JMB

De novo Determination of Protein Structure by NMR using Orientational and Long-range Order Restraints

Jean-Christophe Hus, Dominique Marion and Martin Blackledge*

Institut de Biologie Structurale Jean-Pierre Ebel C.N.R.S. C.E.A. 41, rue Jules Horowitz 38027 Grenoble Cedex, France

Orientational and novel long-range order restraints available from paramagnetic systems have been used to determine the backbone solution structure of the cytochrome c' protein to atomic resolution in the complete absence of restraints derived from the nuclear Overhauser effect. By exploiting the complementary geometric dependence of paramagnetic pseudocontact shifts and the recently proposed Curie-dipolar cross correlated relaxation effect, in combination with orientational constraints derived from residual dipolar coupling, autorelaxation rate ratios and secondary structure constraints, it is possible to define uniquely the fold and refine the tertiary structure of the protein (0.73 Å backbone rmsd for 82/129 amino acid residues) starting from random atomic Cartesian coordinates. The structure calculation protocol, developed using specific models to describe the novel constraint interactions, is robust, requiring no precise *a priori* estimation of the various interaction strengths, and provides unambiguous convergence based only on the value of the target function. Tensor eigenvalues and their component orientations are allowed to float freely, and are thus simultaneously determined, and found to converge, during the structure calculation.

© 2000 Academic Press

Keywords: structure; orientational restraints; dipolar coupling; paramagnetism; relaxation

*Corresponding author

Introduction

The determination of three-dimensional macromolecular structure by NMR using distances derived from the nuclear Overhauser effect (NOE) is now established as a standard technique for studying small to medium-sized proteins in solution (Wüthrich, 1986). Recent developments in heteronuclear multi-dimensional spectroscopy (Cavanagh *et al.* 1996), in particular the exploitation of cross-correlation between chemical shift anisotropy and dipole-dipole interactions (Pervushin *et al.* 1997), have made simple heteronuclear spectra of high molecular mass systems accessible to study by solution state NMR. Despite these advances, the determination of structure of larger systems using NOE-based methods is complicated by the problems of spin-diffusion, which impair both the identification and precision of geometric data and the spectral complexity, which significantly diminishes the possibility of unambiguous

assignment of interaction partners. More importantly, transverse relaxation times become prohibitively short making complex NMR pulse schemes unrealistically insensitive. The determination of the structure of macromolecules is further limited by the inadequate number of inter-proton dipolar interactions available to define long-range order in non-globular extended or modular systems. For these reasons the development of new methods, which do not depend on the measurement of ^{1}H-^{1}H NOE for the study of macromolecular structure, is of great importance for the application of NMR as a tool for the determination of the structure of larger and more complex systems.

During the last five years novel approaches for measuring orientational information relative to a common molecular reference frame, from residual dipolar coupling due to molecular alignment (Tjandra *et al.*, 1997a) or from heteronuclear relaxation in anisotropically tumbling molecules (Brüschweiler *et al.*, 1995; Tjandra *et al.*, 1997b), have provided new tools for the study of proteins and nucleic acids in the solution state. Weak alignment of proteins, which can exist naturally due to the paramagnetic properties of the molecule (Tolman *et al.*, 1995) or be induced by solvation in

---

Abbreviations used: NOE, nuclear Overhauser effect.
E-mail address of the corresponding author: martin@rmn.ibs.fr

liquid crystal media (Tjandra & Bax, 1997), lipid bicelles (Sanders *et al.*, 1994) or a suspension of filamentous bacteriophage (Hansen *et al.*, 1998), prevents complete averaging of the dipolar interaction while retaining the solution properties necessary for high resolution NMR. The measurement of dipolar couplings under these conditions provides long-range order constraints which, in combination with classical NOE data, have been shown to improve structure determination in multidomain systems and protein-ligand complexes (Tjandra, 1999; Fischer *et al.*, 1999; Olejniczak *et al.*, 1999; Clore & Garrett *et al.*, 1999; Bolon *et al.*, 1999). Nevertheless, the determination of three-dimensional protein structure using only orientational information is severely hindered by the lack of long-range distances that define the topology of the molecule.

It has long been recognised that the presence of paramagnetic centres in macromolecules provides structural information relative to a reference point fixed to the molecular frame (Bertini & Luchinat, 1986). In addition, the orientation (θ_1, ϕ_1) of interatomic vectors relative to the anisotropic magnetic susceptibility, χ, is provided by ^1H-^{15}N residual dipolar coupling (J_{dip}) as described above:

$$J_{dip} \propto \gamma_H \gamma_N \left[\Delta\chi_{ax}(3\cos^2\theta_1 - 1) + \frac{3}{2}\Delta\chi_{rh}(\sin^2\theta_1 \cos 2\phi_1) \right]/r^3_{H-N} \tag{1}$$

Where $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and rhombic components of the susceptibility tensor. The dual dependence of pseudocontact paramagnetic chemical shift (δ_{para}) on the electron-proton distance and the angle (θ_2, ϕ_2) made by the observed nucleus, the electron and χ provides important but highly degenerate structural information:

$$\delta_{para} \propto \left[\Delta\chi_{ax}(3\cos^2\theta_2 - 1) + \frac{3}{2}\Delta\chi_{rh}(\sin^2\theta_2 \cos 2\phi_2) \right]/r^3_{H-e^-} \tag{2}$$

A number of recent studies have indeed shown that δ_{para} can be used in combination with NOE data to refine protein and nucleic acid structures (Gochin & Roder, 1995; Banci *et al.*, 1998; Tu & Gochin, 1999). This important structural degeneracy can be partially raised using a novel source of long-range structural information, recently proposed from the measurement of Curie spin-nuclear spin cross-correlated relaxation $(\Gamma^{Curie,DD})$ (Guéron, 1975; Ghose & Prestegard, 1997; Boisbouvier *et al.*, 1999) whose dependence on the angle $N - H - e^-$ (θ_3) and the nuclear-electron distance r_{H-e^-} (with

B_0 the magnetic field, τ_c the correlation time of the molecule and $\bar{\chi}$ the magnetic susceptibility) provides similar, but essentially different geometric constraints:

$$\Gamma^{Curie,DD} \propto [\bar{\chi}B_0\tau_c P_2(\cos\theta_3)]/r^3_{H-e^-} \tag{3}$$

As mentioned above, heteronuclear relaxation rates in anisotropically tumbling molecules, whose geometric dependence on the orientation (θ_4, ϕ_4) of the relaxation interaction and the diffusion tensor are similar to the dipolar coupling dependence, also provide important orientational information. This complex geometric dependence has been extensively presented elsewhere (Woessner, 1962; Brüschweiler *et al.*, 1995; Tjandra *et al.*, 1997b). As shown in Figure 1, these measurements represent a powerful set of constraints, whose evident complementarity should increase available structural definition.

Here, we propose that the geometric definition resulting from the combination of long-range order and paramagnetic constraints provides the essential Cartesian and orientational definition necessary to determine unambiguously protein structure. Using a real experimental model, we show that protein structure can in fact be calculated *de novo* using only this type of constraint, all of the variables of which are relatively simple to measure in systems containing a paramagnetic centre. While most macromolecules do not naturally possess such properties, the replacement of calcium by paramagnetic lanthanide ions in calcium-binding proteins (Contreras *et al.*, 1999), and the ligation of lanthanide-binding peptides to proteins and nucleic acids may allow this method to be applied more generally to determine the structure of macromolecules. We discuss the relative merits of the approach with respect to more conventional NOE-based structure determination methods.

Results and Discussion

Cytochrome *c′*: a model for spin-labelled proteins

As an experimental model system we have chosen the cytochrome *c′* (Cytc′) from *Rhodobacter capsulatus*, for which experimental data from the above phenomena have been measured in our laboratory (Boisbouvier *et al.*, 1999; Tsan *et al.*, 1999; Tsan 1998) in addition to diamagnetic ^{15}N relaxation data (Tsan, 1998). Cytc′ is a five-ligand heam protein possessing two paramagnetic states for the reduced $(S = 2)$ and oxidised $(S = 5/2)$ Fe, and which is diamagnetic when bound to CO as a sixth ligand. To determine the capacity of long-range paramagnetic and orientational information to unambiguously define protein structure we have used only experimental data from the diamagnetic form and one of the two available paramagnetic states $(S = 2)$. Recent studies have shown that this molecule is monomeric in solution in all three
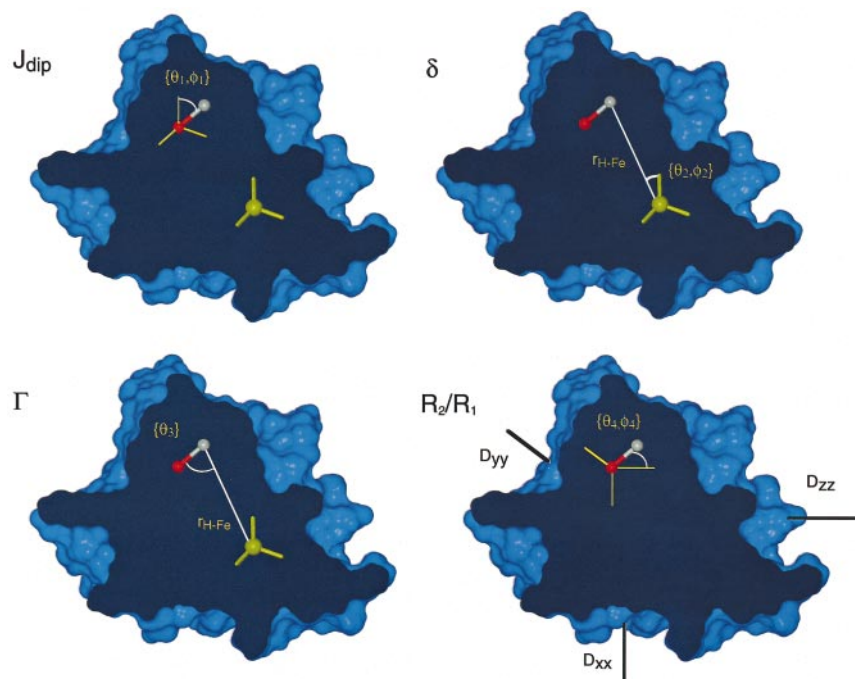
Figure 1. Geometric dependence of the interactions exploited as structural constraints in the calculation. The NH vector is in red/white. The paramagnetic centre and associated magnetic susceptibility tensor χ are in yellow. Top left-hand side. J_{dip}, Residual dipolar coupling, dependent on the covalent distance NH and the angle (θ_1, ϕ_1) made by the vector and χ. Top right-hand side. Pseudocontact shift, dependent on the angle (θ_2, ϕ_2) made between the proton and the tensor and the distance H-Fe. Bottom left-hand side. Curie-dipolar cross-correlation; dependent on the angle θ_3 Fe-H-N and the distance H-Fe. Bottom right-hand side. Autorelaxation rates in anisotropically tumbling molecules, depend on the orientation (θ_4, ϕ_4) with respect to the rotational diffusion tensor, (D_{xx}, D_{yy}, D_{zz}) (co-axiality between csa and dipole-dipole interactions is assumed).

forms (Déméné *et al.*, 2000; Tsan *et al.*, 1999), and in this study we assume that the overall fold of the two forms of the protein is independent of the spin-state of the electron.

Determination of the overall fold

In order to minimise differential dynamic averaging effects, the initial calculation designed to determine the overall fold and define the interaction strengths and orientations only constrains vectors or nuclei present in recognised secondary structural elements. The Cytc′ consists of four helical regions representing 82 of the 129 residues in the molecule, with a central loop of 34 amino acid residues, allowing extensive conformational freedom for possible arrangement of the helices. Spectral overlap and paramagnetic broadening reduces the actual number of long-range constraints used (67 H^α and 81 H^N chemical shifts, 68 $\Gamma^{Curie,DD}$ and 68 J_{dip} values were measured for the $S = 2$ state, and 65 R_2/R_1 ratios measured from the diamagnetic form were also incorporated into the calculation).

Using experimental data from one paramagnetic and one diamagnetic form of Cytc′, a calculation of 285 structures produced a cluster of 13 conformations defined by a threshold energy $E_{lr,thresh}$ of 90 kcal mol^{-1}, whose rmsd for the backbone atoms

of the helical residues (5-47,80-124) compared with the mean coordinates (σ_{hel}) was 0.73($\pm$0.18) Å. The structural definition (Table 1) converges sharply with $E_{lr,thresh}$, and structures selected with a threshold of 110 kcal mol^{-1} have an equivalent dispersion of 6.0 Å. In addition, no structures with a target function $E_{lr,exp} > 100$ kcal mol^{-1} exhibited this fold ($\sigma_{hel} > 5$ Å) while no structure with $E_{lr,exp} < 90$ kcal mol^{-1} is found with $\sigma_{hel} > 1.6$ Å, underlining the ability of the method to identify unambiguously the correct topology.

Simultaneous determination of interaction tensors and structure

Absolute values of the magnetic susceptibility and rotational diffusion tensors define the strength of the measured interactions; it is therefore important to pay particular attention to the treatment of these parameters during the course of the calculation protocol. As described in Methods, we initially estimated the values of the different components of the magnetic susceptibility and rotational diffusion tensors, based on the distribution of experimentally measured parameters (Clore *et al.*, 1998a,b). These methods are, in our case, highly imprecise due to the inadequate sampling available from the limited number of measurable parameters. The estimated component

Table 1. Characteristics of the structural ensemble as a function of threshhold energy $E_{lr,thresh}$, used to define the structures retained in the ensemble

$E_{lr,thresh}$	σ_{hel} [a]	n	E_{phys} [b] (kcal mol⁻¹)	$E_{lr,exp}$ (kcal mol⁻¹)	D_{xx} (μs⁻¹)	D_{yy} (μs⁻¹)	D_{zz} (μs⁻¹)	$\Delta\chi_{ax}$ (10⁻⁸ m³ mol⁻¹)	$\Delta\chi_{rh}$ (10⁻⁸ m³ mol⁻¹)	$\bar{\chi}$ (10⁻²⁶ T⁻² J)
90	0.73 ± 0.18	13[c]	48.6 ± 4.7	83.9 ± 3.0	13.3 ± 0.6	16.8 ± 0.7	21.3 ± 0.6	2.57 ± 0.06	0.81 ± 0.06	23.4 ± 0.4
110	6.0 ± 3.3	21	52.7 ± 8.6	91.0 ± 10.2	13.4 ± 0.8	16.7 ± 0.8	21.3 ± 0.5	2.59 ± 0.11	0.84 ± 0.09	23.2 ± 0.6
150	12.0 ± 4.3	72	79.1 ± 48.3	124 ± 23	13.1 ± 1.1	17.1 ± 1.2	21.4 ± 0.9	2.61 ± 0.17	0.68 ± 0.18	23.1 ± 0.6
250	>12.0	188	92 ± 73	169 ± 44	12.9 ± 1.1	17.3 ± 1.2	21.3 ± 0.9	2.67 ± 0.18	0.62 ± 0.18	22.9 ± 1.0
4000	>12.0	285	614 ± 1010	458 ± 54	12.3 ± 1.3	16.9 ± 1.4	21.8 ± 1.3	2.72 ± 0.19	0.64 ± 0.21	22.9 ± 1.0

All structures were calculated from randomised Cartesian coordinates.
[a] σ_{hel} is the backbone rmsd over residues (5-47, 80-124) in Å.
[b] Includes all covalent and non-bonded interactions calculated for the molecular conformation.
[c] Note that the structural ensembles selected using the different threshold criteria contain all structures with $E_{lr,exp} < E_{lr,thresh}$.

values are therefore only fixed during the regularisation part of the calculation (their orientation remains free throughout). The component amplitudes are then left free to evolve during the high temperature exploratory period ($t = 20$-30 ps). The conformational sampling during the course of a typical successful run, shown in Figure 2, illustrates that the structure can still be far from the native fold at the beginning of the exploratory period, and that during this sampling period, complete structural motifs can fold into the native conformation. The tensor components sample a broad range of values and are simultaneously determined during this period, before converging in the cooling period. Note the particular behaviour of $\Gamma^{Curie,DD}$, whose importance for the organisation of the structural motifs in Cartesian space is illustrated by the stability of the value of the interaction strength $\bar{\chi}$ once the overall fold is defined.

The ensemble of accepted structures shows convergence of both tensor component amplitudes and orientations with lower $E_{lr,thresh}$, again illustrating the ability of the protocol to simultaneously determine structure and define the relevant interaction tensors (Table 1). While the sampling available to tensorial parameters during the calculation implies that the success of the protocol is largely independent of the initial component estimates, we have nevertheless repeated the protocol using significantly different initial estimates (Table 2). The ensemble of structures resulting from these three calculations are indistinguishable in terms of $E_{lr,exp}$, backbone structure, structural dispersion and, most importantly, average tensor values, which in the latter cases are significantly different from the initial estimates. It appears that as long as the global minimum in conformational and tensorial parameter space is well enough defined, as in this case, the method is robust enough to allow for imprecise initial tensor values.

## Folding of the loop region: refinement to high resolution

The structures determined using only constraints present in motifs of secondary structure (EI) were then refined using a restrained MD with additional J_{dip}, δ_{para}, and $\Gamma^{Curie,DD}$ data from the loop region (48-79) for those residues exhibiting negligible internal motion, as determined from heteronuclear NOE measurements (Tsan, 1998). In all, 20, 53 and 23 additional J_{dip}, δ_{para}, and $\Gamma^{Curie,DD}$ constraints were introduced to constrain the loop region. During this stage, the previously determined helices and interaction tensors are fixed for each structure in the EI ensemble. The structures were then all refined with all restraints and all atoms free to move (see Methods). $E_{lr,exp}$ (<115 kcal mol⁻¹) this time extending over the whole molecule, was again used as the only criterion for structure selection. The final refined ensemble (EII, PDB accession code 1eky; Figures 3 and 4 and Table 3), shows that the backbone definition now extends well beyond the secondary structural elements, and that, excluding the N and C termini, only 20 residues of the loop have $\sigma_{local} > 2.0$ Å.

Table 2. Comparison of structural ensembles calculated from different initial tensor estimations

n	σ_{hel} [b]	D_{xx} (μs⁻¹)	D_{yy} (μs⁻¹)	D_{zz} (μs⁻¹)	$\Delta\chi_{ax}$ (10⁻⁸ m³ mol⁻¹)	$\Delta\chi_{rh}$ (10⁻⁸ m³ mol⁻¹)	$\bar{\chi}$ (10⁻²⁶ T⁻² J)
I_1 [a] 15	0.93 ± 0.12	13.1 ± 0.8	17.1 ± 0.9	21.7 ± 0.9	2.55 ± 0.07	0.81 ± 0.06	22.7 ± 0.9
I_2 [a] 22	0.87 ± 0.24	12.7 ± 0.8	17.5 ± 0.9	21.8 ± 0.7	2.51 ± 0.05	0.79 ± 0.05	22.4 ± 0.9
I_3 [a] 32	0.82 ± 0.23	13.0 ± 0.9	17.1 ± 1.0	21.8 ± 0.9	2.56 ± 0.08	0.83 ± 0.08	22.6 ± 0.9

[a] Initial tensor values were: I_1 ($D_{xx} = 15.4$, $D_{yy} = 15.8$, $D_{zz} = 20.4$) μs⁻¹, ($\Delta\chi_{ax} = 2.66$, $\Delta\chi_r = 1.06$)10⁻⁸ m³ mol⁻¹, $\bar{\chi}$=20.8 10⁻²⁶ T⁻² J. I_2 ($D_{xx} = 11.9$, $D_{yy} = 17.8$, $D_{zz} = 22.4$)μs⁻¹, ($\Delta\chi_{ax} = 2.30$, $\Delta\chi_r = 0.66$)10⁻⁸ m³ mol⁻¹, $\bar{\chi}$=22.3 10⁻²⁶ T⁻² J. I_3 ($D_{xx} = 13.6$, $D_{yy} = 16.8$, $D_{zz} = 21.4$)μs⁻¹, ($\Delta\chi_{ax} = 2.81$, $\Delta\chi_r = 0.76$)10⁻⁸ m³ mol⁻¹, $\bar{\chi}$=20.3 10⁻²⁶ T⁻² J.
[b] σ_{hel} is the backbone rmsd over residues (5-47, 80-124) in Å.
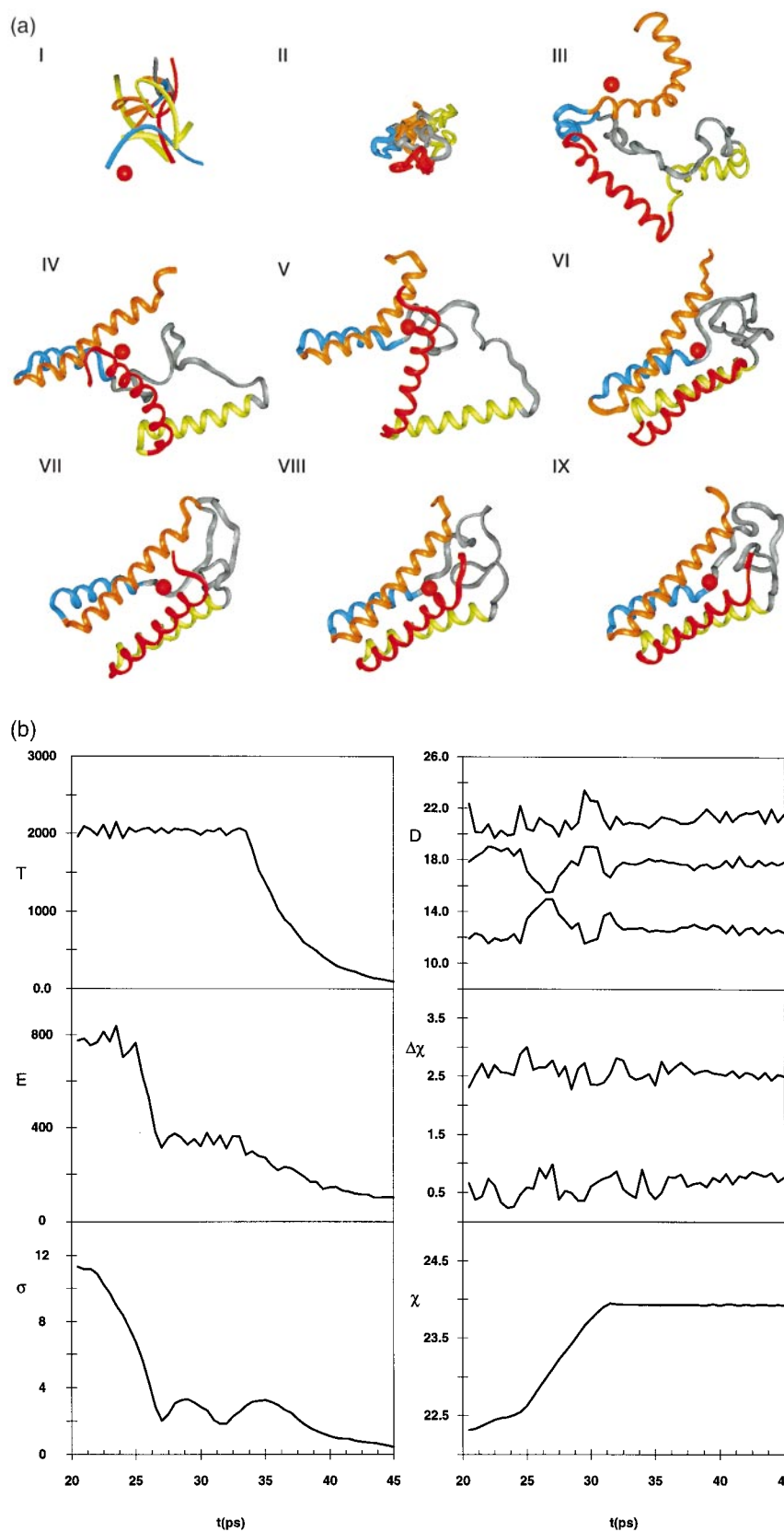
Figure 2. Sampling characteristics of the structure calculation algorithm. (a) Nine conformations taken from a sample successful calculation run from randomised atomic Cartesian coordinates. During the regularisation period (I-III; $t = 0$ ps, 10 ps and 15 ps), all constraints and force field terms are gradually increased. The high temperature exploratory period (IV-VI, $t = 20$ ps, 25 ps and 30 ps) then permits tensors (see (b)) and structure to sample conformational space. Finally the tensors and structure are frozen in the slow cooling period (VII-IX, $t = 35$, 40 and 45 ps). Helices 1-4 are orange, blue, yellow and red, respectively. Fe is shown in red. (b) Evolution of tensorial components, structural sampling and target function during the sampling $t = (20$-$33)$ ps and cooling $t = (33$-$45)$ ps periods of the calculation shown in (a) (stages IV-IX). Top left-hand side Temperature of the calculation (K); middle left, target function $E_{lr,exp}$ in kcal mol^{-1}, bottom left; rmsd of the helical (5-47, 80-124) backbone atoms compared to the final minimised structure from this run (Å). Top right-hand side. Values of the diffusion tensor components D_{xx} (bottom), D_{yy} (middle) and D_{zz} (top) during the same period (μs^{-1}). Middle right-hand side. Values of $\Delta\chi_{ax}$ (top) and $\Delta\chi_{rh}$ (bottom) (10^{-8} m^3 mol^{-1}). Bottom right-hand side, $\Gamma^{curie,DD}$ interaction strength $\bar{\chi}$ (10^{-26} T^{-2} J).

Note that the dispersion of the optimised values of the different constraints used is only slightly greater in the loop region, despite the structural heterogeneity present in the ensemble. The average target function for the loop region is 25($\pm$5) kcal mol^{-1} over 96 loop constraints, compared with 84($\pm$3) kcal mol^{-1} for the 355 constraints in the helical regions.

Figure 3. Final ensemble EII of 28 structures (PDB code 1eky) calculated using long-range paramagnetic and orientational constraints. Helices 1-4 are green, yellow, blue and orange, respectively, the loop region 48-58 in red and the remaining loop (59-79) in grey. The iron is shown in yellow. For clarity 3 C and 3 N-terminal amino acid residues are not shown.
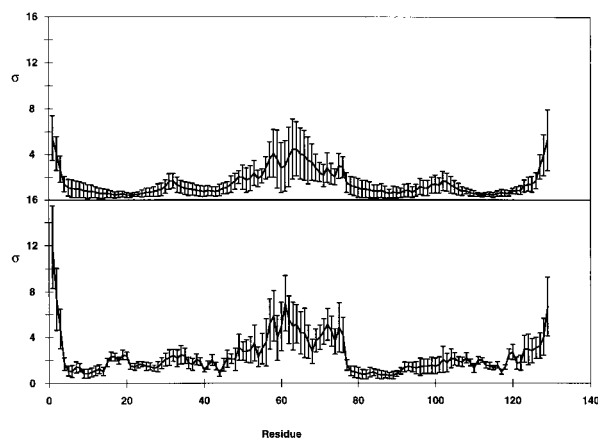
Comparison with Cytc′ in the crystalline state

The structure of Cytc′ is very similar to that determined by X-ray crystallography (Tahirov *et al.*, 1996) (Figure 4(b)). The well-ordered backbone region of the NMR structure (5-55,75-125) superposes to 1.7 Å on the equivalent regions of the crystal structure for the closest structure, and 2.05(±0.26) Å for the ensemble (Figure 5), while the most significant differences exist in regions where the disorder of the NMR ensemble is very

Figure 4. (a) Average backbone rmsd per residue of the ensemble EII compared to the mean coordinates of the ensemble. (b) Average backbone rmsd per residue of the ensemble EII compared to the crystal structure (PDB accession code 1nbb).

high in any case. Note that the tensor component values and orientations determined in this calculation are also similar to those recently estimated (Boisbouvier *et al.*, 1999; Tsan *et al.*, 1999; Tsan 1999) with respect to the crystallographic coordinates.

Characteristics of the search algorithm

It is clear from Table 1 that the algorithm has a low success rate (<5%) compared with analogous NOE-based simulated annealing protocols with which we are familiar. It is at present unclear whether this is due to the inherent underdetermination of the conformational space defined by only 355 experimentally measured structural restraints, or to the calculation strategy which is probably not yet optimised for this type of calculation. It should be noted that the precision of the orientational information would be greatly enhanced if higher spin lanthanides were used as paramagnetic labels instead of $S = 2$ Fe, or more significantly, if molecular alignment using liquid crystal media was available. In general however, the actual success rate of the protocol is of negligible importance as long as the target function allows unambiguous recognition of the overall fold. For this particular case the difference in $E_{lr,exp}$ of converged and non-converged structures is only 10% of the total target function, suggesting that less-precise data could contain multiple minima corresponding to different molecular configurations. This again implies that the precision and completeness of this dataset are
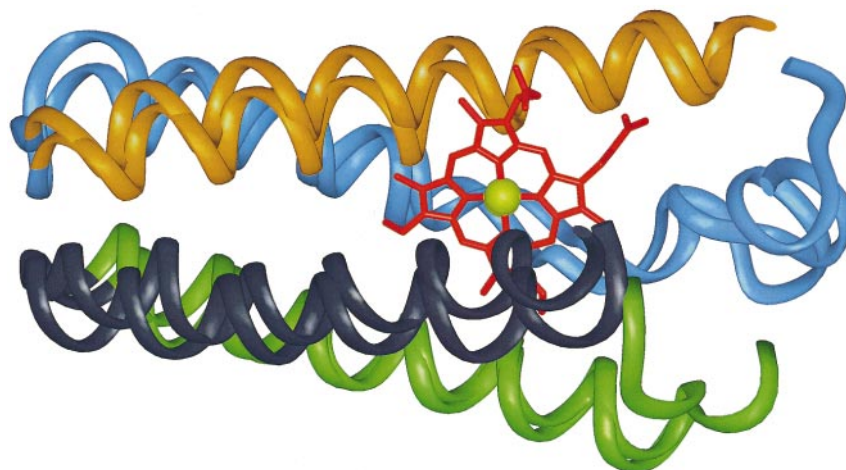
Figure 5. Ribbon representation of the structure closest to the mean coordinates with the backbone conformation of the crystal state structure. Residues 1-30 are orange; 31-60, blue; 75-100, green; and 101-129, blue-black. The final NMR ensemble superposes onto the crystal structure (PDB accession code 1nbb) coordinates to 1.7($\pm$0.2) Å over the helical residues. The ill-determined region 61-74 is not shown on the Figure for clarity.

at the limit of the feasibility of the method, and that stronger paramagnetic effects or liquid crystal-induced alignment would be desirable for more confident identification of the molecular fold.

It is also clear that the high precision of the helical regions contrasts that found for the central part of the long loop; this underlines the importance of combining local structural information, which for one residue may be degenerate, with the known relative geometry found in secondary structural motifs, thus providing a precise definition of the conformation of a group of geometrically correlated vectors. This approach is thus particularly well adapted to the positioning of secondary structural motifs, in particular for predominantly helical proteins, or for the organisation in Cartesian space of any motifs of known tertiary structure in complexes or multi-domain macromolecules. The combination of these techniques with recently proposed methods for identifying hydrogen bonding partners *via* scalar coupling (Cordier & Grzesieck, 1999) will also provide very powerful structural definition irrespective of the secondary structure.

New methods for protein structure determination: perspectives

We have shown that the complementary conformational dependence of pseudocontact paramagnetic chemical shift, Curie-Dipolar cross-correlation, and orientational measurements from molecular alignment or rotational diffusion, allied with identification of secondary structural elements, can provide sufficient structural definition to determine unambiguously backbone protein structure. The presence of a paramagnetic centre provides both purely orientational restraints in the form of residual dipolar coupling, as well as the essential Cartesian information necessary to place the oriented motifs with respect to the molecular frame, and thereby define the overall fold of the molecule. The structure determination protocol described here was made possible by specifically tailoring algorithms (SCULPTOR) incorporated into the program Discover. This allowed us to treat each of the interaction tensors independantly and to decouple the tensor orientation and eigenvalues. Another specific feature of the algorithm is to weight each restraint by the residue-specific experimental uncertainty. For this particular example the algorithm has been shown to be robust, requiring no precise *a priori* estimation of the various interaction strengths. This very important feature of the protocol is illustrated by the fact that both structure, tensor eigenvalues and their component orientations all converge to the same minimum, independent of the initial component estimates.

The implications of this example for the calculation of global folds of proteins in the absence of NOE measurements are clearly very promising, opening new opportunities to study biomolecular

Table 3. Statistics of the final ensemble EII

n	σ_{hel} (Å)	σ^a (Å)	σ^b (Å)	σ^c (Å)	σ^d (Å)	E_{phys} [e] (kcal mol^{-1})	$E_{lr,exp}$ [f] (kcal mol^{-1})
28	0.92 $\pm$ 0.20	1.09 $\pm$ 0.21	1.27 $\pm$ 0.19	1.62 $\pm$ 0.28	1.90 $\pm$ 0.24	61.3 $\pm$ 11.5	107.6 $\pm$ 6.0

[a] All rmsd values are taken from the final ensemble superimposed on the average backbone structure from residue 1-129, this explains the higher value of σ_{hel} compared with those found for table 2. σ_a is the backbone rmsd over residues (5-53, 77-124).
[b] σ_b is the backbone rmsd over residues (5-58, 74-124).
[c] σ_c is the backbone rmsd over residues (4-125).
[d] σ_d is the backbone rmsd over residues (1-129).
[e] Includes all covalent and non-bonded interactions calculated for the molecular conformation.
[f] Including constraints from helical and loop regions.

structure using solution state NMR spectroscopy. In contrast to the measurement of ^{1}H-^{1}H NOE, once the protein backbone has been assigned, the identification of all necessary constraints is unambiguous and requires no further iterative calculation. Calculations performed using only parameters measured from NH vectors show that the methods are applicable to perdeuterated systems, greatly improving relaxation-related spectroscopic considerations for very large (>30 kDa) systems. The high resolution of the structures determined in this study implies that the method can rapidly provide accurate three-dimensional fold of large proteins from the relatively simple and artefact-free NMR experiments necessary for the measurement of the constraints used. The technique requires the presence of a paramagnetic centre, and is therefore particularly appropriate for the study of paramagnetic proteins. Nevertheless the generalisation of the method to the large family of proteins containing a metal-ligation site capable of binding paramagnetic ions (Contreras *et al.* 1999), or to molecules bound to peptides containing lanthanide ions, may be feasible in cases where the structure of the native protein is unperturbed and the paramagnetic centre remains immobile relative to the molecule of interest.

Methods

Details of sample preparation and spectroscopic measurements have been published for the heteronuclear auto-relaxation, assignment of ^{13}C chemical shifts (Caffrey *et al.* 1995), assignment of diamagnetic and paramagnetic H^N and H$^\alpha$ chemical shifts of the $S = 2$ form of the molecule (Tsan *et al.*, 1999), measurement of J_{HN} residual dipolar coupling constants for the $S = 2$ and diamagnetic forms at 600 MHz ^{1}H frequency and measurement of the Curie-dipolar cross correlation effect in the $S = 2$ form (Boisbouvier *et al.* 1999).

Secondary structure

No NOE datum was used in the calculation, and the only additional constraints defined the secondary structural elements (in this case four helices) identified from hydrogen-deuterium exchange experiments and ^{13}C chemical shifts (Spera & Bax, 1991). Standard $(i,i + 4)$ hydrogen bonding (1.8-2.2) Å HN—O, and (2.7-3.3) Å N—O distance restraints and ϕ dihedral angle restraints $-80°$ to $-40°$ were used for this purpose. Weak (0.1 kcal mol^{-1} Å^2) electron-spin nuclear-spin distance restraints ($d < 11$ Å) for protons paramagnetically broadened in the $S = 2$ spectra and similar constraints for non-broadened lines ($d > 10$ Å) were also introduced.

Incorporation of novel restraints

We have programmed each of the novel constraint types in an in-house modified version of Fdiscover (MSI) entitled SCULPTOR (structure calculation using long-range paramagnetic, tensorial and orientational restraints) as explicit target potentials in addition to the classical potential energy function. Potentials of the form:

$$E_{R2/R1} = k_{R2/R1}[R_{2exp}/R_{1exp} - R_{2calc}/R_{1calc}]^2/\sigma_R^2 \quad (4a)$$

$$E_{JDip} = k_{JDip}[J_{exp} - J_{calc}]^2/\sigma_J^2 \quad (4b)$$

$$E_\delta = k_{para}[\delta_{exp} - \delta_{calc}]^2/\sigma_\delta^2 \quad (4c)$$

$$E_{Cur,DD} = k_{Cur,DD}[\Gamma_{exp} - \Gamma_{calc}]^2/\sigma_\Gamma^2 \quad (4d)$$

were incorporated into the AMBER4 force field. The experimentally determined uncertainties σ_{exp} provide a residue-specific weighting depending on the confidence in the measured data, while the various prefactors k_i were scaled so that the total energetic contributions from each different constraint type were approximately matched during the calculation.

Residual dipolar NH couplings measured at a ^{1}H frequency of 600 MHz for the $S = 2$ and $S = 0$ form were combined such that the target coupling is equal to $J_{dip} = (J_{600}^{S = 2} - J_{600}^{S = 0})$. We have preferred using this term rather than equivalent calculations using the term $\Delta J_{dip} = ((J_{600} - J_{400})^{S = 2} - (J_{600} - J_{400})^{S = 0})$ which effectively removes the contribution due to dynamic frequency shift (DFS) (Tjandra *et al.* 1996) , in view of the range of J_{dip} and ΔJ_{dip} compared with their respective experimental errors. This implies the presence of a random error of 10 % of the measured J_{dip} range due to DFS. Curie-dipolar cross-correlated relaxation was characterised using a constant $\bar{\chi}$ defining the strength of the interaction which takes into account the average magnetic susceptibility (Boisbouvier *et al.*, 1999). Potential energy gradients with respect to atomic Cartesian coordinates were calculated numerically for each of the novel terms introduced into the program. The magnetic susceptibility χ, rotational diffusion D and average tensor $\bar{\chi}$ required for the cross-correlation Γ are all represented by a number of additional virtual molecules to facilitate the calculation of energy and gradients associated with each of the novel potentials. Tensor orientation and eigenvalue component amplitude are decoupled entirely, to allow independent manipulation of rotational freedom and interaction strength for the two tensors.

Simulated annealing

All calculations used a simple quartic repulsive term for all non-bond atomic interactions and atomic radii were fixed at 0.825 of their van der Waals values. No non-bond interactions were calculated beyond 5.0 Å and interactions concerning sidechains beyond C$^\beta$ were excluded from the structure calculation to improve structure calculation efficiency and to facilitate reorientation of multi-residue structural elements during the calculation. All calculations used a leapfrog algorithm sequential time-step of 1fs.

Determination of structure from random coordinates

Using SCULPTOR we have developed a simulated annealing structure calculation protocol starting from randomised atomic Cartesian coordinates. The algorithm uses the following approach; covalent, non-bonded and secondary structure contributions starting at 10^{-6} of their maximum values are smoothly increased over a heating period of 20 ps. The weighting of other experimental data is simultaneously increased during this period from 0.001 to 1.0 at $t = 20$ ps, followed by a 10 ps exploratory

period at 2000 K, and a slow cooling period (to 100 K) of 13 ps. During this period the experimental weighting is also increased by a factor of 2. Structures were accepted or rejected only with respect to their agreement with the measured long-range experimental restraint data, defined by the parameter $E_{\text{lr-exp}} = E_{\text{para}} + E_{\text{Curie,DD}} + E_{\text{Jdip}} + E_{\text{R2R1}}$.

In the interests of efficient structure generation, once the overall fold had been unambiguously identified from the structures with the lowest $E_{\text{lr,exp}}$ ($E_{\text{lr,thresh}}$), this ensemble was used as a seed for further annealing calculations starting from the exploratory period at 2000 K. The remainder of the protocol is identical with that described above and again complete freedom is allowed for tensor amplitude and orientation to be determined. The final ensembles derived from these calculations were again selected using $E_{\text{lr,thresh}} = 90$ kcal mol^{-1} and were statistically indistinguishable from the seed structures or from each other.

Initial estimates of the eigenvalues of the rotational diffusion tensors were derived from the raw data with no reference to structure (Clore *et al.* 1998a) using the distribution of R_2/R_1 values. Similar methods, based on the measured coupling constants (Clore *et al.* 1998b), allied to magnetic susceptibility measurements were used to estimate approximate initial values for $\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$ (Tsan *et al.*, 1999). $\bar{\chi}$ was estimated from the theoretical calculation of the interaction strength from physically known parameters, which gave similar values to estimates based on the range of measured values. In order to test the effect of the starting values of interaction strengths, structure calculations were repeated using different initial estimates and the final converged structure and tensor characteristics compared. These estimates are fixed during the initial stages of the calculation and the orientation of the tensor left unrestrained. In order to simultaneously optimise tensor and structure during the exploratory period, both eigenvalues and orientation are left free to float in a large flat-bottomed potential well. The tensors are thus determined only by the correlation between the relevant restraints and the molecular conformation. The interaction strengths are then slowly frozen prior to cooling of the molecular system.

Refinement stage

The helices and tensors were tethered to their determined values during the 10 ps exploratory period at 2000 K, followed by slow cooling to 100 K. Five structures were calculated for each member of the EI ensemble, and the ensemble of structures with the lowest $E_{\text{lr,thresh}}$-values from each initial structure refined. A restrained molecular dynamics at 600 K was performed with all parameters free and using all experimental constraints to relieve local stress in the structure to produce the final ensemble EII. The lowest $E_{\text{lr,thresh}}$-values over the whole molecule were again taken to represent the refined ensemble.

References

Banci, L., Bertini, I., Huber, J. G., Luchinat, C. & Rosato, A. (1998). Partial orientation of oxidized and reduced cytochrome b_5 at high magnetic fields: magnetic susceptibility anisotropy contributions and consequences for protein solution structure determination. *J. Am. Chem. Soc.* **120**, 12903-12909.

Bertini, I. & Luchinat, C. (1986). *NMR of Paramagnetic Molecules in Biological Systems*, Benjamin/Cumings, Inc., Menlo Park, USA.

Bolon, P. J., Al-Hashimi, M. & Prestegard, J. H. (1999). Residual dipolar coupling derived orientational constraints on ligand geometry in a 53 kDa protein-ligand complex. *J. Mol. Biol.* **293**, 107-211.

Boisbouvier, J., Gans, P., Blackledge, M., Brutscher, B. & Marion, D. (1999). Long-range structural information in NMR studies of paramagnetic molecules from electron spin-nuclear spin cross-correlated relaxation. *J. Am. Chem. Soc.* **121**, 7700-7701.

Brüschweiler, R., Liao, X. & Wright, P. E. (1995). Long-range motional restrictions in a multi-domain zinc-finger protein from anisotropic tumbling. *Science,* **268**, 886-889.

Caffrey, M., Simorre, J.-P., Brutscher, B., Cusanovich, M. & Marion, D. (1995). NMR assignment of *Rhodobacter capsulatus c′*, a 28 kDa paramagnetic heme protein. *Biochemistry,* **34**, 5904-5912.

Cavanagh, J., Fairbrother, W., Palmer, A. & Skelton, N. (1996). *Protein NMR Spectroscopy*, Academic Press, San Diego.

Clore, G. M., Gronenborn, A. M., Szabo, A. & Tjandra, N. (1998a). Determining the magnitude of the fully asymmetric diffusion tensor from heteronuclear relaxation data in the absence of structural information. *J. Am. Chem. Soc.* **120**, 4889-4890.

Clore, G. M., Gronenborn, A. M. & Bax, A. (1998b). A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J. Magn. Reson.* **133**, 216-221.

Clore, G. M. & Garrett, D. S. (1999). *R*-factor, free *R*, and complete cross validation for dipolar coupling refinement of NMR structures. *J. Am. Chem. Soc.* **121**, 9008-9012.

Contreras, M. A., Ubach, J., Millet, O., Rizo, J. & Pons, M. (1999). Measurement of one-bond dipolar couplings through lanthanide-induced orientation of a calcium binding protein. *J. Am. Chem. Soc.* **121**, 8947-8948.

Cordier, F. & Grzesiek, S. (1999). Direct observation of hydrogen bonds on proteins by inter-residue $^{3\text{h}}J_{\text{NC}}$ scalar couplings. *J. Am. Chem. Soc.* **121**, 1601-1602.

Déméné, H., Tsan, P., Gans, P. & Marion, D. (2000). NMR determination of the magnetic susceptibility anisotropy of cytochrome *c′* of *Rhodobacter capsulatus* by $^{1}J_{\text{HN}}$ dipolar coupling constant measurements. Characterisation of its monomeric state in solution. *J. Phys. Chem.* **104**, 2559-2569.

Fischer, M. W., Losonczi, J. A., Weaver, J. L. W. & Prestegard, J. H. (1999). Domain orientation and

dynamics in multi-domain proteins from residual dipolar coupling. *Biochemistry,* **38**, 9013-9022.

Ghose, R. & Prestegard, J. H. (1997). Electron spin-nuclear spin cross-correlation effects on multiplet splittings in paramagnetic proteins. *J. Magn. Reson.* **128**, 138-143.

Gochin, M. & Roder, H. (1995). Protein structure refinement based on paramagnetic NMR shifts: Applications to wild-type and mutant forms of cytochrome *c*. *Protein Sci.* **4**, 296-305.

Guéron, M. (1975). Nuclear relaxation in macromolecules by paramagnetic ions: a novel mechanism. *J. Magn. Reson.* **19**, 58-66.

Hansen, M. R., Müller, L. & Pardi, A. (1998). Tunable alignment of biological macromolecules in the magnetic field. *Nature Struct. Biol.* **5**, 1065-1074.

Olejniczak, E. T., Meadows, R. P., Wang, H., Cai, M., Nettesheim, D. G. & Fesik, S. W. (1999). Improved NMR structures of protein/ligand complexes using residual dipolar couplings. *J. Am. Chem. Soc.* **121**, 9249-9250.

Pervushin, K., Riek, B., Wider, G. & Wüthrich, K. (1997). Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl Acad. Sci. USA,* **94**, 12366-12371.

Sanders, C. R., Hare, B. J., Howard, K. P. & Prestegard, J. H. (1994). Magnetically oriented phospholipid micelles as a tool for the study of membrane associated molecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **26**, 421-444.

Spera, S. & Bax, A. (1991). Empirical correlation between protein backbone conformation and C^{α} and C^{β} ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113**, 5490-5492.

Tahirov, T. H., Shintaro, M., Meyer, T. E., Cusanovich, M., Higuchi, Y. & Yasuako, N. (1996). High resolution crystal structure of two polymorphs of cytochrome *c'* from the purple phototrophic bacterium, *Rhodobacter capsulatus*. *J. Mol. Biol.* **259**, 467-479.

Tjandra, N. (1999). Establishing a degree of order: obtaining high resolution NMR structures from molecular alignment. *Structure,* **7**, R205-R211.

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

Tjandra, N., Grzesieck, S. & Bax, A. (1996). Magnetic field dependence of nitrogen-proton J splittings in ^{15}N enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J. Am. Chem. Soc.* **118**, 6264-6272.

Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. & Bax, A. (1997a). Defining long-range order in NMR structure determination of magnetically ordered macromolecules in solution. *Nature Struct. Biol.* **4**, 732-738.

Tjandra, N., Garrett, D. S., Gronenborn, A. M., Bax, A. & Clore, G. M. (1997b). Defining long-range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nature Struct. Biol.* **4**, 443-449.

Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. (1995). Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc. Natl Acad. Sci. USA,* **92**, 9279-9283.

Tsan, P. (1998). Magnetic, structural and dynamic studies of cytochrome *c'* from *Rhodobacter capsulatus*. PhD Thesis, Universite Joseph Fourier, Grenoble, France.

Tsan, P., Caffrey, M., Lawson-Daku, M., Cusanovich, M., Marion, D. & Gans, P. (1999). Unusual contact shifts and magnetic tensor orientation in *Rhodobacter capsulatus c'*: NMR, magnetic susceptibility and EPR studies. *J. Am. Chem. Soc.* **121**, 1795-1805.

Tu, K. & Gochin, M. (1999). Structure determination by restrained molecular dynamics using nmr pseudo-contact shifts as experimentally determined constraints. *J. Am. Chem. Soc.* **121**, 9276-9285.

Woessner, D. E. (1962). Nuclear spin relaxation in ellipsoids undergoing rotational Brownian motion. *J. Chem. Phys.* **37**, 647-654.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, Wiley, New York.