Journal of Biomolecular NMR, **22:** 123–136, 2002.
KLUWER/ESCOM

Efficiency of paramagnetism-based constraints to determine the spatial arrangement of α-helical secondary structure elements

Ivano Bertini[a,*], Marco Longinetti[b], Claudio Luchinat[c], Giacomo Parigi[c] & Luca Sgheri[d]
[a]*CERM and Department of Chemistry, University of Florence, via L. Sacconi, 6, I-50019 Sesto Fiorentino, Italy;* [b]*Dipartimento di Ingegneria Agraria e Forestale, p.le delle Cascine 15, I-50144 Florence, Italy;* [c]*CERM and Department of Agricultural Biotechnology, University of Florence, p.le delle Cascine 28, I-50144 Florence, Italy;* [d]*Istituto di Analisi Globale ed Applicazioni (CNR), via Santa Marta 13A, I-50139 Florence, Italy*

Received 24 September 2001; Accepted 29 November 2001

Key words: α-helical proteins, paramagnetism-based constraints, pseudocontact shifts, quick solution structure determination, residual dipolar couplings

Abstract

A computational approach has been developed to assess the power of paramagnetism-based backbone constraints with respect to the determination of the tertiary structure, once the secondary structure elements are known. This is part of the general assessment of paramagnetism-based constraints which are known to be relevant when used in conjunction with all classical constraints. The paramagnetism-based constraints here investigated are the pseudocontact shifts, the residual dipolar couplings due to self-orientation of the metalloprotein in high magnetic fields, and the cross correlation between dipolar relaxation and Curie relaxation. The relative constraints are generated by back-calculation from a known structure. The elements of secondary structure are supposed to be obtained from chemical shift index. The problem of the reciprocal orientation of the helices is addressed. It is shown that the correct fold can be obtained depending on the length of the α-helical stretches with respect to the length of the non helical segments connecting the α-helices. For example, the correct fold is straightforwardly obtained for the four-helix bundle protein cytochrome b_{562}, while the double EF-hand motif of calbindin D_{9k} is hardly obtained without ambiguity. In cases like calbindin D_{9k}, the availability of datasets from different metal ions is helpful, whereas less important is the location of the metal ion with respect to the secondary structure elements.

Introduction

The large demand of protein structures poses the problem of finding new methodologies for a quicker structure determination. A relevant class of proteins are the paramagnetic metalloproteins, i.e. proteins associated with a paramagnetic metal ion. Other proteins contain a diamagnetic metal ion that can be substituted with a paramagnetic one. Paramagnetic metalloproteins make available paramagnetism-based constraints, which are precious in addition, or in alternative, to the classical NOE constraints. They are the pseudocontact shifts (PCS) (Banci et al., 1996), the residual dipolar couplings (RDC) due to self-orientation of the metalloprotein in high magnetic fields (Tolman et al., 1995), and the cross correlation between dipolar relaxation and Curie relaxation (CCR) (Boisbouvier et al., 1999), plus constraints based on contact interactions, which are specific for a given system (Bertini et al., 2001a,c). Nuclear relaxation enhancements are also precious constraints based on the metal ion–nucleus distance (Bertini et al., 2001a). The development of new protocols to obtain the correct protein fold from paramagnetism-based constraints is thus a quite promising field of research (Banci et al., 1998a,b; Turner et al., 1998; Hus et al., 2000, 2001; Bertini et al., 2001c,e). Paramagnetism-based constraints provide basically long-range (with respect to NOEs) information only, that may result

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

of moderate utility in defining the elements of secondary structure, but may be very relevant in their assembling to give rise to the whole protein fold. Furthermore, such constraints are orientation constraints, as they depend on the relative positions of nuclei, or pairs of nuclei, in the molecular frame. The most commonly used programs for structure determination (e.g., DYANA (Güntert and Wüthrich, 1991; Güntert et al., 1997; Banci et al., 1998a), X-PLOR (Brunger, 1992; Brunger et al., 1998)) have been optimized to converge with the classical NOE constraints, and may not handle with the same efficiency the structural information contained in this new type of constraints, that do not simply consist of upper and/or lower distance limits between atoms. Therefore, as already outlined (Hus et al., 2000), the performance of these programs with paramagnetic constraints may not necessarily be a good indicator of the intrinsic strengths or weaknesses of the latter.

With this in mind, we deemed useful to build an exploratory algorithm to assess the potentiality of different sets of simulated paramagnetic constraints using a deterministic rather than a statistical approach. In the present work, attention is concentrated on α-helical metalloproteins, for which it is assumed that the number, length and location of the α-helices are known, for instance through the use of the chemical shift index (Hus et al., 2000, 2001). A procedure is then developed to determine the relative position of metal ion(s) and α-helices, modeled as rigid structures. The idea of modeling proteins in terms of rigid subdomains or fragments has recently being exploited in connection with the use of residual dipolar couplings induced by external anisotropic media (Fischer et al., 1999; Mollova et al., 2000; Delaglio et al., 2000; Meiler et al., 2000, 2001; Fowler et al., 2000; Dosset et al., 2001). In the present work, we start from rigid α-helix structures and simulate data from one or more of the following categories of paramagnetism-based constraints: pseudocontact shifts (PCS), residual dipolar coupling (RDC), and cross correlation between dipolar coupling and Curie relaxation (CCR). The additional effect of paramagnetic relaxation constraints is also evaluated. The algorithm allows for the use of additional datasets obtained by using another metal ion, either in the same or in a different position. The way this procedure behaves in various simulated cases is instructive on the intrinsic information content of paramagnetic constraints. In favorable cases, this simple procedure as such is able to reproduce the correct protein fold.

Theoretical background

Pseudocontact shifts

The pseudocontact shifts (PCS), which arise in the presence of magnetic anisotropy (Kurland and McGarvey, 1970), are given by the following equation (Kemple et al., 1988):

$$\delta_i^{pcs} = \frac{1}{12\pi r_i^3}\left[\Delta\chi_{ax}\frac{2z_i^2 - x_i^2 - y_i^2}{r_i^2} + \frac{3}{2}\Delta\chi_{rh}\frac{x_i^2 - y_i^2}{r_i^2} + \chi_{xy}\frac{6x_i y_i}{r_i^2} + \chi_{xz}\frac{6x_i z_i}{r_i^2} + \chi_{yz}\frac{6y_i z_i}{r_i^2}\right], \quad (1)$$

where χ_{ij} are the components of the symmetric magnetic susceptibility tensor χ of the metal,

$$\Delta\chi_{ax} = \chi_{zz} - \frac{\chi_{xx} + \chi_{yy}}{2},\ \Delta\chi_{rh} = \chi_{xx} - \chi_{yy},$$

r_i is the distance between the atom i and the metal ion and x_i, y_i, z_i are the coordinates of the atom i in a frame where the metal ion is at the origin. Alternatively Equation 1 can be written in the metal frame, via a rotation defined by three Euler angles. In this case, only the first two terms are retained, i.e., the pseudocontact shifts depend on the two values of magnetic susceptibility anisotropy.

Residual dipolar couplings

Residual dipolar couplings (RDC) are due to the induced partial orientation in high magnetic field caused by magnetic anisotropy. This prevents the dipolar coupling energies from averaging to zero for all the pairs of atoms of the protein. The equation describing this effect is similar to that of PCS (Tolman et al., 1995; Tjandra et al., 1996; Vold and Prosser, 1996; Bax and Tjandra, 1997). However, the metal nucleus distance is not present in the equation. In particular, due to the presence of partial orientation induced by the paramagnetic magnetic susceptibility anisotropy, the ^1J of a pair of nuclei AB experience a residual dipolar coupling given by

$$RDC\ (Hz) = -\frac{1}{4\pi}\frac{B_0^2}{15kT}\frac{\gamma_A\gamma_B\hbar}{2\pi r_{AB}^3}$$

$$\left[\Delta\chi_{ax}\frac{2\,(z_A - z_B)^2 - (x_A - x_B)^2 - (y_A - y_B)^2}{r_{AB}^2}\right.$$

$$\left.+\frac{3}{2}\Delta\chi_{rh}\frac{(x_A - x_B)^2 - (y_A - y_B)^2}{r_{AB}^2}\right], \qquad (2)$$

where B_0 is the applied magnetic field, γ_A and γ_B are the magnetogyric ratios of nuclei A and B, respectively, T is the temperature, k is the Boltzmann constant, $\hbar$ is the Plank constant divided by 2π, r_{AB} is the distance between A and B and x_A, y_A, z_A and x_B, y_B, z_B are the Cartesian coordinates of the two nuclei in the reference frame of the χ tensor. Such experimental data can be obtained, for several nuclear pairs in the protein backbone, such as ^{15}N-^{1}H of the backbone amides, $^{13}C^{\alpha}-{}^{1}H^{\alpha}$ and $^{13}C^{\alpha}$-$^{13}C'$ of the protein backbone.

Curie-dipole-dipole cross correlation

Cross correlation between various interactions involving nuclear spins causes a difference in the nuclear linewidths (Goldman, 1984; Farrar and Quintero-Arcaya, 1985, 1987). One of these cross correlation effects is that between the dipole-dipole interaction and the Curie spin relaxation (CCR) (Ghose and Prestegard, 1997). The difference in the linewidth (in Hz) of the two components for, e.g., the H nucleus of the HN doublet, calculated in the limit of an isotropic χ tensor, is given by

$$\Delta\nu_i = \frac{\mu_0}{4\pi}\frac{B_0\gamma_H^2\gamma_N\hbar\chi}{10\pi^2}\left(4\tau_c + \frac{3\tau_c}{1+\omega_H^2\tau_c^2}\right)\frac{3\,\langle \boldsymbol{r}_i, \boldsymbol{r}_{HN}\rangle^2 - r_i^2 r_{HN}^2}{2r_i^5 r_{HN}^5}, \qquad (3)$$

where r_i is the H-metal distance, r_{HN} is the proton-nitrogen distance, $\langle , \rangle$ is the scalar product, μ_0 is the permeability of vacuum, g_e is the electron g factor, τ_c is the correlation time modulating the relaxation process and ω_H is the proton Larmor frequency.

Program implementation

Solution degeneracy

The relative position of the metal ion M with respect to any rigid protein backbone can be determined from a single dataset *via* existing fitting procedures (Banci et al., 1996, 1997) by using Equation 1 with $x_i - x_M$, $y_i - y_M$, $z_i - z_M$ instead of x_i, y_i, z_i. We assume here that existing information from, e.g. the chemical shift index, permits the identification of the location and length of α-helical elements of secondary structure along a given protein sequence. Although the connectivities between each consecutive α-helix are not known, we can treat each α-helix separately as a rigid body and, using a proper dataset, the position of the metal ion can be found with respect to each of them. However, it is only possible to determine the direction of the axes of the magnetic tensor up to reflections, due to the fact that the formulas describing the dependence of PCS, RDC and CCR (Equations 1–3) contain only the squares of the coordinates of the atoms. For a single α-helix there are thus $2^3 = 8$ possible choices of the axes of the magnetic tensor, that become 4 after choosing a right hand system for the principal axes of the tensor χ. Let n be the number of α-helices of the molecule. Then there are 4^n possible ways of assembling the relative position of the α-helices, by superimposing the χ tensors determined separately from each α-helix. Since the axes of the χ tensor can be arbitrarily selected for the first α-helix, *from each single dataset* 4^{n-1} *solutions are generated*. The spatial positions of the atoms in each of the solutions give the same values for PCS, RDC or CCR (Equations 1–3), so there is no way to distinguish them.

From the point of view of the metal, once a random orientation for the axes of the magnetic susceptibility tensor is chosen, each atom has exactly 8 possible locations satisfying Equations 1–3 for the same values of the constraints. However only 4 of them maintain the chirality of the α-helix, so again we find 4^n possible ways of assembling the structure of the molecule, and 4^{n-1} *different rigid configurations of the n* α*-helices.*

In case only one dataset is available, i.e., measurements are performed in the presence of one paramagnetic metal ion only, and if the protein is constituted, for example, by 4 α-helices, there are thus 64 indistinguishable solutions, as far as the paramagnetism-based constraints are concerned. In order to distinguish the correct one, other considerations must be introduced.

For example, solutions where there is co-penetration among different helices, or where two consecutive helices are too far with respect to the number of residues connecting them, must be discarded. This will reduce considerably the number of allowed structures. However, the remaining degeneracy, if any, can be removed only by introducing some distance constraints between nuclei belonging to different helices, i.e., by using a few long range NOEs.

From the preceding considerations, it is clear that additional datasets may become very useful to select the correct protein structure among the 4^{n-1} solutions. If two different metals are substituted in a single location, any degeneracy is already resolved, in principle, when the magnetic susceptibility tensors of the two metals do not share any principal axis (not collinear). A similar condition for RDC constraints is stated in (Ramirez and Bax, 1998; Al-Hashimi et al., 2000). In fact, for each atom of any α-helix, two different tensors permit the existence of only 2 positions, opposite with respect to the metal, but only one maintains the proper chirality of the α-helix. The same condition also holds when the metals are in two different sites. Although the condition is formally the same, in this case parallel tensors are allowed, provided the vector joining the two metals is not along a principal axis of any tensor. Therefore, there is in principle only one configuration that guarantees a good agreement with both datasets. In conclusion, *for datasets arising from two metals with not collinear tensors only one configuration is possible.* Of course, this is true if the datasets are not affected by errors. Otherwise, in principle more than one solution can be found. A procedure must be thus developed to determine the best geometrical solutions with respect to the given data.

Computational strategy

To test the feasibility of the use of paramagnetic constraints only to determine the correct reciprocal orientation of rigid α-helical secondary structure elements within a protein of unknown structure a program has been developed that calculates the reciprocal positions of the α-helices and of the metal ion(s). This goal is achieved by performing a fit of paramagnetism-based constraints artificially generated from a given protein structure (d^{obs}). The fit permits to find the relative position of the α-helices with respect to the metal(s). The α-helices used in the calculations are rigid polypeptides built from the known primary sequence of the selected protein and the dihedral angles defining the ideal α-helix geometry using standard molecule-building programs. In the fit, a target function is used to determine the agreement between calculated and 'experimental' data.*

The fitting procedure is divided into two stages. The first stage is applied to single datasets, the second is applied when more than one dataset is available, which is the case when more than one metal ion is considered. The first stage of the procedure is in itself a two-step process. In the first step (Figure 1A) each α-helix in turn is kept fixed in the laboratory system, and a minimization procedure determines a suitable location for the metal and the values of the magnetic susceptibility tensor. Three variables are needed to fix the spatial position of the metal, and five more variables to represent the magnetic susceptibility tensor. Therefore, the minimization is carried over eight variables. At the end of this step, a rigid helix-tensor structure is obtained. The second step of the first stage of the procedure (Figure 1B) consists of superimposing the tensors belonging to each helix-tensor structure. This is achieved by defining a coordinate system centered on the metal (metal coordinate system). Then a rigid motion is determined, bringing the principal axes of each tensor of the helix-tensor structures to coincide with the metal coordinate system axes. The rigid motion (a composition of a translation and a rotation) is obtained by finding the diagonal form of the tensor. The normalized eigenvector matrix is the rotation matrix needed, and the three eigenvalues

*The function that performed best in our tests is a weighted normalized form of least squares:

$$\begin{aligned} TF = {} & w_{pcs} \sum_{\alpha-\text{helices}} n_{\alpha,pcs} \sum_{\text{atoms}} (d_{pcs}^{\text{obs}} - d_{pcs}^{\text{cal}})^2 \\ & + w_{rdc} \sum_{\alpha-\text{helices}} n_{\alpha,rdc} \sum_{\text{atoms}} (d_{rdc}^{\text{obs}} - d_{rdc}^{\text{cal}})^2 \\ & + w_{ccr} \sum_{\alpha-\text{helices}} n_{\alpha,ccr} \sum_{\text{atoms}} (d_{ccr}^{\text{obs}} - d_{ccr}^{\text{cal}})^2 \end{aligned} \tag{4}$$

where the internal sums are carried over the atoms of each α-helix and the external sums are carried over the α-helices composing the protein. d_{pcs}^{obs}, d_{rdc}^{obs}, d_{ccr}^{obs} are the observed values, d_{pcs}^{cal}, d_{rdc}^{cal}, d_{ccr}^{cal} are the values calculated from the protein structure, $n_{\alpha,pcs}$, $n_{\alpha,rdc}$, $n_{\alpha,ccr}$ are normalization factors, given by the reciprocal of the sum of the squares of the measured data of each α-helix, w_{pcs}, w_{rdc}, w_{ccr} are weights that depend on the estimated precision of the measurements. The normalization factors are introduced to scale the contribution to the target function of each α-helix, because different α-helices may generate values with different orders of magnitude. To deal with the excessively high data values, the square may be filtered with a ramp filter. When the fit is performed on more than one dataset, the errors corresponding to each dataset are summed. Note that this form of the target function allows us to extract the error relative to any subset of the α-helices and/or to any subset of the datasets.

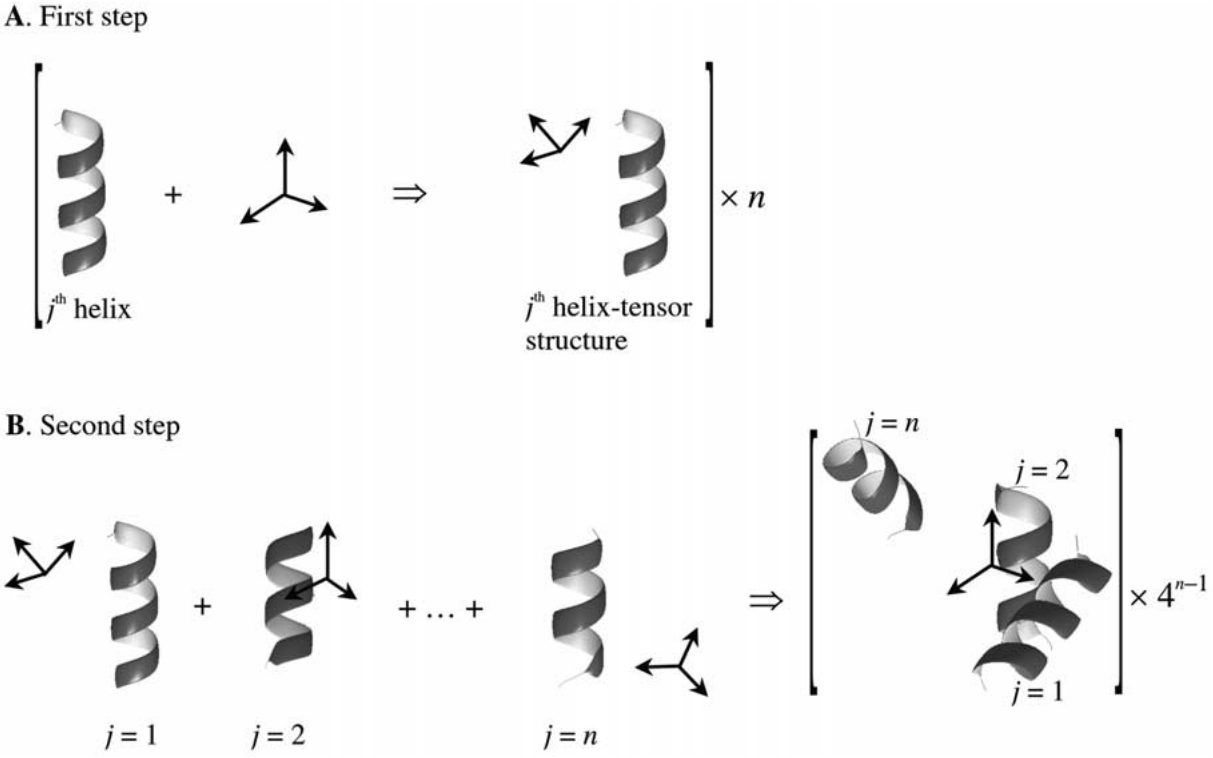

Figure 1. (A) Sketch of the first step of the fist stage of the fitting program. For each helix, the position of the metal and of the magnetic susceptibility tensor is fit in the helix frame. (B) Sketch of the second step of the fist stage of the fitting program. All helices are moved in the magnetic susceptibility tensor frame centered on the metal ion, and, from this initial position, a global fit is performed. In this way, one of the 4^{n-1} solutions is found.

are the principal values of the magnetic susceptibility tensor. The five tensor values are therefore substituted by three Euler angles representing the rotation and by $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, defined as in Equation 1. The translation vector is simply defined by the metal coordinates taken with opposite sign.

One of the 4^{n-1} solutions (as already discussed) is thus found. For datasets consisting of exact values, the other solutions can be easily generated by applying axial symmetries to the tensor axes of the helix-tensor structures, before the roto-translation is applied. For non-exact values, different $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ parameters can be obtained from the fit to the different helices. Therefore, a global fit is performed, with a total number of variables equal to $6n+2$, where n is the number of α-helices in the molecule. The position of each α-helix, as a rigid body, is in fact determined by 6 variables in a fixed reference frame. Again, one of the 4^{n-1} solutions is obtained.

In the second stage of the procedure, multiple datasets, each originating from a different metal ion, are considered. As discussed above, the first dataset requires a total of $6n + 2$ variables. Each metal ion introduces as additional variables only the position and orientation of the new tensor with respect to the first. Therefore, each additional dataset introduces only the variables related to the new tensor. These can be either 5 variables (a rotation and two values for the magnetic susceptibility anisotropy) in case the new tensor refers to a different metal in the same site as the first, or 8 variables (a translation, a rotation and two values for the magnetic susceptibility anisotropy) in case it refers to a metal in a different site. A minimization is then carried over this new set of variables, as explained below.

As the target function, due to the nature of the constraints, has many local minima, the second stage of the program is successful only if the starting values for the minimization process are close to the correct values. The search for the correct starting values is, indeed, an essential step in the search for the global minimum. To discuss this point, let us restrict, for simplicity, to the case of two datasets representing two metals in the same position. From the first stage, two

sets of $6n + 2$ variables are obtained, each set representing a rigid configuration of the α-helices with respect to its metal tensor. We know, however, that from a single configuration 4^{n-1} different symmetric positions may be generated, without any change in the target function. If the observed values are exact, only one of these positions of the first dataset coincides with one of the 4^{n-1} different positions of the second dataset (see Figure 2). This is the unique solution, and the only one having zero value of the target function for both datasets. All other configurations will have a target function larger than zero, as they have different sets of symmetries for each of the datasets, provided that the magnetic susceptibility tensors have different orientations. The trivial solution is to perform a pairwise comparison of the 4^{n-1} solutions from the first dataset with the 4^{n-1} solutions from the second dataset. Using a matching strategy (see Appendix A), the number of possible configurations that have to be inspected is quite high: $n \times 8^n$. This number grows exponentially with the number of datasets. However, since the contributions to the target function from the different helices are independent from one another, the problem reduces to a set of $n \times 8n$ cases (see Appendix A), as the best match for the whole protein is found by merging the best matches of the single α-helices. This number cannot be further reduced without running the risk of bypassing the best match, though in this set of cases there are exactly $2n$ elements with an exact match.

When an error is added to better simulate real experimental values, we still need to analyze the symmetric positions of the α-helices to find the best starting values of the fitting parameters, as described in Appendix A. The number of cases leading to exact matches is known a priori in the case of exact data. When a small error is added this situation is preserved. The identification of this $2n$ cases is a good independent test of the fact that this property is not destroyed by the error and (in principle) that the solution may be reliable. A minimization process starting from the best values (with a polytope type algorithm) then determines a candidate minimum. The program then uses the minimal value of the target function to select a subset of all the possible initial matches, and finds new minima by performing minimization processes. The final result is a set of possible configurations, ordered by increasing target function value.

Results and discussion

Simulations on protein models

The program was tested through simulation of two protein models: the oxidized *E. coli* cytochrome b_{562} (Arnesano et al., 1999), a heme protein containing a low spin Fe(III) ion, and calbindin D_{9k} (Bertini et al., 2001b), a protein which contains two calcium ions that can be substituted by lanthanides(III) in either or both binding sites (Linse et al., 1987; Kretsinger, 1980; Vogel et al., 1985). Both proteins contain four α-helices. The starting assumption is that the location and length of the α-helical elements of secondary structure along the protein sequences are known. The following procedure was used:

(1) Ideal α-helices were constructed from their known sequences using commercial programs;

(2) they were positioned in the same configuration as in the real proteins, among them and with respect to metal ion(s). Such structure configurations, constituted by ideal α-helices only, represent our protein models;

(3) PCS, RDC and CCR relative to the N, H^N, C^α and H^α atoms were then calculated to create dataset(s) by employing the programs FANTASIAN (Banci et al., 1996, 1997), FANTAORIENT (Banci et al., 1998b) and a new program (FANTACROSS) written to calculate CCR* (Bertini et al., 2001d). Realistic values for the magnetic susceptibility tensor parameters were selected. For cytochrome b_{562} the magnetic susceptibility anisotropy values of low spin iron(III) derivative was used (Arnesano et al., 1999), while for calbindin D_{9k} the magnetic susceptibility tensors of the Ce(III), Yb(III) and Dy(III) derivatives were used (Bertini et al., 2001b);

(4) calculated dataset(s) and modeled α-helices were provided to the program developed in this work to back-calculate the relative positions of the helices and of the metal ion(s), which is the required solution;

(5) such solution was compared with the starting protein model, and the RMSD between the two was calculated.

Subsequently, step 3 was performed by generating the constraints from the actual protein structures (cytochrome b_{562} and calbindin D_{9k}) rather than from the ideal helical models, and the calculations in steps 4 and 5 were repeated. The latter procedure simu-

*The programs FANTASIAN, FANTAORIENT and FANTACROSS are available at the web site: www.postgenomicnmr.net

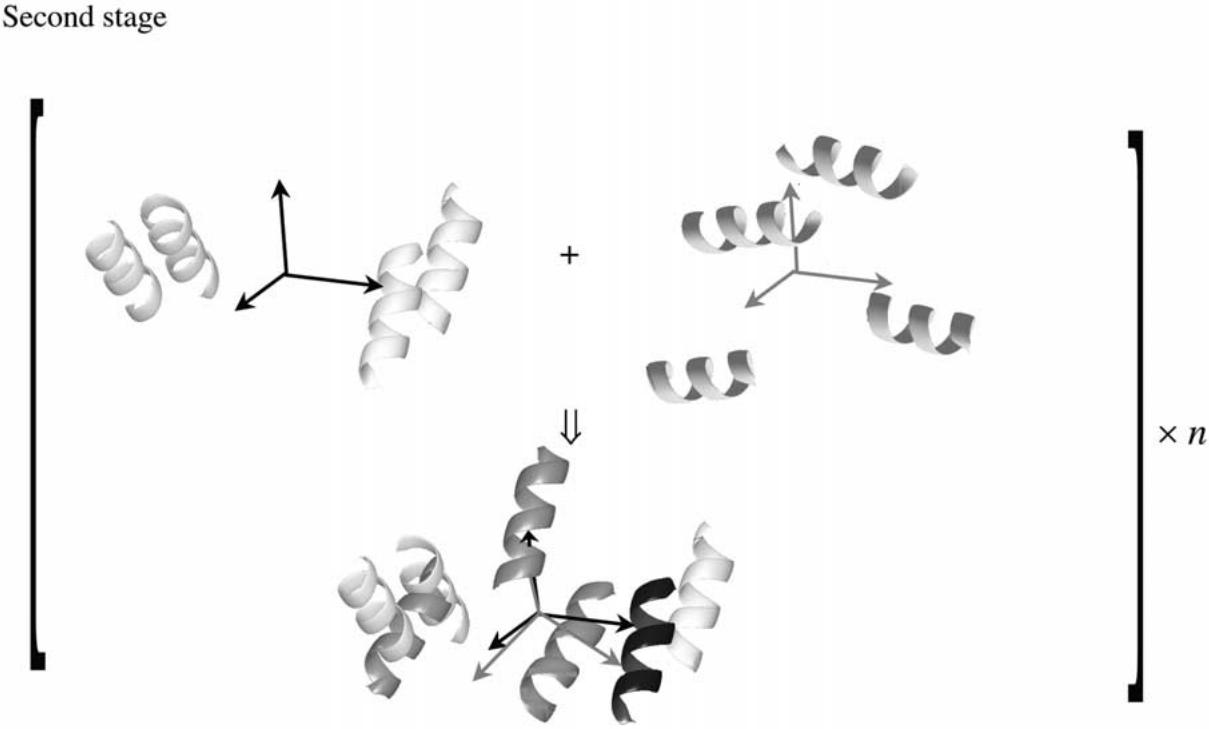

Figure 2. Sketch of the second stage of the fitting program. With two datasets, each helix can be located in four positions from the fit of the first dataset and in four different positions from the fit of the second dataset. Therefore, after all positions are referred to the same reference system, 8 positions must be inspected for each helix. For exact datasets, two out of the eight positions coincide (dark helix) and provide the solution. Differently, global fits must be performed.

lates a possible use of the present program to actually determine a protein fold.

It is expected that for a four-helix bundle protein, which is the case of cytocrome b_{562}, it is possible to obtain the correct structure. Indeed, the structure of the four helix bundle protein cytochrome c' was recently obtained with experimental data in the absence of NOE constraints (Hus et al., 2000; unpublished results from our laboratory). For calbindin, this task is expected more difficult to be achieved (Bertini et al., 2001e).

Cytochrome b_{562}

For the cytochrome b_{562} and the dataset (constituted by PCS, RDC and CCR) generated from the ideal helical model, the program finds one of the 64 possible configurations, and the others are generated by symmetry operations as described above. The 64 solutions were analyzed to remove structures with co-penetrations among α-helices and with inter-helix distances too long with respect to the maximum length that can be covered by the aminoacids between consecutive α-helices.*

Only 2 configurations have no co-penetrations and, among them, only one is acceptable as far as the inter-helix chain length is concerned. Violations for all other configurations are so large that they can be safely rejected. The RMSD between the model used to generate the dataset and the obtained configuration is equal to zero, as expected for a dataset generated with zero error.

Therefore, this first test shows that the program is indeed capable of assessing, in a deterministic way, the intrinsic ability of a set of paramagnetic constraints to univocally generate the correct protein fold. The result of the test also shows that *the cytochrome b_{562} topology (and, by extension, the 4-helix bundle topology)*

*Co-penetrations have been defined by modeling each α-helix as a cylinder, with principal directions defined by the inertia tensor and radius of 3 Å, as the position of the lateral chains are not known. The choice of 3 Å for the helix radius is a conservative one, as the real radius may be larger in the presence of bulky side chains. However, a value as small as possible was selected, as it is expected that the most probable configuration is that with no co-penetrations but with the closest lateral chains among α-helices.

Table 1. Value of the RMSD between the modeled structure of cytochrome b_{562}, used to generate a dataset containing PCS and RDC, and the solution provided by the program, for increasing value of the absolute and percent components of the error affecting the dataset

Error PCS		Error RDC	RMSD
abs	%	abs	(Å)
0	0	0	0.00
±0.05	±4	±0.2	0.35
±0.10	±8	±0.4	0.61
±0.20	±8	±0.4	0.65

can be arrived at with the sole use of paramagnetic constraints, in the presence of secondary structure information. To further test the strength of these constraints, realistic errors were then added in the dataset, with a percent component and an absolute component. The best-fit solution was obtained and the 64 possible configurations with the same value of target function were generated. Co-penetrations and inter-helix chain length selected again one acceptable solution only. Again, all symmetric configurations are rejected with confidence, as violations are really large. Table 1 shows that the RMSD between the selected solution and the protein model remains acceptable. In fact for errors of up to ±0.2 ppm plus ±8% for PCS, and up to ±0.4 Hz for RDC, the RMSD is smaller than 0.65 Å. Therefore, the correct solution is always obtained. CCR with errors of ±0.2 Hz ± 10% (i.e., of the order of the experimental ones) do not affect the obtained solution appreciably, if all PCS and RDC are present, and they were thus removed. If RDC are also removed and calculations are performed only with PCS constraints, the correct solution is found only for errors in PCS below ±0.05 ppm plus ±4% (the RMSD in this limiting case being 2.8 Å). In the same way, the correct solution is found for datasets containing both PCS and RDC of either N and H^N or C^α and H^α only if errors are kept small.

An even more stringent test for the strength of the constraints can be carried out by using real helices from the actual cytochrome b_{562} structure (Arnesano et al., 1999) to calculate the values of PCS and RDC in the datasets. Real helices may be in fact somewhat distorted with respect to the ideal geometry. The four helices were again modeled as ideal by the knowledge of the theoretical values of the dihedral angles of the α-helices domains. The program succeeded in finding

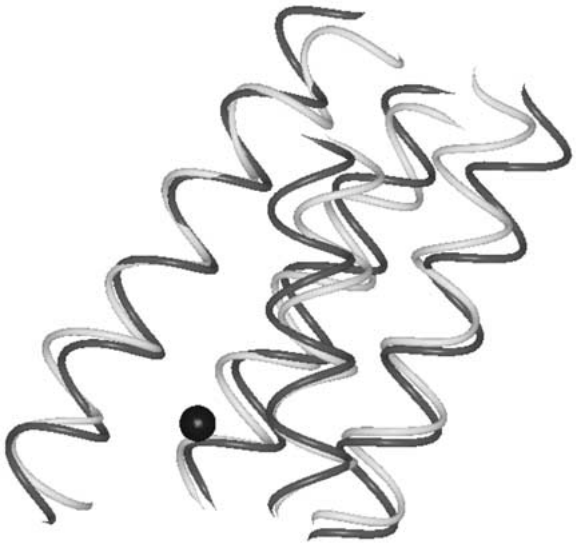

Figure 3. Superposition of the real helices of cytochrome b_{562} (light gray line), used to calculate PCS and RDC, to the solution obtained from the program (dark gray line) by fitting such data. RMSD = 1.51 Å.

the correct relative positions of the four helices with respect to the metal ion *as a unique acceptable solution*, in the sense that co-penetrations and inter-helix chain length selected only one structure among the 64 symmetric configurations obtained from the best fit solution, with smallest target function value, and thus basically the fold of the protein is again obtained (see Figure 3). In Table 2, the values of the RMSD between the protein solution structure and the calculated structure are reported for increasing values of the error introduced in the calculated values of PCS and RDC. The protein fold is reconstructed with sufficient precision even for errors larger than the experimental ones. PCS were calculated for H^N, N, C^α and H^α signals and RDC for NH and CH pairs. We noted however that if only the N and H^N or the C^α and H^α atoms were considered, no best-fit solution is acceptable in both cases. The above test performed with constraints generated from a real protein structure demonstrates that, although not designed for this purpose, in favorable cases *the present program can be used as such to univocally determine the protein fold.*

Calbindin D_{9k}

For the calbindin model the situation is strikingly different. In case one dataset only is considered, and the magnetic susceptibility tensor is chosen to be that of any of the three lanthanides, Ce(III), Yb(III)

Table 2. Value of the RMSD between the real structure of cytochrome b_{562}, used to generate a dataset containing PCS and RDC, and the solution provided by the program, for increasing value of the absolute and percent components of the error affecting the dataset

Error PCS		Error RDC	RMSD
abs	%	abs	(Å)
0	0	0	1.51
±0.025	±2	±0.1	1.32
±0.05	±4	±0.2	1.40
±0.10	±8	±0.4	1.56
±0.20	±8	±0.4	1.67

or Dy(III), 18 configurations have no co-penetrations and, among them, there are 12 acceptable configurations as far as the inter-helix chain length is concerned. This solution degeneracy is ascribed to the fact that in calbindin the helices are much shorter than in cytochrome b_{562} and inter-helix chains are much longer. This confirms the already noted difficulty of experimentally obtaining the correct protein fold of calbindin without a minimal use of NOEs (Bertini et al., 2001e). To confirm this, calculations repeated with as few as three distance constraints between atoms of different α-helices (with accuracy of ±2 Å) are enough to select the correct solution among the 12.

The situation is even worse for non exact PCS and RDC values. If datasets are affected by an error of ±0.1 ppm plus ±4% on PCS and ±0.4 Hz on RDC, in fact, in case the magnetic susceptibility of Ce(III) is considered, the program provides 5 configurations that do not even include the correct one, whereas if the larger magnetic susceptibility of Yb(III) or Dy(III) is used, the correct solution is present among 13 or 12 acceptable configurations, respectively.* All the above calculations show that *the calbindin topology is difficult to be obtained from paramagnetic constraints only*, as more than one acceptable solution is always found.

It becomes thus interesting to check whether degeneracy can be removed by considering datasets relative to two metal ions in the same position or in different positions. Calbindin is an ideal protein for such a test, as it is really possible either to substitute more than one metal ion in the same site, or in different sites. As expected from theoretical considerations, *for exact datasets, one solution only is always obtained, which, in any case, is the correct one*. The behavior of the fitting procedure for non exact datasets is thus analyzed in order to estimate: (i) the maximum error allowed by the system for providing the correct solution, (ii) the strength of the fit. In fact, errors influence the values of the many local minima of the target function, among which the global minimum may not even coincide with the 'correct' solution. The amplitude of the errors was increased up to values larger than those experimentally expected. The acceptable solutions (i.e., after discarding solutions with co-penetrations or inter-helix chain length violations) are reported in Table 3. Configurations are reported for increasing value of the target function. In some cases one configuration only is acceptable, in other cases they can be many, depending on the error and on the magnetic susceptibility tensors. The data show that the program does succeed in finding the correct solution, but the whole approach is weak, as the correct solution has a target function only slightly smaller than those of completely wrong configurations. The relevancy of this fact increases by increasing the amplitude of the error introduced in the datasets. If PCS and RDC values of the N-H pair only or of the C-H pair only are present with large errors in the datasets, the program often does not succeed in finding the correct solution. It was also checked that inclusion of CCR with errors of ±0.2 Hz ± 10% does not improve the quality of the solutions appreciably, if all PCS and RDC are present.

If datasets are calculated using the real protein structure as such, and again ideal α-helices, modeled according to the theoretical dihedral angles expected in α-helices domains, are provided to the program to be located according to the protein fold, the program succeeds in finding the correct solution only for Yb(III) and Dy(III) ions substituted in the same position, with RMSD = 2.6 Å. This further worsening is

*The large number of allowed configurations without co-penetrations is also due to the fact that, for a metal ion that occupies a lateral position in a protein fold, the symmetries of the susceptibility tensor axes distribute the helices over a wider space than for a metal ion in the center of the protein. This reduces the probability of co-penetrations. Therefore, we made the same calculations with the metal fictitiously located in the center of the protein. For an exact dataset and for the magnetic susceptibility of Ce(III), 6 configurations have no co-penetrations and, among them, there are 3 acceptable configurations as far as the inter-helix chain length is concerned (to be compared with the 12 solutions obtained in the lateral metal ion case). For a non exact dataset, the correct location is always found among 4, 8 or 3 configurations, for the magnetic susceptibility of Ce(III), Yb(III) or Dy(III), respectively. This shows that indeed a lateral position of the metal ion makes the solution degeneracy larger.

Table 3. Value of the RMSD between the modeled structure of calbindin, used to generate datasets containing PCS and RDC, and the best-fit solutions provided by the program, ordered for increasing value of the target function, for increasing value of the absolute and percent components of the error affecting the datasets. Couples of datasets have been used among those calculated for Ce(III), Yb(III) and Dy(III) in the same or in different sites

Ion 1	Ion 2	Error abs	PCS %	Error RDC abs	TF	RMSD (Å)
Metals in the same site						
Ce(III)	Yb(III)	±0.025	±2	±0.1	0.038	0.048
		±0.05	±4	±0.2	0.153	0.098
		±0.1	±8	±0.4	0.607	0.202
		±0.2	±8	±0.4	1.45	0.361
					1.81	∼11
Ce(III)	Dy(III)	±0.025	±2	±0.1	0.038	0.065
					0.051	∼8
					0.059	∼12
					0.074	∼14
		±0.05	±4	±0.2	0.154	0.135
					0.167	∼8
					0.177	∼12
					0.192	∼14
					0.194	∼13
					0.200	∼14
		±0.1	±8	±0.4	0.609	0.280
					0.623	∼8
					0.624	∼8
					0.638	∼12
					0.646	∼15
					0.647	∼14
					0.650	∼13
		±0.2	±8	±0.4	1.610	0.387
					1.631	∼8
					1.631	∼8
					1.635	∼13
					1.647	∼14
					1.653	∼13
					1.653	∼12
Metals in different sites						
Ce(III)	Ce(III)	±0.025	±2	±0.1	0.498	0.312
		±0.05	±4	±0.2	1.398	0.614
					1.665	∼10
		±0.1	±8	±0.4	2.514	1.011
		±0.2	±8	±0.4	3.517	∼9
Ce(III)	Yb(III)	±0.025	±2	±0.1	0.042	0.072
		±0.05	±4	±0.2	0.170	0.155
		±0.1	±8	±0.4	0.663	0.349
					0.903	∼8
		±0.2	±8	±0.4	1.84	∼8
Ce(III)	Dy(III)	±0.025	±2	±0.1	0.036	0.073
		±0.05	±4	±0.2	0.149	0.152
		±0.1	±8	±0.4	0.588	0.320
					0.977	∼8
		±0.2	±8	±0.4	1.343	0.596
					1.567	∼8
					1.601	∼11

expected to be due to deformations of real helices with respect to the ideal ones, as it is discussed below.

A set of calculations has been performed to analyze the possible contribution of relaxation constraints to discriminate between the correct solution and other solutions with similar target function. For the calbindin D_{9k} system, it appears that the contribution to the target from the violations of relaxation constraints is relatively small, and it is not necessarily the smallest for the correct solution.

Helix length and distortion

Despite the striking difference in behavior between the two protein models, for both cytochrome b_{562} and calbindin D_{9k} a relevant contribution to the RMSD value is provided by the fact that helices are not ideal and often distorted in some way (see Figure 3). Even if such distortions correspond to differences in the position of the atoms up to $0.5-1$ Å, i.e., they are very small, PCS and RDC are sensitive enough constraints to cause the appearance of new minima that can be even lower than that corresponding to the correct structure. Therefore, the program may not be able to find the correct solution. It is found that helix distortion in cytochrome b_{562} is tolerable, whereas it is not in calbindin. As already discussed, solution degeneracy in calbindin is ascribed to the fact that the helices are much shorter than in cytochrome b_{562}, and inter-helix chains are much longer.

It could thus be instructive to investigate what happens when helix length is reduced from that of cytochrome b_{562} to that of calbindin. To do so, the first and last aminoacids are thus progressively removed from each modeled helix of cytochrome b_{562}, as major distortions often occur at the heads and tails of the helices. Table 4 shows that even for multiple reductions of the helix length and simultaneous increase of the inter-helix chains, one acceptable solution only is obtained. Such solution is obtained by fitting exact dataset calculated for the modeled protein. However, as soon as the helix length and inter-helix chains are of the same order as in calbindin, the same degeneracy is obtained than that calculated for calbindin. This shows that solution degeneracy indeed depends on helices length with respect to inter-helix chain length. Table 4 also reports the number of structures with small violations, that can thus be acceptable if small distortions of helices were allowed. Such value is thus an indication of the strength of the obtained acceptable solution(s).

Table 4. Number of acceptable structures, as far as the co-penetrations and the inter-helix chain length is concerned, and of structures with small violations, obtained from the program by fitting the exact PCS and RDC values of one dataset. For cytochrome b_{562} calculations have been performed by progressively removing the first and last aminoacids from all helices and thus simultaneously increasing the inter-helix chains length

	Acceptable structures	Structures with small violations
Cytochrome b_{562}		
Whole helices	1	0
helices-2 a.	1	0
helices-4 a.	1	1
helices-6 a.	1	5
helices-8 a.	4	13
helices-10 a.	14	13
Calbindin		
Whole helices	12	15

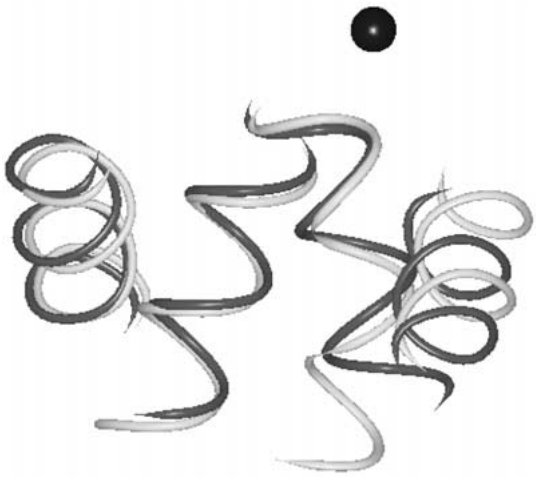

Figure 4. Superposition of the real helices of calbindin D_{9k} (light gray line), used to calculate PCS and RDC in the presence of Dy(III) or Yb(III) in the same site, from which the values relative to the first and last aminoacids of each helix have been removed, to the solution obtained from the program (dark gray line) by fitting such data. RMSD = 1.96 Å.

A similar calculation was performed for calbindin. Here PCS and RDC (CCR are not used) of the first and last aminoacids calculated for the real helices are removed from the dataset. Surprisingly, if the ions with the largest magnetic susceptibility are used (i.e., Dy(III) and Yb(III)), the structure with the smallest target function is now closer to the correct structure, with RMSD smaller than 2 Å (see Figure 4). This shows that a major requirement for the program to be able to indicate the correct protein fold is that real helices are close to ideality.

Concluding remarks

In summary, the above theoretical considerations and the simulations performed with the present program indicate that:

(i) PCS, RDC and CCR provide enough constraints to obtain the correct position of one helix in the reference frame of a metal ion;

(ii) in case a metalloprotein is constituted by n helices, their relative orientation is not uniquely determined by PCS, RDC and CCR constraints, if all related to the same metal ion, but there are 4^{n-1} possible configurations. In order to select the correct structure, helix co-penetrations and inter-helix length violations can be inspected. If this is not enough to remove all degeneracy, a unique solution can be obtained (a) by substituting two different paramagnetic metal ions in the same location, or (b) by employing two paramagnetic metal ions in different locations, or (c) by acquiring a few long range distance constraints. Paramagnetic relaxation constraints are not particularly helpful;

(iii) in the particular case of four helices, which are constituents of several common protein folds, strikingly different behaviors are shown by cytochrome b_{562} and calbindin D_{9k}. Crucial elements to obtain the correct fold are helix length and helix ideality. In fact, for cytochrome b_{562}, the correct fold of the α-helices is determined even by using the paramagnetic constraints relative to one metal ion only, and it is obtained also by generating PCS and RDC from the real protein structure; for calbindin, datasets relative to two different metal ions (or two metal ions in different sites) are needed to remove any degeneracy and to obtain the correct structure. However, the fits in the presence of non-exact datasets are weak, and further problems arise when PCS and RDC are generated from the real protein structure, due to the non ideality of helices.

As a side result, the present calculations also indicate that the different orientation of the susceptibility tensor for different lanthanides, even if relatively small (Bertini et al., 2001b), is large enough to allow the use of one lanthanide to select the correct solution among those allowed by another lanthanide, even in

the presence of large errors affecting the input data. It is also shown that lanthanides with large magnetic susceptibility have a somewhat better discriminating power among the various solutions.

In conclusion, the approach described here and the developed program has allowed us to put on more quantitative basis the intrinsic information content of paramagnetic constraints, for α-helical proteins, as a function of the protein topology. It has been also shown that, in favorable cases, the present program can by itself generate the proper protein fold. This result is encouraging towards further attempts at developing more advanced programs for solution structure determination of paramagnetic proteins with optimized use of paramagnetic constraints. Such programs could be precious especially for large molecular weight proteins, where the classical approach cannot be successfully applied.

Acknowledgements

Theoretical discussions with Professor Giorgio Talenti have been very useful. This work has been supported by Murst ex 40%, Italy, the European Union, contract HPRI-CT-1999-00009, contract HPRN-CT-2000-00092, contract QLG2-CT-1999-01003 and contract BIO4-CT98-0156, CNR, Progetto Finalizzato Biotecnologie, Italy, contract 990039349, Italy and CNR, Comitato nazionale per le scienze chimiche, Italy, contract 970113349.

Appendix A

The first stage of the numerical procedure consists in determining, for each dataset D^k, a set of rigid motions (a_j^k, t_j^k), $j = 1, \ldots, n$ which bring the α_j from the lab system to the M^k metal system, and two values $\Delta\chi_{ax}^k$ and $\Delta\chi_{rh}^k$. This stage of the procedure does not show particular difficulties if the error is kept to a reasonable level.

The second stage of the procedure consists in joining data coming from different datasets. Merging the rigid motions in a single setting involves a first arbitrary choice. A second arbitrary choice has to be made to find, in the setting already determined, the best matching positions for the α_j. First we have to define a set of transformations to describe the multiple dataset setting, denoted by capital letters. Small letters stand instead for data coming from stage one. Let $(A_j^1, T_j^1) = (a_j^1, t_j^1)$ be transformations from the lab frame to the metal system M^1. The relative positions of the α_j are already defined by (A_j^1, T_j^1). The tensor M^k may be represented by means of a rigid motion (A^{1k}, T^{1k}) bringing the metal system M^1 in the metal system M^k. In other words, an atom belonging to α_j with position vector x in the lab frame has coordinates $A_j^1(x - T_j^1)$ in the system M^1, and has coordinates:

$$A^{1k}(A_j^1(x - T_j^1) - T^{1k}) \tag{A1}$$

in the system M^k. Two or more metal ions may share the same location (this is an a-priori piece of information), so not every T^{1k} is defined independently. The set of variables are hence $6n + 2$ for the first dataset, while each successive dataset D^k adds 5 variables (a rotation and two coefficients) if M^k is in an already defined location, or 8 variables (a rotation, a translation and two coefficients) if M^k is in a new position. Let X_j be the set of coordinates of the ordered atoms of α_j in the reference system. Let Z_j^k be the set of coordinates of the same atoms in the system M^k. The situation is described by the following graph:

$$\left.\begin{array}{l} X_1 \overset{(A_1,T_1)}{\rightarrow} Z_1^1 \\ X_2 \overset{(A_2,T_2)}{\rightarrow} Z_2^1 \\ \cdots \\ X_n \overset{(A_n,T_n)}{\rightarrow} Z_n^1 \end{array}\right\} \overset{(A^{1k},T^{1k})}{\rightarrow} \left\{\begin{array}{l} Z_1^k \\ Z_2^k \\ \cdots \\ Z_n^k \end{array}\right. .$$

Now we have to select the best matching configurations, taking into account symmetries. In case we have noisy data, the problem becomes to find the best matching positions for α_j, considering all sets of possible symmetries. Since data coming from stage one is already the output of a minimization process, the best matching positions for α_j may be identified with the lowest possible value of the TF. Note that the target function is modular, and can be split in the sum:

$$TF = \sum\nolimits_k TF^k = \sum\nolimits_{j,k} TF_j^k, \tag{A2}$$

where TF_j^k is the target function of α_j and D^k, and $TF^k = \sum_j TF_j^k$. For the sake of simplicity, let us suppose there are only two datasets present. There are arbitrary choices here, the strategy we selected is the following.

When two metals share the same location, in principle $t_j^1 = t_j^2$. With experimental data they may be different, however this is seldom a problem because

in practice the t_j^k are very well determined by the first stage. Then we can define T_j^1 by either t_j^1 or t_j^2, the one corresponding to the dataset with minimal target function value. We then choose a set of n possible definitions for A^{12} in the following way. Fix an index i. Being $t_i^1 \approx t_i^2$, an obvious choice to have coherent representations a_i^1 and a_i^2 is

$$A^{12}a_i^1 = a_i^2. \tag{A3}$$

The choice of the index i is arbitrary, so there are n possible choices for $A^{12} = a_i^2(a_i^1)^*$, where $(a_i^1)^*$ is the inverse of a_i^1. As a final step we define $A_j^1 = a_j^1$, $j = 1, \ldots, n$ to find a complete set of initial values. However, we must consider axial symmetries s_τ, $\tau = 1, \ldots, 3$, with respect to the τ coordinate axis of the metal system. Let s_0 be the identity. Any symmetric choice $A_j^1 = s_\tau a_j^1$ does not change the value of TF^1 in Equation A2 because s_τ is a symmetry applied to the metal system M^1. However this is not a symmetry for the metal system M^2, so the value of TF^2 changes. This accounts for the 4 possible choices for $A_j^1 = s_\tau a_j^1$, $\tau = 0, \ldots, 3$. We also consider the 4 symmetries with respect to the metal system M^2. We need to find A_j^1 such that, when plugged in Equation A1, the resulting transformation is a symmetry with respect to M^2, i.e.:

$$A^{12}A_j^1(x - T_j^1) = s_\tau a_j^2(x - T_j^1). \tag{A4}$$

Solving Equation A4 we find: $A_j^1 = (A^{12})^* s_\tau a_j^2$. In conclusion we get the following 8 choices for A_j^1:

$$\begin{cases} A_j^1 = s_\tau a_j^1, & \tau = 0, \ldots, 3 \\ A_j^1 = (A^{12})^* s_\tau a_j^2, & \tau = 0, \ldots, 3\,. \end{cases} \tag{A5}$$

Following this strategy, the program analyzes these 8 possible choices for the location of each α_j. The contributions of each α_j to TF are independent (see Equation A2), so the best configuration of the molecule is determined by joining the best locations of the α_j found separately. In this way, only $8n$ (instead of 8^n) different α-helix locations must be inspected to find the best configuration of the molecule for a fixed A^{12}. We have n possible choices for A^{12} in Equation A3, so $8n^2$ is the total number of comparisons needed to select the best match out of the set of $n \times 8^n$ initial configurations.

In the case of two metals in different locations, Equations A4 and A5 do not change. Of course in this case t_j^1 and t_j^2 are different. In this case we have the following definitions:

$$A^{12} = a_i^2(a_i^1)^*, \qquad T^{12} = a_i^1(t_i^2 - t_i^1), \quad i = 1, \ldots, n.$$

$$\begin{cases} A_j^1 = s_\tau a_j^1, & T_j^1 = t_j^1 \\ & \tau = 0, \ldots, 3, \\ A_j^1 = (A^{12})^* s_\tau a_j^2, & T_j^1 = t_j^2 - (s_\tau a_j^2)^* A^{12} T^{12} \\ & \tau = 0, \ldots, 3, \end{cases}$$

for the transformations involved in the selection strategy.

References

Al-Hashimi, H.M., Valafar, H., Terrell, M., Zartler, E.R., Eidsness, M.K. and Prestegard, J.H. (2000) *J. Magn. Reson.*, **143**, 402–406.

Arnesano, F., Banci, L., Bertini, I., Faraone-Mennella, J., Rosato, A., Barker, P.D. and Fersht, A.R. (1999) *Biochemistry*, **38**, 8657–8670.

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

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

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.

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

Bertini, I., Luchinat, C. and Parigi, G. (2001a) *Solution NMR of Paramagnetic Molecules*, Elsevier, Amsterdam.

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., Kowalewski, J., Luchinat, C. and Parigi, G. (2001d) *J. Magn. Reson.*, **152**, 103–108.

Bertini, I., Donaire, A., Jimenez, B., Luchinat, C., Parigi, G., Piccioli, M. and Poggi, L. (2001e) *J. Biomol. NMR*, **21**, 85–98.

Boisbouvier, J., Gans, P., Blackledge, M., Brutscher, B. and Marion, D. (1999) *J. Am. Chem. Soc.*, **121**, 7700–7701.

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

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.

Delaglio, F., Kontaxis, G. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 2142–2143.

Dosset, P., Hus, J.C., Marion, D. and Blackledge, M. (2001) *J. Biomol. NMR*, **20**, 223–231.

Farrar, T.C. and Quintero-Arcaya, R.A. (1985) *Chem. Phys. Lett.*, **122**, 41–45.

Farrar, T.C. and Quintero-Arcaya, R.A. (1987) *J. Phys. Chem.*, **91**, 3224–3228.

Fischer, M.W., Losonczi, J.A., Weaver, J.L. and Prestegard, J.H. (1999) *Biochemistry*, **38**, 9013–9022.
Fowler, B.A., Tian, F., Al-Hashimi, H.M. and Prestegard, J.H. (2000) *J. Mol. Biol.*, **304**, 447–460.
Ghose, R. and Prestegard, J.H. (1997) *J. Magn. Reson.*, **128**, 138–143.
Goldman, M. (1984) *J. Magn. Reson.*, **60**, 437–452.
Güntert, P. and Wüthrich, K. (1991) *J. Biomol. NMR*, **1**, 447–456.
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.
Hus, J.C., Marion, D. and Blackledge, M. (2001) *J. Am. Chem. Soc.*, **123**, 1541–1542.
Kemple, M.D., Ray, B.D., Lipkowitz, K.B., Prendergast, F.G. and Rao, B.D.N. (1988) *J. Am. Chem. Soc.*, **110**, 8275–8287.
Kretsinger, R.H. (1980) *CRC Crit. Rev. Biochem.*, **8**, 119–174.
Kurland, R.J. and McGarvey, B.R. (1970) *J. Magn. Reson.*, **2**, 286–301.
Linse, S., Brodin, P., Drakenberg, T., Thulin, E., Sellers, P., Elmden, K., Grundstrom, T. and Forsén, S. (1987) *Biochemistry*, **26**, 6723–6735.
Meiler, J., Blomberg, N., Nilges, M. and Griesinger, C. (2000) *J. Biomol. NMR*, **16**, 245–252.
Meiler, J., Prompers, J.J., Peti, W., Griesinger, C. and Bruschweiler, R. (2001) *J. Am. Chem. Soc.*, **123**, 6098–6107.
Mollova, E.T., Hansen, M.R. and Pardi, A. (2000) *J. Am. Chem. Soc.*, **122**, 11561–11562.
Ramirez, B.E. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 9106–9107.
Tjandra, N., Grzesiek, S. and Bax, A. (1996) *J. Am. Chem. Soc.*, **118**, 6264–6272.
Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 9279–9283.
Turner, D.L., Brennan, L., Chamberlin, S.G., Louro, R.O. and Xavier, A.V. (1998) *Eur. Biophys. J.*, **27**, 367–375.
Vogel, H.J., Drakenberg, T., Forsén, S., O'Neil, J.D. and Hofmann, T. (1985) *Biochemistry*, **24**, 3870–3876.
Vold, R.R. and Prosser, P.S. (1996) *J. Magn. Reson. Ser. B*, **113**, 267–271.