Uniqueness and degeneracy in the localization of rigid structural elements in paramagnetic proteins

2002 J. Phys. A: Math. Gen. 35 8153

(http://iopscience.iop.org/0305-4470/35/39/302)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **35** (2002) 8153–8169 PII: S0305-4470(02)33291-8

Uniqueness and degeneracy in the localization of rigid structural elements in paramagnetic proteins

Marco Longinetti[1], **Giacomo Parigi**[2] **and Luca Sgheri**[3]

[1] Dipartimento di Ingegneria Agraria e Forestale, Piazzale delle Cascine, 15, I-50144 Florence, Italy
[2] CERM and Department of Agricultural Biotechnology, University of Florence, p. le delle Cascine 28, I-50144 Florence, Italy
[3] Istituto per le Applicazioni del Calcolo (CNR), Via di S Marta 13/A, I-50139 Florence, Italy

E-mail: longinetti@unifi.it, parigi@cerm.unifi.it and luca@iaga.fi.cnr.it

Received 31 January 2002, in final form 11 July 2002
Published 17 September 2002
Online at stacks.iop.org/JPhysA/35/8153

Abstract
The uniqueness problem in the localization of some rigid structural elements is studied using constraints available for proteins containing paramagnetic metal ions. The degeneracy arising with a single set of data is investigated, and uniqueness is restored using multiple magnetic tensors. An efficient numerical strategy to deal with multiple datasets is presented.

PACS numbers: 87.14.Ee, 87.15.−v, 02.40.Dr
Mathematics Subject Classification: 92E10, 65K05, 51N20

1. Introduction

The availability of genomic data has created a need for rapid and efficient determination of the three-dimensional structures of the corresponding proteins. It is in fact commonly accepted that each protein has a unique fold. No theoretical method is presently available to obtain the fold from the knowledge of the sequence of amino acids which constitute the protein. It has been estimated that about one-third of all proteins contain at least one metal ion inside. A new class of NMR structural constraints can be obtained for proteins containing paramagnetic metal ions: the paramagnetism-based constraints [1]. They are: paramagnetic relaxation rates [2], contact shifts [2], pseudocontact shifts [3], self-orientation residual dipolar couplings [4] and cross correlations (between Curie relaxation and dipolar relaxation) [5]. In this paper we focus only on pseudocontact shifts and residual dipolar couplings, because among the paramagnetic constraints they provide the best source of information as structural constraints in terms of the number of data available and accuracy of measurements.

The *pseudocontact shifts* (PCS) δ_i^{pcs} arise in the presence of an anisotropic magnetic susceptibility tensor as the rotational average of the dipolar coupling between the magnetic moment of the unpaired electron(s) and the magnetic moment of the resonating nuclei (in the metal-centred point-dipole–point-dipole approximation) [2]. They depend on the magnetic susceptibility tensor χ and on the atomic coordinates according to the following equation [6]:

$$\delta_i^{pcs} = \frac{C_{pcs}}{r_i^5}\left[\left(\chi_{zz} - \frac{\mathrm{tr}(\chi)}{3}\right)\left(2z_i^2 - x_i^2 - y_i^2\right) + (\chi_{xx} - \chi_{yy})\left(x_i^2 - y_i^2\right) + 4\chi_{xy}x_i y_i + 4\chi_{xz}x_i z_i + 4\chi_{yz}y_i z_i\right] \tag{1.1}$$

where C_{pcs} is a constant,

$$\chi = \begin{pmatrix} \chi_{xx} & \chi_{xy} & \chi_{xz} \\ \chi_{xy} & \chi_{yy} & \chi_{yz} \\ \chi_{xz} & \chi_{yz} & \chi_{zz} \end{pmatrix}$$

is the *magnetic susceptibility tensor* of the metal ion, (x_i, y_i, z_i) are the differences between the coordinates of atom i and the coordinates of the metal ion, and $r_i = \sqrt{x_i^2 + y_i^2 + z_i^2}$.

The *residual dipolar couplings* (RDC) δ_{ab}^{rdc} are due to the induced partial orientation in high magnetic field caused by the anisotropy of the magnetic susceptibility tensor. This prevents the dipolar coupling energies from averaging to zero for all the pairs of atoms of the protein. They depend on the magnetic susceptibility tensor χ and on the atomic coordinates according to the following equation [4]:

$$\delta_{ab}^{rdc} = \frac{C_{rdc}}{r_{ab}^5}\left[\left(\chi_{zz} - \frac{\mathrm{tr}(\chi)}{3}\right)\left(2z_{ab}^2 - x_{ab}^2 - y_{ab}^2\right) + (\chi_{xx} - \chi_{yy})\left(x_{ab}^2 - y_{ab}^2\right) + 4\chi_{xy}x_{ab}y_{ab} + 4\chi_{xz}x_{ab}z_{ab} + 4\chi_{yz}y_{ab}z_{ab}\right] \tag{1.2}$$

where C_{rdc} is a constant, (x_{ab}, y_{ab}, z_{ab}) are the differences between the coordinates of selected pairs of atoms a and b, and $r_{ab} = \sqrt{x_{ab}^2 + y_{ab}^2 + z_{ab}^2}$. RDC are usually measured for the NH pairs and for the $C^\alpha H^\alpha$, $C^\alpha C^\beta$, $C^\alpha C$, CH^α pairs in ^{13}C-enriched samples.

For many metalloproteins it is possible to substitute the metal ion contained inside with a different one. Furthermore, some proteins contain two or more locations where a paramagnetic metal ion can be found. In these cases more than one set of PCS and RDC can be obtained, as different metal ions determine different paramagnetic susceptibility tensors [7, 8]. The removal of the metal ion present in the binding site may cause conformational modifications. These should be, however, limited by substituting the metal ion with a different one, having the same charge [9].

PCS and RDC can be used to determine the protein structure. The components of the tensor χ and the coordinates of protein atoms must thus be obtained by using equations (1.1) and (1.2) from the values of δ_i^{pcs} and δ_{ab}^{rdc}. This problem cannot be solved in general without further assumptions, because usually the number of unknowns is much larger than the number of data. A possible approach to reduce the number of unknowns is to consider protein *rigid structural elements*, or rigid fragments. By rigid structural elements we mean all protein fragments for which the structure is known. They can be (i) protein domains with three-dimensional structure already obtained in previous studies, in cases with conformational ambiguity due to the lack of NOE [10] between the domains of multidomain proteins, (ii) elements of the secondary structure (α-helix or β-sheet), (iii) the tetrahedrally arranged atoms centred on the C^α atom of single amino acids, or the CONH peptide planes. The idea of modelling the proteins in terms of rigid structural elements has recently been exploited in connection with the use of residual dipolar couplings induced by external anisotropic media [11–19]. These rigid structural elements may easily be modelled in an arbitrary reference frame from the knowledge of all atom chemical bonds and dihedral angles. Using this approach, the problem reduces to the determination of the relative position of rigid structural elements.

The problem of finding the relative position of several rigid structural elements through the use of paramagnetism-based constraints has been recently addressed within the frame of experimental structural genomic projects carried out at the Centre of Magnetic Resonance (CERM) of the University of Florence. The determination of the spatial positions of α-helices, considered as rigid structures, has been studied in [20] through the use of the following paramagnetic data: pseudocontact shifts, residual dipolar couplings and Curie dipole–dipole cross correlations.

In this paper we provide a detailed mathematical analysis of the problem of assembling rigid structural elements through the use of PCS and RDC. In section 2.1 the mathematical model is presented, and the uniqueness and degeneracy of the solution using data obtained in the presence of a single metal ion are analysed. In section 2.2 a complete mathematical proof is given of how uniqueness can be restored using data obtained in the presence of different metal ions substituted in the same or in different binding sites. As already mentioned, the idea of restoring the uniqueness using multiple metal ions has already been exploited in the cited literature. However, the precise conditions have not been fully stated, and a mathematical proof is missing. Section 2 of this paper fills this gap. We then give a mathematical basis for a numerical approach that has already been successfully implemented in [20] (section 3) and finally we show how this approach works using simulated data (section 4).

2. Mathematical analysis

2.1. Mathematical model and single metal problem

We denote the rigid structural elements by α_j, $j = 1, \ldots, n$. Let $x_{i,j}$ be the position vector of the ith atom of α_j in an arbitrary reference system (the lab frame). $X_j = \{x_{i,j}\}$ is the representation of α_j in the lab frame. The representations X_j, $j = 1, \ldots, n$ are known. The aim is to reconstruct the spatial positions of the α_j with respect to a metal ion M contained in the protein, but not belonging to any α_j. The *metal system* is a privileged Cartesian system with origin at M and axes coinciding with the principal directions of the tensor χ of M. It is easy to see from (1.1) and (1.2) that the values of δ_i^{pcs} and δ_{ab}^{rdc} do not depend on the trace of χ, but only on the five parameters defining the anisotropic part of χ, which is $\chi - (\mathrm{tr}(\chi)/3)I$. Therefore we assume $\mathrm{tr}(\chi) = 0$ because it cannot be determined from (1.1) and (1.2). The metal system is a good choice for representing the relative spatial positions of α_j. With this choice, χ is in diagonal form. The dependence of δ_i^{pcs} and δ_{ab}^{rdc} on χ is only in the magnetic susceptibility anisotropy coefficients [21]:

$$\Delta\chi_{ax} = \lambda_3 - \frac{\lambda_1 + \lambda_2}{2} \qquad \Delta\chi_{rh} = \lambda_1 - \lambda_2$$

where $\lambda_1, \lambda_2, \lambda_3$ are the eigenvalues of χ in increasing order. The representation of α_j in the metal system is found by applying a rigid motion R_j to X_j. R_j is the composition of a translation t_j and a rotation a_j, and $R_j(X_j) = a_j(X_j - t_j)$ is the location of α_j with respect to M. Any rotation matrix a_j may be represented by three Euler angles [22]. With a slight abuse we will call a_j both the 3×3 rotation matrix and the three Euler angles. The translation t_j is represented by a vector in $\mathbb{R}^3$, the location of the metal ion in the lab frame being t_j. The rigid motion R_j is then represented by (a_j, t_j), depending on six parameters. This rigid motion is all we need to reconstruct the spatial position of α_j with respect to the metal system. The only other values that are needed to fully reconstruct the tensor χ are the anisotropy coefficients $\Delta\chi_{ax}, \Delta\chi_{rh}$. Our unknowns are then the $6n + 2$ variables: (a_j, t_j), $j = 1, \ldots, n$

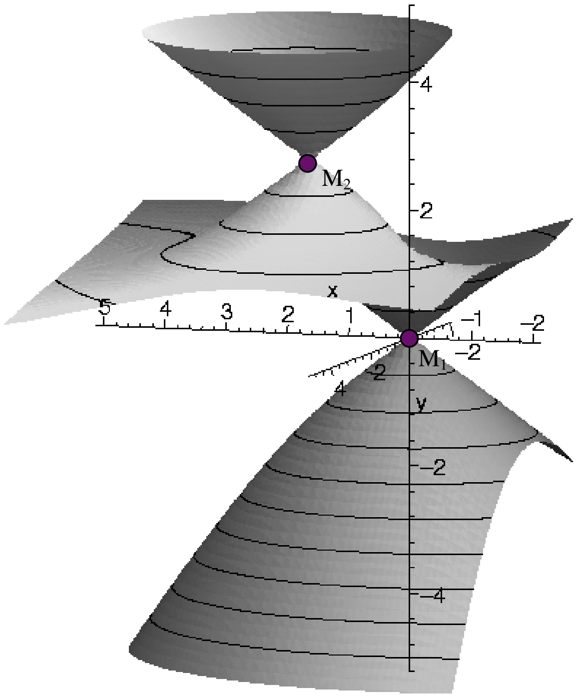

Figure 1. A manifold for which the uniqueness is lost. If all atoms (for which PCS is known) lie on this manifold, it is not possible to distinguish the two locations M_1 and M_2 for the position of the metal ion.

and $\Delta\chi_{ax}$, $\Delta\chi_{rh}$. We will call these variables (and also a set of values taken by these variables) a *configuration*.

On the other hand, the available measurements are values $\tilde{\delta}_i^{pcs}$ and $\tilde{\delta}_{ab}^{rdc}$. We suppose that the problem of assignment is already solved, i.e. we know to which atoms of the chemical structure of α_j these values correspond. As already mentioned in the introduction, there may be measurements relative to more than one metal (a different metal ion and/or a different binding site). We will call a *dataset* the measurements relative to both a single metal ion and a single location.

If a configuration is known, the α_j can be positioned in the metal system, and $\Delta\chi_{ax}$, $\Delta\chi_{rh}$ are known, therefore it is possible to compute δ_i^{pcs} and δ_{ab}^{rdc} from formulae (1.1) and (1.2). A *solution* is a configuration such that $\tilde{\delta}_i^{pcs} = \delta_i^{pcs}$ and $\tilde{\delta}_{ab}^{rdc} = \delta_{ab}^{rdc}$ for every measurement in the dataset.

The tensor χ and the translations t_j may be found by (1.1) and (1.2), if enough measurements $\tilde{\delta}_i^{pcs}$ and $\tilde{\delta}_{ab}^{rdc}$ are available. First we can determine χ from (1.2) solving a linear system because RDC do not depend on the position of the metal. Once χ is known, we can use (1.1) to uniquely determine t_j, unless the following situation occurs. Given any vector $\bar{t}_j \neq t_j$ we can explicitly determine a non-empty manifold in $\mathbb{R}^3$. If all atoms (for which $\tilde{\delta}_i^{pcs}$ is known) lie on this manifold, t_j and $\bar{t}_j$ are both solutions for the location of the metal ion. Uniqueness is then lost. Figure 1 is a picture of such a manifold. The appendix contains a proof in terms of elementary geometry that these manifolds are always non-empty.

We will always suppose that both χ and the translations t_j are uniquely determined. The problem of positioning the α_j is however not fully solved. In fact the tensor χ identifies the metal system only up to reflections of the coordinate axes. For each α_j an orientation choice has to be made. The first choice is arbitrary, and corresponds to a global choice in the orientation of the metal system. The subsequent choices influence, however, the relative positions of the α_j. In terms of transformations, suppose that any rigid motion (a_j, t_j) contained in a configuration is replaced by $(s_\tau a_j, t_j)$, where s_τ is a 180° rotation with respect to the τ coordinate axis, $\tau = 1, 2, 3$. In the metal system the rotations s_τ are represented by diagonal matrices with 1 in element $\tau\tau$, and -1 in the remaining diagonal elements, thus changing only the signs of two coordinates. It follows that the values computed from (1.1) and (1.2) do not change, because (1.1) and (1.2) contain only the squares of the coordinates when χ is in diagonal form. It is convenient to include also the identity s_0. We call these rotations s_τ, $\tau = 0, \ldots, 3$ *axial symmetries* because they are also symmetries with respect to the τ axis. We only consider the reflections that are rotations, otherwise a mirror copy of α_j is obtained. The mirror copy has opposite chirality, so it cannot be accepted. For each α_j, $j = 1, \ldots, n$, there are therefore four possible choices which can be combined to obtain 4^n possibilities. They reduce to 4^{n-1} different combinations because the global choice of orientation is arbitrary.

2.2. Multiple metals problem

Data from different metal ions can be available [8]. This suggests considering multiple datasets to find a unique configuration. Let D^k, $k = 1, \ldots, m$ be datasets, corresponding to metal ions M^k. Each D^k produces a solution $C^k = \left\{\left(a_j^k, t_j^k\right), \Delta\chi_{ax}^k, \Delta\chi_{rh}^k\right\}$ with the axial symmetry ambiguity seen above. However, these configurations must be assembled to find a global solution, so they must be *mutually consistent*. In other words, for each k_1, k_2 there must exist a rigid motion $R^{k_1k_2}$ such that $\left(a_j^{k_2}, t_j^{k_2}\right) = R^{k_1k_2}\left(a_j^{k_1}, t_j^{k_1}\right)$, $\forall j$. A *solution* $\Gamma = \{C^k, k = 1, \ldots, m\}$ of the multiple metals problem is then a set of mutually consistent solutions of the single metal problems.

Two global solutions may represent the same molecule. This is stated in the following definition.

Definition 2.1. *$\Gamma = \{C^k, k = 1, \ldots, m\}$ and $\tilde{\Gamma} = \{\tilde{C}^k, k = 1, \ldots, m\}$ are equivalent if:*

(i) $\Delta\chi_{ax}^k = \Delta\tilde{\chi}_{ax}^k$ *and* $\Delta\chi_{rh}^k = \Delta\tilde{\chi}_{rh}^k$, $k = 1, \ldots, m$.
(ii) $t_j^k = \tilde{t}_j^k$, $k = 1, \ldots, m$, $j = 1, \ldots, n$.
(iii) For every $k = 1, \ldots, m$ *there exists an axial symmetry* s^k *such that* $\tilde{a}_j^k = s^k a_j^k$, $\forall j$.

Remarks. Conditions (i) and (ii) imply that the anisotropy coefficients and the locations of the metal ions are correctly identified. Condition (iii) is necessary because of the reflection ambiguity of the tensors χ^k. For each k there is an arbitrary choice for the orientation of the tensor. This choice corresponds to the choice of s^k of condition (iii). The symmetry s^k does not depend on the rigid structural elements α_j because the two solutions represent the same physical molecule.

Definition 2.2. *Let M^1 and M^2 be two metal systems. We say that the metal systems are collinear if at least one of the lines identified by the coordinate axes of M^1 coincides with one of the lines identified by the coordinate axes of M^2.*

Remarks. The collinearity of M^1 and M^2 is not changed if an axial symmetry is applied to any of the metal systems (i.e. if a different orientation choice is made). This is why definition 2.2 is given in terms of lines and not of axes. The coincidence of the origin of the metal systems

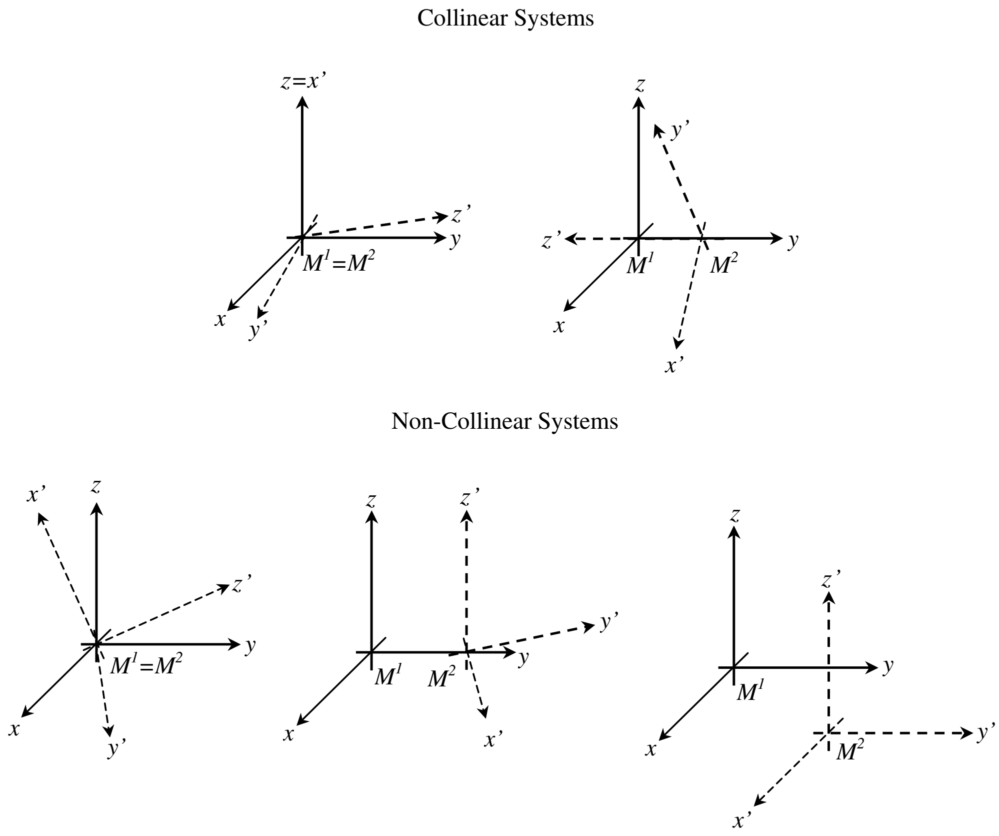

Figure 2. Collinear and non-collinear systems.

is not relevant for definition 2.2. The geometrical meaning is however slightly different (see figure 2 for examples):

(a) If M^1 and M^2 are in the same location, they are not collinear if and only if their tensors have no eigenvectors in common. This condition has already been stated in [23] for the RDC case.
(b) If M^1 and M^2 are in different locations, they are not collinear if and only if the distance vector M^1M^2 is not an eigenvector of both the metal tensors.

Theorem 2.1. *Let* Γ *be a solution of the multiple metals problem, with at least two non-collinear metal systems* M^{k_1} *and* M^{k_2}. *Then,* Γ *is unique up to equivalent sets of configurations.*

The proof of the theorem needs the following lemmas.

Lemma 2.1. *Let* s_τ *and* s_σ *be any two axial symmetries. Let* $T_{\tau\sigma}$ *be the* 16 *matrices with elements* $(T_{\tau\sigma})_{ij} = (s_\tau)_{ii}(s_\sigma)_{jj}$. *Then* $T_{\tau\sigma}$ *are represented by the following classification:*

(a) T_{00} *is the matrix with all elements equal to* 1.
(b) $T_{0\sigma}$, $\sigma \neq 0$ *is the matrix with elements* 1 *in column* σ, *and with elements* -1 *in the other columns.*
(c) $T_{\tau 0}$, $\tau \neq 0$ *is the matrix with elements* 1 *in row* τ, *and with elements* -1 *in the other rows.*

(d) *$T_{\tau\sigma}$, $\sigma \neq 0$, $\tau \neq 0$ is a matrix with* 1 *in element* (τ, σ), -1 *in the other elements of row* τ *and column* σ, *and* 1 *in the remaining elements.*

Proof. An axial symmetry s_τ is represented by a diagonal matrix with elements ± 1, and with an odd number of elements equal to +1. The proof follows by elementary exhaustive computation of all the 16 matrices $T_{\tau\sigma}$. □

Lemma 2.2. *Let* $\{e_1, e_2, e_3\}$ *be the canonical basis of* $\mathbb{R}^3$. *A rotation matrix R transforms* e_i *in* $\pm e_j$ *if and only if the matrix representing R has two elements equal to* 0 *in the same column or in the same row.*

Proof. Let r_{ij} be the elements of the matrix R. Then $Re_i = \pm e_j$ is equivalent to $r_{ij} = \pm 1$. The column and the row vectors of a rotation matrix have Euclidean norm 1, so $r_{ij} = \pm 1$ implies that the other elements of row i and column j are 0. □

Remark. The rotations connecting two tensors having an eigenvector in common are represented by matrices R having the previous property. The rotations not satisfying this property are called *oblique*. Definition 2.2 implies that the rotation connecting two non-collinear metal systems with the same origin is oblique.

Lemma 2.3. *Let R and* $\tilde{R}$ *be two given oblique rotations satisfying* $\tilde{R}s_\sigma = s_\tau R$. *Then* s_σ *and* s_τ *are uniquely determined.*

Proof. In terms of matrix elements the identity $\tilde{R} = s_\tau R s_\sigma$ is equivalent to

$$\tilde{r}_{ij} = r_{ij}(s_\sigma)_{ii}(s_\tau)_{jj} \qquad i, j = 1, \ldots, 3.$$

The matrix $T_{\tau\sigma}$ with elements $(T_{\tau\sigma})_{ij} = (s_\tau)_{ii}(s_\sigma)_{jj}$ is one of the 16 matrices of lemma 2.1. The element $(T_{\tau\sigma})_{ij}$ can be determined from the previous equation if and only if $r_{ij} \neq 0$. From lemma 2.2 it follows that the maximal number of vanishing elements in R is three, with no more than one element equal to 0 in the same column and in the same row. Therefore, all the elements of $T_{\tau\sigma}$ are determined except for the three undetermined elements corresponding to the zero elements of R. Let T be any of the $96 = 6 \times 16$ matrices with undetermined elements obtained by $T_{\tau\sigma}$. A complete classification of T may be derived from the classification of $T_{\tau\sigma}$ in lemma 2.1. The indices τ and σ can be identified by T, as shown by the following argument:

(A) If all the determined elements of T are equal to 1, then T agrees with a matrix (with undetermined elements) of case (a) in lemma 2.1, and it does not agree with cases (b)–(d). Hence $\tau = 0$ and $\sigma = 0$.

(B) If T has two columns containing elements -1 and the remaining elements are 1, then T agrees with a matrix (with undetermined elements) of case (b) in lemma 2.1. Any matrix coming from the remaining cases (a), (c) and (d) does not satisfy this property, so $\tau = 0$, and σ is the index of the column containing 1.

(C) If T has two rows containing elements -1 and the remaining elements are 1, then with a similar argument as in case (B), $\sigma = 0$, and τ is the index of the row containing 1.

(D) If the only elements of T equal to -1 fill a single row and a single column, except the crossing element which is 1 (if determined), then T agrees with case (d) of lemma 2.1. This property is not verified by matrices (with undetermined elements) coming from the other cases, so τ and σ are the indices of the crossing element.

The same argument applies of course if the number of vanishing elements of T is less than three. Then τ and σ, and so s_τ and s_σ, can be uniquely determined. $\square$

Proof of theorem. Let $\Gamma = \{C^k, k = 1, \ldots, m\}$ and $\tilde{\Gamma} = \{\tilde{C}^k, k = 1, \ldots, m\}$ be two solutions of the same multiple metals problem. We will show that Γ and $\tilde{\Gamma}$ are equivalent. Properties (i) and (ii) follow from the uniqueness of the magnetic tensor and location of the metal ion as shown in section 2.1. To prove (iii), assume for simplicity that $k_1 = 1$ and $k_2 = 2$. Let $C^1 = \left\{\left(s_{\tau_j} a_j^1, t_j^1\right), \Delta\chi_{ax}^1, \Delta\chi_{rh}^1\right\}$ and $\tilde{C}^1 = \left\{\left(\tilde{s}_{\tau_j} \tilde{a}_j^1, t_j^1\right), \Delta\chi_{ax}^1, \Delta\chi_{rh}^1\right\}$. Both C^1 and $\tilde{C}^1$ are solutions of D^1, hence

$$\tilde{a}_j^1 = s_j^1 a_j^1 \tag{2.1}$$

for suitable symmetries s_j^1. Similarly, by considering C^2 and $\tilde{C}^2$

$$\tilde{a}_j^2 = s_j^2 a_j^2. \tag{2.2}$$

Since C^1 and C^2 are mutually consistent, there exists R^{12} such that

$$\left(a_j^2, t_j^2\right) = R^{12}\left(a_j^1, t_j^1\right). \tag{2.3}$$

Analogously, there exists $\tilde{R}^{12}$ such that

$$\left(\tilde{a}_j^2, t_j^2\right) = \tilde{R}^{12}\left(\tilde{a}_j^1, t_j^1\right). \tag{2.4}$$

Suppose the location of M^1 and M^2 is the same. Then $t_j^1 = t_j^2$, so R^{12} and $\tilde{R}^{12}$ are rotations, and we will drop the translation part of the notation. Substituting (2.3) and (2.4) in (2.2) yields $\tilde{R}^{12}\tilde{a}_j^1 = s_j^2 R^{12} a_j^1$. Substituting (2.1) in the previous formula gives $\tilde{R}^{12} s_j^1 a_j^1 = s_j^2 R^{12} a_j^1$, and so $\tilde{R}^{12} s_j^1 = s_j^2 R^{12}$. The rotation matrices R^{12} and $\tilde{R}^{12}$ are oblique because M^1 and M^2 are not collinear. They do not depend on j, so from lemma 2.3 $s_j^1 = s^1$ and $s_j^2 = s^2$ are the uniquely determined symmetries that fulfil property (iii) because of (2.1) and (2.2). This concludes the proof of theorem 2.1 in the case the location of M^1 and M^2 is the same.

Suppose the location of M^1 is different from that of M^2. Then $R^{12} = (A^{12}, T^{12})$ and $\tilde{R}^{12} = (\tilde{A}^{12}, \tilde{T}^{12})$. From (2.3) and (2.4) we have

$$\begin{aligned} a_j^2\left(x - t_j^2\right) &= A^{12}\left(a_j^1\left(x - t_j^1\right) - T^{12}\right) \\ \tilde{a}_j^2\left(x - t_j^2\right) &= \tilde{A}^{12}\left(\tilde{a}_j^1\left(x - t_j^1\right) - \tilde{T}^{12}\right) \qquad \forall x \in \mathbb{R}^3. \end{aligned}$$

From the previous equation, we obtain

$$A^{12} = a_j^2\left(a_j^1\right)^* \qquad \tilde{A}^{12} = \tilde{a}_j^2\left(\tilde{a}_j^1\right)^* \tag{2.5}$$

where an asterisk denotes the inverse of a rotation. Then

$$T^{12} = a_j^1\left(t_j^2 - t_j^1\right) \qquad \tilde{T}^{12} = \tilde{a}_j^1\left(t_j^2 - t_j^1\right).$$

Applying (2.1) we get

$$\tilde{T}^{12} = s_j^1 T^{12}. \tag{2.6}$$

The distance vector T^{12} represents the coordinates of M^2 in the metal system M^1. Since M^1 is not collinear with M^2, T^{12} is not an eigenvector of both tensors. Switching M^1 and M^2 if necessary, we can assume that T^{12} is not an eigenvector of the tensor of M^1. It follows that T^{12} is a vector with at least two non-vanishing components. This suffices to uniquely identify from (2.6) the axial symmetry s_j^1. Thus $s^1 = s_j^1$ does not depend on j. Eliminating $a_j^1, a_j^2, \tilde{a}_j^1, \tilde{a}_j^2$ from (2.1), (2.2), (2.5) we get $\tilde{A}^{12} s_j^1 = s_j^2 A^{12}$. Since s_j^1 does not depend on j, the same holds for $s_j^2 = s^2$. Then s^1 and s^2 are the uniquely determined symmetries that fulfil property (iii) because of (2.1) and (2.2). This concludes the proof of theorem 2.1. $\square$

3. Numerical approach

3.1. Introduction

Here we describe a strategy that has been successfully used in [20] to reconstruct the positions of the α_j with respect to the metal ions using experimental data. This has been done by minimizing a *target function* (TF) having the following expression:

$$TF = \sum_i \left(\delta_i^{pcs} - \tilde{\delta}_i^{pcs}\right)^2 + \sum_{ab} \left(\delta_{ab}^{rdc} - \tilde{\delta}_{ab}^{rdc}\right)^2. \tag{3.1}$$

This formula is only a model; the actual formula used in computations is more complicated, including filters and multiple-level normalization terms. The target function is modular, and can be split into the sum:

$$TF = \sum_k TF^k = \sum_{j,k} TF_j^k \tag{3.2}$$

where TF_j^k is the target function relative to α_j and the dataset D^k, and $TF^k = \sum_j TF_j^k$. We have already shown in section 2.2 that if the measurements are exact, there is a unique solution Γ up to equivalent sets of configurations. By continuity, if the experimental error is small enough, i.e. $\tilde{\delta}_i^{pcs}$ and $\tilde{\delta}_{ab}^{rdc}$ are close enough to δ_i^{pcs} and δ_{ab}^{rdc}, the point $\tilde{\Gamma}$ where TF reaches its absolute minimum is close to Γ. The question of how large the experimental error can be while preserving the fact that $\tilde{\Gamma} \sim \Gamma$ is not trivial. It will be addressed in the remark following proposition 2.3.

3.2. Numerical solution of the multiple metals problem

For each dataset D^k, a numerical solution may be determined using any standard minimization technique. This configuration, giving the minimum of the TF, is represented by rigid motions $\left(a_j^k, t_j^k\right)$, $j = 1, \dots, n$ which bring the α_j from the lab system to the M^k metal system, and by two values $\Delta\chi_{ax}^k$ and $\Delta\chi_{rh}^k$. The critical point of this approach is joining these configurations in a single setting, and then determining the best match to be used as a starting point for a local minimization technique. A possible strategy is introduced in this section, its properties are described in section 3.3. In the following we will use lower case letters to denote the rotations and translations found with a minimization relative to a single metal, and capital letters to denote the transformations needed to implement the multiple metals setting.

3.2.1. Multiple metals setting. We have used the following variables to describe a set of configurations, as presented by the following diagram:

$$\left.\begin{array}{l} \alpha_1 \xrightarrow{(A_1^1, T_1^1)} \\ \alpha_2 \xrightarrow{(A_2^1, T_2^1)} \\ \quad \cdots \\ \alpha_n \xrightarrow{(A_n^1, T_n^1)} \end{array}\right\} M^1 \left\{\begin{array}{l} \xrightarrow{(A^{12}, T^{12})} M^2 \\ \quad \cdots \\ \xrightarrow{(A^{1m}, T^{1m})} M^m. \end{array}\right.$$

Let D^k, $k = 1, \dots, m$ be the datasets involved. Fix an arbitrary D^1. Then the positions of α_j with respect to M^1 are defined by rigid motions $\left(a_j^1, t_j^1\right)$. Let $\left(A_j^1, T_j^1\right) = \left(a_j^1, t_j^1\right)$. Now consider any other dataset D^k. The relative positions of the α_j are already defined by $\left(A_j^1, T_j^1\right)$, and cannot be changed since only one correct spatial arrangement exists. The metal tensor M^k must be determined, since it does not depend on M^1. The tensor M^k may be represented by means of a rigid motion (A^{1k}, T^{1k}) bringing the metal system M^1 to 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

$$A^{1k}\big(A_j^1(x - T_j^1) - T^{1k}\big) \tag{3.3}$$

in the system M^k (see the diagram). Formula (3.3) holds even for $k = 1$ by defining A^{11} as the identity matrix and T^{11} as the zero vector. Two or more metal ions may be substituted in the same location (this is an *a priori* piece of information), so there may be some constraints on the T^{1k}. This suggests the following ordering of the D^k. We can suppose that the D^k relative to a single location are grouped together, so we can define $G_l = \{D^{l_1}, D^{l_1+1}, \ldots, D^{l_2}\}$. G_l is the group of datasets relative to a single location index l, $l = 1, \ldots, L$, $1 \leqslant L \leqslant m$. For each dataset D^k, $k > 1$, there is a rotation A^{1k}. For each group G_l, $l > 1$, there is a different translation, though for consistency the T^{1k} will still be identified by the dataset index. For each $D^s \in G_l$, T^{1s} is the same vector. There are hence $6n + 2$ variables for the first dataset, and each successive dataset D^k adds five variables (a rotation and two coefficients) if M^k is in an already defined location, or eight variables (a rotation, a translation and two coefficients) if M^k is in a new position.

3.2.2. Selection of the best matching configurations. Suppose that the rigid motions obtained from the single metal minimizations are found with exact data. The corresponding configurations can be superimposed if and only if, for each α_j and for each dataset, a consistent choice of the symmetries s_τ is selected, as shown by the proof of theorem 2.1. In the case of noisy data, the problem is to find the best matching positions for α_j, considering all possible symmetries. There are arbitrary choices here; the strategy we select is the following.

(i) *Selection of the t_j^k.* The first step consists in selecting a subgroup of the t_j^k. Let $D^{i_1}, D^{i_2} \in G_l$. Then in principle $t_j^{i_1} = t_j^{i_2}$, since the datasets are relative to the same location. 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 single metal minimization. Let $D^k \in G_l$. Then we substitute t_j^k with $t_j^{i_l}$ where i_l is the index relative to the $D^s \in G_l$ having minimal target function value.

(ii) *Definition of the transformations (A^{1k}, T^{1k}).* For each $k > 1$, we choose n definitions for (A^{1k}, T^{1k}) in the following way. Fix an index i. Imposing the consistency of the configurations on the single α_i and applying (3.3), we get

$$A^{1k}\big(a_i^1(x - t_i^1) - T^{1k}\big) = a_i^k\big(x - t_i^k\big)\,. \tag{3.4}$$

Solving (3.4) gives

$$A^{1k} = a_i^k\big(a_i^1\big)^* \qquad T^{1k} = a_i^1\big(t_i^k - t_i^1\big) \tag{3.5}$$

where $\big(a_i^1\big)^*$ is the inverse of a_i^1. The choice of the index i is arbitrary, so there are n possible choices for (A^{1k}, T^{1k}).

(iii) *Definitions of A_j^1, T_j^1.* As a final step we have to define A_j^1, T_j^1, $j = 1, \ldots, n$ to find a complete set of initial values. Note that, due to step (i), the t_j^1 may be different from the values coming from stage one. We must now consider symmetries, because by section 2.1 the a_j^k are defined up to symmetries. Fix an α_j and consider $s_\tau a_j^1$. This does not change TF^1, but it changes TF^k, $k > 1$. Replacing a_j^1 with $s_\tau a_j^1$ in (3.4) does not preserve the equality sign, because the rotations $A^{1k} s_\tau a_j^1$ and $s_\tau a_j^k$ do not coincide even when $A^{1k} = a_i^k\big(a_i^1\big)^*$ is defined with $i = j$. Equating the two rotations would imply $a_j^k(a_j^1)^* s_\tau = s_\tau a_j^k(a_j^1)^*$ but in general the symmetry s_τ does not commute with the rotation $a_j^k(a_j^1)^*$ unless $\tau = 0$. This accounts for four possible choices for $A_j^1 = s_\tau a_j^1$, $\tau = 0, \ldots, 3$. We also consider the four symmetries with respect to D^k. We need to find A_j^1 such that, when substituted in (3.4), the resulting transformation is a symmetry with respect to M^k, i.e.

$$A^{1k}\big(A_j^1\big(x - T_j^1\big) - T^{1k}\big) = s_\tau a_j^k\big(x - t_j^k\big). \tag{3.6}$$

Solving (3.6) we find $A_j^1 = (A^{1k})^* s_\tau a_j^k$, $T_j^1 = t_j^k - \big(s_\tau a_j^k\big)^* A^{1k} T^{1k}$. We thus consider the following eight choices for A_j^1, T_j^1:

$$\begin{cases} A_j^1 = s_\tau a_j^1 & T_j^1 = t_j^1 & \tau = 0, \ldots, 3 \\ A_j^1 = (A^{1k})^* s_\tau a_j^k & T_j^1 = t_j^k - \big(s_\tau a_j^k\big)^* A^{1k} T^{1k} & \tau = 0, \ldots, 3. \end{cases} \tag{3.7}$$

3.3. Mathematical properties of the merging strategy

In this section some mathematical properties of the numerical approach are presented. The first two propositions justify the arbitrary choices made in the merging strategy. The rather technical proposition 2.3 has a practical consequence explained in the final remark of this section. For simplicity, we will assume in the following that the metal ions are in the same location, that is $T^{1k} = 0$.

Let us define $P_\tau^k = \big\{\big(s_{\tau_1} a_i^k\big)\big(s_{\tau_2} a_i^1\big)^*, i = 1, \ldots, n, \tau_1, \tau_2 = 0, \ldots, 3\big\}$ as the set of transformations from M^1 to M^k which can be found by combining the rotations of stage one with symmetries. Let $P^k = \big\{a_i^k\big(a_i^1\big)^*, i = 1, \ldots, n\big\}$ be the subset of P_τ^k, which is found when symmetries are neglected.

Proposition 3.1. *For each choice of A^{1k} in P_τ^k there is a suitable choice of A^{1k} in P^k giving the same value of TF.*

Proof. For simplicity we drop the subscript from s_τ and merge any adjacent group of symmetries. Let $A_j^1 = s a_j^1$ as in (3.7*a*) and let $A^{1k} = s a_i^k\big(s a_i^1\big)^*$. Then TF^1 is evaluated on $s a_j^1\big(x - t_j^1\big)$ and TF^k on $s a_i^k\big(s a_i^1\big)^* s a_j^1\big(x - t_j^1\big)$, i.e. on $s a_i^k\big(a_i^1\big)^* s a_j^1\big(x - t_j^1\big)$. Now take $A^{1k} = a_i^k\big(a_i^1\big)^* \in P^k$. With this choice TF^1 does not change, and TF^k is evaluated on $a_i^k\big(a_i^1\big)^* s a_j^1\big(x - t_j^1\big)$. The missing final symmetry does not change the value of TF^k, so the value of TF is the same. Now let $A_j^1 = (A^{1k})^* s a_j^k$ as in (3.7*b*), and let $A^{1k} = s a_i^k\big(s a_i^1\big)^*$. Then TF^1 is evaluated on $s a_i^1\big(a_i^k\big)^* s a_j^k\big(x - t_j^1\big)$ and TF^k on $A^{1k}(A^{1k})^* s a_j^k = s a_j^k$. Choosing $A^{1k} = a_i^k\big(a_i^1\big)^* \in P^k$ does not change TF^k. In fact, TF^1 is evaluated on $a_i^1\big(a_i^k\big)^* s a_j^k\big(x - t_j^1\big)$ and again the missing final symmetry does not change the value. □

The previous proposition justifies the use of P^k in (3.5), thus reducing the possible choices for the transformations A^{1k}.

The first dataset D^1 in our strategy has the role of the base dataset. The choice of D^1, however, does not affect the values of the *TF*, as shown by proposition 3.2. Let D^1 and D^2 be two datasets, and suppose B_j^2, B^{21} are the transformations defined taking D^2 as the base dataset. More precisely

$$B^{21} = a_i^1\big(a_i^2\big)^* \qquad i = 1, \ldots, n$$

and

$$\begin{cases} B_j^2 = s_\tau a_j^2 & \tau = 0, \ldots, 3 \\ B_j^2 = (B^{21})^* s_\tau a_j^1 & \tau = 0, \ldots, 3. \end{cases} \tag{3.8}$$

Proposition 3.2. *For each choice of A_j^1 and A^{12} there is a suitable choice of B_j^2 and B^{21} giving the same value of TF.*

Proof. Let $A_j^1 = sa_j^1$ as in (3.7*a*) and $A^{12} = a_i^2(a_i^1)^*$. Then TF^1 is evaluated on $sa_j^1(x - t_j^1)$ and TF^2 on $a_i^2(a_i^1)^* sa_j^1(x - t_j^1)$. Take $B^{21} = a_i^1(a_i^2)^*$ and $B_j^2 = (B^{21})^* sa_j^1$ as in (3.8*b*). Then TF^1 is evaluated on $B^{21}(B^{21})^* sa_j^1(x - t_j^1) = sa_j^1(x - t_j^1)$ and TF^2 on $B_j^2(x - t_j^1) = a_i^2(a_i^1)^* sa_j^1(x - t_j^1)$, so TF is the same. In the same way if A_j^1 is drawn from (3.7*b*) B_j^2 should be defined as in (3.8*a*). □

In the presence of exact data, there is always a choice of s_τ in (3.7*a*) and one in (3.7*b*) such that $TF = 0$, as shown by the following proposition.

Proposition 3.3. *Suppose $TF_j^k = 0$ on $sa_j^k(x - t_j^k)$. Then both when A_j^1 is defined as in (3.7a) and as in (3.7b) there is a point where $TF = 0$.*

Proof. The fact that $TF_j^k = 0$ on $sa_j^k(x - t_j^k)$ implies that each α_j is correctly positioned (up to a symmetry) both by D^1 and by D^2. This means that $t_j^1 = t_j^2$ and there exists a single rotation R, up to symmetries, bringing the correctly positioned structural elements in the system M^1 to the correctly positioned structural elements in the system M^2. Then $sa_j^2 = Rsa_j^1 \ \forall j$, so that $R = sa_j^2(sa_j^1)^*$ and $R \in P_\tau^2$. In other words, $A^{12} = a_i^2(a_i^1)^*$ (again up to symmetries) does not depend on the particular index i chosen. Fix an index j and take A_j^1 as in (3.7*a*). Then by hypothesis, $\forall s_\tau \ TF_j^1 = 0$ on $s_\tau a_j^1(x - t_j^1)$. We can replace $A^{12} = a_i^2(a_i^1)^*$ with $sa_j^2(sa_j^1)^*$. This means that TF_j^2 is evaluated on

$$sa_j^2(a_j^1)^* s^* s_\tau a_j^1(x - t_j^1). \tag{3.9}$$

The symmetry s^* depends on j, but we can choose s_τ to be s^* so that $s^* s_\tau$ cancels and (3.9) reduces to $sa_j^2(x - t_j^1) = sa_j^2(x - t_j^2)$ and here, by hypothesis, $TF_j^2 = 0$. So far we have shown that for a fixed j when $A_j^1 = s_\tau a_j^1$ there is a choice of the symmetry s_τ such that $TF_j^1 + TF_j^2 = 0$. This argument can be repeated for all the α_j since s_τ can be chosen independently for each j, thus obtaining a set of symmetries for which $TF = 0$. A similar proof can be carried out when A_j^1 is as in (3.7*b*). In this case, for each choice of s_τ we have that TF_j^2 is evaluated on $A^{12}(A^{12})^* s_\tau a_j^2(x - t_j^1) = s_\tau a_j^2(x - t_j^2)$ and hence is 0. By substituting A^{1k} with $sa_j^2(sa_j^1)^*$ we find that TF_j^1 is evaluated on $sa_j^1(a_j^2)^* s^* s_\tau a_j^2(x - t_j^2)$, so there is again a choice of s_τ such that $TF_j^2 = 0$. □

Remark. The previous proposition shows a coupling between the correct symmetry of (3.7*a*) and the correct symmetry of (3.7*b*). If the experimental error is small, by continuity arguments, it is possible to trace the correct symmetry as that having the smallest *TF* value. Because of the coupling, this is true both for (3.7*a*) and (3.7*b*). Tracing this agreement is an independent way of checking whether the experimental error may allow a reliable reconstruction of the positions of the protein structural elements. We develop this argument in the example of section 4.

4. An example

To test the efficiency of the numerical procedure introduced in the previous section on a synthetic model, a Fortran program has been developed [20]. Three rigid protein structural elements (or substructures) are assumed to be known. To show that the program works

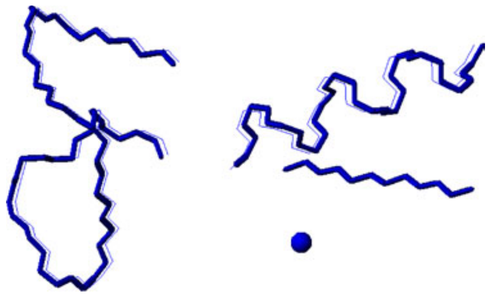

Figure 3. True model (thick line) and reconstruction (thin line) of three substructures. The dot represents the position of the metal ion.

independently on the type and on the length, we have selected an α-helix (with 12 amino acids), a β-sheet (with four amino acids) and a coil fragment of 19 amino acids. The three substructures and a metal ion were placed in a suitable way as reported in figure 3.

The datasets included PCS data for N, NH, C, C^α, H^α, C^β (if applicable), and RDC data for the N–NH and C^α–H^α couples, using two paramagnetic metal ions (Dy^{3+} and Yb^{3+}) substituted in the same binding site with relative magnetic tensor orientation as experimentally found in [8]. The magnetic susceptibility anisotropies were $\Delta\chi_{ax} = 34.7 \times 10^{-32}$ m^3 and $\Delta\chi_{rh} = -20.3 \times 10^{-32}$ m^3 for Dy^{3+}, and $\Delta\chi_{ax} = 8.26 \times 10^{-32}$ m^3 and $\Delta\chi_{rh} = -5.84 \times 10^{-32}$ m^3 for Yb^{3+}. The datasets were generated by using (1.1) and (1.2) and the atom coordinates of the known substructures.

With exact data the program succeeded in providing the correct tensors and relative substructure orientations perfectly. We simulated an experimental error by adding an absolute and a relative component, with uniform distribution. For PCS we used an error of ± 0.5 ppm $\pm 10\%$ and for RDC an error of ± 0.5 Hz, which are reasonable estimates of experimental measurement errors. To analyse the quality of the reconstruction, we compared the smallest target function solution found by the program and the model used to generate the datasets. We got a root-mean-square deviation (RMSD) for the backbone of 0.4 Å. We then raised the error to ± 0.8 ppm $\pm 16\%$ for PCS and ± 0.8 Hz for RDC, still getting a good reconstruction with an RMSD of 0.4 Å. This shows the stability of the solution, thus making us confident that the program is efficient if the substructures are known exactly. With a larger error the program succeeded in reconstructing the position of the α-helix and of the coil, but not of the β-sheet. This should not be surprising, since the β-sheet is the substructure composed of the smallest number of residuals.

Of course there is no *a priori* way of predicting the accuracy of the final reconstruction. A tempting way is to analyse the agreement of data coming from single metal problems. This can be done during the selection of the best matching configuration. For each substructure, we considered the corresponding TF_j^k values obtained by using (3.7), $k = 1, 2$, $j = 1, 2, 3$. We defined ρ_j^k as the quotient of the best and the second best values of TF_j^k. Figure 4 shows a plot of the harmonic mean ρ_j of ρ_j^1 and ρ_j^2 (divided by the number of residuals) versus the experimental error. We point out that there is a jump in ρ_2 corresponding to wrong reconstruction of the β-sheet.

To test the behaviour of the program with non-exact substructure models, we added an error on the coordinates of the atoms, still using for PCS an error of ± 0.5 ppm $\pm 10\%$ and for RDC an error of ± 0.5 Hz. The atoms were moved randomly in a ball of radius 0.4 Å centred on the correct position. The 20 solutions with the smallest values of the TF were compared

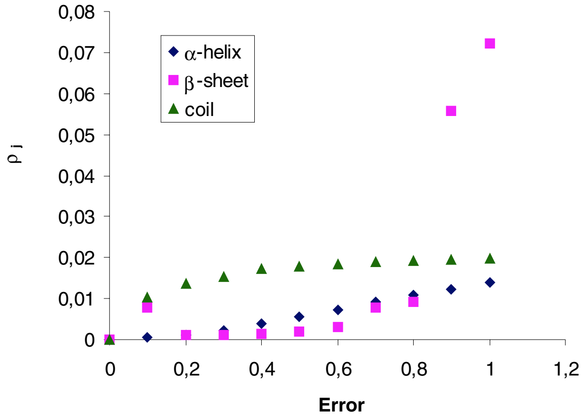

Figure 4. Test of the matching configurations.

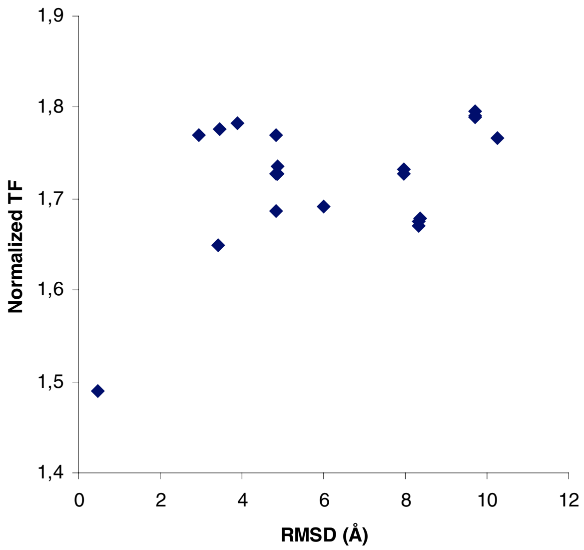

Figure 5. Normalized TF versus RMSD in the reconstructed solutions.

with the correct structure by calculating their RMSDs. The solution with the smallest TF has an RMSD equal to 0.5 Å and is shown in figure 3. Figure 5 reports the plot of the normalized TF as a function of the RMSD, showing that the solution with the smallest TF and smaller RMSD is well separated with respect to incorrect solutions with larger TF and RMSD. This is the upper limit for the error in the atom coordinates we managed to achieve in our example, still getting a good reconstruction.

Applications of the present approach to real proteins (cytochrome b_{562} and calbindin D_{9k}) have been performed and reported in [20]. In the latter, it was shown that the relative position of the four α-helices constituting each protein can be determined using theoretical α-helical models and data calculated from the real substructures.

5. Conclusions

This paper is aimed at providing a rigorous mathematical analysis of the commonly accepted idea that it is possible to reconstruct the position of rigid structural elements using paramagnetic data (PCS and RDC) only. We have shown that there are some conditions to be fulfilled to avoid degeneracy in the determination of the paramagnetic tensor and of the position of the metal ions. These conditions are in practice always verified; however, they may cause numerical instabilities close to these exceptional cases. We have then focused on the well-known symmetry problem in the determination of the principal axes of the paramagnetic tensor. We stated in theorem 2.1 the precise conditions to remove this degeneracy using multiple metal ions. When the tensors are almost collinear (a condition which is often met in practice), experimental errors cause problems in the numerical reconstruction. A strategy to overcome these difficulties has been presented in section 3, based on the best agreement amongst the multiple solutions of each single metal problem.

The length of the rigid structural elements is a key point to obtain a good reconstruction due to the presence of experimental errors, as shown by the example of section 4. We feel that this strategy may also be useful when the rigid structural elements are shortened to single amino acids, although problems may arise if their correct position is not found due to the small number of error-affected data.

Acknowledgments

We wish to thank Professors Ivano Bertini and Claudio Luchinat of CERM (University of Florence), and Professor Giorgio Talenti of the IAC Institute of the CNR for fruitful discussions.

Appendix. Analysis of the non-uniqueness situation for the location of the metal

Fix a metal ion M_1 at the origin and let $(\tilde{x}, \tilde{y}, \tilde{z})$ be a second location M_2 for the same metal ion. We assume that χ has already been obtained using RDC data. Then the anisotropy coefficients $\Delta\chi_{ax}$, $\Delta\chi_{rh}$ are known and we can use the diagonal form of the tensor. Let $r_1 = \sqrt{x^2 + y^2 + z^2}$ and $r_2 = \sqrt{(x - \tilde{x})^2 + (y - \tilde{y})^2 + (z - \tilde{z})^2}$ be the distances from the locations of the metal. Using (1.1), an atom in position (x, y, z) will give the same δ_i^{pcs} with respect to M_1 and M_2 if and only if (x, y, z) satisfy the following equation:

$$\frac{\Delta\chi_{ax}(2z^2 - x^2 - y^2) + \Delta\chi_{rh}(x^2 - y^2)}{r_1^5} = \frac{\Delta\chi_{ax}(2(z - \tilde{z})^2 - (x - \tilde{x})^2 - (y - \tilde{y})^2) + \Delta\chi_{rh}((x - \tilde{x})^2 - (y - \tilde{y})^2)}{r_2^5}. \tag{A1}$$

Equation (A1) represents a manifold in $\mathbb{R}^3$. If all atoms in α_j belong to this manifold, there is no uniqueness for the position of the metal. To prove that (A1) is not empty, let

$$k = r_1^2 / r_2^2 \qquad k \in [0, +\infty]. \tag{A2}$$

Each point in $\mathbb{R}^3$ defines a unique value k in the previous formula. Using (A2) in (A1) yields

$$k^{5/2}[\Delta\chi_{ax}(2z^2 - x^2 - y^2) + \Delta\chi_{rh}(x^2 - y^2)] = \Delta\chi_{ax}(2(z - \tilde{z})^2 - (x - \tilde{x})^2 - (y - \tilde{y})^2) + \Delta\chi_{rh}((x - \tilde{x})^2 - (y - \tilde{y})^2). \tag{A3}$$

In the case $k = 1$, (A3) is the plane of the points with equal distance from M_1 and M_2. If $k \neq 1$, (A3) is a quadric whose type depends on $\Delta\chi_{ax}$, $\Delta\chi_{rh}$. Rewriting (A2) and (A3) in normal form we get

$$\begin{cases} \left(x + \frac{k}{1-k}\tilde{x}\right)^2 + \left(y + \frac{k}{1-k}\tilde{y}\right)^2 + \left(z + \frac{k}{1-k}\tilde{z}\right)^2 = \frac{k}{(1-k)^2}(\tilde{x}^2 + \tilde{y}^2 + \tilde{z}^2) & \text{(A4}a\text{)} \\ c_x\left(x - \frac{\tilde{x}}{1-k^{5/2}}\right)^2 + c_y\left(y - \frac{\tilde{y}}{1-k^{5/2}}\right)^2 + c_z\left(z - \frac{\tilde{z}}{1-k^{5/2}}\right)^2 = \frac{k^{5/2}}{(1-k^{5/2})^2}(c_x\tilde{x}^2 + c_y\tilde{y}^2 + c_z\tilde{z}^2) & \text{(A4}b\text{)} \end{cases}$$

where

$$\begin{cases} c_x = \Delta\chi_{rh} - \Delta\chi_{ax} \\ c_y = -\Delta\chi_{rh} - \Delta\chi_{ax} \\ c_z = 2\Delta\chi_{ax}. \end{cases}$$

Since $\mathrm{tr}(\chi) = 0$ and λ_3 is the largest eigenvalue, $\Delta\chi_{ax} = 3\lambda_3/2 > 0$, so c_z is positive. Moreover at least one of c_x, c_y is negative, so (A4*b*) represents a family of hyperboloids. The number (one or two) of sheets depends on the sign of $\Delta = c_x\tilde{x}^2 + c_y\tilde{y}^2 + c_z\tilde{z}^2$. In the case $\Delta = 0$, the hyperboloids degenerate to a family of cones with vertex on the line l containing the vector M_1M_2, and with l on their surfaces. The parametric equations of l are

$$\begin{cases} x = t\tilde{x} \\ y = t\tilde{y} \\ z = t\tilde{z}. \end{cases}$$

The intersections of l and the surfaces defined by (A4*a*) are obtained for $t_a^{\pm} = \frac{k\pm\sqrt{k}}{1-k}$, and those of l and (A4*b*) for $t_b^{\pm} = \frac{1\pm\sqrt{k}}{1-k^{5/2}}$. We want to prove that, for a fixed k, (A4*a*) and (A4*b*) intersect in a curve. Suppose $k \in [0, 1]$. It is enough to show that $t_a^- < t_b^{\pm} < t_a^+$ for at least one of the two values of $t_b^{\pm}$ because (A4*a*) is compact and (A4*b*) is not. Solving for k we determine an interval $[\bar{k}, 1]$, $\bar{k} \sim 0.147$. In the case $k > 1$, by changing the roles of r_1 and r_2, we get the reciprocal interval $[1, 1/\bar{k}]$. Then if $k \in [\bar{k}, 1/\bar{k}]$ (A4*a*) and (A4*b*) intersect in a curve and every element of (A3) is a non-empty manifold.

References

[1] Bertini I, Luchinat C and Piccioli M 2001 Paramagnetic probes in metalloproteins. Turning limitations into advantages *Methods Enzymol.* **339** 314–40

[2] Bertini I, Luchinat C and Parigi G 2001 *Solution NMR of Paramagnetic Molecules* (Amsterdam: Elsevier)

[3] Banci L, Bertini I, Bren K L, Cremonini M A, Gray H B, Luchinat C and Turano P 1996 The use of pseudocontact shifts to refine solution structures of paramagnetic metalloproteins: Met80Ala cyano-cytochrome c as an example *J. Biol. Inorg. Chem.* **1** 117–26

[4] Tolman J R, Flanagan J M, Kennedy M A and Prestegard J H 1995 Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution *Proc. Natl Acad. Sci. USA* **92** 9279–83

[5] Boisbouvier J, Gans P, Blackledge M, Brutscher B and 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** 7700–1

[6] Kurland J R and McGarvey B R 1970 Isotropic NMR shifts in transition metal complexes: the calculation of the Fermi contact and pseudocontact terms *J. Magn. Reson.* **2** 286–301

[7] Allegrozzi M, Bertini I, Janik M B L, Lee Y-M, Liu G and Luchinat C 2000 Lanthanide induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 Å from the metal ion *J. Am. Chem. Soc.* **122** 4154–61

[8] Bertini I, Janik M B L, Lee Y-M, Luchinat C and Rosato A 2001 Magnetic susceptibility tensor anisotropies for a lanthanide ion series in a fixed protein matrix *J. Am. Chem. Soc.* **123** 4181–8

[9] Allegrozzi M, Bertini I, Janik M B L, Lee Y-M, Liu G and Luchinat C 2000 Lanthanide-induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 Å from the metal ion *J. Am. Chem. Soc.* **122** 4154–61

[10] De Alba E and Tjandra N 2002 NMR dipolar couplings for the structure determination of biopolymers in solution *Progr. NMR Spectrosc.* **40** 175–97

[11] Fischer M W, Losonczi J A, Weaver J L and Prestegard J H 1999 Domain orientation and dynamics in multidomain proteins from residual dipolar couplings *Biochemistry* **38** 9013–22

[12] Mollova E T, Hansen M R and Pardi A 2000 Global structure of RNA determined with residual dipolar couplings *J. Am. Chem. Soc.* **122** 11561–2

[13] Delaglio F, Kontaxis G and Bax A 2000 Protein structure determination using molecular fragment replacement and NMR dipolar couplings *J. Am. Chem. Soc.* **122** 2142–3

[14] Meiler J, Blomberg N, Nilges M and Griesinger C 2000 A new approach for applying residual dipolar couplings as restraints in structure elucidation *J. Biomol. NMR* **16** 245–52

[15] Fowler B A, Tian F, Al-Hashimi H M and Prestegard J H 2000 Rapid determination of protein folds using residual dipolar couplings *J. Mol. Biol.* **304** 447–60

[16] Dosset P, Hus J C, Marion D and Blackledge M 2001 A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings *J. Biomol. NMR* **20** 223–31

[17] Meiler J, Prompers J J, Peti W, Griesinger C and Bruschweiler R 2001 Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins *J. Am. Chem. Soc.* **123** 6098–107

[18] Tian F, Valafar H and Prestegard J H 2001 A dipolar coupling based strategy for simultaneous resonance assignment and structure determination of protein backbones *J. Am. Chem. Soc.* **123** 11791–6

[19] Hus J C, Marion D and Blackledge M 2001 Determination of protein backbone structure using only residual dipolar couplings *J. Am. Chem. Soc.* **123** 1541–2

[20] Bertini I, Longinetti M, Luchinat C, Parigi G and Sgheri L 2002 Efficiency of paramagnetic-based constraints to determine the spatial arrangement of α-helical secondary structure elements *J. Biomol. NMR* **22** 123–36

[21] Kemple M D, Ray B D, Lipkowitz K B, Prendergast F G and Rao B D N 1998 The use of lanthanides for solution structure determination of biomolecules by NMR: evaluation of methodology with EDTA derivatives as model systems *J. Am. Chem. Soc.* **110** 8275–87

[22] Landau L D and Lifschitz E M 1976 *Mechanics* (Oxford: Pergamon)

[23] Al-Hashimi H M, Valafar H, Terrel M, Zartler E R, Eidsness M K and Prestegard J H 2000 Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings *J. Magn. Reson.* **143** 402–6