pcs (Ln = Ce, Pr, Nb, Eu, Sm) from the structure is seen in the second metal binding loop, which shows a peak in the local r.m.s.d. of above 1.2 Å. As expected, the exclusion of medium- and long-range pcs (Ln = Er, Yb, Tb, Dy, Ho, Tm) causes a less drastic increase of r.m.s.d. in the Ln^{3+} binding loop (Figure 4C). The sphere of action of each individual class of paramagnetism-based constraints can be further clarified by comparing the local r.m.s.d. values when one class only is inserted (Figures 5A–C).

We should finally comment on the contribution provided by the identification of lanthanide-bound residues obtained by using direct detection ^{13}C experiments (Bertini et al., 2001a). This is the first time that these constraints are considered in the solution structure calculation of a metalloprotein. Constraints derived by metal coordination significantly reduce the local r.m.s.d. in the binding loop (Figure 5D), which decreases from a maximum value of about 2.1 Å to about 1.5 Å. Overall, their effect is quite similar to that obtained by the employment of T_1 constraints although, at variance with the latter, they also contribute to the overall improvement of general statistics, as shown in Table 4.

Inspection of the diagrams of the r.m.s.d. per residue for the backbone protons of the CaLnCb families, obtained with different groups of constraints, shows an interesting behavior of the Ca^{2+} binding loop (loop I) of the protein: while the exclusion of either pcs or rdc in two different runs leads to a modest and comparable worsening in the quality of the structure (Figures 3A and 3B) in that region, the exclusion of both groups of paramagnetism-based constraints yields a drastic increase of the r.m.s.d. per residue (Figure 5C). This demonstrates a useful complementarities of rdc and pcs constraints. It should be remembered that no assumptions have been made about the coordination of Ca^{2+} in the first metal binding loop. This loop is a good example of intrinsic scarcity of diamagnetic constraints due to the presence of an NMR-silent region of the protein.

In Table 4 is also reported the r.m.s.d. between each average structure obtained with subsets of constraints and the average final structure. The exclusion of a subset of constraints from calculations not only leads to a family with a higher r.m.s.d., but also

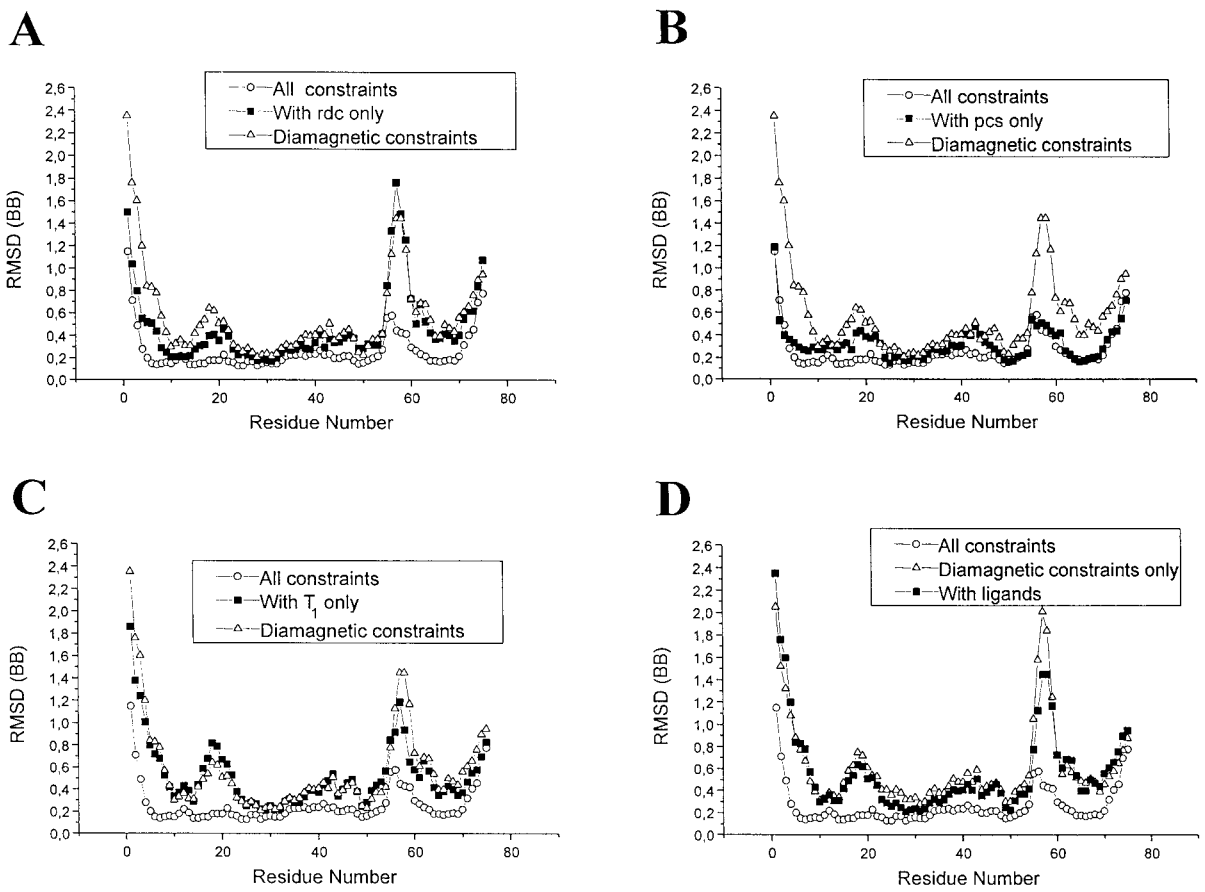

Figure 5. Backbone r.m.s.d. for (A) para vs. dia and para with rdc only; (B) para vs. dia and para with pcs only; (C) para vs. dia and para with T_1 only; (D) para vs. dia and para with ligands only.

bears to a mean structure different from the one obtained with all constraints. This means that a subset of constraints affects not only the precision, but also the accuracy of the final family (Zhao and Jardetzky, 1994). In fact it is implicit that, if the accuracy was not affected by the inclusion of subsets of constraints, the mean structures obtained with or without them would have been identical. The above analysis has shown that paramagnetism-based constraints are certainly valuable in structure refinement, as they show a certain degree of complementarity, and their variety (and tunability by the use of different metal ions) can be exploited to design the best experimental strategy for a given system. We address now the issue of their suitability *as substitutes for NOE constraints*. To do so, several calculations have been performed by retaining all paramagnetism-based constraints and by removing increasing numbers of NOEs.

The results are illustrated in Figure 6. The first calculation has been performed by removing all but long range NOEs. This decreased the total number of NOE constraints from 1793 to 244. The resulting family has target function values in the range 0.8–1.1, i.e.,

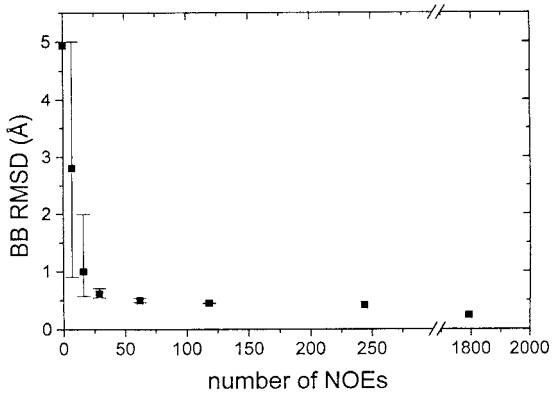

Figure 6. Backbone r.m.s.d. for the structure obtained using all the paramagnetism-based constraints and an increasing number of NOEs.

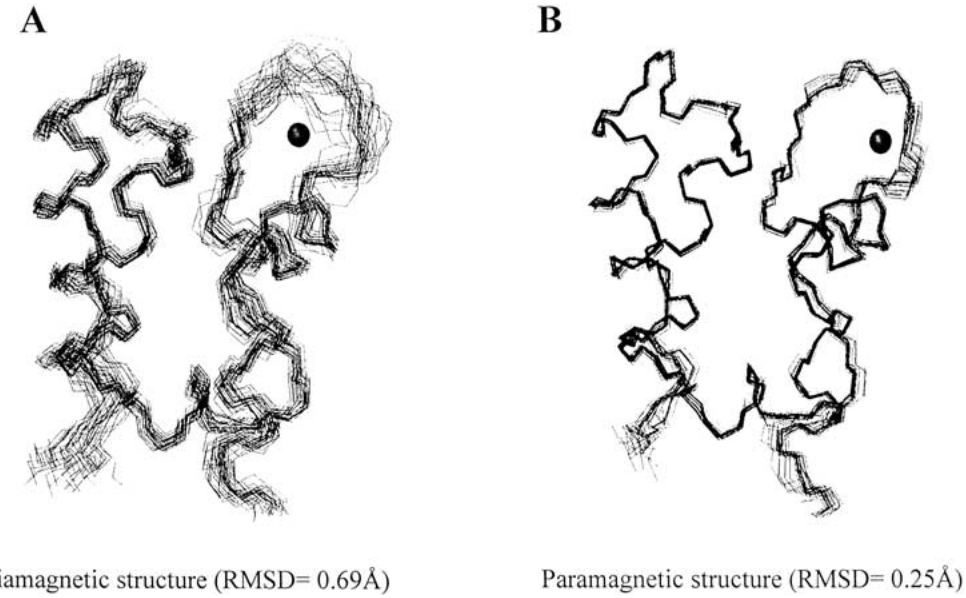

Figure 7. Family of 30 conformers obtained with diamagnetic (A) and paramagnetism-based (B) constraints.

sensibly smaller than the reference solution structure family, and backbone r.m.s.d. of 0.41 Å (compared to 0.25 Å of the best structure). By removing further long range NOEs at random, the target function remains small and the backbone r.m.s.d. increases, but the quality of the structure, as judged from the r.m.s.d., remains acceptable until the NOEs are reduced to very small numbers. It is a striking result that the r.m.s.d. remains below 1 with the use of as few as 29 NOEs. The error bars in Figure 6 give an idea of the variability of the results when different sets of NOEs are randomly selected. When all NOEs are removed, however, the r.m.s.d. jumps to above 5 Å, and the family bears only a vague resemblance to the original high resolution family. r.m.s.d. values of 0.91 Å are found even when the NOEs are reduced to 7, if they are chosen to involve atoms from all the four helices in the protein. This is not surprising, since the fact that only few distances are needed to position helices with respect to each other was already known in literature, as in the case of the 'heuristic refinement method' (Altman and Jardetzky, 1989; Brinkley et al., 1998).

It should be recalled that, besides NOE constraints, the diamagnetic constraints include dihedral angles, which are crucial in determining the secondary structure elements (four helices in the present case). These constraints have been retained in all calculations shown in Figure 6. This choice is based on the consideration that the Chemical Shift Index (CSI) may provide dihedral angles as well. To answer the question of their relative importance with respect to paramagnetism-based constraints in yielding a reasonable well resolved family with just 7 NOEs, the same calculations was repeated without paramagnetism-based constraints. Again, a r.m.s.d. larger than 4 Å was obtained in all cases.

Conclusions

Until the present work, structures of paramagnetic proteins obtained so far showed a somewhat larger r.m.s.d. on the region encompassing the metal ion. Such a behavior depends on the balance between the loss of information due to the presence of the paramagnetic center and the gain of information due to the same paramagnetic center. We show here that advantages may effectively overcome drawbacks. This is reflected in the actual values of local r.m.s.d. Overall, the above data put into evidence how synergistic can be the role of the different contributions arising from paramagnetism.

To summarize, the overall effect on the solution structure, solved with a given set of diamagnetic constraints, of the introduction of paramagnetism-based constraints for a protein containing a paramagnetic ion, is shown in Figure 7. For the first time, pcs and rdc were merged with short range information, arising from the identification of lanthanide ligands and

from T_1 values (Bertini et al., 2001a). The increase in resolution in the close proximity of the paramagnetic center as a consequence of paramagnetism-based constraints is a well known and expected feature. On the other hand, the increase of resolution in backbone regions like the N- and C-terminal sites, usually loosely characterized, is clear from Figure 7. The resolution for the diamagnetic protein CaCaCb is of course higher than that obtained for CaCeCb (Figure 7A) by using diamagnetic constraints only.

Conversely, the paramagnetism-based constraints alone produce an unresolved structure (r.m.s.d. of about 5). However, the concomitant use of paramagnetism-based constraints, angular constraints (including intra-helix hydrogen bonds) and a few long range NOEs is able to provide reasonable results (Figure 6). Since the angular constraints can be essentially obtained through the chemical shift index, a few long range constraints, which may well depend on the particular nature of the protein, provide acceptable results. The present research represents, *inter alia*, a hint toward solution structure determinations without NOEs.

Acknowledgements

This research has been financially supported under the RTD projects ('FIND Structure' contract QLG2-CT-1999-01003) and ('Transient NMR' contract HPRI-CT-1999-50006). Support to AD under the 'Access to Infrastructure' programme (PARABIO LSF contract HPRI-CT-1999-00009) and to BJ ('Marie Curie Training Site' contract HPMT-CT-2000-00137) are also acknowledged. AD thanks the 'Generalitat Valenciana' (Conselleria de Educacion y Ciencia) for a grant.

References

Allegrozzi, M., Bertini, I., Janik, M.B.L., Lee, Y.-M., Liu, G. and Luchinat, C. (2000) *J. Am. Chem. Soc.*, **122**, 4154–4161.

Altman, R.B. and Jardetzky, O. (1989) *Meth. Enzymol.*, **177**, 218–246.

Archer, S.J., Ikura, M., Torchia, D.A. and Bax, A. (1991) *J. Magn. Reson.*, **95**, 636–641.

Banci, L., Bertini, I., Bren, K.L., Cremonini, M.A., Gray, H.B., Luchinat, C. and Turano, P. (1996) *JBIC*, **1**, 117–126.

Banci, L., Bertini, I., Gori Savellini, G., Romagnoli, A., Turano, P., Cremonini, M.A., Luchinat, C. and Gray, H.B. (1997) *Proteins Struct. Funct. Genet.*, **29**, 68–76.

Banci, L., Bertini, I., Huber, J.G., Luchinat, C. and Rosato, A. (1998a) *J. Am. Chem. Soc.*, **120**, 12903–12909.

Banci, L., Bertini, I., Cremonini, M.A., Gori Savellini, G., Luchinat, C., Wüthrich, K. and Güntert, P. (1998b) *J. Biomol. NMR*, **12**, 553–557.

Bax, A. and Tjandra, N. (1997) *J. Biomol. NMR*, **10**, 289–292.

Bax, A. and Wang, A.C. (1995) *J. Am. Chem. Soc.*, **117**, 1810–1813.

Bertini, I., Capozzi, F., Luchinat, C., Piccioli, M. and Vila, A.J. (1994) *J. Am. Chem. Soc.*, **116**, 651–660.

Bertini, I., Couture, M.M.J., Donaire, A., Eltis, L.D., Felli, I.C., Luchinat, C., Piccioli, M. and Rosato, A. (1996) *Eur. J. Biochem.*, **241**, 440–452.

Bertini, I., Lee, Y.-M., Luchinat, C., Piccioli, M. and Poggi, L. (2001a) *ChemBioChem*, **2**, 550–558.

Bertini, I., Janik, M.B.L., Lee, Y.-M., Luchinat, C. and Rosato, A. (2001b) *J. Am. Chem. Soc.*, **123**, 4181–4188.

Bertini, I., Luchinat, C. and Piccioli, M. (2001c) *Meth. Enzymol.*, **339**, 314-340.

Bertini, I., Janik, M.B.L., Liu, G., Luchinat, C. and Rosato, A. (2001d) *J. Magn. Reson.*, **148**, 23–30.

Bertini, I., Luchinat, C., Parigi, G. and Walker, F.A. (1999) *JBIC*, **4**, 515–519.

Brinkley, J.F., Altman, R.B., Duncan, B.S., Buchanan, B.G. and Jardetzky, O. (1998) *J. Chem. Inf. Comput. Sci.*, **28**, 194–210.

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

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

Cordier, F. and Grzesiek, S. (1999) *J. Am. Chem. Soc.*, **121**, 1601–1602.

Eccles, C., Güntert, P., Billeter, M. and Wüthrich, K. (1991) *J. Biomol. NMR*, **1**, 111–130.

Gagne', R.R., Tsuda, S., Li, M.X., Chandra, M., Smillie, L.B. and Sykes, B.D. (1994) *Protein Sci.*, **3**, 1961–1974.

Granot, J. (1982) *J. Magn. Reson.*, **49**, 257–270.

Güntert, P. and Wüthrich, K. (1991) *J. Biomol. NMR*, **1**, 447–456.

Güntert, P., Braun, W. and Wüthrich, K. (1991) *J. Mol. Biol.*, **217**, 517–530.

Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.

Hus, J.C., Marion, D. and Blackledge, M. (2000) *J. Mol. Biol.*, **298**, 927–936.

Johansson, C., Brodin, P., Grundstrom, T., Thulin, E., Forsén, S. and Drakenberg, T. (1990) *Eur. J. Biochem.*, **187**, 455–460.

Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) *J. Magn. Reson.*, **89**, 496–514.

Kay, L.E., Marion, D. and Bax, A. (1989) *J. Magn. Reson.*, **84**, 72–84.

Kechuan, T. and Gochin, M. (1999) *J. Am. Chem. Soc.*, **121**, 9276–9285.

Kurland, R.J. and McGarvey, B.R. (1970) *J. Magn. Reson.*, **2**, 286–301.

La Mar, G.N. and de Ropp, J.S. (1993) In *Biological Magnetic Resonance*, Vol. 12, Berliner, L.J. and Reuben, J., Eds., Plenum Press, New York, NY, pp. 1–78.

Linse, S., Brodin, P., Drakenberg, T., Thulin, E., Sellers, P., Elmden, K., Grundstrom, T. and Forsén, S. (1987) *Biochemistry*, **26**, 6723–6735.

Malmendal, A., Carlström, G., Hambraeus, C., Drakenberg, T., Forsén, S. and Akke, M. (1998) *Biochemistry*, **37**, 2586–2595.

Reif, B., Hennig, M. and Griesinger, C. (1997) *Science*, **276**, 1230–1233.

Salzmann, M., Wider, G., Pervushin, K., Senn, H. and Wüthrich, K. (1999) *J. Am. Chem. Soc.*, **121**, 844–848.

Shaka, A.J., Barker, P.B. and Freeman, R. (1985) *J. Magn. Reson.*, **64**, 547–552.

Tjandra, N., Grzesiek, S. and Bax, A. (1996) *J. Am. Chem. Soc.*, **118**, 6264–6272.

Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) *Nat. Struct. Biol.*, **4**, 732–738.

Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 9279–9283.

Vold, R.R. and Prosser, P.S. (1996) *J. Magn. Reson. Ser.* **B113**, 267–271.

Vuister, G.W. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 7772–7777.

Wilkens, S.J., Xia, B., Volkman, B.F., Weinhold, F., Markley, J.L. and Westler, W.M. (1998) *J. Phys. Chem.*, **102**, 8300–8305.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.

Wüthrich, K. (1996) In *Encyclopedia of Nuclear Magnetic Resonance*, Grant, D.M. and Harris, R.K., Eds., John Wiley & Sons, Chichester, pp. 932–939.

Yamazaki, T., Tochio, H., Furui, J., Aimoto, S. and Kyogoku, Y. (1997) *J. Am. Chem. Soc.*, **119**, 872–880.

Zhao, D. and Jardetzky, O. (1994) An assessment of the precision and accuracy of protein structures determined by NMR. *J. Mol. Biol.*, **239**, 601–607.