J.Sci.Soc.Thailand, 24(1998) 101-122

STRUCTURE DETERMINATION OF PARAMAGNETIC PROTEINS THROUGH NMR

PORNTHEP SOMPORNPISUT

Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok, 10330, Thailand

(Received May 18, 1998)

ABSTRACT

Now it is possible to determine the three-dimensional structure of paramagnetic proteins in solution through the state-of-the-art NMR techniques. The paramagnetic effects enhance the nuclear relaxation rate giving rise to undetectable NMR signals, thus NMR study of paramagnetic systems has been ignored. Nevertheless, such systems have been continuously studied by some NMR communities to eventually show that in a favorable case, the problems can be overcome. The strategy to determine the structure of paramagnetic metalloproteins will be presented together with the highlights of data used for determining the three-dimensional structure of high-potential iron-sulfur proteins (HiPIPs) and ferricytochrome c. The strategy consists of the NMR conventional approach and an approach to paramagnetic systems. The quality of the structures determined is comparable to that of the diamagnetic proteins of the same size.

1. INTRODUCTION

The three-dimensional structure of a protein is exquisite and exploitable. It contains a wealth of information leading to the understanding of the relationship between the structure and its biological function. One of the most well known methods in three-dimensional structure determination of proteins is X-ray crystallography. However, the requirement to have the crystallized protein limits the application of X-ray crystallography. In addition, one might think that the crystal structure generally represents a conformation in the solid state. They are all static pictures, with little dynamic information. Instead, a protein is not rigid but extremely flexible. In order for a protein to function, the protein can move and rearrange in various conformations. Such fluctuations depend on time scales and thermodynamic factors which could be responsible for the protein activity. A possible strategy for protein investigation is to determine an average structure in solution first, and to analyze afterwards the protein dynamics.

NMR is a spectroscopic technique that allows us to study the structural and dynamic information of macromolecules of biological interest at atomic level. It is the only technique that is well established in the structure determination of proteins in solution. The advantage of the technique is that a more complete picture of the molecule can be obtained. The availability of high field NMR spectrometers associated with an improvement of the methodology nowadays allows us to achieve the three-dimensional solution structure of proteins up to 30000 MW with resolution comparable to the X-ray structure of the same protein.

The use of NMR for paramagnetic proteins is not as common as for diamagnetic proteins. A paramagnetic protein is characterized by a protein containing at least one metal ion (metalloprotein) with at least one unpaired electron. Unpaired electron(s) can interact with nuclei; that of the metal itself; those of the atoms covalently bonded with the metal; and those

of the atoms located around the metal. The metalloprotein refers to a protein with the metal ion incorporated with any non-amino acid groups, the so called "the prosthetic group". The prosthetic group plays a crucial role as the active center of the protein. For example, the iron-sulfur cluster in ferredoxin, or heme in cytochrome c are the active centers responsible for electron transport in the biological process[1]. A metalloprotein can exist in more than one electronic structure according to the oxidation state of the metal ion. Here, the term "paramagnetic" refers to only protons that have the significant short nuclear relaxation time and/or resonance shifted with respect to the usual proton. Protons experiencing the coupling of the unpaired electron give rise to large nuclear relaxation rates. One of the most relevant effects on resonating nuclei is that the couplings between the electronic and nuclear spins enhance the nuclear relaxation rate, and hence broaden NMR linewidth. As a result, the signal is not detectable. This influence seems to hamper applications of NMR in paramagnetic systems.

However, understanding how the unpaired electron interacts with the resonating nucleus is a challenge subject from the theoretical point of view. NMR is still a powerful tool for investigating this specific topic. During the last decade, extensive investigations have been made, including an effort in achieving the solution structures of paramagnetic proteins. The NMR community has demonstrated that for some favorable cases, the problem can be overcome. Eventually, the first solution structure of a paramagnetic protein was reported in 1994[2]. The quality of the solution structures of paramagnetic metalloproteins is comparable to that of diamagnetic proteins of the same size.

Here, the goal of this report is to give a conclusive theme of research in determining the solution structure of paramagnetic metalloproteins through NMR spectroscopy. This consists of the strategies used to obtain results, and briefly the results themselves. Extensive details are described by Wüthrich, Ernst, Bax and in Volumes 176, 177 and 239 of Methods in Enzymology which are excellent sources for information on protein NMR[3-6].

2. METHODOLOGY

2.1 Classical approach to the solution structure of protein *via* NMR

Wüthrich and coworkers have established a systematic method for determining a three-dimensional structure of proteins in solution. It usually relies on the comparison of COSY (Correlated SpectroscopY)[7-10] and TOCSY (Total Correlated SpectroscopY)[11-13] 2D spectra on one side, in which cross peaks originate from scalar connectivities between two nuclei mostly protons and NOESY (Nuclear Overhauser Effect SpectroscopY)[14,15] 2D spectra on the other side in which cross peaks arise from dipolar connectivities. The schematic representation of Fig. 1 summarizes the individual steps of a protein structure determination via NMR.

Non labile protons in each amino acid that constitutes the protein gives rise to a characteristic pattern on 2D NMR spectra called "the spin pattern or the spin systems"[3]. Not all spin patterns of the 20 common amino acids are unique but can be identified through COSY and TOCSY spectra recorded with the sample in deuterated solvent. The COSY and TOCSY cross peaks provide only intra-residue connectivities via the scalar couplings (*through-bond* connectivities). The next step is to perform the sequence-specific assignments by comparing COSY and TOCSY spectra with NOESY spectra recorded in aqueous solution to identify dipolar connectivities involved amide backbone protons ($H\alpha_i$-HN_{i+1}, HN_i-HN_{i+1}, Fig.2). In the sequential assignment process, cross peaks in NOESY spectra are identified based on the *through-space* connectivity between an amino acid being considered and the adjacent ones in the primary sequence. As an example, Figure 3 shows the fingerprint region (the region of a 2D

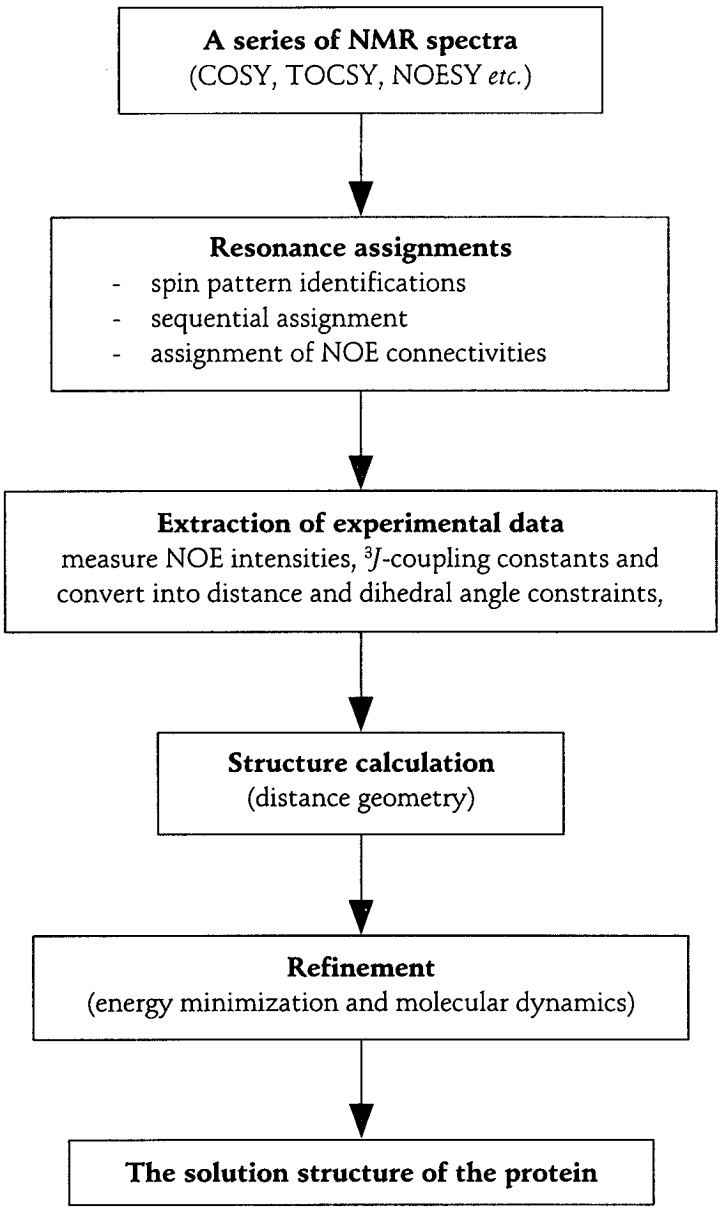


Fig.1 Flowchart illustrating the individual steps of the structure determination of a protein through NMR.

spectrum containing cross peaks between *NH* and *Hα* protons which extends from 7.8 to 9.2 ppm in one direction and from 4.8 to 3.7 ppm in the other direction) of a NOESY spectrum showing the sequential resonance assignments for a small segment of backbone protons of a protein[16].

Based on the analysis of ^{1}H NMR 2D spectra, this approach is suitable for proteins with a molecular weight less than 12 kDa[17]. To tackle larger proteins, the addition of new dimensions to experiments is typically carried out to increase the spectral resolution. Such experiments require at least one or two heteronuclei. In general, ^{13}C and ^{15}N are suitable for protein NMR, even if the natural abundance of these isotopes is very low (1.1% ^{13}C and 0.37% ^{15}N). The key to carrying out 3 or more dimensional NMR experiments is to enrich the protein with these heteronuclear isotopes. The technique called isotopically labelling technique has opened a new strategy to perform sequence specific assignment for proteins up to 30 kDa. Example of 3D heteronuclear experiments is ^{1}H-^{1}H-^{15}N 3D NOESY-HMQC (Heteronuclear Multiple Quantum Correlation)[18,19] spectra in which cross peaks represent a correlation between three nuclei: two protons via through-space connectivity, one of which is connected to a ^{15}N nucleus.

Although the process of the sequential resonance assignments allows us to establish connections between the ^{1}H spin pattern of an amino acid residue and its adjacent one throughout the protein sequence. The sequential connectivities alone are not enough to identify the secondary (α-helices and β-sheet) and tertiary structures of proteins. Therefore, further analysis is essential to obtain the whole protein structure. Fortunately, homonuclear ^{1}H-^{1}H NOE (1D NOE, 2D NOESY or 3D NOESY) experiments contain all these structure information. The basic idea behind these experiments is that if a pair of protons is close to each other, less than 5 Å apart in space, it will give rise to dipolar connectivity. Extensive structural studies reporting short ^{1}H-^{1}H distances in protein structures have demonstrated that the existence of the secondary structure elements gives rise to characteristic NOE patterns: inter-residue connectivities of $d_{NN(i,i+1)}$, $d_{\alpha N(i,i+3)}$ and $d_{\alpha\beta(i,i+3)}$ are observed in the presence of α-helices while those of $d_{\alpha N(i,i+1)}$ and of $d_{\alpha N(i,j)}$, $d_{NN(i,j)}$ and $d_{\alpha\alpha(i,j)}$ between two polypeptide chains are detected in β-sheets. Figure 4 shows the secondary structure elements identified through NOESY connectivities. As a consequence of the spatial folding of proteins providing plenty of short distances between non-neighboring inter-residual protons, this type of the dipolar "long range" connectivities is the most important information to determine the tertiary structure.

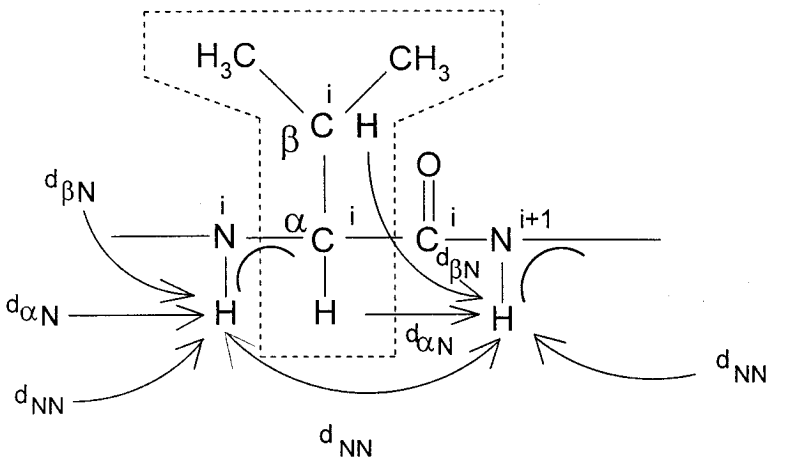


Fig.2 Polypeptide segment with indication of the spin patterns (inside dash lines), and the sequential NOE connecvitities (arrows).

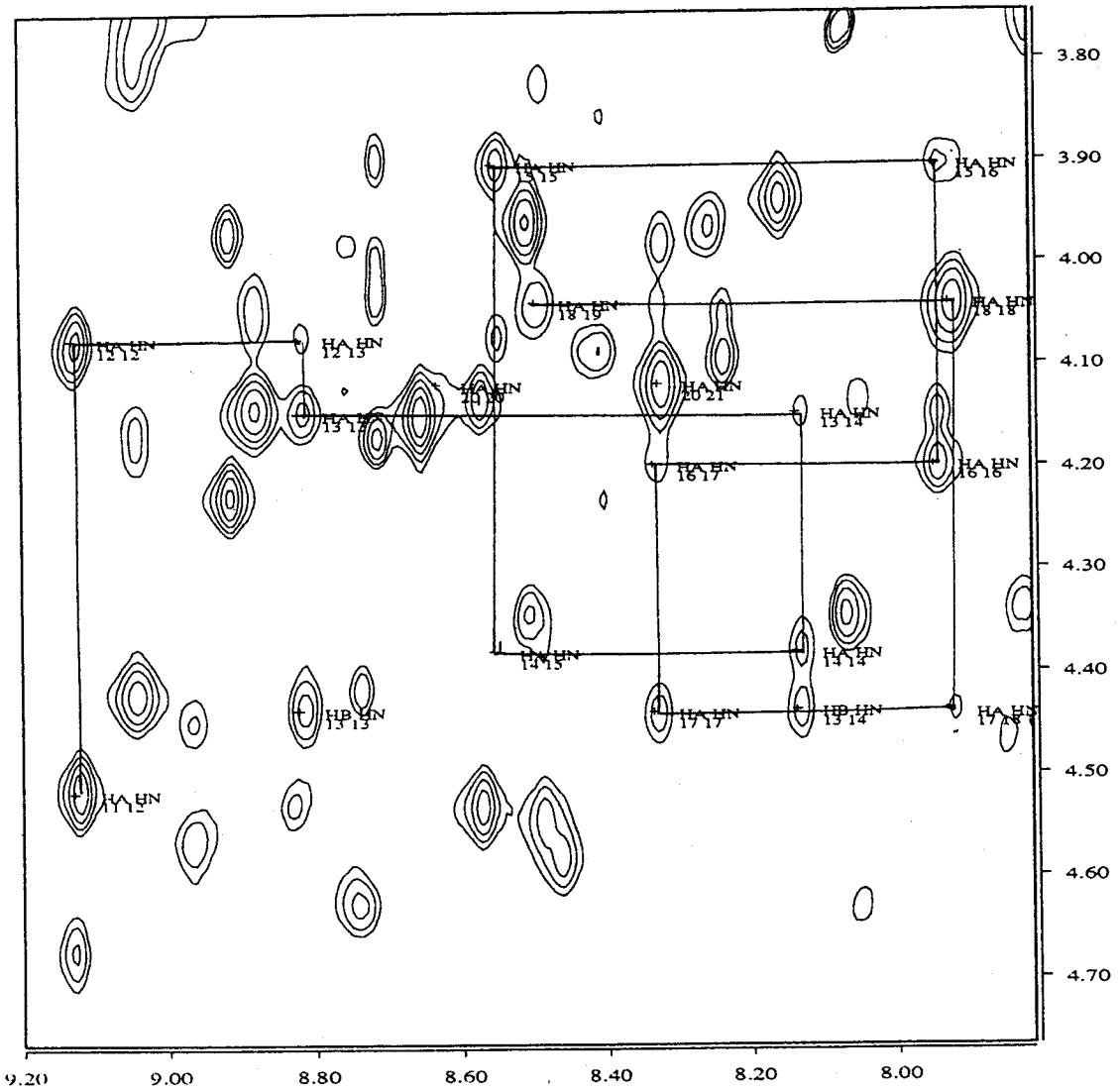


Fig.3 The sequential assignment for residues 11-19 observed from NOESY spectrum of the reduced HiPIP from *C. vinosum*, recorded at 300 K, 0.05 M phosphate buffer in aqueous solution (mixing time 100 ms).

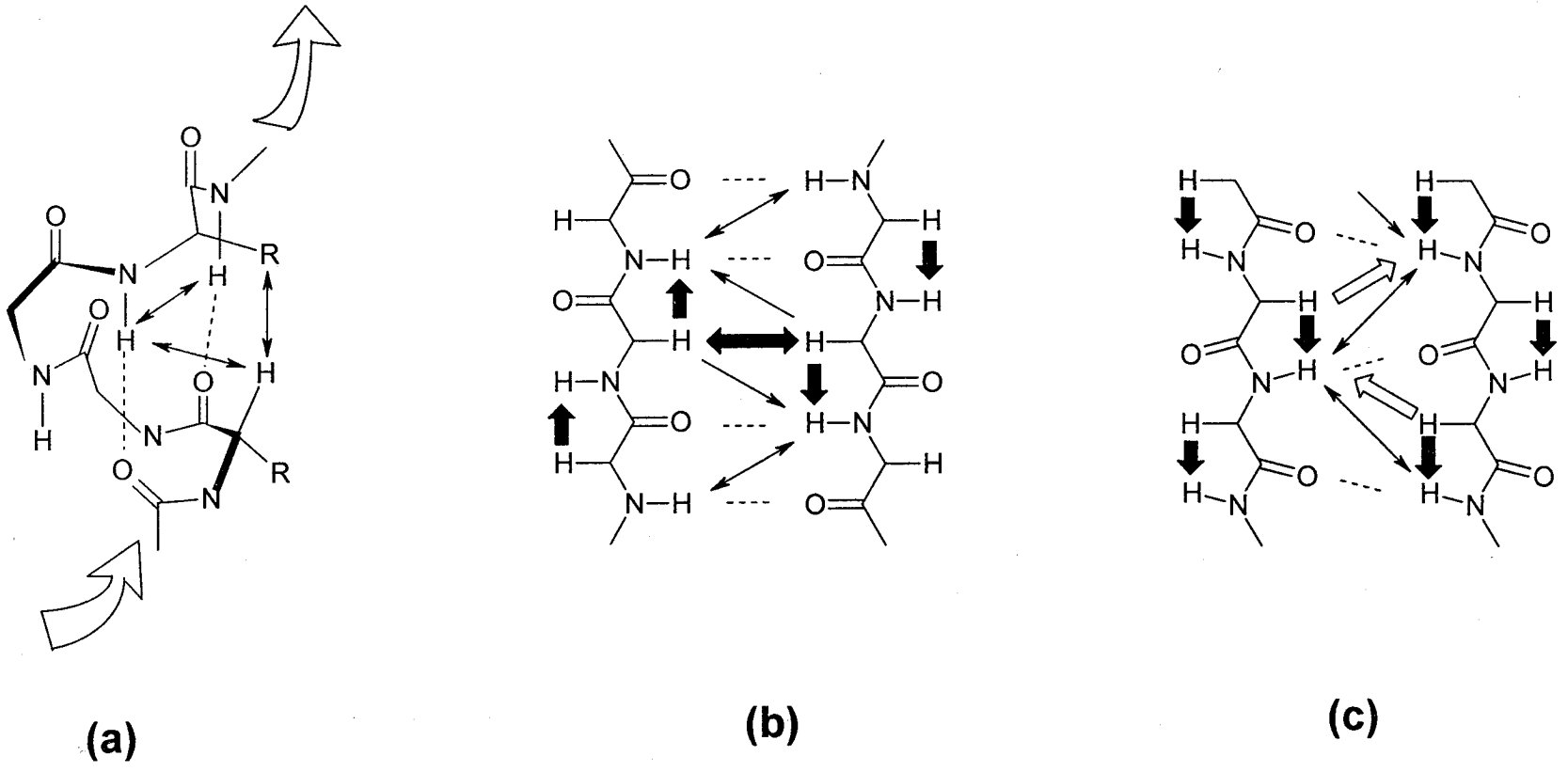


Fig.4 Short ^{1}H-^{1}H distance (arrows) in regular secondary structure found in proteins (a) α-helix, (b) antiparallel β-sheet, (c) parallel β-sheet.

2.1.1 How to obtain NMR constraints for structure calculation

The following step after the resonance assignments is an extraction of NMR data from the spectra. NOEs and the ^{3}J scalar coupling constants are the most common NMR constraints used in the structure calculation. The intensities of assigned NOESY cross peaks are measured and converted into inter-proton distances. How can the intensities of dipolar connectivities be converted into inter-proton distances? Several approaches have been proposed. One of the most successful approaches is a calibration of NOE intensity *versus* the corresponding upper distance limit in such a way that dipolar connectivities giving rise to NOE intensities are classified as follows:

i) intra-residue

ii) sequential

iii) medium range

iv) long range backbone

v) long range[20].

The calibration curve is derived from appropriate reference distances of protons at fixed distance (r) with the observed intensities of the corresponding cross peaks (V). The obtained correlation is then used to convert the intensity of NOESY cross peaks into estimates of distances constraints. The empirical form of the correlation is given as follows:

$$V = \frac{C}{r^n} \quad (1)$$

where C is a constant and the values of n are typically in the range from 4 to 6. Obviously, $n = 6$ for rigid segments in the protein. Exponent $n < 6$ reproduces the better relation between the intensities and distances for protons that involve the more mobile segments.

This is the easiest but also the less accurate approach. Nevertheless, the strength of NMR structure determination is based on the ability to increase the number of assigned NOEs. The adequate NOEs (approximately 1000-1500 NOE constraints for a protein of 100 residues) usually lead to the correct folding. More precise methods to calculate distance constraints are available (relaxation matrix analysis[21,22]) but will not be outlined here. Since their advantages with respect to the semiquantitative ones can be appreciated only from the theoretical point of view.

The second type of constraints obtained from COSY or NOESY spectra is the dihedral angle constraints (ϕ) derived from the $^3J_{HNH\alpha}$ scalar coupling constants. This geometrical relationship established by Karplus is defined as:

$$^3J_{HNHa} = A\cos^2\phi + B\cos(\phi - 60) + C \quad (2)$$

Where A, B and C are constants. In the structural calculation, each dihedral angle constraints holds a boundary range of the angle involving a scalar coupling of three bonds in proteins[23]. A successive strategy for taking an advantage of dihedral constraints during the structure calculation has been described earlier by Güntert[24].

A hydrogen bond can be likewise used as a constraint with some careful considerations as follows: first there must be a clear indication of a regular secondary structure in the protein, and if the slow exchange solvent-amide proton has been observed in D_2O and if in successive calculation the hydrogen bond involving that amide proton has been observed in more than 70% of the obtained structures.

In addition, the stereospecific or diastereotopic protons that can be assigned based on comparison between the model and NOESY spectra can be used as an experimental constraints during the stage of structure calculation. Such constraints can substantially improve the resolution of the structure.

2.1.2 Distance geometry calculations

The goal of the distance geometry calculation (DG) is to produce one or more molecular structures that satisfy a set of input constraints[25]. The distance geometry algorithm is in some way unique in that it does not need any initial approximate models. It can therefore be regarded as an unbiased method of determining structure. In a DG program such as DIANA[20], the initial calculation is to build molecular structures (defined by a set of random numbers that are associated with protein conformation) which are completely denatured or similar to a polypeptide chain of the primary sequence. Then a mathematical function called the target function, that is a sum of the deviation between the experimental data and the value measured in the calculated structure is minimized in such a way that bond lengths and bond angles are kept fixed and each rotatable bond characterized by a dihedral angle is varied during the minimization[20].

There is a general agreement that the quality of the determined structure is judged with the two criterions: a small values for both 1) the target function and 2) the root mean square deviation (RMSD). The former shows a satisfactory correlation between the structure and experimental constraints, and the latter illustrates a well-defined group of calculated structures, usually called a family of structures, since the DG calculations does not generate a single structure, but a family of structures instead.

2.1.3 Structural refinement

Apparently the algorithm used for distance geometry neglects the energy of protein structures. The calculations described above convert only geometric constraints (distances, dihedral angles, Van der Waals radii and the primary sequence) into molecular coordinates. The restrained energy minimization (REM) or restrained molecular dynamics calculations (RMD) are regularly performed on a set of structures obtained from the DG calculation. The potential energy function defined by molecular mechanics is usually derived as the sum of the potential terms[5]:

$$V_{total}=V_{bond}+V_{angle}+V_{dihed}+V_{vdW}+V_{coulomb}+V_{NMR} \qquad (3)$$

where V_{bond}, V_{angle}, and V_{dihed} are harmonic potentials for keeping bond lengths, bond angles and dihedral angles at their equilibrium values V_{vdW} is the Lennard-Jones potential describing the van der Waal interactions, consisting of the repulsive and attractive forces, and $V_{coulomb}$ describes the coulombic or electrostatic interactions. The last term, V_{NMR} representing the mixed linear harmonic potential, takes the NMR experimental constraints into consideration. Now it is possible to calculate the energy of the protein and minimize the energy.

REM usually finds the protein in a local minimum of the potential energy surface, but in many cases could not lead to the accurate structures. RMD is employed with the aim not only to refine the structure but also to further investigate dynamics properties of proteins. The idea relies on the evaluation of the potential energy and the kinetic energy (through the velocities associated with each particle) which allows us to solve Newton's equation of motion for small intervals of time. An MD trajectory can be regarded as a set of pictures comprising a sequence of the fluctuation of the nuclear coordinates of proteins at a certain temperature.

If it can be visualized as a short animation, it would be very amusing and instructive to realize how the protein in solution behaves.

2.2 Approach to paramagnetic metalloproteins

2.2.1 NMR spectra of paramagnetic metalloproteins

The two relevant consequences in the presence of one or more unpaired electrons to NMR spectra are much wider chemical shifts and much shorter nuclear relaxation times. Such contributions are called "hyperfine shift" and "hyperfine broadening", respectively. The electronic magnetic moment is orders of magnitude larger than (658 times) that of the proton[26], therefore the nuclear spin transitions are induced effectively by fluctuating magnetic fields generated from unpaired electron spins. The coupling between a magnetic dipole of unpaired electrons and another magnetic dipole of a resonating nucleus is called hyperfine coupling.

The first goal is to attempt in optimizing acquisition and processing NMR parameters to detect the signals affected by the paramagnetic center, whenever it is possible. 1D NOE's are found to be extremely useful in paramagnetic systems[27-30]. If the chemical shifts of paramagnetic signals are located outside the diamagnetic region (0-12 ppm) and are well resolved, the 1D technique is useful to detect only the signal(s) which respond to the irradiated one. In this case, 1D NOE difference spectra allow us to detect NOE on either slow or fast relaxing nuclei. This makes 1D NOE superior in terms of connectivity intensities with respect to 2D NOESY[31].

The 2D NOESY[32,33] spectra should be optimized to detect cross-peaks between paramagnetic signals and between paramagnetic and diamagnetic signals. For diamagnetic systems, the treatment of the time dependence of magnetization during 2D NMR experiments is usually suitable to allow the magnetization to completely relax back to thermal equilibrium during the entire pulse sequence, before the detection pulse. This is not for the case of paramagnetic systems where the magnetization can, partially or totally, return to equilibrium during the 2D-pulse sequence, thus causing the disappearance of the cross-peaks. As a result, the relaxation aspect should be concerned in such a way that the interval times (during preparation, evolution (t_1), mixing (τ_m) and detection (t_2) periods) in 2D NMR pulse (Fig.5) sequences must be set up appropriately to detect cross-peaks involved fast relaxing nuclei.

The general strategy is to use standard pulse sequences and optimize the evolution time, t_1 and the acquisition time, t_2, which should not exceed the spin-spin relaxation time, T_2. Long acquisition times, t_2, are not necessary. As the transverse magnetization disappears from the xy plane, T_2 is usually very short. After the signal has decayed only noise is acquired. Thus, the best compromise should be in between short interval times for maintaining magnetization before the detecting pulse and simultaneously long enough to allow magnetization or coherence to be transferred from one set of spins to the other. The mixing parameter (τ_m) should be of the order of the spin-lattice relaxation time, t_1, of the signal of interest. This strategy can apply to any 2D (NOESY, COSY, TOCSY, NOE-NOESY) and 3D (NOESY-NOESY, NOESY-TOCSY, NOESY-HMQC, TOCSY-HMQC)[31].

Once the spectra optimized for detecting dipolar connectivities involving paramagnetic signals are recorded, the resonance assignment process is carried out based on their connectivities and spin patterns. A structural model such as X-ray structure of a corresponding protein can be used to find connectivity between short distance protons.

2.2.2 Structure calculation of paramagnetic proteins

The 1D NOE's obtained by saturating a hyperfine shifted signal will be converted into

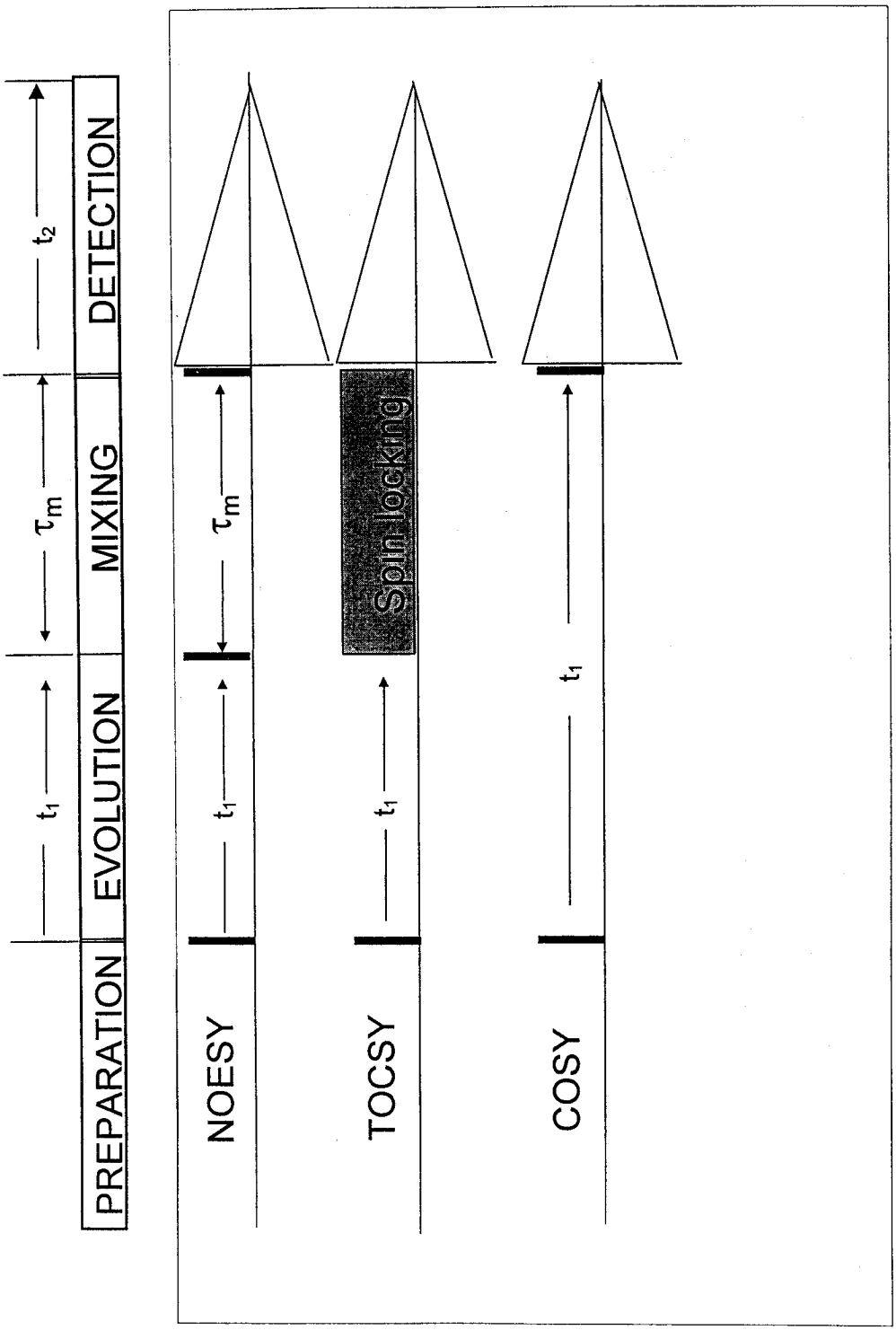


Fig.5 General experimental scheme for 2D NOESY, TOCSY and COSY pulse sequence.

distances using Eq. (1) and the same strategy (in section 2.2.1) can obtain the calibration parameter. The constant C should be estimated in such a way that upper distance limits are calculated, which will be used as input for calculations.

For the NOESY cross-peaks, again, intensities will be converted into upper distance limits by means of a calibration procedure. One has to keep in mind that the enhancements of the nuclear relaxation rate due to the paramagnetic center can decrease NOE intensities or, for the worst case, reduce the number of NOE cross-peaks. Hence, if the paramagnetic effect is neglected, the estimated distance is larger than the actual distance, thus resulting in a poor resolution of the regions surrounding the paramagnetic center. Nevertheless, this problem may be ignored, if the quality of the structure does not suffer too much as a result of adequate number of NOE's. However, a procedure to obtain experimentally paramagnetic constraints has been proposed (nuclear relaxation rate enhancements and pseudocontact shift[34,35]). This relies on a connection between the metal and the protons of the protein.

The program is typically used for calculating the three dimensional structure of non-metalloprotein. Therefore, it is necessary to include the prosthetic group in such a calculation. In the distance geometry program, DIANA, the problem has been resolved by including two new amino acids in the amino acid library that are used for introducing, respectively, the cluster 4Fe-4S present in the high-potential iron-sulfur protein (HiPIP) (Fig.6a) and the heme group present in cytochrome (Fig.6b). In the cluster 4Fe-4S, it is necessary to introduce a new residue to make an additional bond called "link" for maintaining the geometry of the cluster. The distance and the angle of the bond are taken from the average X-ray crystal structures of analogous systems. In the case of cytochrome, the new residue constitutes of the histidine bound to the heme iron.

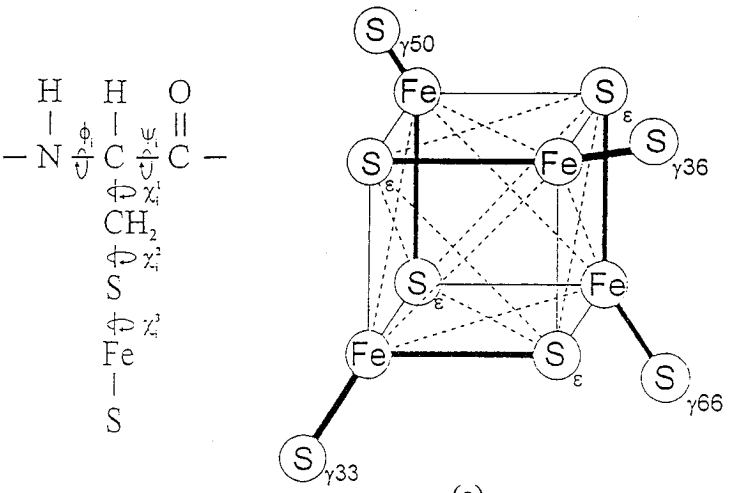
In the refinement stage, the energy minimization and molecular dynamic simulation employed by the program AMBER is particularly important to implement parameters used for the paramagnetic center. It should be noted that for molecular dynamics simulations performed on metalloproteins the definition of the geometric properties of the cofactors (*e.g.* iron-sulfur clusters, heme moieties, etc.) is still based on X-ray derived parameters.

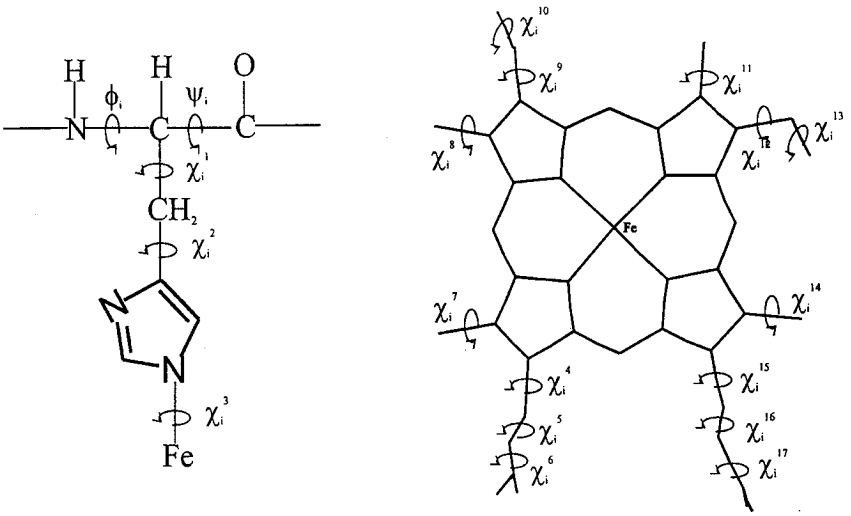
3. APPLICATIONS

The following part will outline the themes of research related to the proteins studied. Firstly, the description of the solution structure of the reduced high-potential iron-sulfur proteins (HiPIP) from *Chromatium vinosum* is given[16] and following with the solution structure of the oxidized form of the same protein[36]. They are generally found in photosynthetic bacteria. The last three sections deal with the work on cytochromes. It begins with the solution structure of oxidized *Saccharomyces cerevisiae* iso-1-cytochrome *c*[37]. The solution structure of Met80Ala mutant of *Saccharomyces cerevisiae* iso-1-cytochrome *c* is presented in the next section[38]. In the last section, the structure of a fully oxidized tri-heme cytochrome c_7 or cytochrome $c_{551.1}$ from *Desulfuromonas acetoxidans* which contains three paramagnetic centers, is reported[39]. The three-hemes cytochrome is a case example to demonstrate the efficiency of NMR techniques to obtain the solution structure of paramagnetic metalloproteins, also when no other structural information on the system is available.

3.1 NMR solution structure of reduced HiPIP from *C. vinosum*

The work started with the assignment of the reduced HiPIP from *C. vinosum*. In fact, almost 80% of ^{1}H resonances of the protein had been previously assigned through conventional





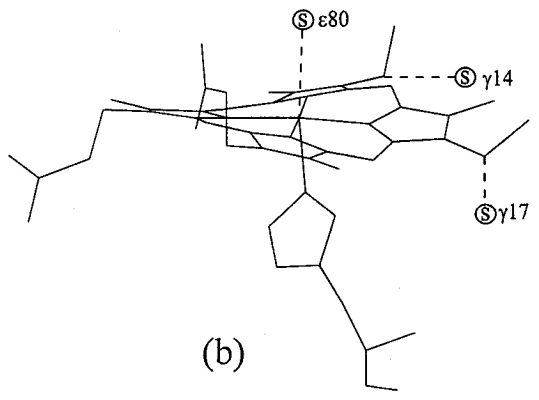


Fig.6 The geometric definition of the 4Fe-4S cluster (6a) and the heme (6b) groups was implemented in the distance geometry library containing individual amino acid residues.

2D NMR spectra (COSY, TOCSY and NOESY)[40]. By taking advantage of the data in the literature, the extension of the ^{1}H resonance and the new ^{15}N resonance assignments were achieved through the 2D homonuclear experiments (NOESY and TOCSY) and through 2D and 3D heteronuclear experiments (2D ^{1}H-^{15}N HMQC and 3D ^{1}H-^{1}H TOCSY/^{1}H-^{15}N-HMQC and 3D ^{1}H-^{1}H NOESY/^{1}H-^{15}N-HMQC). The heteronuclear spectra were used to assign the amide protons and to perform the sequential assignments. The analysis of these spectra has helped to extend the assignment up to 85% of the total protons (83 spin systems out of 85 amino acid residues were assigned; only Leu65 and Ser79 remain completely unassigned) and to 90% of the ^{15}N signals (79 ^{15}N signals were identified; five unassigned ^{15}N signals were prolines). In addition, ^{15}N resonances of most side-chains were also assigned. Since the protein is paramagnetic, TOCSY, NOESY and 1D NOE optimized to detect the fast relaxing signals were recorded to obtain the connectivities involving the hyperfine signals.

1147 out of 1489 (1461 NOESY cross-peaks, 28 1D NOE connectivities) distance constraints, which correspond to 13.4 NOE per residue, were found to be relevant and used in the DG calculations. 57 out of 126 stereospecific assignments were identified. A family of the 15 structures obtained through DG calculations ($TF < 0.9$ Å^2) was refined by REM and RMD. The RMSD from a family of 15 structures obtained after RMD calculations are 0.62±0.12 Å for backbone atoms, and 1.19±0.18 Å for all heavy atoms (see figure 7). Coordinates of the RMD family of the structure and of the RMDw structure have been deposited in the Brookhaven Protein Data Bank (ID codes are 1HRQ for the RMD family and 1HRR for the RMDw structure). The RMD structure with lowest energy has been subjected to restrained molecular dynamics in water (RMDw), with explicit solvent. The comparison of the RMDw structure with the X-ray structure reveals only few differences. The general folding and the elements of secondary structure (2 α-helixes from Ala12-Ala16 and Arg28-Ala32) are maintained in solution. All seven side-chains of aromatic amino acids (Tyr19, His42, Phe48, Trp60, Phe66, Trp76, Trp80), that form a compact hydrophobic core around the cluster and have been proposed to play a role in the process of electron transfer, are unaltered, except a small deviation for the aromatic rings of His42 and Phe66 between the X-ray structure of oxidized form and the solution structure of the reduced protein.

3.2 NMR solution structure of oxidized HiPIP from *C. vinosum*

The solution structure of the oxidized HiPIP from *C. vinosum* has been solved using the same approach as in the reduced protein. From the previous report, the ^{1}H assignment of the oxidized HiPIP[41] have been extended up to 85% of the total protons (154 proton resonances and 17 residues additionally assigned) and to 80% of the ^{15}N resonances using 1D NOE, 2D TOCSY and NOESY, and 2D ^{1}H-^{15}N HMQC. 84 out of 85 amino acids were assigned (only Ser79 remained completely unassigned). 1D NOE experiments provided 39 connectivities around the paramagnetic center[42,43]; a total of 1537 (1489 2D NOESY connectivities and 39 1D NOE connectivities) distance constraints were collected and 1142 constraints were found to be relevant for the DIANA calculations. Despite the fact that the oxidized HiPIP (S=1/2) is more paramagnetic than the reduced protein (S=0), The number of assigned protons was slightly more resonances in the oxidized case than those in the reduced case. The T_1 of protons are longer for the more paramagnetic oxidized species presumably because the electron relaxation rates are different.

The DG calculations were carried out by the program DIANA. The strategy of the structural calculations is identical to that in the reduced protein. The program GLOMSA yields 56 stereospecific assignments. The 15 structures obtained through DG calculations (TF

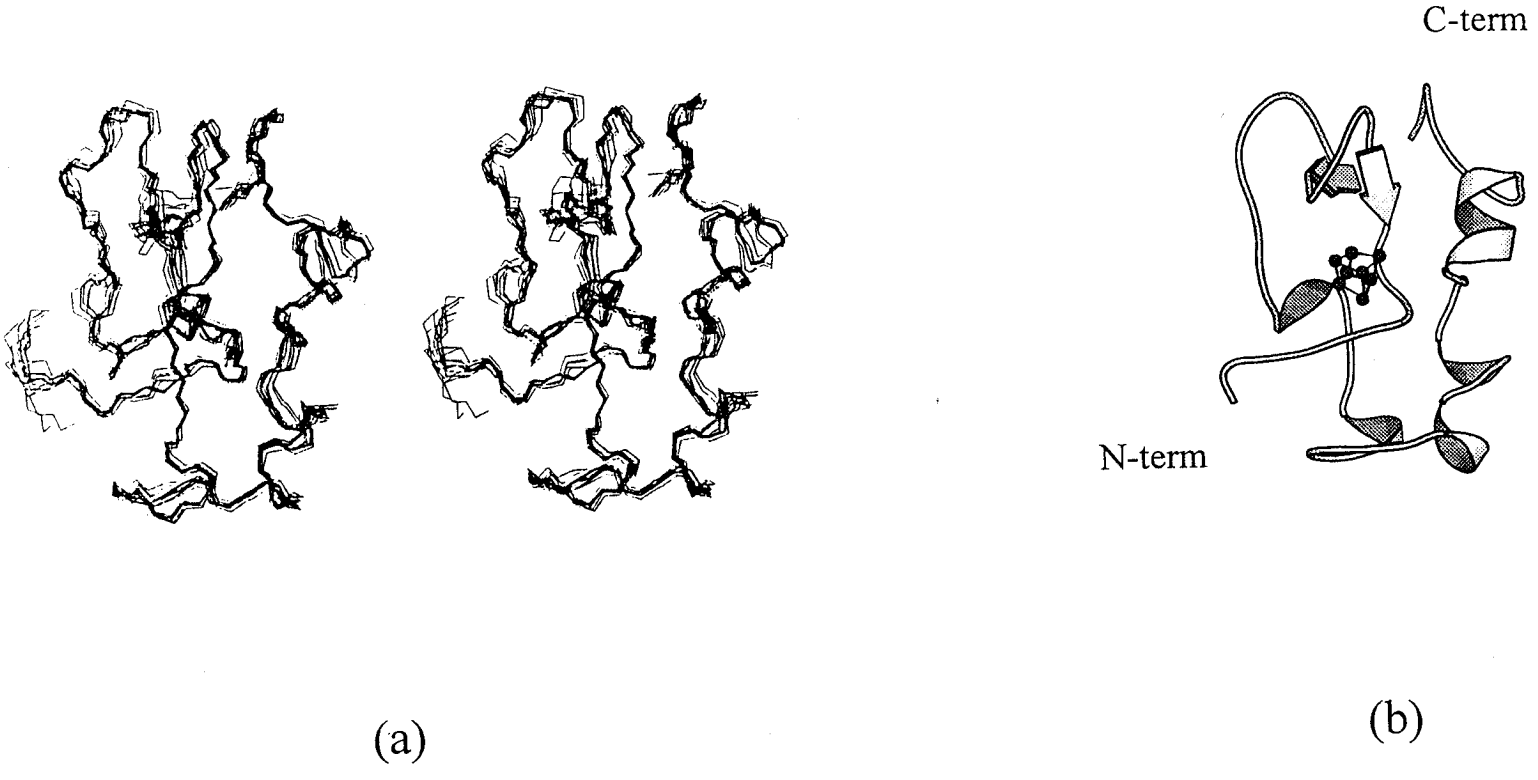


Fig.7 (a) Stereo drawing of the 15 structures and (b) schematic representation of the reduced HiPIP from *C. vinosum*. The figure was produced with Molscript[49]. The cluster is shown in ball-and-stick, N-term and C-term are the N- and C-termini, respectively.

< 0.65 Å^2) was refined by REM and RMD. The RMSD values for the family of 15 structures obtained after RMD calculations are 0.57±0.14 Å for backbone atoms, and 1.08±0.16 Å for all heavy atoms. The RMD structure with lowest energy has been subjected to restrained molecular dynamics in water, with explicit solvent.

3.3 NMR solution structure of oxidized cytochrome *c*

Yeast cytochrome *c* (cyt *c*) cannot be enriched in ^{15}N because of the lack of an adequate expression system. In contrast to *C. vinosum* HiPIP, only ^{1}H NMR experiments were used to assign the protons and identify the dipolar connectivities. The size of the protein (108 amino acids) is at the limit of the applicability of homonuclear NMR techniques. The proton assignment of the oxidized protein was previously published[44]. The assignments have extended it up to 77% of the proton resonances. NOESY and TOCSY with the short τ_m and 1D NOE were used to observe the connectivities involving with the heme resonances, the signal of His18 and Met80. Only 3 residues (Thr-5, Glu-4 and Gly83) remained unassigned.

The assignment yields 1361 out of a total of 1676 meaningful distance constraints. Of them, 1356 constraints were taken from NOESY and five 1D NOEs were obtained by saturating the methyl group of Met80, one Hγ proton of Met80, and the Hε1 and Hδ1 protons of the proximal histidine. The number of experimental NOE constraints per residue corresponds to 15.5 as input and to 12.6 accepted NOEs per residue. These values are comparable to those used for solution structure calculations of other oxidized and reduced cytochromes[38,45,46].

For the structural calculation, the new residue called HEM, which contains the heme and histidine coordinates, was taken from the X-ray structure of the oxidized cytochrome c and was introduced in the DIANA library. During the calculation, upper and lower distance constraints were included to maintain the covalent bond between heme iron and the sulfur of the two cysteines (Cys14 and Cys17) and a sulfur atom of the methionine (Met80). Descriptions of the strategy used for the calculation have already been mentioned, therefore it will not be addressed here. 42 stereospecific assignments were obtained with the program GLOMSA; and 14 hydrogen bond constraints, involving slowly exchanging amide protons consistently present in the initially calculated DG structures, were introduced as further constraints in the final stages of the structure calculations.

The RMSD from the average structure for the family of 20 structures are 0.63±0.06 Å for backbone atoms and 1.10±0.08 Å for all the heavy atoms.

An algorithm which allows us to use the pseudocontact shift (δ_{pc}) constraints in DG structure calculations has been reported previously[47]. The procedure for using the pseudocontact shift constraints in the final stage of the DG structural calculations consists of: (i) determination of the tensor anisotropy parameters from a set of 15-20 structures with the lowest target function (This calculation was performed by the program FANTASIA); (ii) recalculation of the new structures using NOE and pseudocontact shift constraint; (iii) re-determination of the tensor anisotropy parameters from the new structures, and so on until convergence. The magnetic susceptibility tensor parameters derived from these calculations were then used in a modified version of the DIANA program, called PSEUDIANA[47].

The pseudocontact shifts are obtained by subtracting from the ^{1}H shifts of the oxidized protein those of the reduced protein. The chemical shifts of the assigned protons belonging to Cys14, Cys17, His18, Met80 and heme group were not taken into consideration because of the contact shift contribution. From this analysis, 256 of δ_{pc} constraints have been obtained.

The RMSD values with respect to the mean structure for the family obtained from

PSEUDIANA ($TF < 7.12$ Å^2) are 0.58±0.07 and 1.05±0.07 Å for the backbone and all heavy atoms, respectively.

Finally, from the best 48 structures of PSEUDIANA, an energy minimization with a package called PSEUDO-REM[48] was performed. A family of the best 20 structures obtained from PSEUDO-REM with RMSD of 0.58±0.08 and 1.05±0.10 Å and target function in the 3.58-4.10 Å^2 range is obtained. These values are quite satisfactory because of a decreasing of the target function and maintaining the RMSD values. The family of the structure is shown in figure 8.

3.4 NMR solution structure of the oxidized mutant cytochrome c

The 1D and 2D spectra (COSY, TOCSY, NOESY, NOE-NOESY, 1D NOE) were used to assign the proton resonances. 82% of the protein protons have been identified. 106 residues have been assigned. Thr-5 and Glu-4 remained unassigned.

For the structure calculations, 1428 meaningful NOE constraints (8 from 1D NOE constraints) out of a total of 1842 measured NOEs were taken into consideration by DIANA. The number of constraints corresponds to 17.0 NOEs per residue as input for the DG calculation of which 13.1 constraints per residue were found to be relevant. The heme and the bound cyanide were included in the DG calculations through the addition of an artificial amino acid to the residue library used by the program DIANA. A distance of 1.85 Å between the iron and the carbon of cyanide was set, as that found in the X-ray structure of the cyanide adduct of cytochrome *c* peroxidase48. The artificial residue, denoted by HES, consists of a histidine residue, whose N2 was connected to the heme skeleton through links with the four pyrrole nitrogens (upper distance limit 2.9 Å). The "special covalent bonds" were also used to define the link between the heme and Cys14 and Cys17; the upper and lower distance limits between the cysteine sulfur and the -carbon of the corresponding heme thioether substituent were set to 2.1 and 1.9 Å respectively.

17 hydrogen bond constraints (the H-acceptor distance being less than 2.4 Å and the angle H-donor-acceptor less than 35°) and 60 stereospecific assignments were used in the final stage of the calculations. The RMSD values at the stage of restrained energy minimization of 17 structures obtained from distance geometry calculations are 0.68±0.11 Å and 1.32±0.14 Å for the backbone and heavy atoms, respectively.

3.5 NMR solution structure of the cytochrome c_7

The spin patterns of the amino acid residues have been assigned through conventional NMR methods. The sequential assignment was performed using the presaturation and WATERGATE NOESY and TOCSY experiments. COSY, TOCSY and NOESY experiments tailored to detect connectivities involving the hyperfine shifted signals allows us to identify six heme-bound histidine residues and heme resonances. All histidine-ring protons cannot be located except the two of the four up-field shifted signals in 1D reference spectrum. Based on 1D NOE experiments, those two signals were assigned to the H1 of His30 and the H2 of His17. Despite the influence of three paramagnetic centers of this small protein, 67 out of 68 residues which corresponds to 82% of the total proton resonances have been assigned. Ala1 remained unassigned.

From 1142 NOEs, 793 constraints have been found meaningful. 13 hydrogen-bond constraints were used. H-bond constraints are included whenever NH resonance is still detectable in D_2O maps (*i.e.* there is an experimental evidence of slow exchange) and the H-bond between

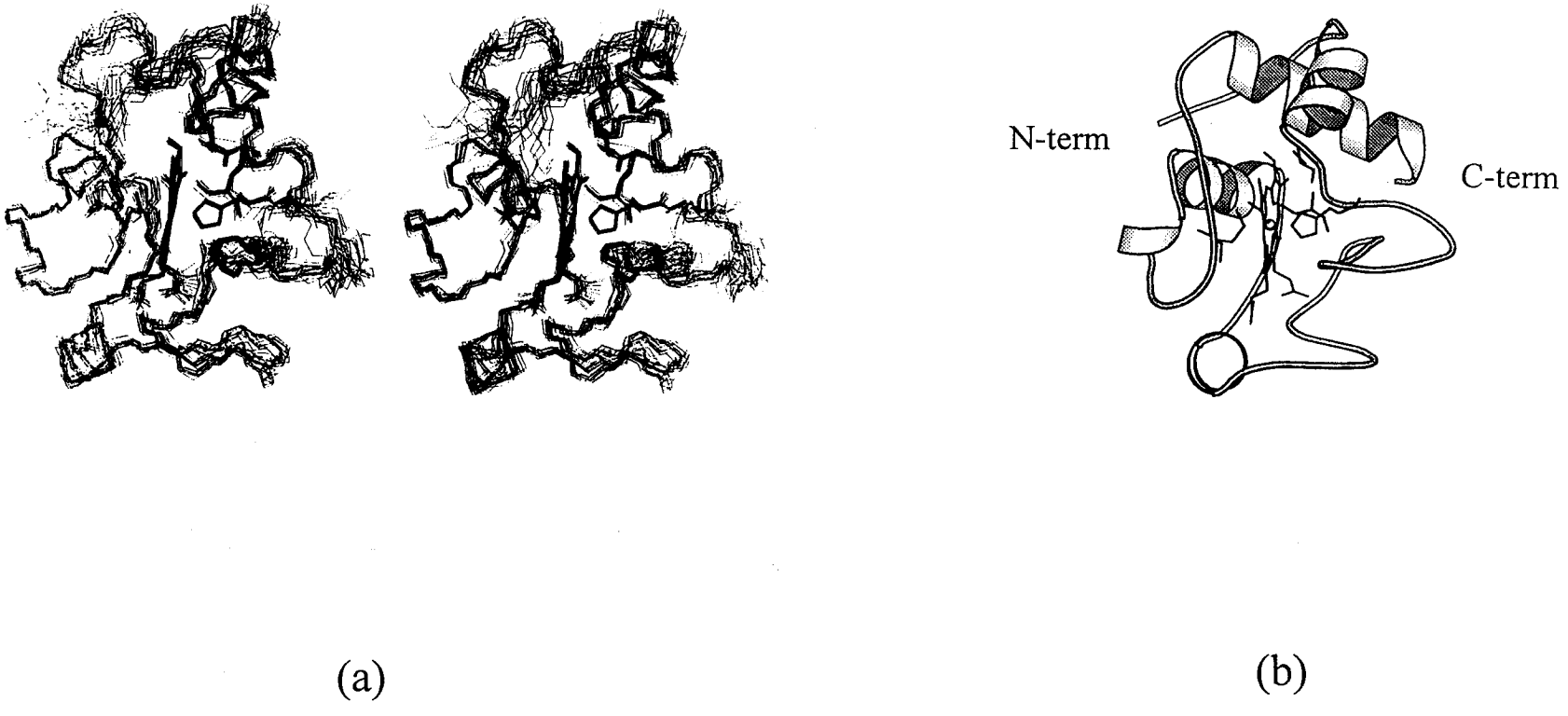


Fig.8 (a) Stereo drawing of the 20 structures and (b) schematic representation of the oxidized cytochrome c. Some relevant residues and the heme group is shown in wireframe, N-term and C-term are the N- and C-termini, respectively.

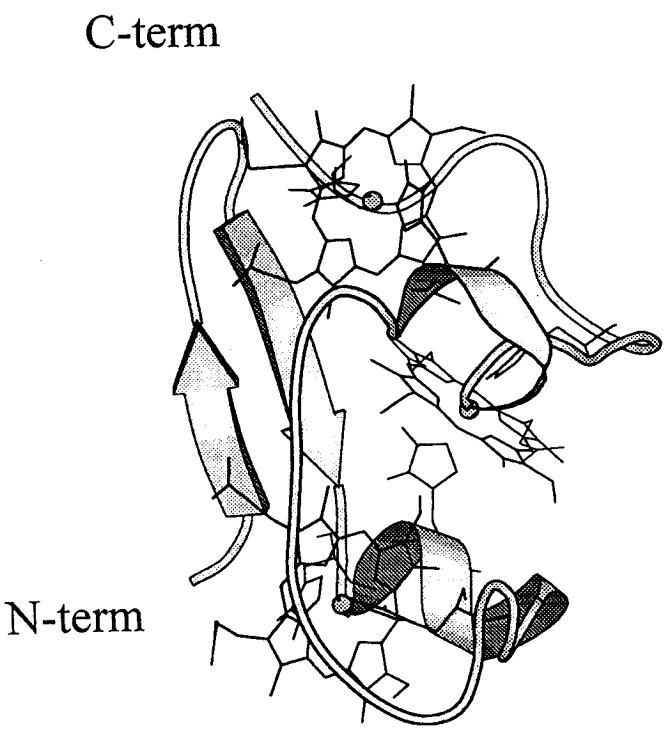


Fig.9 Schematic representation of the tri-heme cytochrome c_7.

Table 1. Summary of the different features of the two NMR approaches to structural studies.

Features	Proteins	
	Diamagnetic	**Paramagnetic**
Chemical shift dispersion	0-15 ppm	Much larger depending on the strength of paramagnetism
Relaxation times (T1 and T2)	100 ms - 1 s	A few milliseconds
NMR techniques	2D COSY, TOCSY NOESY with standard parameters	1D NOE, 2D COSY, TOCSY, NOESY optimized to detect fast relaxing signals
Assignment	Conventional procedure	Based on comparison with the available model
Structure calculation (DG) & refinement (REM, RMD)	Using standard constraints to complete the 3D structure	Using paramagnetic constraints (the pseudo contact shifts etc.) to refine the resolution of the active site

Table 2. NMR experiments, constraints and final data relative to the solution structures determined.[a]

Type of information	Solution structures of the protein determined				
	Red.*Cv.*	**Ox.*Cv.***	**Cyt *c***	**Ala80**	**C_7**
No. residues.	85	85	108	108	68
Prosthetic group	$[Fe_4S_4]^{2+}$	$[Fe_4S_4]^{3+}$	*c*-type heme	*c*-type heme	Three *c*-type hemes
%assigned ^1H	85	85	77	82	83
%assigned ^{15}N	93	80	-	-	-
2D NOESY	1461(1119)[c]	1498(1103)[c]	1671(1356)[c]	1834(1418)[c]	1142(793)[c]
1D NOE	28	39	5	8	13
Dihedral angles	-	-	-	-	21
H-bonds	-	-	14	17	13
Stereospecific[d]	57	56	42	60	44
Structure calculation methods	DG, REM, RMD, RMDw	DG, REM, RMD, RMDw	DG[e], REM[e]	DG, REM	DG
No. structures	15	15	20	17	20
RMSD for BB(Å)	0.42±0.15	0.39±0.11	0.58±0.08	0.68±0.11	0.70±0.22
RMSD for HA(Å)	0.90±0.17	0.74±0.11	1.05±0.10	1.32±0.14	0.98±0.19
TF($Å^2$)	0.56	0.51	4.10	0.33	0.41
Total energy (kJ mol-1)	-6648.0	-6609.0	-4556.2	-6712.6	-
1PDB code	1hrq, 1hrr	1neh	1yic	1fhb	1cfo

[a] Red.*Cv* and Ox.*Cv* are the reduced and oxidized high-potential iron-sulfur proteins from *Chromatium vinosum* respectively, Cyt *c* is the widetype cytochrome *c* and Ala80 is the mutant cytochrome *c* and C_7 is the tri-heme cytochrome (further details can be seen from the text).

[b] all determined structure were used 1D NOE difference spectra, COSY, TOCSY, NOESY recorded on Bruker MSL 200, DRX 500 and AMX 600 spectrometers operating, repsectively, at 200.13, 500.13 and 600.13 MHz proton Larmor frequency at Unversity of Florence. Most of the dipolar connectivities were taken from the 100 ms NOESY in H2O solution at 300-303 K.

[c] The number of meaningful constraints is reported in brackets.

[d] Number of stereospecific assignments of diastereotopic protons and methyl groups.

[e] The paramagnetic constraints were used in DG and REM.

this amide protons and a potential acceptor atoms is present in at least 70% of the calculated structures. 21 experimentally evaluated dihedral angle constraints, obtained analyzing the ^3J(NH-H) coupling constants were also introduced. A distance geometry family of 20 structures (TF <0.41 Å^2) has been obtained through the program DIANA. RMSD values of the family with respect to the average structure, of 0.70±0.22 and 0.98±0.19 Å for the backbone and all heavy atoms, respectively (figure 9).

CONCLUDING REMARKS

Signals in the NMR spectra of paramagnetic metalloproteins can be distinguished into (1) diamagnetic, and (2) paramagnetic signals. The strategy to achieve the solution structure of paramagnetic proteins can be considered as a combination of the two approaches, the conventional approach and an approach to obtain additional information extracted from the influence of paramagnetism. The former is used to obtain structural information on diamagnetic parts of the protein and the latter is carried out to get information on the electronic distribution around the paramagnetic center and to use distance constraints to refine the structure in solution of the active site. The latter approach has improved significantly the resolution of the structure. The schematic representation of the two different NMR approaches to structural studies of paramagnetic metalloproteins is summarized in Table 1.

From the structural information and relevant experimental data of the protein determined summarized in Table 2, it appears that the determination of the solution structure of paramagnetic metalloproteins is in favorable cases possible and results are as good as those obtained in the case of diamagnetic proteins, even though it requires some more care and work. NOE's are still the most important data required to obtain a good structure for both diamagnetic and paramagnetic proteins. Some other constraints may be needed to obtain a satisfactory resolution in the regions close to the paramagnetic center.

ACKNOWLEDGEMENTS

The author would like to express appreciation to Prof. I. Bertini and his research group at the Florence Laboratory of Relaxometry and Magnetic Resonance on Paramagnetic Metalloproteins and to Dr. Bates of Chulalongkorn University for reading the manuscript. Rachadapisek Sompoj Endowment is gratefully acknowledged.

REFERENCES

1. Bertini, I., Gray, H.B., Lippard, S.J., and Valentine, J.S. (1994) Bioinorganic Chemistry, Univerisity Science Books, USA.
2. Banci, L., Bertini, I., Eltis, L.D., Felli, I. C., Kastrau, D. H. W., Luchinat, C., Piccioli, M., Pierattelli, R., and Smith, M. (1994) *Eur. J. Biochem.* **225**, 715-725.
3. Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids, Wiley, New York.
4. Ernst, R.R., Bodenhausen, G., and Wokaun, A. (1987) Principles of Nuclear Magnetic Resonance in one and two dimensions, Oxford University Press, London.
5. Scheek, R.M., van Gunsteren, W.F., and Kaptein, R. (1989) *Methods Enzymol.* **177**, 204-218.
6. Case, D.A., Dyson, H.J., and Wright, P.E. (1994) *Methods Enzymol.* **239**, 392-416.
7. Jeener, J. (1971) AMPERE INTERNATIONAL SUMMER SCHOOL (Abstract).
8. Aue, W.P., Bartholdi, E., and Ernst, R.R. (1976) *J. Chem. Phys.* **64**, 2229-2235.
9. Bax, A., Freeman, R., and Morris, G. (1981) *J. Magn. Reson.* **42**, 164-168.

10. Wemmer, D.E. (1989) *Conc. Magn. Reson.* **1**, 59-72.
11. Braunschweiler, L. and Ernst, R.R. (1983) *J. Magn. Reson.* **53**, 521-528.
12. Bax, A. and Davis, D.G. (1985) *J. Magn. Reson.* **65**, 355-360.
13. Davis, D.G. and Bax, A. (1985) *J. Am. Chem. Soc.* **107**, 2820-2821.
14. Wider, G., Macura, S., Kumar, A., Ernst, R.R., and Wüthrich, K. (1984) *J. Magn. Reson.* **56**, 207-234.
15. Macura, S., Wüthrich, K., and Ernst, R.R. (1982) *J. Magn. Reson.* **47**, 351-357.
16. Banci, L., Bertini, I., Dikiy, A., Kastrau, D.H.W., Luchinat, C., and Sompornpisut, P. (1995) *Biochemistry* **34**, 206-219.
17. Wagner, G. (1990) *Progr. Nucl. Magn. Reson. Spectrosc.* **22**, 101-139.
18. Bax, A., Griffey, R.H., and Hawkins, B.L. (1983) *J. Am. Chem. Soc.* **105**, 7188-7190.
19. Bax, A., Griffey, R.H., and Hawkins, B.L. (1983) J. Magn. Reson. **55**, 301-315.
20. Güntert, P., Braun, W., and Wüthrich, K. (1991) *J. Mol. Biol.* **217**, 517-530.
21. Borgias, B., Gochin, M., Kerwood, D.J., and James, T.L. (1990) *Progr. Nucl. Magn. Reson. Spectrosc.* **22**, 83-100.
22. Boelens, R., Koning, T.M.G., Van der Marel, G.A., Van Boom, J.H., and Kaptein, R. (1989) *J. Magn. Reson.* **82**, 290-308.
23. Karplus, M. (1963) *J. Am. Chem. Soc.* **85**, 2870-2871.
24. Güntert, P. and Wüthrich, K. (1991) *J. Biomol. NMR* **1**, 447-456.
25. Blumenthal, L.M. (1970) Theory and Applications of Distance Geometry, New York.
26. Koenig, S.H., Prodell, A.G., and Kusch, P. (1952) *Phys. Rev.* **88**, 191.
27. Johnson, R.D., Ramaprasad, S., and La Mar, G.N. (1983) *J. Am. Chem. Soc.* **105**, 7205-7206.
28. Ramaprasad, S., Johnson, R.D., and La Mar, G.N. (1984) *J. Am. Chem. Soc.* **106**, 5330-5335.
29. Unger, S.W., Lecomte, J.T.J., and La Mar, G.N. (1985) *J. Magn. Reson.* **64**, 521-526.
30. Banci, L., Bertini, I., Luchinat, C., and Piccioli, M. (1991) in NMR and biomolecular structure (Bertini, I., Molinari, H., and Niccolai, N., eds) pp. 31-60, VCH.
31. Banci, L., Bertini, I., and Luchinat, C. (1994) *Methods Enzymol.* **239**, 485-514.
32. Jeener, J., Meier, B.H., Bachmann, P., and Ernst, R.R. (1975) *J. Chem. Phys.* **71**, 4546-4553.
33. Kumar, A., Ernst, R.R., and Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun.* **95**, 1104.
34. Bertini, I., Donaire, A., Felli, I.C., Luchinat, C., and Rosato, A. (1996) *Magn. Reson. Chem.* **34**,948-950.
35. Hestenes, M.R. and Stiefel, E. (1952) *J. Res. Nat. Bureau Standards* **49**, 409-436.
36. Bertini, I., Dikiy, A., Kastrau, D.H.W., Luchinat, C., and Sompornpisut, P. (1995) *Biochemistry* **34**, 9851-9858.
37. Banci, L., Bertini, I., Bren, K.L., Gray, H.B., Sompornpisut, P., and Turano, P. (1997) *Biochemistry* **36**, 8992-9001.
38. Banci, L., Bertini, I., Bren, K.L., Gray, H.B., Sompornpisut, P., and Turano, P. (1995) *Biochemistry* **34**, 11385-11398.
39. Banci, L., Bertini, I., Bruschi, M., Sompornpisut, P., and Turano, P. (1996) *Proc. Natl. Acad. Sci. USA* **93(25)**, 14396-14400.
40. Gaillard, J., Albrand, J.-P., Moulis, J.-M., and Wemmer, D.E. (1992) *Biochemistry* **31**, 5632-5639.
41. Nettesheim, D.G., Harder, S.R., Feinberg, B.A., and Otvos, J.D. (1992) *Biochemistry* **31**, 1234-1244.
42. Bertini, I., Capozzi, F., Ciurli, S., Luchinat, C., Messori, L., and Piccioli, M. (1992) *J. Am. Chem. Soc.* **114**, 3332-3340.
43. Banci, L., Bertini, I., Carloni, P., Luchinat, C., and Orioli, P.L. (1992) *J. Am. Chem. Soc.* **114**, 10683-10689.
44. Gao, Y., Boyd, J., Williams, R.J.P., and Pielak, G.J. (1990) *Biochemistry* **29**, 6994-7003.
45. Qi, P.X., Di Stefano, D.L., and Wand, A.J. (1994) *Biochemistry* **33**, 6408-6417.
46. Baistrocchi, P., Banci, L., Bertini, I., Turano, P., Bren, K.L., and Gray, H.B. (1996) *Biochemistry* **35**, 13788-13796.
47. Banci, L., Bertini, I., Bren, K.L., Cremonini, M.A., Gray, H.B., Luchinat, C., and Turano, P. (1996) *JBIC* **1**, 117-126.
48. Edwards, S.L. and Poulos, T.L. (1990) J. Biol. Chem. 265, 2588-2592.
49. Banci L., Bertini I., Savellini G.G., Romagnoli A., Turano P., Cremonini M.A., Luchinat C., Gray H.B. (1997) *Proteins. Struct. Funct.* **29**, 68-76.
50. Kraulis, P.J. (1991) *J. Appl. Crystallograph.* **24**, 946-950.

บทคัดย่อ

ปัจจุบันมีความเป็นไปได้ในการหาโครงสร้างสามมิติของโปรตีนชนิดพาราแมกเนติกส์ในสารละลายด้วยเทคนิคเอ็นเอ็มอาร์สมัยใหม่ อิทธิพลของพาราแมกเนติกส์ไปขยายอัตราการรีแลกซ์ของนิวเคลียร์เป็นผลให้สัญญาณเอ็นเอ็มอาร์ไม่สามารถตรวจพบได้ ฉะนั้นการศึกษาด้วยเทคนิคเอ็นเอ็มอาร์ของระบบพาราแมกเนติกส์จึงถูกทอดทิ้ง อย่างไรก็ตามระบบดังกล่าวได้รับการศึกษามาอย่างต่อเนื่องโดยกลุ่มวิจัยทางเอ็นเอ็มอาร์บางกลุ่ม จนกระทั่งสามารถแสดงให้เห็นว่าในบางกรณีสามารถเอาชนะอุปสรรคเหล่านี้ไปได้ แบบแผนการหาโครงสร้างของพาราแมกเนติกส์เมทัลโลโปรตีนจะถูกนำเสนอร่วมกับตัวอย่างที่เป็นข้อมูลที่ใช้ในการหาโครงสร้างสามมิติของพวกไอออนซัลเฟอร์โปรตีนชนิดศักย์สูง (HiPIPs) และพวกเฟอร์ริไซโตโครม ซี แบบแผนการนั้นประกอบด้วยเทคนิคเอ็นเอ็มอาร์แบบดั้งเดิม และแบบที่ใช้สำหรับระบบพาราแมกเนติกส์ คุณภาพของโครงสร้างที่หาได้นี้ทัดเทียมกับโครงสร้างของโปรตีนชนิดไดอะแมกเนติกส์ที่มีขนาดใกล้เคียงกัน