J|A|C|S
ARTICLES
Published on Web 04/18/2002

Paramagnetically Induced Residual Dipolar Couplings for Solution Structure Determination of Lanthanide Binding Proteins

Renato Barbieri,[†] Ivano Bertini,*[,†] Gabriele Cavallaro,[†] Yong-Min Lee,[†] Claudio Luchinat,[‡] and Antonio Rosato[†]

Contribution from the Magnetic Resonance Center and Department of Chemistry, University of Florence, Via Luigi Sacconi 6, 50019, Sesto Fiorentino, Italy, and Magnetic Resonance Center and Department of Agricultural Biotechnology, University of Florence, Via Luigi Sacconi 6, 50019, Sesto Fiorentino, Italy

Received January 4, 2002

Abstract: Lanthanides may substitute calcium in calcium-binding proteins, such as, for instance, EF-hand proteins. Paramagnetic lanthanides are capable of orienting the protein in high magnetic fields to an extent similar to that obtained by using orienting devices, and each lanthanide orients according to its magnetic susceptibility tensor. Here, Ce^{3+}, Tb^{3+}, Dy^{3+}, Ho^{3+}, Er^{3+}, Tm^{3+}, Yb^{3+} in the C-terminal site of calbindin D_{9k} have been investigated. Such systems provide $^{1}H-^{15}N$ residual dipolar couplings (rdc) which can be used for solution structure determinations. Within the frame of optimizing the use of residual dipolar couplings for efficient solution structure determination, it is proposed here to use a number of lanthanides (e.g., >2) to obtain the orientations of the internuclear vectors with respect to an arbitrary reference system. This is facilitated by the independent knowledge of the magnetic susceptibility anisotropy tensor of each metal, obtained from the analysis of the pseudocontact shifts. A further module of the program PARAMAGNETIC-DYANA, called RDCDYANA-ANGLES, is developed to efficiently incorporate such rdc-derived orientations, instead of the rdc themselves, as constraints in the solution structure calculation. This strategy is absolutely general and can be extended to any other pair of dipole−dipole coupled nuclei. The effect of mobility is also assessed. In principle, information on the mobility can be obtained with a number of lanthanide ions >5, or by combining a smaller number of lanthanide ions with a few orienting devices.

Introduction

Molecules with significant magnetic susceptibility anisotropy partially align at high magnetic fields,[1−5] resulting in an incomplete rotational averaging of several interspin interactions.[6,7] This incomplete averaging can give rise to new observables in solution-state NMR spectroscopy. Residual dipolar couplings (rdc) constitute one such observable.[8,9] Globular diamagnetic proteins of the size typically amenable to NMR studies seldom have sufficient magnetic anisotropy to give rise to significant rdc's. The use of orienting media has been shown to be a practical way to induce partial alignment of the latter diamagnetic systems in a magnetic field to an extent suitable for the accurate determination of rdc's.[9]

The relationship between magnetic anisotropy tensor parameters and the actual rdc at a given magnetic field for, for example, a dipole−dipole coupled $^{15}N-^{1}H$ pair, can be expressed as a linear function of five tensor parameters defined in an arbitrary, external reference system.[10−13] For the present purpose, it can be written in the following form:

$$\mathrm{rdc}(\theta,\phi) = K\left(\frac{A_1}{2}G_1 + \frac{A_2}{2}G_2 + A_3G_3 + A_4G_4 + A_5G_5\right) \quad (1)$$

where the five parameters A_1, A_2, ..., A_5 describe the nondiagonal magnetic susceptibility tensor of the metal (χ) in the arbitrary reference system chosen ($A_1 = \chi_{zz} - {}^1/_3\mathrm{Tr}(\chi)$; $A_2 = \chi_{xx} - \chi_{yy}$; $A_3 = \chi_{xy}$; $A_4 = \chi_{xz}$; $A_5 = \chi_{yz}$), the geometric factors G_1-G_5 are

* To whom correspondence should be addressed. Tel.: +39 055 4574272. Fax: +39 055 4574271. E-mail: bertini@cerm.unifi.it.
[†] Magnetic Resonance Center and Department of Chemistry, University of Florence.
[‡] Magnetic Resonance Center and Department of Agricultural Biotechnology, University of Florence.

(1) Lohman, J. A. B.; Maclean, C. *Chem. Phys. Lett.* **1978**, *58*, 483−486.
(2) Domaille, J. P. *J. Am. Chem. Soc.* **1980**, *102*, 5392−5393.
(3) Bothner-By, A. A.; Domaille, J. P.; Gayathri, C. *J. Am. Chem. Soc.* **1981**, *103*, 5602−5603.
(4) Bothner-By, A. A.; Gayathri, C.; Van Zijl, P. C. M.; Maclean, C. *J. Magn. Reson.* **1984**, *56*, 456−462.
(5) Bothner-By, A. A.; Gayathri, C.; Van Zijl, P. C. M.; Maclean, C.; Lai, J.-J.; Smith, K. M. *Magn. Reson. Chem.* **1985**, *23*, 935−938.
(6) Van Zijl, P. C. M.; Ruessink, B. H.; Bulthuis, J.; Maclean, C. *Acc. Chem. Res.* **1984**, *17*, 172−180.
(7) Bothner-By, A. A. In *Encyclopedia of Nuclear Magnetic Resonance*; Grant, D. M., Harris, R. K., Eds.; John Wiley and Sons: Chichester, 1996; pp 2932−2938.
(8) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9279−9283.
(9) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114.
(10) Saupe, A. *Z. Naturforsch.* **1964**, *19a*, 161−171.
(11) Emsley, J. W. In *Encyclopedia of Nuclear Magnetic Resonance*; Grant, D. M., Harris, R. K., Eds.; Wiley: London, 1996; pp 2788−2799.
(12) Moltke, S.; Grzesiek, S. *J. Biomol. NMR* **1999**, *16*, 121−125.
(13) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098−6107.

defined as $G_1 = (3\cos^2\theta - 1)$, $G_2 = (\sin^2\theta\cos 2\phi)$, $G_3 = (\sin^2\theta\sin 2\phi)$, $G_4 = (\sin 2\theta\cos\phi)$, $G_5 = (\sin 2\theta\sin\phi)$, and the constant K is given by

$$K = -\frac{1}{4\pi}\frac{B_0^2}{5kT}\frac{\gamma_H\gamma_N h}{4\pi^2 r_{HN}^3}$$

where r_{HN} is the length of the N−H bond, and all the other symbols have the usual meaning. In the geometric factors G_1−G_5, θ and ϕ are the polar coordinates describing the orientation of the N−H bond vector in the external reference system. If the reference system is made coincident with the principal axes of the χ tensor, only the first two of the five terms in square brackets are retained.

The potential of rdc's for solution structure determination of biological macromolecules has stimulated a number of researchers worldwide to develop new methods for their accurate determination (recently reviewed in ref 14) and for their application in structure calculation protocols.[15−18] Rdc's have also been suggested as markers of structure quality through a disagreement factor R.[19]

The efficient use of rdc's in solution structure calculation protocols is still not a trivial task. Structure calculation methods are available where a pseudopotential has been defined on the basis of the deviation between back-calculated and experimental rdc's.[16−18,20] However, this straightforward approach may pose convergence problems in simulated annealing protocols due to the complicated form of the corresponding energy surface.[17,20,21] Several attempts at translating the geometric information contained in rdc's into constraints that are more tractable by simulated annealing protocols are available.[12,17,22] Also, the combined use of rdc's measured on the same protein experiencing different partial alignments (i.e., using different orienting media), which in principle can greatly reduce the degeneracy of the solutions,[23] does not yield a direct advantage from the computational point of view if different sets of rdc's are directly introduced as constraints.

In parallel, structural determination methods have been developed on the basis of the direct determination of peptide plane orientation from multiple sets of rdc's relative to various internuclear vectors.[22,24−26] These data can be measured simultaneously in the same sample. A single alignment system could, in principle, be sufficient, but in practice the use of more than one alignment system is preferred.[25,26] Even in these integrated approaches, the problem remains of optimally adapting the structural information provided by different alignments for use in structure solving programs.

An attractive strategy would be that of using rdc's to determine directly the angular orientation of the internuclear vector defined by the two dipole−dipole coupled nuclei in an arbitrary reference frame, that is, to determine the θ and ϕ parameters in eq 1. To do this reliably, it is necessary to measure the rdc's for a given dipolar interaction for more than two noncollinear alignment conditions. A higher number of experimental data enhances the reliability of the results. With the necessary experimental data at hand, and by knowing the principal components and orientation of the alignment tensors, it is possible to derive the orientation of the internuclear vector in an arbitrary reference system. The need for a number of independent alignment systems reduces somewhat the attractiveness of this approach, as this requires measurements on different samples. On the other hand, in the presence of dynamic processes, several independent orienting systems may become useful to reveal mobility and to have information on it. In the presence of mobility, the five trigonometric functions in eq 1 are substituted by their time-averaged value, and become five different unknowns.[13] Therefore, resorting to several (>5) different orienting systems becomes advisable at least to check the possibility that mobility may make some θ and ϕ values, as well as the rdc themselves, meaningless.

Our lab has been long involved in the study of methodologies for solution structure determination of paramagnetic metalloproteins.[27−37] These systems often feature a very high (as compared with that of the large majority of globular diamagnetic proteins) molecular magnetic anisotropy, which is largely due to the paramagnetic metal ion(s) contribution. Rdc's of the same order of those measured in the presence of external orienting devices can be obtained in the presence of several paramagnetic metals. Indeed, this kind of experimental approach to the measurement of rdc's was proposed very early[8] and successfully applied in a number of cases,[20,38−48] even to solution structure

(14) Brunner, E. *Concepts Magn. Reson.* **2001**, *13*, 238−259.
(15) Tjandra, N.; Omichinski, J. G.; Gronenborn, A. M.; Clore, G. M.; Bax, A. *Nat. Struct. Biol.* **1997**, *4*, 732−738.
(16) Clore, G. M.; Gronenborn, A. M.; Tjandra, N. *J. Magn. Reson.* **1998**, *131*, 159−162.
(17) Meiler, J.; Blomberg, N.; Nilges, M.; Griesinger, C. *J. Biomol. NMR* **2000**, *16*, 245−252.
(18) Sass, H. J.; Musco, G.; Stahl, S. J.; Wingfield, P. T.; Grzesiek, S. *J. Biomol. NMR* **2001**, *21*, 275−280.
(19) Clore, G. M.; Garrett, D. S. *J. Am. Chem. Soc.* **1999**, *121*, 9008−9012.
(20) Hus, J. C.; Marion, D.; Blackledge, M. *J. Mol. Biol.* **2000**, *298*, 927−936.
(21) Meiler, J.; Peti, W.; Griesinger, C. *J. Biomol. NMR* **2000**, *17*, 283−294.
(22) Skrynnikov, N. R.; Kay, L. E. *J. Biomol. NMR* **2000**, *18*, 239−252.
(23) Ramirez, B. E.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 9106−9107.
(24) Mueller, G. A.; Choy, W. Y.; Skrynnikov, N. R.; Kay, L. E. *J. Biomol. NMR* **2000**, *18*, 183−188.
(25) Hus, J. C.; Marion, D.; Blackledge, M. *J. Am. Chem. Soc.* **2001**, *123*, 1541−1542.
(26) Tian, F.; Valafar, H.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123*, 11791−11796.
(27) Banci, L.; Bertini, I.; Eltis, L. D.; Felli, I. C.; Kastrau, D. H. W.; Luchinat, C.; Piccioli, M.; Pierattelli, R.; Smith, M. *Eur. J. Biochem.* **1994**, *225*, 715−725.
(28) Banci, L.; Bertini, I.; Bren, K. L.; Gray, H. B.; Sompornpisut, P.; Turano, P. *Biochemistry* **1995**, *34*, 11385−11398.
(29) Banci, L.; Bertini, I.; Gray, H. B.; Luchinat, C.; Reddig, T.; Rosato, A.; Turano, P. *Biochemistry* **1997**, *36*, 9867−9877.
(30) Bentrop, D.; Bertini, I.; Cremonini, M. A.; Forsén, S.; Luchinat, C.; Malmendal, A. *Biochemistry* **1997**, *36*, 11605−11618.
(31) Bertini, I.; Donaire, A.; Luchinat, C.; Rosato, A. *Proteins: Struct., Funct., Genet.* **1997**, *29*, 348−358.
(32) Aono, S.; Bentrop, D.; Bertini, I.; Donaire, A.; Luchinat, C.; Niikura, Y.; Rosato, A. *Biochemistry* **1998**, *37*, 9812−9826.
(33) Bertini, I.; Kurtz, D. M., Jr.; Eidsness, M. K.; Liu, G.; Luchinat, C.; Rosato, A.; Scott, R. A. *JBIC, J. Biol. Inorg. Chem.* **1998**, *3*, 401−410.
(34) Im, S.-C.; Liu, G.; Luchinat, C.; Sykes, A. G.; Bertini, I. *Eur. J. Biochem.* **1998**, *258*, 465−477.
(35) Bertini, I.; Luchinat, C.; Rosato, A. *Inorg. Chim. Acta* **2000**, *297*, 199−205.
(36) Bertini, I.; Ciurli, S.; Dikiy, A.; Fernández, C. O.; Luchinat, C.; Safarov, N.; Shumilin, S.; Vila, A. J. *J. Am. Chem. Soc.* **2001**, *123*, 2405−2413.
(37) Banci, L.; Bertini, I.; Cavallaro, G.; Luchinat, C. *JBIC, J. Biol. Inorg. Chem.* **2002**, *7*, 416−426.
(38) Banci, L.; Bertini, I.; Huber, J. G.; Luchinat, C.; Rosato, A. *J. Am. Chem. Soc.* **1998**, *120*, 12903−12909.
(39) Volkman, B. F.; Wilkens, S. J.; Lee, A. L.; Xia, B.; Westler, W. M.; Berger, R.; Markley, J. L. *J. Am. Chem. Soc.* **1999**, *121*, 4677−4683.
(40) Biekofsky, R. R.; Muskett, F. W.; Schmidt, J. M.; Martin, S. R.; Browne, J. P.; Bayley, P. M.; Feeney, J. *FEBS Lett.* **1999**, *460*, 519−526.
(41) Contreras, M. A.; Ubach, J.; Millet, O.; Rizo, J.; Pons, M. *J. Am. Chem. Soc.* **1999**, *121*, 8947−8948.
(42) Déméné, H.; Tsan, P.; Gans, P.; Marion, D. *J. Phys. Chem. B* **2000**, *104*, 2559−2569.
(43) Arnesano, F.; Banci, L.; Bertini, I.; van der Wetering, K.; Czisch, M.; Kaptein, R. *J. Biomol. NMR* **2000**, *17*, 295−304.
(44) Bertini, I.; Felli, I. C.; Luchinat, C. *J. Biomol. NMR* **2000**, *18*, 347−355.

determination,[20,38,47] although the use of external orienting devices has gained such a popularity that paramagnetic systems have been essentially forgotten in the recent literature.[49]

A number of noncollinear alignment systems can be obtained with relative ease by substituting different lanthanide ions into calcium-binding sites. In this work, we selectively substitute seven different paramagnetic lanthanide ions into the C-terminal calcium binding site of calbindin D_{9k}. These metals induce different orientations of the protein in the magnetic field,[50] analogous to what can be obtained with different orienting media, due to their different electronic configurations. It is true, however, that the same coordination geometry tends to make the magnetic susceptibility tensor axes similar, so that the different orientations due to lanthanides are not as random as would have been desirable. The contributions to the rdc values due to the lanthanides are measured for backbone amide $^1H-^{15}N$ moieties in all seven systems. The tensor parameters (A_1-A_5 in eq 1) are independently provided by the analysis of the pseudocontact shifts induced by each metal ion.[50] From the latter parameters, the θ and ϕ angles defining the orientation of the internuclear N−H vectors in the laboratory frame are obtained by fitting eq 1 to all the rdc sets simultaneously. Such rdc values are very well reproduced by the fit procedure, and the calculated N−H vector orientations are in good agreement with the NMR structure obtained only with NOE, dihedral angle, and pcs constraints, the error being related to the accuracy of the *A* values. A new module implemented in the program DYANA[51] is then developed to use directly as constraints in solution structure calculations the orientations of each N−H vector defined by the θ and ϕ values. The whole program, which contains a number of modules targeted to paramagnetic molecules, is called PARAMAGNETIC-DYANA.[52] The simulated annealing algorithm is shown to converge rapidly, and the resulting structure is of improved quality. The effect of mobility is also addressed. The present structure refinement procedure is not limited to calcium-binding proteins, but can also be applied to other diamagnetic proteins through the engineering of calcium-binding sites.[45,48]

Materials and Methods

Sample Preparation. Protein expression[53] and purification[54] of both the Ca^{2+} and the apo form of bovine calbindin D_{9k} were performed as reported. A Pro43 → Met43 (P43M) mutant representing the minor A form was prepared to avoid any conformational heterogeneity due to cis−trans isomerization as found for the wild-type protein.[53,55] The expression system was a gift of Prof. S. Forsén. Uniformly ^{15}N-labeled P43M calbindin D_{9k} was obtained from M9 minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source. NMR samples were prepared by dissolving the lyophilized protein in 550 μL of 90% H_2O/10% D_2O to final 2.0 mM protein solutions. The pH was adjusted to 6.0 by means of 0.1 M NaOH or 0.1 M HCl. Lanthanide-containing calbindin D_{9k} samples (LnCaCb) were obtained by titrating the dicalcium form with 0.02 M solutions of analytical grade $LnCl_3$ (Sigma, Aldrich) up to 1 equiv. Under these conditions, the lanthanide ions selectively substitute the calcium ion in the C-terminal site. Titrations were followed by 2-D $^1H-^{15}N$ HSQC spectroscopy. The samples were kept at 4 °C between measurements.

Determination of Rdc Values. NMR spectra were acquired at 300 K on Bruker AVANCE 800 and 700, and DRX 500 spectrometers operating at proton Larmor frequency of 800.13, 700.13, and 500.13 MHz, respectively. One-bond $^1H-^{15}N$ coupling constants were measured either by fitting a series of $^1J_{NH}$-modulated HSQC spectra as described in ref 56 or by using the IPAP method.[57] For TmCaCb, the coupling constants were measured with both methods; no systematic differences were observed between the two sets of data.

Rdc values for each LnCaCb derivative were determined by subtracting from the coupling constant measured for the backbone amide moiety of each residue the coupling constant measured for the same residue and at the same magnetic field in either Ca_2Cb or CaLaCb. ^1H and ^{15}N resonance assignments for all derivatives were taken from our previous work.[50] With this procedure, only the paramagnetic contribution to the rdc is retained, while the scalar coupling constant and the contribution to the partial orientation of the system due to the diamagnetic polypeptide chain cancel out. In addition, the contribution to the splitting due to the diamagnetic dynamic frequency shift[56] is also canceled, while the paramagnetic contribution of the dynamic frequency shift due to the cross-correlation between Curie−dipole and dipole−dipole relaxation[58,59] is retained. The theoretical contribution of the latter paramagnetic dynamic frequency shift to the measured ^1J values has been calculated using available equations.[58,59] It is found that in the present case this effect is expected to be at most marginally significant (see also Discussion).

Implementation of Vector Orientation Constraints in PARAMAGNETIC-DYANA. A new module RDCDYANA-ANGLES was written as described below to incorporate the θ and ϕ angles as such in structure calculations using PARAMAGNETIC-DYANA.[52] The *i*th pair of θ and ϕ values univocally defines the orientation of the *i*th N−H bond, which in Cartesian space is equivalently defined by the unit vector $\mathbf{v0}_i$, with coordinates $x0$, $y0$, $z0$ given by the well-known transformation rule:

$$x0 = \sin\theta\cos\phi$$

$$y0 = \sin\theta\sin\phi$$

$$z0 = \cos\theta$$

The unit vector $\mathbf{v}_i$, which represents the actual direction in the structure of the *i*th N−H bond in the same reference system, was then defined as

$$\mathbf{v}_i = (\mathbf{r}_H - \mathbf{r}_N)_i/d_{NH}$$

where $\mathbf{r}_H$ is the vector of the coordinates of H, $\mathbf{r}_N$ is the vector of the coordinates of N, both defined with respect to the external axis system,

(45) Ma, C.; Opella, S. J. *J. Magn. Reson.* **2000**, *146*, 381−384.
(46) Gaponenko, V.; Dvoretsky, A.; Walsby, C.; Hoffman, B. M.; Rosevear, P. R. *Biochemistry* **2000**, *39*, 15217−15224.
(47) Bertini, I.; Janik, M. B. L.; Liu, G.; Luchinat, C.; Rosato, A. *J. Magn. Reson.* **2001**, *148*, 23−30.
(48) Feeney, J.; Birdsall, B.; Bradbury, A. F.; Biekofsky, R. R.; Bayley, P. M. *J. Biomol. NMR* **2001**, *21*, 41−48.
(49) Prestegard, J. H.; Valafar, H.; Glushka, J.; Tian, F. *Biochemistry* **2001**, *40*, 8677−8685.
(50) Bertini, I.; Janik, M. B. L.; Lee, Y.-M.; Luchinat, C.; Rosato, A. *J. Am. Chem. Soc.* **2001**, *123*, 4181−4188.
(51) Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol.* **1997**, *273*, 283−298.
(52) The source codes of the modules PSEUDYANA, RDCDYANA-ORIENT, and that of the presently developed module RDCDYANA-ANGLES are available at www.postgenomicnmr.net. They are implemented in the program PARAMAGNETIC-DYANA, which can be obtained through www.postgenomicnmr.net by those who already have a licensed version of DYANA (ETH, Zurich, see ref 51). The programs FANTASIAN and FANTALIN can be freely downloaded from www.postgenomicnmr.net.
(53) Brodin, P.; Grundstrom, T.; Hofmann, T.; Drakenberg, T.; Thulin, E.; Forsén, S. *Biochemistry* **1986**, *25*, 5371−5377.
(54) Johansson, C.; Brodin, P.; Grundstrom, T.; Thulin, E.; Forsén, S.; Drakenberg, T. *Eur. J. Biochem.* **1990**, *187*, 455−460.
(55) Chazin, W. J.; Kördel, J.; Drakenberg, T.; Thulin, E.; Brodin, P.; Grundstrom, T.; Forsén, S. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 2195−2198.
(56) Tjandra, N.; Grzesiek, S.; Bax, A. *J. Am. Chem. Soc.* **1996**, *118*, 6264−6272.
(57) Ottiger, M.; Delaglio, F.; Bax, A. *J. Magn. Reson.* **1998**, *131*, 373−378.
(58) Ghose, R.; Prestegard, J. H. *J. Magn. Reson.* **1997**, *128*, 138−143.
(59) Bertini, I.; Kowalewski, J.; Luchinat, C.; Parigi, G. *J. Magn. Reson.* **2001**, *152*, 103−108.

and d_{NH} is the N−H bond distance, which is fixed in RDCDYANA-ANGLES. Thus, the ith constraint is fulfilled when the $\mathbf{v}_i$ vector is parallel to the corresponding $\mathbf{v0}_i$ vector. This requirement was expressed by the following pseudopotential, U_{TP}:

$$U_{TP} = W_{TP}\sum_i w_i[1 - (\mathbf{v0}_i \cdot \mathbf{v}_i)^2] \quad (2)$$

which was added to the global PARAMAGNETIC-DYANA target function. W_{TP} is a global weighting factor, while w_i is the weighting factor for the ith constraint. Both are provided as an input by the user, together with θ and ϕ values. U_{TP} is a function with two equivalent minima, corresponding to the two possible orientations (0 and 180°) of $\mathbf{v}_i$ with respect to $\mathbf{v0}_i$. For the integration of the equations of motion in RDCDYANA-ANGLES, explicit expressions for the torques about the rotatable bonds, that is, the gradients of the potential energy with respect to torsion angles, are required. The derivative of the pseudopotential U_{TP} with respect to the kth torsion angle, α_k, has the following form:

$$\partial U_{TP}/\partial \alpha_k = -2W_{TP}\sum_i w_i(\mathbf{v0}_i \cdot \mathbf{v}_i)[(\mathbf{v0}_i \times \mathbf{v}_i)\cdot \mathbf{e}_k] \quad (3)$$

where $\mathbf{e}_k$ is the unit vector directed along the kth rotatable bond. Sample calculations with artificially generated vector orientations showed the desired convergence properties. In all the calculations performed here, 200 random structures were annealed in 12 000 steps, setting W_{TP} equal to 2.0.

Results

Calculation of N−H Vector Orientations. In an arbitrary reference system, the rdc induced by the partial molecular alignment in the magnetic field due to the magnetic anisotropy of the metal ion can be written as in eq 1. Magnetic susceptibility tensor parameters can be obtained independently by fitting the pseudocontact shifts induced by the paramagnetic metal ion on the nuclei of the protein[50] using the program FANTASIAN.[60] For the present purpose, we have written a new version of FANTASIAN, called FANTALIN, that uses the linear version of the pseudocontact shift equation,[61,62] analogous to the linear version of the rdc equation (eq 1). In this way, the values of A_1−A_5 were obtained from the solution structure family previously determined without rdc constraints. Such a family (starting structure hereafter) was obtained with as many as 1666 constraints from pseudocontact shift data and 1893 NOE and torsion angle constraints.[63] The values of A_1−A_5 were then directly used as constants in eq 1 to fit the experimental rdc values for each NH group, by using θ and ϕ as adjustable parameters. The average θ and ϕ values from the starting structure were taken as initial estimates for the fitting.[64] The input rdc data are available as Supporting Information. Different lanthanides broaden the proton line widths beyond detection at different distances from the metal ion itself. Therefore, for many residues, less than seven rdc data could be obtained. Only residues for which data were available from at least three out of the seven lanthanide derivatives were used. Furthermore, N−H moieties experiencing S values for the ^{15}N nucleus smaller than 0.85[65] were discarded at this stage, as is customary when using rdc's for solution structure determination.[15] The agreement between best-fit and experimental rdc values is quite good, indicating a good consistency among the seven sets of rdc data.

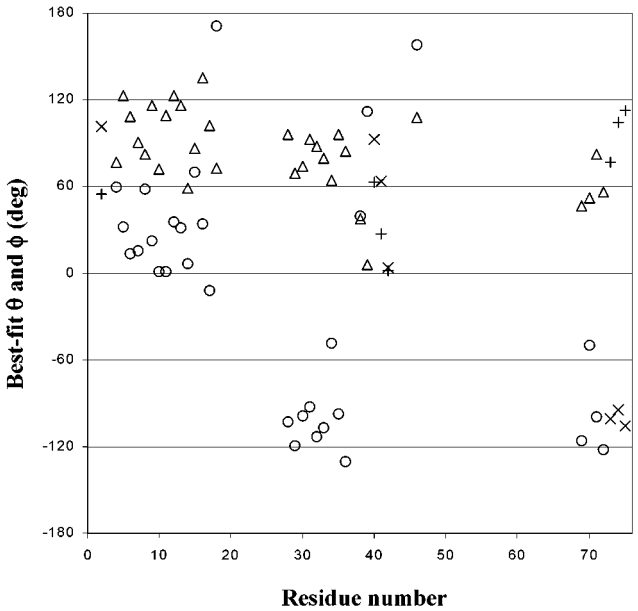

Figure 1. Plot of best-fit θ (△) and ϕ (○) values on a per-residue basis. The observed pattern correlates to the presence of helical structure; that is, residues in the same helix tend to have similar θ and ϕ values, as a result of the approximate collinearity of the backbone N−H vectors in helices. The θ (+) and ϕ (×) values for the mobile residues are also shown.

Figure 1 shows the best-fit θ and ϕ values on a per-residue basis. It can be readily observed that the θ and ϕ values are not randomly scattered, but follow a pattern. The observed pattern correlates to the presence of helical structure; that is, residues in the same helix tend to have similar θ and ϕ values, as a result of the approximate collinearity of the backbone N−H vectors in helices. Consequently, also the N−H vectors of residues in parallel helices (e.g., helix 2, spanning residues 25−35, and the C-terminal helix) have similar θ and ϕ values. This observation already provides a qualitative indication that the orientations calculated through the present procedure are reasonable.

Solution Structure Calculations. The program DYANA[51] calculates three-dimensional protein structures through molecular dynamics in torsion angle space (TAD), using a fast recursive algorithm to integrate the equations of motion.[66] The degrees of freedom are exclusively torsion angles, since the covalent structure parameters (bond lengths, bond angles, chiralities, and planarities) are kept fixed at their optimal values during the calculation. For TAD in DYANA, the role of the potential energy is carried out by the target function, which measures the violations of the experimental constraints in the calculated structure. The modules PSEUDYANA[52,67] and RDCDYANA-ORIENT[38,52] had been developed earlier by us in an integrated package of DYANA called PARAMAGNETIC-DYANA to exploit pseudocontact shifts and rdc, respectively, as structural

(60) Banci, L.; Bertini, I.; Gori Savellini, G.; Romagnoli, A.; Turano, P.; Cremonini, M. A.; Luchinat, C.; Gray, H. B. *Proteins: Struct., Funct., Genet.* **1997**, *29*, 68−76.
(61) Kemple, M. D.; Ray, B. D.; Lipkowitz, K. B.; Prendergast, F. G.; Rao, B. D. N. *J. Am. Chem. Soc.* **1988**, *110*, 8275−8287.
(62) Forsberg, J. H.; Delaney, R. M.; Zhao, Q.; Harakas, G.; Chandran, R. *Inorg. Chem.* **1995**, *34*, 3705−3715.
(63) Bertini, I.; Donaire, A.; Jimenez, B.; Luchinat, C.; Parigi, G.; Piccioli, M.; Poggi, L. *J. Biomol. NMR* **2001**, *21*, 85−98.
(64) Extensive tests showed that the final results are insensitive to the choice of the starting point.
(65) Kordel, J.; Skelton, N. J.; Akke, M.; Palmer, A. G., III; Chazin, W. J. *Biochemistry* **1992**, *31*, 4856−4866.
(66) Abe, H.; Braun, W.; Noguti, T.; Go, N. *Comput. Chem.* **1984**, *8*, 239−247.
(67) Banci, L.; Bertini, I.; Cremonini, M. A.; Gori Savellini, G.; Luchinat, C.; Wüthrich, K.; Güntert, P. *J. Biomol. NMR* **1998**, *12*, 553−557.

Table 1. Selected Parameters Describing the Solution Structure Obtained with All Available NOE, Dihedral Angle, and pcs Constraints, without or with θ and ϕ Constraints

	RMSD backbone (Å)[a]	RMSD all heavy atoms (Å)[a]	TF θ and ϕ constraints (Å^2)	TF all other constraints (Å^2)
without θ and ϕ constraints	0.40 ± 0.07	0.86 ± 0.07	12.1 ± 0.5[b]	2.0 ± 0.1
with θ and ϕ constraints (nonmobile residues)	0.40 ± 0.08	0.86 ± 0.09	2.8 ± 0.4	4.2 ± 0.4
with θ and ϕ constraints including mobile residues	0.32 ± 0.06	0.84 ± 0.08	3.2 ± 0.3	4.1 ± 0.3

[a] From the mean structure, calculated for residues 2−75. [b] Including mobile residues.

constraints together with distance and torsion angle constraints. Following the same approach, we have developed a new module of the package, RDCDYANA-ANGLES, which allows one to include vector orientation constraints rather than rdc in structure calculations by PARAMAGNETIC-DYANA (see Materials and Methods). The difference between the present and the older modules is not trivial, as now we are dealing with constraints defined with respect to an absolute external axis system. It is noteworthy that, at variance with the other constraints used in DYANA and its various modules, the use of vector orientation constraints forces the molecule to orient with respect to the fixed reference frame in which the input θ and ϕ values are defined. Reorientation of the whole system has no effect on the fulfillment of the other constraints, which deal with internal coordinates only.

The fact that this approach permits the use of a single pseudopotential for each N−H vector orientation (eq 2), rather than as many pseudopotentials for each N−H as the number of rdc data sets, together with the fact that each of the rdc pseudopotentials admits an infinite number of minima (see also Discussion in ref 17), suggests that the use of vector orientations, rather than rdc values, should enhance the efficiency of the calculation itself, improving its convergence and, therefore, increasing the percentage of acceptable conformers out of the starting random structures. In other words, the information contained in all the rdc sets is condensed into one vector orientation constraint, whose degeneracy is reduced to the minimum, that is, 2-fold. Apart from the form of eq 3, which depends on the specific method implemented in DYANA, that is, torsion angle dynamics, any other program which performs constrained molecular dynamics to determine protein structure, like for instance X-PLOR,[68] should benefit from this approach.

The presently obtained 31 N−H vector orientations could be indeed introduced successfully along with all other available constraints into calculations directly starting from random structures using a constant weight throughout the computation, with no convergence problems. These constraints thus appear to be well behaved in calculations, as expected. Selected parameters describing the results of the calculations run with vector orientation constraints together with all other available constraints are compared in Table 1 with those obtained for the starting structure family. The inclusion of the new constraints (except those relative to residues experiencing mobility, see later) does not result in an improvement of the RMSD within the resulting family. This is not surprising, as the starting

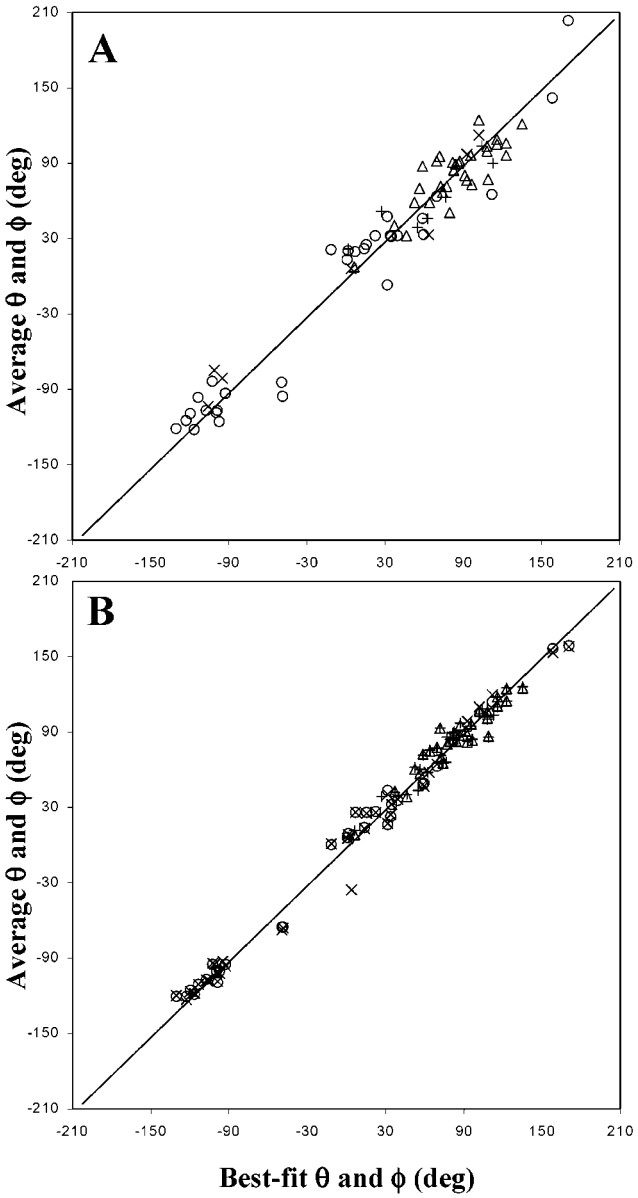

Figure 2. (A) Plot of the average θ (△) and ϕ (○) values in the NMR structure family calculated without θ and ϕ constraints versus the corresponding values derived from the fit of rdc data. The + and × symbols refer to mobile residues. The line represents the best fit resulting from a linear regression performed over the data sets for the nonmobile residues (slope, 0.95 ± 0.03; average deviation, 19°; R^2, 0.949). (B) Plot of the average θ (△) and ϕ (○) values in the NMR structure family calculated with θ and ϕ constraints versus the corresponding values derived from the fit of rdc data. The line represents the best fit resulting from a linear regression performed over the data sets for the nonmobile residues (slope, 0.99 ± 0.02; average deviation, 12°; R^2, 0.988). The results obtained by including in RDCDYANA-ANGLES the θ (+) and ϕ (×) values for the mobile residues are also shown.

structure is a very refined structure[63] even in the absence of vector orientation constraints. The target function for all the other constraints is higher in the calculations run with the new constraints. However, this increase is by a factor of 2, whereas the agreement between the input vector orientations and the corresponding values in the final family of conformers is improved by a factor of 4 with respect to the calculation performed without these constraints (Table 1). Figure 2 shows a plot of the average θ (△) and ϕ (○) values in the starting structure (A) and in the new structure (B) versus the corre-

(68) Brunger, A. T. *X-PLOR Manual Version 3.1. A System for X-ray Cristallography and NMR*; Yale University Press: New Haven, CT, 1992.

sponding best-fit values. Linear regressions performed over the data sets of Figure 2 respectively yield a slope of 0.95 ± 0.03 (A) and 0.99 ± 0.02 (B), with an average deviation of 19° (A) and 12° (B) and a correlation coefficient, *R*, equal to 0.949 (A) and 0.988 (B). Comparison of Figure 2B with Figure 2A shows the improvement in the agreement of the resulting structure with the θ and ϕ values. This is achieved with only a minimal impairment in the agreement with the other restraints in the target function, thus leading to an improvement of the overall quality of the structure. A further estimate of the accuracy may be given by the agreement with the X-ray structure; if the potential U_{TP} is calculated with respect to the $\mathbf{v0_i}$ orientations of the N−H bond vectors in the X-ray structure, then its value (setting W_{TP} equal to 2.0) decreases from 6.68 in the starting structure to 4.27 in the new structure, thus showing a notably better agreement with the X-ray structure. On the other hand, the appearance of the Ramachandran plot does not substantially change with the introduction of the new constraints, being very good even in the starting family of conformers (88.9% of the residues in the most favored regions + 11.1% in additional allowed regions).

Discussion

In the Results section, we have described in detail an approach aimed at obtaining structural information from multiple sets of rdc data that can be easily implemented into structural calculation protocols. This information coincides with the orientation of the vector connecting the two nuclei dipolarly coupled (backbone ^{1}H−^{15}N bonds in the present example) in an arbitrary reference frame (the orientation being defined by the values of θ and ϕ in eq 1). Such orientations can be used as constraints for the corresponding internuclear vectors in solution structure determinations. It is to be noted that the approach described in this work takes advantage of a distinct feature of paramagnetically aligned systems, that the magnetic susceptibility of the system, and thus its degree of alignment, can be directly and independently measured from pseudocontact shift data.[50] On the contrary, the alignment induced by orienting media can only be derived from calculations,[69] or evaluated from the same rdc data which are then used in calculations. The present formulation of the problem results in a pseudopotential for calculations (eq 2) which is dependent only on the angle between the input value and the corresponding actual orientation of the vectors. The mathematical form of the pseudopotential is thus much simpler than that used in the case of direct structure calculations against rdc data, in turn resulting in a very good performance of the calculation program as compared with previous protocols for the use of rdc constraints. It is important to stress that the present approach also constitutes a more thorough exploitation of the removal of degeneracy implied by the simultaneous use of multiple alignment conditions.[23]

As already mentioned, in the presence of internal motion, eq 1 contains the motional averages of the spherical harmonics, rather than directly the spherical harmonics.[13] In a recent publication, molecular dynamics (MD) simulations on ubiquitin have been used to quantitatively assess the impact of the motional averaging on the structural parameters extracted from simulated rdc's following an approach analogous to the one employed here.[13] It was shown that for regions where the amplitude of internal molecular motions is small (e.g., regions with regular secondary structure), the fit of the simulated data yields θ and ϕ values which are within 2° from the average values observed over the MD trajectory.[13] The deviation was larger in loop regions, where errors up to 5° are predicted.[13] Of course, only motions which could be sampled in the MD simulation, that is, subnanosecond motions, were taken into account when calculating the above deviations.[13] In the present case, correction of the rdc data for the experimental values of the model-free *S* (as long as $S > 0.85$), which accounts for subnanosecond motions, indeed did not improve significantly the agreement between the θ and ϕ values obtained from the rdc analysis and the corresponding values obtained from the NMR structure (not shown). This can be taken as an indication that at the present level of refinement of the solution structure, the factors limiting the agreement are most likely to be the uncertainty on tensor parameters, and/or the accuracy of the structure itself, rather than mobility.

As is customary in the use of rdc values for structural purposes,[15] residues with model-free *S* values smaller than a threshold (0.85 in this case) have been left out from the present analysis. These seven residues are at the N-terminal (1−3), C-terminal (72−75), as well as at the southern loop (40−45) connecting the two EF-hands. Of course, this could be done because a model-free analysis was available for the present systems. In its absence, all these residues would have been retained. Moreover, other residues that may be mobile on a time scale longer than the nano-picosecond range, but still fast with respect to the rdc values (shorter than tens of milliseconds), may also be present and have their rdc values altered by mobility. In principle, one could take advantage of the availability of an extended set of rdc values to obtain further information on mobility of both types.[13] To explore the information content of the present data to this end, we have back-calculated all five geometric functions G_1−G_5 in eq 1, rather than the two θ and ϕ angles, for the 21 residues for which at least six rdc data were available. From the geometric factors, an order parameter S^2_{rdc} can be calculated:[13]

$$S^2_{rdc} = \frac{3}{4}\left(\frac{1}{3}G_1^2 + G_2^2 + G_3^2 + G_4^2 + G_5^2\right) \qquad (4)$$

Some of the residues that were discarded as mobile from the analysis (i.e., residues 42, 74, 75) indeed turned out to have an S^2_{rdc} smaller than 0.5. A quantitative analysis of the type of motion involved from the knowledge of the five *G* functions[13] is outside the scope of the present work, and should be better attempted together with rdc measurements involving also other pairs of nuclei. Nevertheless, it is already apparent that the availability of several rdc datasets may become a precious tool to address motional problems.

Finally, a comment is due to the possible use as structural constraints of rdc data on a protein whose mobility properties are unknown, thus including several rdc data on mobile residues. We can simulate this situation by calculating θ and ϕ values from eq 1 for all the previously discarded residues (1−3, 40−45, 73−75) and then using them in RDCDYANA-ANGLES calculations. First of all, perhaps not surprisingly, the fits are reasonably good, and the angle values are in reasonable agreement with the average values measured on the starting structure family (+ and ×'s in Figure 2A). Second, RD-

(69) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791−3792.

CDYANA-ANGLES calculations still converge quite well, the target function in the resulting structure family is only marginally higher, and the θ and ϕ values from the new family are in better agreement with the input data (+ and ×'s in Figure 2B). Third, the RMSD of the family decreases sizably (Table 1), especially at the N- and C-terminal ends and at the loop region, that is, where the new angular information has been introduced. While a decrease of RMSD in mobile regions may be regarded as an overrefinement, it is reassuring to notice that in all cases the θ and ϕ values are close to the average values in the starting structure family. In other words, it appears that in the presence of a redundant set of rdc values, the use of angular constraints for mobile residues yields some kind of average of the θ and ϕ values which is always reasonable, that is, relatively unbiased with respect to the true average θ and ϕ values defining the center of the motion. Further support to this conclusion comes from a comparison with the θ and ϕ values for these residues obtained from X-ray data. The agreement is also very good (data not shown), except for residue 43 (the site of the Pro−Met mutation) and for the following residue, which obviously has to adapt to the preceding nonconservative mutation.

A further point to address in the present context is the possible presence of paramagnetic dynamic frequency shifts (DFS).[58] In the present work, rdc's were estimated by subtracting the splittings measured on the diamagnetic sample from those measured on the paramagnetic sample at the same field (i.e., either 500 or 800 MHz, see Supporting Information). The differential splittings obtained in this way contain, besides the rdc, the paramagnetic DFS contribution, if present. If this is true, the differential splittings measured at 800 and 500 MHz for the same lanthanide, the latter properly scaled by a factor of $800^2/500^2$, are expected to differ because the paramagnetic DFS contribution is essentially field independent. However, the two sets of data coincide within the experimental errors, properly propagated. Therefore, it can be concluded that paramagnetic DFS does not affect the present data analysis, and, accordingly, the differential splittings contain only the rdc.

Conclusions

We have shown that it is possible to exploit lanthanide ions to induce partial orientation of calcium-binding proteins immersed in strong magnetic field resulting in different alignment tensors depending on the particular lanthanide used. By measuring the pseudocontact shifts induced by the paramagnetic center, the magnetic susceptibility parameters of the metal ion can be independently and reliably derived. The latter parameters are then used to perform a simultaneous fit of all the rdc data available for each internuclear vector (up to seven different experimental data per NH moiety in the present work), yielding the orientation of the vector in an arbitrary reference frame. This constitutes useful structural information, which can be straightforwardly introduced in structure calculation algorithms, as done in the present work. In addition, information on mobility of the residues can be obtained from back-calculation of the G_1-G_5 geometric factors. Indeed, the present approach appears to be rather insensitive to the presence of internal molecular mobility on any time scale.

The method requires the knowledge of the A_1-A_5 sets of tensor parameters for each paramagnetic derivative. Therefore, it constitutes a structure refinement tool, and it is used as such in the present work. However, it should be kept in mind that the tensor parameters can be optimized as the structure refinement progresses, in an iterative fashion. This is commonly done for the magnetic anisotropy tensor when using pcs as constraints in metalloproteins.[28−30] The latter constraints usually result in a fast convergence of the tensor parameters.

In summary, in this work it has been shown that it is possible to obtain a unified, fully internally consistent structural picture from paramagnetic and diamagnetic structural data. In particular, the use of information from different paramagnetic effects (pseudocontact shifts and rdc's) permits the full exploitation of the experimental data, while minimizing artifacts due to minor structural errors.

Acknowledgment. Thanks are expressed to Prof. Sture Forsén for providing us with the expression system for calbindin D_{9k}. Financial support of the EU through contract nos. HPRI-CT-1999-50006 and QLG2-CT-1999-01003 is gratefully acknowledged. This work has been also partly supported by MURST COFIN 2001 and CNR (contract no. 99.00393.49). The web site www.postgenomicnmr.net has been implemented under EC contract HPRI-CT-1999-40005.

Supporting Information Available: Two tables reporting all of the rdc data used for fittings (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA025528D