Università degli Studi di Firenze

DOTTORATO DI RICERCA IN
Scienze Chimiche

CICLO XXV

COORDINATORE Prof. Andrea Goti

Theory and Applications of Nuclear Magnetic Resonance of Biomolecules in solution and in the solid state

-

Teoria ed Applicazioni della Risonanza Magnetica Nucleare di Biomolecole in soluzione ed allo stato solido

Settore Scientifico Disciplinare CHIM/03

Dottorando
Dott. Ravera Enrico

_____
(firma)

Tutore
Prof. Luchinat Claudio

_____
(firma)

Anni 2010/2012

Abstract

This thesis describes the application of Nuclear Magnetic Resonance (NMR), with particular reference to paramagnetic effects, to the characterization of structural and dynamical features of some biomolecules. The research reported is framed into a critical conspectus of the state of the art of biomolecular NMR.

Motions in solution give rise to sets of averaged experimental observables. The reconstruction of the motion involved in the averaging is not a trivial task: a method is developed to obtain accurate information on dynamics from average data. Paramagnetism-based restraints obtained in solution are used to study the dynamics of two-domain proteins calmodulin and matrix-metalloproteinase 1 (MMP1).

In the solid state any molecule is surrounded by a number of neighbors: thus if the molecule is paramagnetic, it will sense a contribution to paramagnetic observables also from crystal neighbors. To recover this information an algorithm is developed and implemented. Such contribution is explicitly included in the calculations as restraints obtained in the solid state and allows for the refinement of the structure of MMP12 and of its crystalline environment at the same time.

Solid state NMR is gaining interest because it allows one to overcome some limitations (especially in size) of solution state NMR, and has recently seen a profound development. Anyway, sample preparation in solid state NMR has not completely followed up: usually microcrystalline preparations are employed to achieve high resolution, yet protein crystallization is a highly unpredictable process and simpler approaches are sought for. We developed a new technique, that we refer to as "sedimented solutes NMR" (SedNMR), that allows for the observation in the solid state of soluble macromolecules without the need for crystallization, lyophilization or freezing. A theoretical description of the process is provided and, from this, a number of experiments are proposed.

Dedication

This thesis is dedicated to the memory of Professor Dr. Ivano Bertini (1940-2012), great teacher, advisor, mentor and source of inspiration.
His neverending contributions to Biological Inorganic Chemistry and to Nuclear Magnetic Resonance will echo for a long time.

Acknowledgements

The first person to acknowledge is my beloved one Caterina Gheri.
I would then like to thank my colleagues and coworkers:
Ms. Linda Cerofolini (CERM), Dr. Björn Corzilius (MIT), Dr. Frank Engelke (Bruker Biospin), Mr. Lucio Ferella (CERM), Dr. Gregg B. Fields (Torrey Pines), Dr. Marco Fragai (Unifi), Dr. Angelo Gallo (CERM), Mr. Gianluca Gallo (Unifi), Prof. Dr. Carlos F. G. C. Geraldes (Univ. Cohimbra) Dr. Andrea Giachetti (CERM), Dr. Leonardo Gonnelli (Unifi), Prof. Dr. Robert G. Griffin (MIT), Mr. Benno Knott (Bruker Biospin), Ms. Magda Korsak (Giotto Biotech), Mr. Enrico Luchinat (CERM), Dr. Jiafei Mao (CERM, now BMRZ at Frankfurt), Dr. Vladimir K. Michaelis (MIT), Dr. Malini Nagulapalli (CERM), Mr. David Osen (Bruker Biospin), Prof. Dr. Giacomo Parigi (Unifi), Ms. Susanne Penzel (MIT), Dr. Maxim V. Petoukhov (EMBL Hamburg), Prof. Dr. Bernd Reif (TUM) Dr. Mauro Rinaldelli (CERM), Ms. Camilla Rosa (CERM), Prof. Dr. Antonio Rosato (Unifi), Dr. Albert (Andy/Bert) Smith (MIT), Dr. Dmitri I. Svergun (EMBL Hamburg), Mr. João M.C. Teixeira (CERM and Univ. Cohimbra) Prof. Dr. Paola Turano (Unifi).

Prof. Dr. Roberta Pierattelli deserves special mention, for introducing me to bioinorganic chemistry and NMR, for mentoring and continuously pushing me.
Finally, I wish to acknowledge Prof. Claudio Luchinat for his guidance, supervision and encouragement from the beginning of my Master's work and for taking the role of supervisor at the time of passing of Prof. Dr. Bertini.

Contents

List of Figures

Chapter 1

Introduction

1.1 Preface

1.1.1 Integrated Structural Biology - reconciliation

Since the resolution of the very first structures of biological macromolecules [1,2] biology has come to rely ever-increasingly on the structural information. Since then, a number of different techniques have been proposed to address the structural problem from different perspectives and at different levels and resolutions. And since then all these methods have taken different pathways. Nowadays the complexity of biological systems is understood as a feature and not as a problem, thus reductionist approaches are no longer sustainable and great efforts should be devoted to put back together this very diverse contributions, to build a unified view of biology, from the molecular level to the cell and, eventually, to the whole organism.

The sequencing of the human genome [3] (and of the genomes of hundreds of other organisms) has disclosed the amino acid sequences of all the proteins encoded therein. The knowledge of their three-dimensional structure (i.e. of their fold) is an important piece of information required in order to understand their functions.

But biomolecules are inherently moving structures, that often act in complexes rather than themselves, so a crucial step in understanding the mechanism of action of a number of proteins is to add a fourth dimension (time) to

the picture and explore conformational flexibility. [4] Yet, detailed information on mobility may prove difficult or impossible to access experimentally. Proteins may be composed by two or more domains that have well-defined structure, connected by flexible linkers, for which no structural information is available. In many instances interdomain mobility is critical to the function of the protein.
The experimental techniques able to solve protein folds at atomic resolution are X-ray diffraction and NMR spectroscopy. The former provides 3D structures in protein crystals, the latter in solution, crystals and disordered aggregates. X-ray crystallography is currently routine and proves an extremely fast method, [5] but it may not be fully informative: crystals may not form or, if a crystal is formed, only one frozen conformation is observed that might be different from what is found in solution. [6–12]
On the other hand, NMR techniques have long been used to obtain precious information on the mobility of the investigated systems. Of course, it is easy to understand that a complete description of such motions will never be obtained, because the number of experimental data is far smaller than the number of unknowns to be determined.

1.1.2 The contribution of Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) is a technique that has intrinsic atomic resolution, because it is based on the interaction of the nuclear spin with the electromagnetic radiation, in the presence of an external magnetic field.
NMR was introduced by Isidor Rabi in 1938. [13] Felix Bloch and Ed Purcell proved the technique for liquids and solids in 1946. [14, 15]
Nuclei are extremely sensitive to the electronic structure of the molecule, thus NMR was immediately promoted from an exotic toy for physicists to a capital tool for chemists. This is clearly testified by the name of the main observable in NMR: the Chemical Shift.
Any magnetic nucleus in equivalent sites of a molecule contributes to the spectrum. Complication in the spectrum arises already at moderate-size

molecules and becomes more and more severe in the case of proteins. The spectral overlap prevents extraction of site-specific information. To solve this issue, and to encode even more information into the spectrum, two-dimensional NMR was introduced. Two dimensional NMR was proposed as a gedankenexperiment by Jean Jeener in 1971 at the Basko Polje Ampere School and was then translated to reality by Richard R. Ernst in 1976 [16]. By 2DNMR, and by multidimensional NMR in general, it is possible to encode in the spectra the information that is needed to reconstruct the structure of the biomolecules. This possibility was demonstrated by Kurt Wütrich in 1981. [17] Such structural information is not readily available as in crystallography or in microscopy, but rather requires a complicated analysis to decode the distance informations that can be encoded in the spectra. [18]
For X-ray crystallography the presence of heavy atoms can be regarded as a blessing from the above, [19–23] while the biological NMR community needs to overcome, besides the several intrinsic limitations of the technique, the presence of paramagnetic metals in the molecule, which may be a source of severe line broadening. Anyway, It has been increasingly recognized that the spectral complication due to the electron-nucleus interactions can provide useful structural information. [24]
A large fraction of the enzymes use metals for catalysis [25], over 30% of the proteins with known structure bind metals, and a rough estimate of the fraction of proteins able to bind metal is set to about 50% of the total proteins [26], and a significant number of these are paramagnetic, so the issue is not an irrelevant one.

1.1.3 Dynamic Nuclear Polarization - DNP

The sensitivity of a spectroscopic technique is related to the energy separation between the levels between which transitions are observed. NMR is based on nuclear spin transitions, that are of the order of 0.01 cm^{-1} (0.03 at the highest available fields), while $k_BT \simeq 200\, cm^{-1}$ at room temperature. It is easy to calculate that approximately only 1 molecule out of 10000 will respond to NMR (see appendix B.1).

Dynamic Nuclear Polarization is a method that can be used to improve NMR sensitivity, by increasing the difference in population of the nuclear spin energy levels beyond Boltzmann distribution via the electron spin polarization. The maximum theoretical enhancement is $\frac{\gamma_e}{\gamma_n}$ (i.e.: 658 for ^1H, 2617 for ^{13}C and 6490 for ^{15}N, to mention the more common nuclei for biological NMR). DNP was proposed by Al Overhauser in 1953 [27] and demonstrated by Tim Carver and Charlie Slichter [28] on ammonia solutions of alkali metals.
In biological applications the electron spins are introduced as radicals (TEMPOL, TEMPONE, TEMPAMINE). [29–32] Recently, biradicals as TOTAPOL or SPIROPOL have been introduced to exploit the intramolecular electron-electron dipolar coupling to drive the nuclear transitions in the solid state. [33–35] For biological applications the protein is usually dispersed in a glycerol containing medium that forms an amorphous glass upon freezing and prevents the radical to segregate.

1.2 Solution studies

1.2.1 Paramagnetic NMR

Many of the relevant effects that the presence of paramagnetic metals has on the behavior of the protein nuclei have to do with the metal magnetic susceptibility (see later on). Most of these effects have been well known since the 50s and 60s, but a few have been described more recently. In any case, the theoretical and instrumental advances in NMR have promoted even the more exotic effects from the rank of curiosities to that of useful structural tools [24, 36, 37].

Pseudocontact Shift

The pseudocontact shift (PCS) is the effect on the nuclear chemical shift of the average residual dipolar interaction between the nucleus and the electron spin magnetic moment (see B.2.2). PCS are dependent on the position of the

nucleus in the frame provided by the susceptibility anisotropy of the metal center. In principle, if three different susceptibility anisotropy tensors from three different metals were known, the pseudocontact shift would provide enough information for locating all the protein atoms. In practice, the experimental indetermination is far too large to obtain an acceptable structure without the use of additional information.
The pseudocontact shift is a contribution to the nuclear energy levels and has the same symmetry of chemical shift with respect to spin and space operations. Thus it is preserved under Magic Angle Spinning in the solid state.

Residual Dipolar Coupling

Magnetic anisotropy induces difference in energy between different molecular orientation with respect to the magnetic field. Partial orientation in a high magnetic field can thus be induced by magnetic anisotropy.Magnetic anisotropy thus prevents the dipolar coupling energies from averaging to zero for all the pairs of atoms of the protein. Residual dipolar couplings (RDC, see B.2.2) in paramagnetic compounds, like PCS's, can be used as structural constraints. They can provide information on the relative orientation of internuclear pairs and, since they do not depend on the distance of the pairs from the metal ion, they are extremely long-range reporters of structural information.

Paramagnetic Relaxation Enhancements

In general, nuclei experience fluctuating fields, whose contribution to the spectrum averages to zero on the timescale of the NMR experiment. These fluctuations are not negligible anyway, because their overall effect is to perturb the distribution of populations (T_1) or dephase coherences (T_2). In the presence of unpaired electrons, the nuclear spins sense (a) a random fluctuation of the electron dipolar field (Solomon, Bloembergen and Morgan mechanism) and (b) a dipolar interaction with the average magnetic moment of the electron that "deforms" the magnetic field as sensed by the nuclei (Curie-

Spin mechanism). See appendix B.3.
As compared to PCS and RDC the averaging to PRE in multidomain systems is more complex: indeed it may be the movement of the domain itself to modulate PRE. The discussion is part of this thesis (2.2.3).

1.2.2 Small Angle X-Ray Scattering

Small-angle scattering (SAS) both of X-Rays and neutrons [38] is a universal low-resolution method to study native particles in solution and to analyze structural changes in response to variations of external conditions. SAS needs monodisperse solutions of purified macromolecules, but, normally, does not require special sample preparation. The scattering of X-Rays and neutrons yields information about the overall shape of the macromolecule, and, thanks to the recent progress in the analysis methods, particle shapes can be reconstructed from the SAS data ab initio. [39]

1.3 Solid State studies

1.3.1 Solid State NMR

Magic Angle Spinning

As already mentioned (1.1.2), nuclei are extremely sensitive reporters of the electronic and nuclear environment they sense. Such environment is usually anisotropic. The interactions that the nucleus sense (chemical shift and dipolar coupling for spins $S = 1/2$) usually span a large spread of values upon reorientation of the molecule with respect to the magnetic field. When working in solution all the anisotropic interactions are averaged due to random tumbling of the molecules, and the nucleus sense only the isotropic value. In the solid state the molecules do not tumble freely (with a few notable exceptions in highly symmetric compounds such as adamantane and fullerenes). [40]

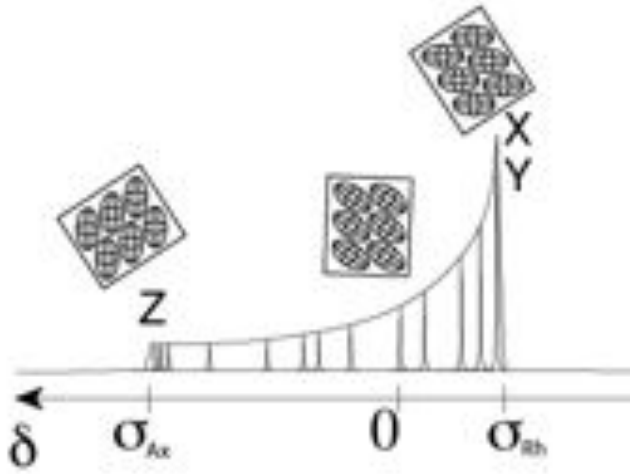

Figure 1.1: Axial CSA distribution. Crystallites corresponding to the extreme values (Z and XY) and to zero (at the magic angle 54 °73 ')

In a crystal, all the equivalent nuclei in different molecules sense the same orientation with respect to the magnetic field[1], thus the line is narrow and centered at some value, generally different from the isotropic one. If the crystal is rotated, the value changes. When the crystal is powdered, all the crystallites have different orientations, thus all the different frequencies are present at the same time and show up in the spectrum. This is summarized in figure 1.1 for the case of Chemical Shielding Anisotropy and discussed in deep detail by M. J. Duer. [41]
The appearence of the spectra is thus rather complicated also in the simple case of two-carbon compounds (1.2).

[1]assuming a single molecule per asymmetric unit

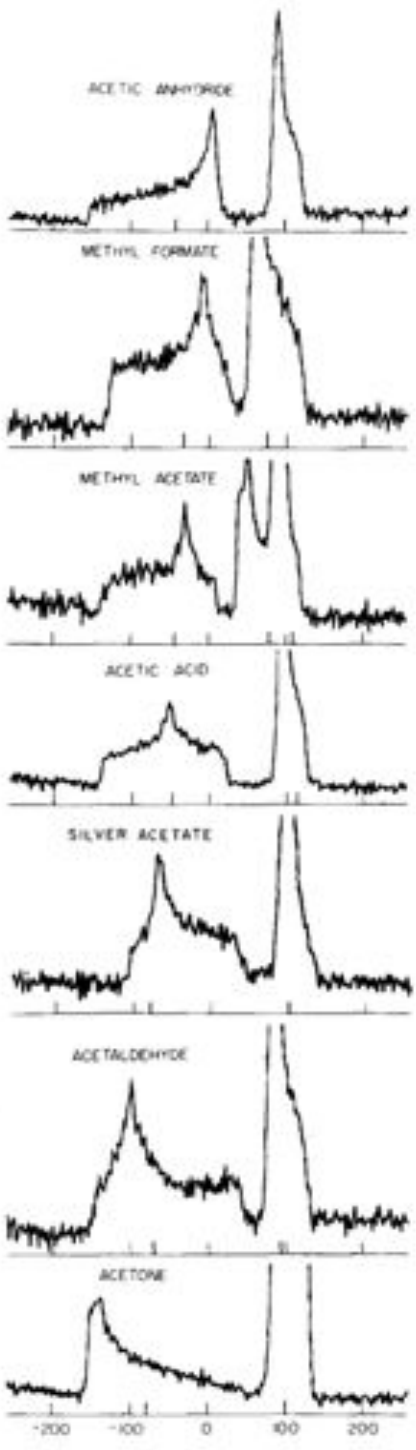

Figure 1.2: Proton-enhanced ^{13}C spectra of polycrystalline compounds containing carbonyl groups. The low-field peaks are from the carbonyl and the high-field ones from methyl groups. [42]

To partially solve this problem a coherent averaging can be introduced by rotating the sample at the magic angle, $\arccos(1/\sqrt{3}) = 54.73$. [43–46] This method is referred to as Magic Angle Spinning (MAS). To obtain efficient averaging, the sample should spin at a rate larger than the frequency of interactions. The maximum achievable rotation rate available in commercial probes is 70 kHz. The size of interactions is reported in table 1.1.

Interaction	**Frequency (kHz)**
^{13}C CSA	6÷ 22 (at 20 T field)
^{13}C-^{13}C DD	6
^{13}C-^{15}N DD	2
^{1}H-^{13}C DD	23
$^{1}H_{\alpha}$-$^{1}H_{N}$ DD	50
$^{1}H_{ali}$-$^{1}H_{ali}$ DD	210

Table 1.1: Frequency of interactions encountered in solid state NMR of biological solids

From the last line in table 1.1 it is immediately clear that proton detection in solid state NMR is severely hampered by a dipolar coupling that is so large that it cannot be removed by magic angle spinning. The discussion of the effect of dipolar coupling in the MAS spectrum of proton can be found in Duer. [41]

A brief summary of notable application and history of Solid State NMR

Until very recently, [47–53] solid state NMR has thus relied on exclusively heteronuclear detection.
Recent developments in sample preparation [47, 52, 54] and in theoretical background [55–60] have brought solid state NMR to compete with the stat-of-the-art solution NMR for the determination of structure [52, 61–64] and

dynamics [53, 65–67] in micro- to nanocrystalline protein preparations. Pioneering studies in the use of MAS in biological NMR have been made by the Griffin group at MIT, that contributed to the biological understanding of rhodopsin. [68–71] Soon the topic extended to cover other relevant targets. In particular, it is worth to mention the work by Griffin, [72] Meredith [73] and Tycko, who first provided the structure of a plaque-forming amyloid fibril in 2000, [74] because they undisclosed the key feature of solid state NMR: the ability to address at the atomic level biological questions about systems that lack long range order, such as amyloid fibrils, [75–90] or insoluble aggregates [91–94] for which X-ray crystallography is of limited use.

Sample preparation in solid state NMR

In order for MAS to be effective in the coherent averaging of the nuclear interactions it must have frequency higher than the interaction that it is meant to average, and at the same time incoherent molecular motions should be slower than in order not to interfere with it, [95] but even very large molecules have reorientational correlation times that are around 1-2 orders of magnitude faster than the usual spinning rates (3.1.2). [96] Thus MAS-NMR requires the proteins to be immobilized by some mean.
Over the years, different possibilities have been proposed and used, among them lyophilization [69, 97, 98] and freezing [98, 99] are the most notable. Freezing in the presence of cryoprotectants is one of the most common approaches to DNP sample preparation. [100] However lyophilized or frozen samples usually give rise to severely broadened spectra. [101, 102] The mechanism for this broadening is not yet fully understood. [103] Well ordered samples usually provide improved quality over other sample preparations. [101]
The use of crystalline preparations in solid state NMR dates back to the early '80, where the crystals of paramagnetic proteins were suspended in a saturated solutions and magnetically aligned. [104–106] Since then, the usual crystallization approaches for x-ray crystallography [5] have been extended to NMR, although smaller crystals are required for the latter. [101]

Chapter 2

Paramagnetic NMR for protein dynamics

2.1 Maximum Occurrence of conformations from average data

2.1.1 Contributions of this thesis

During the research work described in this thesis, I have contributed to this field of research through the following papers:

- "Conformational space of flexible biological macromolecules from average data" where the concept of Maximum Occurrence is proposed. [107]
- "MaxOcc: a web portal for Maximum Occurence Analysis", where the implementation of the algorithm in a user-friendly interface is reported and the results obtained from ensemble averaging are discussed. [108]
- "Paramagnetic relaxation enhancements for the characterization of the conformational heterogeneity in two-domain proteins", where PRE are used as restraints and their averaging is discussed [109]

The MO analysis on MMP1 is under consideration in a peer-reviewed international journal.

2.1.2 Ensemble averaging

In the case of flexible systems composed of entities of known, rigid structure, as it can be the case for protein complexes or multidomain proteins, the entity to which the metal ion is not bound experiences different positions and orientations with respect to the metal-bound entity. Once the magnetic susceptibility anisotropy tensor has been determined, PCSs and RDCs can be calculated for any hypothetical conformation. Possible ensembles of conformations can be determined by selecting the families of conformations providing averaged PCSs and RDCs in agreement with the experimental data. The same approach can be used for PREs only if the conformational averaging occurs on time scales longer than the rotational correlation time but shorter than the nuclear transverse relaxation times (see previous section). The rate of interconversion must be larger than the differences in chemical shifts corresponding to the different states. Differently, the motional modulations of PCS cause additional exchange contribution to the transverse relaxation rates, as shown for the case of the cobalt(II)-substituted two-domain protein PA0128, where the observed rate enhancements have been ascribed to a concerted domain motion on a time scale of a few milliseconds. [110] This was shown by ^{15}N relaxation rate measurements performed with relaxation-compensated CPMG sequences, which resulted dependent on the delays between π pulses due to the exchange process. It is clear that the limited number of experimental data resulting from averaging do not contain enough information to retrieve the ensemble itself. It can in fact be easily proven that multiple different conformational ensembles can reproduce any set of experimental data corresponding to systems with extensive conformational variability. [108] Therefore, no reliability can be granted to any member of the reconstructed ensembles of being really sampled by the system. In the case of proteins with restricted internal motions, ensemble average approaches have been used to recover the protein conformational variability using NMR data. [111–115] Paramagnetic restraints have also been used in this respect. [116, 117] Ensemble averaging can actually provide a clear indication of the presence of internal mobility and its extent, but only if the

smallest number of conformations that is sufficient to achieve fair agreement with the experiment is used. [118] The use of a redundant number of conformations in the ensemble can in fact introduce noise through the inclusion of counterbalanced conformations. On the opposite extreme, ensemble averaging approaches have also been applied for the characterization of intrinsically disordered proteins, using PREs arising from paramagnetic spin labels together with other NMR data. [119, 120] The resulting ensembles, containing a very large number of conformations, have been analyzed to obtain information on long range structural features. PRE-derived distances were also incorporated as ensemble-averaged restraints in molecular dynamics simulations to characterize the free energy landscape (expected to correlate with the statistical weights) of the protein α-synuclein as a function of the gyration radius and the solvent exposed surface area. [121] The quantity and quality of the experimental restraints that are needed to extract parameters providing an accurate representation of the system are however still debated. [122, 123] Paramagnetic spin labels have been used in several works through alternative attachment in different points of proteins, to detect the presence of conformational ensembles which include low weight conformations together with the known predominant conformation of the system: the paramagnetic broadening affecting the residues in close contact, although transiently, with the spin labels can in fact provide information on conformational variability. This is due to the sixth power dependence of the distance between the nucleus and the paramagnetic center, which causes very large PREs at short distances, so that conformations with nuclei close to the paramagnetic center, although sampled only for a short time, can be detected. For instance, in the case of the complex between cytochrome c and cytochrome c peroxidase, the available crystal structure is not consistent with the experimental paramagnetic enhancements observed for nuclei that, according to the crystal structure, should be far from the paramagnetic center. The PRE data can only be explained if $\approx$30% of the lifetime is spent by the complex in an ensemble of different conformations. [124–126] The latter can be visualized by mapping the areas around the spin labels giving rise to PREs and those not giving rise to PREs. In another study, spin labels were used to assess the presence of a

compact conformation, with a weight as low as 5%, in rapid exchange with the conformation seen in the crystal structure of the apo maltose-binding protein, where the amino-terminal and carboxy-terminal domains adopt an extended conformation. [127] Analogously, an equilibrium between open and closed conformations was found to occur for the tandem RNA recognition motif domains RRM1-RRM2, even in the absence of RNA ligands representing the polypyrimidine tract, [128] and Mn^{2+}-PREs indicated the presence of low weight conformations in a protein-DNA complex [129] with a known predominant structure. PREs induced by spin labels revealed conformational heterogeneity in the complex between the phosphocarrier protein HPr and the N-terminal domain of enzyme I. A single structure of this complex had been obtained through NMR using NOEs and RDCs induced by orienting media. [130] This structure, as well as any single rearrangement of the protein positions, could not account for the observed PREs, unless complemented by an ensemble of conformations amounting in total to 10% of the weight. [117, 131] These conformations indicate that the distribution of HPr molecules is qualitatively correlated with the negative electrostatic potential isosurface of the N-terminal domain of enzyme I. In the complex between cytochrome c and adrenodoxin, the lack of significant intermolecular PREs (expected from the paramagnetic FeS cluster contained in adrenodoxin) and PCSs (expected by changing the oxidation state of iron in cytochrome c) has been ascribed to extensive averaging among very different conformations. [132] The simulations showed in fact that cytochrome c must sample a large area of the adrenodoxin surface to reduce PCSs to insignificant values, in agreement with the large area of interaction identified by the chemical shift perturbation. This result was confirmed by the small RDC values measured for adrenodoxin when a lanthanide tag was attached to cyctochrome c. [133] In order to recover information on the structural ensemble experienced by a system composed of rigid bodies, in our laboratory, we have developed a rigorous approach that can be used to analyze averaged data resulting from the internal mobility of systems, in the absence of any predefined model for the overall structure and dynamics. [107] This approach, described in detail in the next section, does not provide a probability distribution for the pro-

tein conformations, but the largest possible weight that any conformation can have and still be compatible with the experimental data. This measure is called the Maximum Occurrence. Although there is no direct proof that Maximum Occurrence is related to the actual probability, strong experimental evidence (vide infra) is available in this sense. Furthermore , this is the only kind of information that can be reliably obtained from average data, given the redundancy of the solutions to the problem.

For many metalloproteins it is possible to substitute more than one paramagnetic metal ion in the same site [134, 135], without significant changes in the conformation of the protein. When metal substitution is not possible or inefficient, it is possible to anchor rigidly ligands for paramagnetic metal ions. [136–146]

2.1.3 Theoretical Background

A theorically exact method for deriving the MO value has been described using RDC data only. [147] Due to their nature, the dipolar interaction between nuclear pairs do depend only upon the relative orientation of nuclear pairs in the common frame defined by the magnetic susceptivity tensor $\boldsymbol{\chi}$. They do not depend on the distance of the pair itself from the metal ion, as described by the equation

$$\delta_{ab}^{rdc} = \frac{C_{rdc}}{r_{ab}^5} \mathbf{p}_{ab}^{\dagger} \boldsymbol{\chi} \mathbf{p}_{ab} \tag{2.1}$$

where C_{rdc} is a constant, $\mathbf{p}_{ab} = (x_{ab}, y_{ab}, z_{ab})$ is the position vector in the same reference system as the tensor $\boldsymbol{\chi}$ and u_{ab} is the difference between the u coordinate of the selected pair of atoms a and b.
Since the isotropic part of the magnetic susceptibility tensor does not influence the RDCs (i.e. such interactions average out to zero under isotropic tumbling), we may only consider the anisotropic part of the tensor, that has null trace. As noted elsewhere [148], data from different metal ions can be combined to remove some non-uniqueness issues.

The RDCs can monitor the motions that occur on a timescale faster than, or of the order of, $10^{-2}\,s$, i.e.: all longer than those timescales that are relevant for protein motions. Under these conditions singly averaged RDC values are obtained. These can be expressed in the form of

$$\bar{\delta}^{rdc}_{ab} = \frac{C_{rdc}}{r^5_{ab}} \int_{SO_3} (\mathbb{R}\mathbf{p}^{\dagger}_{ab})\boldsymbol{\chi}(\mathbb{R}\mathbf{p}_{ab}) dp(\mathbb{R}) \tag{2.2}$$

The orientation of one domain with respect to $\boldsymbol{\chi}$ can be represented by the rotation matrix $\mathbb{R}$, with an unknown probability measure p in the set of rotations SO_3. Using the mean magnetic susceptibility tensor $\bar{\boldsymbol{\chi}}$, defined by

$$\bar{\chi}_{ij} = \int_{SO_3} (\mathbb{R}^* \boldsymbol{\chi} \mathbb{R})_{ij} dp(\mathbb{R}) \tag{2.3}$$

the formula (2.2) can be recast in the form

$$\bar{\delta}^{rdc}_{ab} = \frac{C_{rdc}}{r^5_{ab}} \mathbf{p}^{\dagger}_{ab} \bar{\boldsymbol{\chi}} \mathbf{p}_{ab} \tag{2.4}$$

Clearly, the mean RDC of any pair can be calculated from $\bar{\boldsymbol{\chi}}$ that, in turn, can be evaluated from a number of mean RDC of one protein domain. Specifically, five values are enough if the mean RDC are exact, a larger number is needed in numerical approaches if errors are present.
In the case of a metalloprotein, $\boldsymbol{\chi}$ moves rigidly with the atoms of the domain containing the binding site. These can be used to evaluate the tensor itself. On the other hand, pairs belonging to a different domain reflect the movement of that domain with respect to that containing the binding site. Therefore, for this domain, we measure a mean tensor, averaged over all the orientations sampled by the moving domain.
From two extreme cases, it is already clear that this mean tensor contains information about the motion of the moving domain:

- Suppose the protein domain does not move, then $p = \delta_{\mathbb{R}_0}$ is a Dirac function for a particular rotation $\mathbb{R}_0$ and $\bar{\boldsymbol{\chi}} = \mathbb{R}_0^* \boldsymbol{\chi} \mathbb{R}_0$ has the same eigenvalues as $\boldsymbol{\chi}$;

- Suppose that all the orientations of the protein domain are equally likely, then $p(\mathbb{R}) = const. \forall\, \mathbb{R}$ and $\bar{\boldsymbol{\chi}} = 0$.

It is often very instructive to compare the effects of an averaged (panel B in Fig.2.1) with respect to a non-averaged (panel A in Fig.2.1) tensor on the distribution of RDC values: the ratio between the two tensors can be regarded as a generalized order parameter and can provide qualitative insights on the kind of motions that are involved. It is also to be noticed that the reorientation of the domain cannot be fully isotropic because of steric limits, resulting in a non-null average tensor (panel C in Fig.2.1).

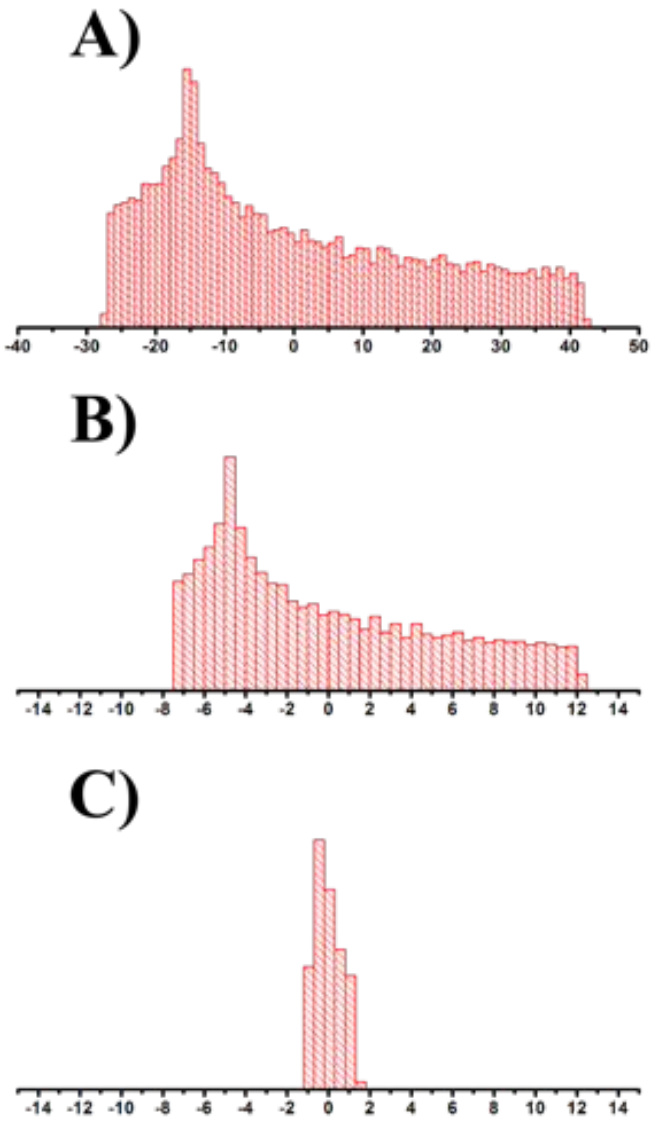

Figure 2.1: RDC distributions A) calculated in the limit of no movement, B) experimentally determined for MMP1 (generalized order parameter ~ 0.3 and C) calculated from a large ensemble of sterically allowed conformations.

Given a finite set of structures every magnetic susceptibility tensor can be expressed as:

$$\bar{\boldsymbol{\chi}} = \sum_i p_i \mathbb{R}_i^* \boldsymbol{\chi} \mathbb{R}_i \tag{2.5}$$

Let, for a set k of n metal ions

$$V_k = \left\{ \bar{\boldsymbol{\chi}}^k : \bar{\boldsymbol{\chi}}^k = \sum_i p_i \mathbb{R}_i^* \boldsymbol{\chi}^k \mathbb{R}_i, \sum_i p_i = 1, p_i \geq 0 \right\} \tag{2.6}$$

be the set of mean magnetic susceptibility tensor, then p_{max} is defined as [147]

$$p_{max} = sup \left\{ 0 \leq t < 1 : P(t) = \frac{\bar{\boldsymbol{\chi}}^{1,\ldots,m} - t\mathbb{R}_i^* \boldsymbol{\chi}^{1,\ldots,m} \mathbb{R}_0}{1-t} \in V^{1,\ldots,m} \right\} \tag{2.7}$$

From the RDC data sets, a solution can be found analytically: in a previous work [149], the orientational space for Calmodulin was sampled and the value of $p_{max}(\mathbb{R})$ was accurately calculated. As already mentioned, this quantity is an upper bound of the real probability $p(\mathbb{R})$. Unfortunately, there is no mathematical proof that the shape of $p_{max}(\mathbb{R})$ resembles that of the true $p(\mathbb{R})$. Yet the numerical tests gave clues that the most favoured orientations R assumed by one domain with respect to the other lay amongst those with a large value of $p_{max}(\mathbb{R})$. [150]

Since the PCSs depend on the distance between the selected atoms and the metal, they can discriminate between both orientation and spatial position of the protein domain. Here the disadvantage appeares that, being r_a not a constant, the data structure of RDCs that allowed the calculation of the mean tensor is lost. Thus the MO approach is still applicable but the theory underneath has not yet come to an analytical solution of the problem.

Independently from, but in strict analogy with equation (2.7), an analytical approach to obtaining p_{max} from PRE data was proposed. [124]

The equation there proposed can be generalized as follows: Let us assume that we have a number i of data to represent a conformation. According to the definition of MO we have to find the best completing ensemble. Let us

assume also that this ensemble is made only by one conformation represented by the data c_i^e and that $c_i^{min} \leq c_i^e \leq c_i^{max}$, i.e: that the ensemble has a finite compensation capacity with respect to the selected conformation.
In the mathematical form

$$o_i = pc_i + (1-p)c_e \tag{2.8}$$

This equation can be recast in the following form, from which we note that p_{max} is reached when the ensemble cannot compensate anymore for the selected conformation:

$$p_{max} = \min_i \left\{ \frac{o_i - c_i^{lim}}{c_i - c_i^{lim}} \right\} \tag{2.9}$$

where $c_i^{lim} = \begin{cases} c_i^{min},\ c_i > o_i \\ c_i^{max},\ c_i < o_i \end{cases}$

2.1.4 The MO profile

Let us assume that we have a single datum to represent a conformation. According with what defined above (2.8), the discrepancy between the calculated and the experimental value for this single datum can be evaluated as a function of the weight that the structure can have.

$$TF(p_s) = \min_{c_e}\{(c-o)^2\}. \tag{2.10}$$

In the simple, yet generic, case of $o = 0$,

$$TF(p_s, c_s) = \min_{c_e}\{p_s^2(c_s - c_e)^2 + 2p_s(c_s c_e - c_e^2) + c_e^2\}. \tag{2.11}$$

This is intuitively monotonically increasing. To extend this simple reasoning to an arbitrary number of correlated data is rather involved. yet, the take-home message is that the profiles of $\min\{T.f.(w)\}$ are overall monothonically increasing, as view in figure 2.2. By this consideration, any local stationary point is not possible, and if found during any calculation, it just points out

that a local minimum of the multidimensional topology was found and that the minimization procedure did not find the proper solution. The maximum allowed probability is then defined as

$$p_{max} = \{p : Tf(p)/Tf(0) \geq 1.1\}\,. \tag{2.12}$$

2.1.5 Computational aspects

The Algorithm for MO calculation

In a previous implementation, used for instance in [149], any conformation were generated during the calculations by rotation/translation of the moving domain and the relative data were calculated at each step. While this was still reasonable for PCS and RDCs data, the calculation of SAXS was excessively time-consuming, thus, as a part of this thesis, a new algorithm was developed for faster MO analysis. The calculation of MO proceeds through the steps that are summarized in figure 2.2. A large pool of conformations is generated, ideally covering all the accessible conformations without steric clashes (1). From this pool one conformation is selected and assigned a weight (2). Then, an ensemble is sought to fit the experimental data together with the selected conformation (3). To minimize the number of the conformations required to fit , the conformations in the ensemble have adjustable weight. This procedure is repeated increasing the weight of the selected conformer and searching for new ensembles, until no more ensembles can be found to fit the data (4).

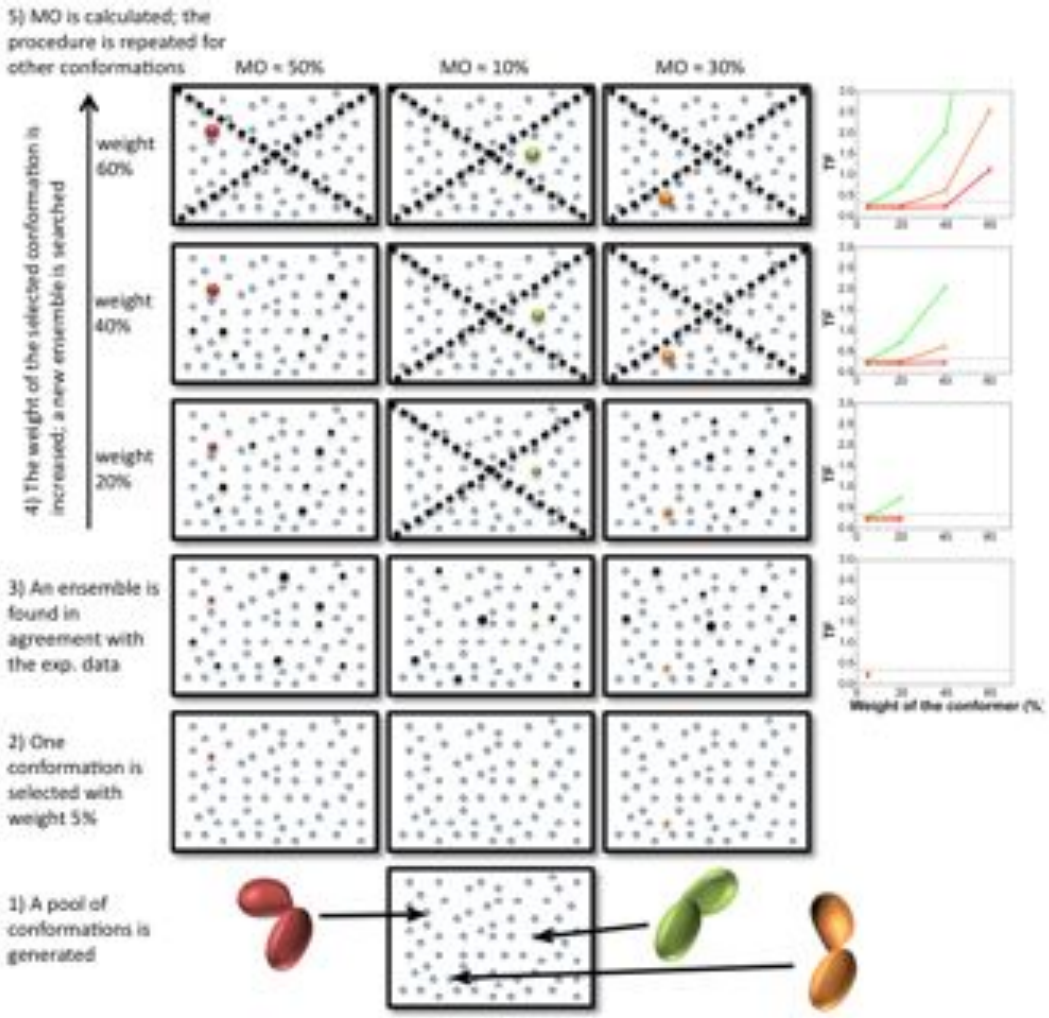

Figure 2.2: Pictorial view of MO calculations

Ensemble Optimization

Ensembles are built according to the protocol based on the EOM approach by Bernadó et al. [151]. Ensembles contain one predefined structure at a given weight. Structures are represented by arrays of numbers representing the corresponding experimental data. Such numbers are averaged over the

ensemble

$$v_{calc} = w_{pred} v_{pred} + \sum_{i=1,}^{N} w_i v_i \tag{2.13}$$

and compared with the experimental data to define a Target function of the form reported below (2.14). The weight of all the structures belonging to the ensemble is adjusted to give the best possible agreement via Conjugate Gradients [152] minimization. This ensures that a limited number of structures can be used also to fit RDCs. As described in the flowchart 2.3, the ensembles are iteratively built through several steps of simulated annealing. [107, 108]

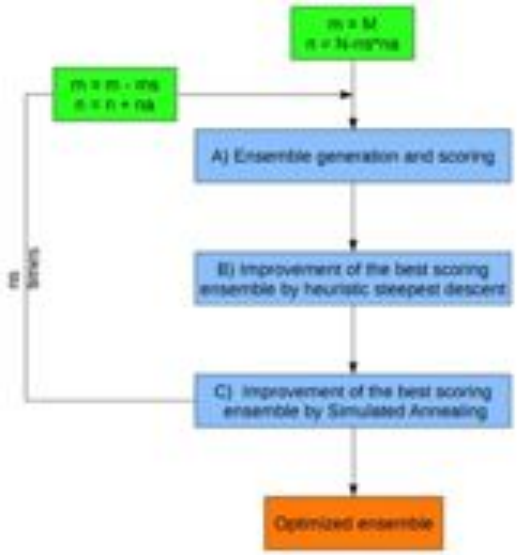

Figure 2.3: Current Algorithm for the Ensemble Optimization in the `MaxOcc` program

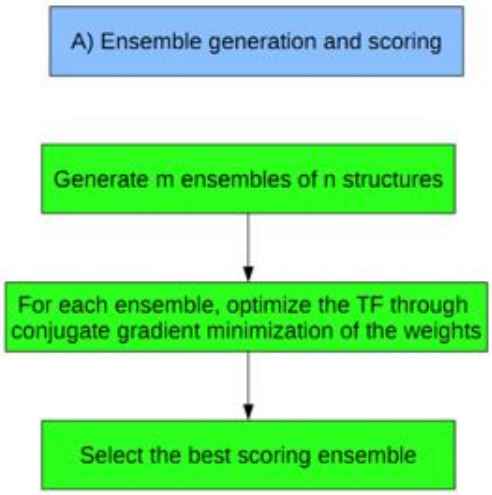

Figure 2.4: Description of the generation and initial selection of the ensembles used for the calculations process (step A).

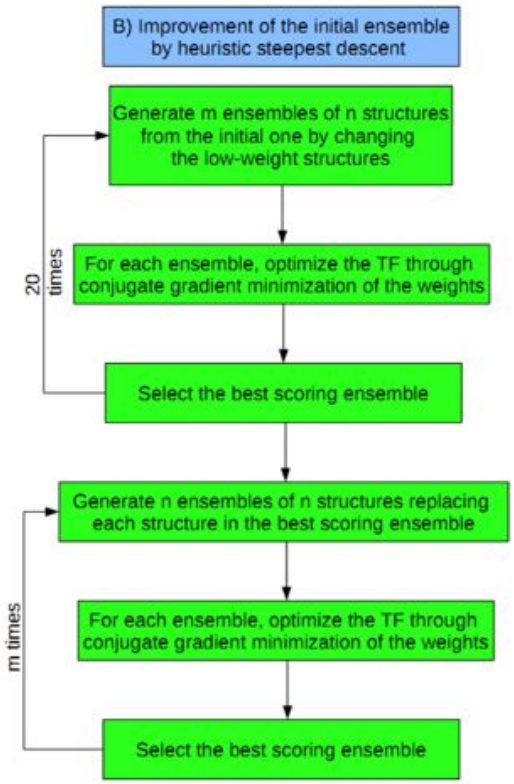

Figure 2.5: The best ensemble is optimized through several steps of heuristic steepest descent(step B).

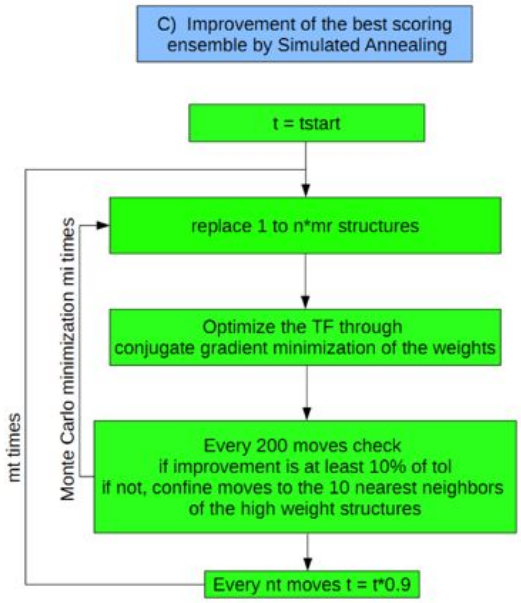

Figure 2.6: The best ensemble is optimized through several steps of simulated annealing, where the moves are preferentially confined to the nearest neighbors of the high-weight conformations (step C).

Each step, labeled with a capital letter in 2.3 is explained in deeper detail as follows:

1. The number m of ensembles that are initially calculated is set to M (400). Each ensemble comprises one selected conformation with a fixed weight and n other conformations randomly selected from the entire RanCh pool. This number n is calculated from the maximum number of structures that are required in the final ensemble, N, the number of steps ns, and the number of structures added at each step, na (6).

2. For every ensemble, the discrepancy to the experimental data is minimized by changing the weights of the randomly selected conformations through a conjugate gradient minimization (400). The ensembles are thus sorted according to their target function.

3. The best fitting ensemble is replicated m times and structures with low

weight are randomly replaced with structures belonging to the pool generated with RanCh. This is done twenty times or until no improvement is achieved. The best fitting ensemble is the starting point for m steps of heuristic steepest descent, during which the n completing structures composing the ensemble are randomly replaced.

4. The minimization proceeds with a simulated annealing schedule of mi (10) steps comprising mt (2000) random moves each, during which the structures of the completing ensemble are again randomly replaced with structures belonging to the RanCh pool. Differently from the previously published version, if the ensemble does not improve over all the random moves taken at the same temperature, the program switches to an informed version of the algorithm in which the moves are confined to the 10 nearest neighbors of the conformations having higher weights.

5. When the minimum temperature is reached, the ensemble is replicated in order to have a total number of ensembles decreased by ms (11) with respect to step 1) to reduce the computational burden and a number na (4) of new conformations are included to each ensemble. The whole schedule is repeated again (from step 2), until the number of conformations within the ensembles matches the maximum number N allowed by the user (50).

Each multidomain protein is represented by the structures of the rigid domains (in PDB format) and a chain of C_α's. One reference domain is held fixed, while the others are moved through rotation of the flexible linkers. Not all the conformations are allowed for linkers: the C_α trace cannot clash and the angles between the C_α cannot exit a quasi-ramachandran plot. [153] This search is accomplished using RanCh (Random Chain generator) from the EOM package [151] that was adapted to comply with the web implementation of MaxOcc. [108].

As mentioned above, the program performs the minimization of a combined

target function of form

$$TF = a_{PCS}TF_{PCS} + a_{RDC}TF_{RDC} + a_{PRE}TF_{PRE} + a_{SAXS}TF_{SAXS} \quad (2.14)$$

being

$$TF_{NMR} = \frac{\sum_{i=1}^{N_{NMR}} [(v_{i,exp} - v_{i,calc})]^2}{\sum_{i=1}^{N_{NMR}} v_{i,exp}^2} \quad (2.15)$$

$$TF_{SAXS} = \frac{1}{N_{scattering}} \sum_{i=1}^{N_{scattering}} \left[\frac{I_{exp}(s_i) - cI_{calc}(s_i)}{err(s_i)} \right]^2 \quad (2.16)$$

where c is a scaling factor calculated as

$$c = \sum_{i=1}^{N_{scattering}} \left[\frac{I_{exp}(s_i) I_{calc}(s_i)}{err(s_i)^2} \right] \left[\frac{I_{calc}(s_i)}{err(s_i)^2} \right]^{-1} \quad (2.17)$$

.

Overall T.F. The a_{PCS}, a_{PRE}, a_{RDC} and a_{SAXS} in (2.14) are arbitrary values to make all corresponding contribution to the target function comparable. To make the minimization with the weights by conjugate gradients as simple as possible, every i-th confromation is assigned its own weight w_i, instead of letting one have $w_N = 1 - \sum_{i=1}^{N-1} w_i$. Thus normalization must be accomplished through a term of the form

$$TF_w = 1000 \left(1 - \sum_{i=1}^{N-1} w_i \right)^2 \quad (2.18)$$

that is added to the overall target function.
To sample a reasonable amount of the conformational space, MO must be evaluated over hundreds of conformations. [107] The number of conformations to be scored varies with the quality of the data and is probably inversely proportional to the mobility of the system. We systematically evaluated how many conformations are needed to obtain a satisfactory representation of the conformational space of calmodulin. The evaluation was carried on

according to the following approach: MO was evaluated over the first 2000 conformations out of 50000 and the values plotted as a histogram. Then 4 groups comprising 100, 200, 300, 400 and 500 conformations out of the 2000 were randomly selected and the same histogram of MO values was prepared 2.7. If the number of conformations provides an appropriate description of the MO distribution, the histogram will not change significantly upon increase, as it is apparent on increasing above 400 conformations.

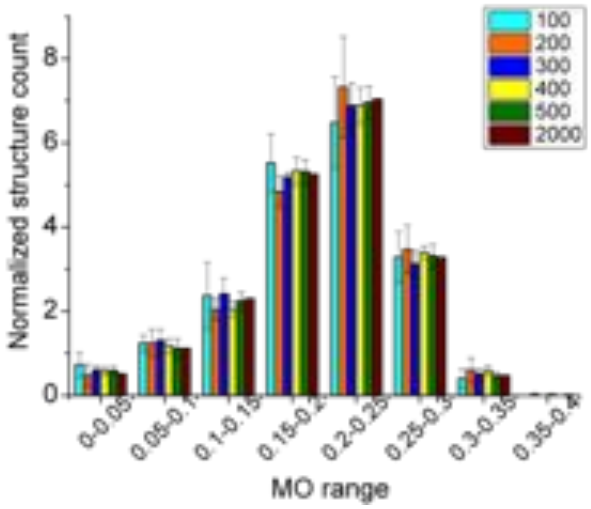

Figure 2.7: Comparison of the histograms of MO for 2000 calmodulin conformations and subsets comprising different number of conformations

Since every structure can be scored independently, the calculations can be performed through grid-distributed computing. For this reason, the program for the calculation has been developed on the e-NMR grid infrastructure [107] and then implemented in a user-friendly web interface (available at http://py-enmr.cerm.unifi.it/access/index/maxocc, under We-NMR). [108] The web interface is currently able to handle PCS, RDC, PRE and SAXS.

2.2 Calmodulin

2.2.1 Calmodulin biochemisty

Calmodulin (CALcium MODULated proteIN,CaM), is a ubiquitous calcium binding protein, involved in many physiologically relevant events linked to the sudden raise of calcium concentration in cells. Furthermore, this protein is a paradigm case in structural biology [135]. CaM is a two-domain protein belonging to the large family of EF-hand proteins. It contains 148 amino acid residues, organized into two domains of ∼70 aa each and connected by a short linker. Each domain is made up of two helix-loop-helix motifs (EF-hand motifs). Calmodulin binds up to four calcium (II) ions in the four EF-hand loops: one of them with a higher affinity ($K_{Diss} \sim 10^{-7}M$) so that it is able to bind even at the low concentrations in a resting cell, and more weakly in the other ($K_{Diss} \sim 10^{-6}M$).Upon metal binding, this protein undergoes a change in conformation to recognize, bind and activate, a number of proteins and enzymes.Early X-ray data [154, 155] reported the four-calcium $(Ca_2)_N(Ca_2)_C$CaM form to have a dumbbell shape, with helix 4, the last helix of the N-terminal domain, and helix 5, the first helix of the C-terminal domain, together with the interdomain linker, forming a long continuous helical structure (Fig.2.8).

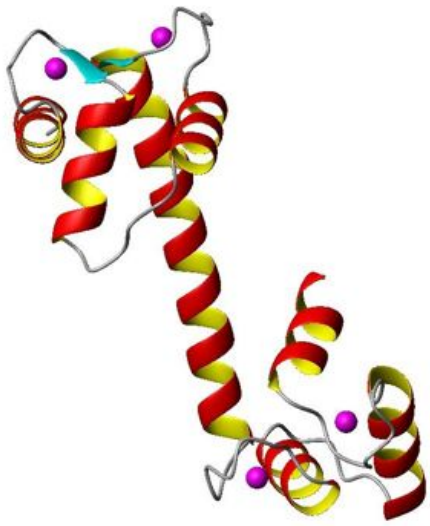

Figure 2.8: Crystal structure of Calmodulin [PDB 1CLL]

The fully elongated structure was consistent also with some X-ray scattering data. [156, 157]
On the other hand it was soon recognized that the NMR properties of the four-calcium form of CaM in solution were inconsistent with the rigid dumbbell shape observed in the early X-ray work, and that the central part of the helix loses its helical character and allows for reciprocal reorientation of the two domains (Fig.2.9) [6].

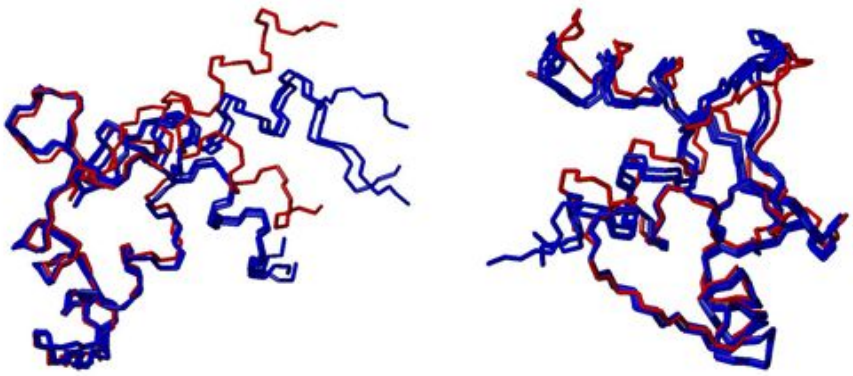

Figure 2.9: Comparison between the X-ray structure [1CLL] (red) and the solution NMR structures [1J7O, 1J7P] of the two domains of CaM

Molecular dynamics simulations confirmed the flexibility of the linker between the two domains. [158, 159]
An extended model-free analysis characterized the relative motions as occurring on a time scale of about 3 ns, with a squared order parameter of 0.7 with respect to the overall tumbling of the molecule, at room temperature [160]. When temperature was raised to 40 °C, a larger interdomain motion was observed, as a result of a doubling of the random coil residues in the central linker [161]. Disorder in the central part of the interdomain helix was observed in an X-ray structure at 1.0 Å resolution [162]. It has also been shown that native calcium-loaded CaM can crystallize in the closed conformation (Fig.2.10) [163].

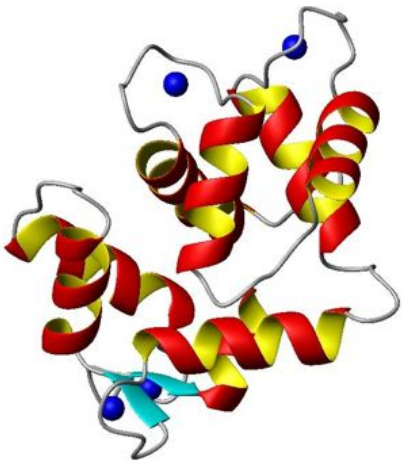

Figure 2.10: Crystal structure of compact Calmodulin [PDB 1PRW]

The p_{max} approach was applied to Calmodulin before the research for this thesis started. Through the analysis of PCS and RDC from three paramagnetic metals it was suggested that partially closed conformation may provide significant contribution to the real ensemble, with $p_{max} = 0.35$ (Fig.2.11). [149]

Figure 2.11: Structures with $p_{max} = 0.35$ obtained by combining RDC and PCS data

2.2.2 Joint use of SAXS and paramagnetic NMR

The SAXS data alone were found to be in fair agreement with a single CaM conformation, however, NMR data ruled out such possibility. The RDC measured for the C-terminal domain of CaM are reduced of about a factor of 10 with respect to what is expected in the case of the rigid adducts. [135] PCS, RDC and SAXS data were then used in conjunction to score 400 conformations on the basis of their maximum occurrence. As reported in figure 2.12, the crystal structures as well as the canonical closed conformations can contribute only up to 15% of the conformational ensemble, while structures with MO as large as 45% are present.

With these results we not only confirmed that CaM can undergo a very large mobility but also that two regions of the conformational space may be more sampled than the others. These regions correspond to partially bent structures, on the way towards the canonical closing position and on the opposite side. [107]

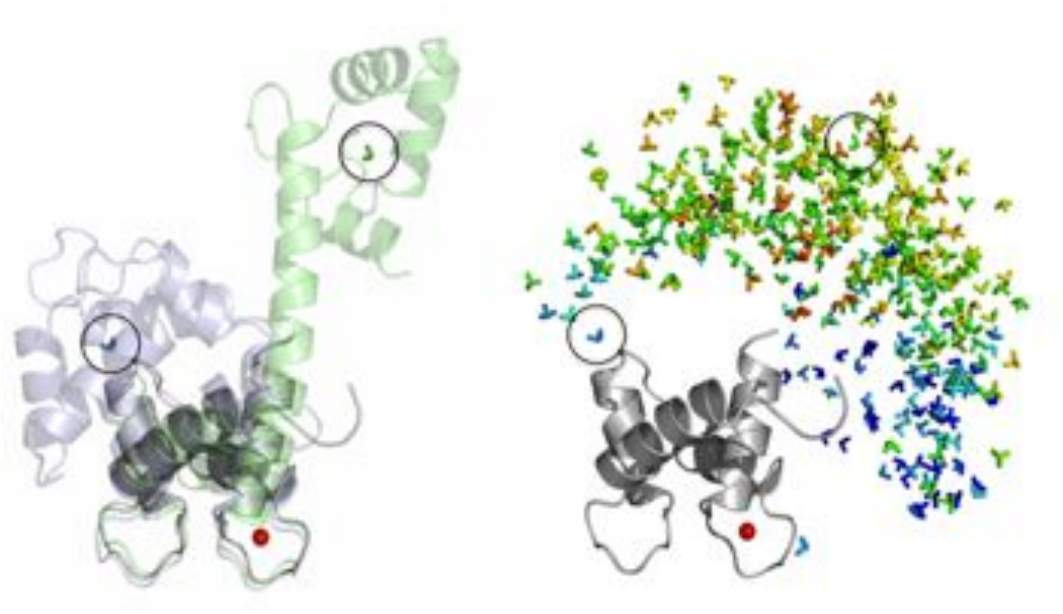

Figure 2.12: Pictorial representation of the MO of calmodulin conformations obtained with PCS, RDC and SAXS. Conformations are represented as a set of three orthogonal vectors (orientation cartesian tensor) rotated according to the rotation of the C-terminal domain with respect to the crystallographic position, centered at the center of mass of the C-terminal domain and color-coded according to their MO from blue (less than 5%) to red (more than 40%). In panel a) the two crystallographic structures (1CLL and 1PRW) are represented both as 3D model and with the tensor representation. In panel b) the MO of 400 conformations is represented. The crystallographic structures are marked with a circle.

Simulations

To understand how MO is related to the actual probability and how a MO distribution can map the original probability distribution, a series of synthetic tests was performed. A synthetic dataset is generated from an ensemble comprising 10 structures differing of about 30°in rotation from each other (black in figure 2.13). The MO of the same 400 conformations of figure 2.12 is calculated. It is evident that the conformations with high MO (bold in

figure 2.13) are nicely overlapped with the original ensemble.

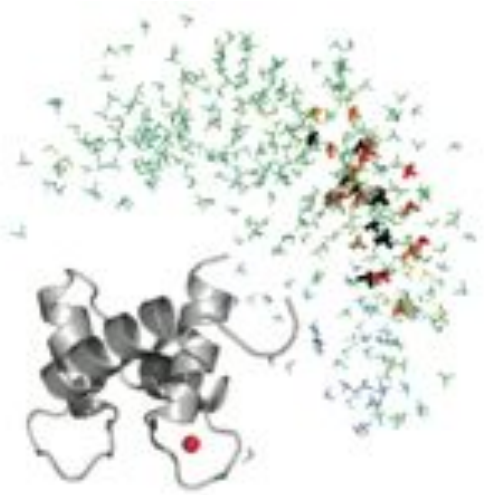

Figure 2.13: MO calculation for 400 CaM conformations with a synthetic dataset. The starting ensemble is represented in black. The conformations are represented as described for figure 2.12 and color coded from blue (less than 5%) to red (more than 40%). For the sake of clarity conformations with higher MO (more than 28%) are bold.

Contribution of the different experimental datasets

By the use of the web interface MO calculations were performed on CaM using different combinations of experimental datasets: RDC+PCS, SAXS and the three together, over the same number of conformations (300). [108] The MO distribution obtained from the 300 structures in the three cases is shown in figure 2.14. The MO values calculated from SAXS alone are sizably larger than the values calculated from the paramagnetic data. However, the different conformations are still discriminated according to the largest weight that they can have. In addition, when SAXS data are included together with the paramagnetic restraints, the number of conformations with the largest MO values in the distributions decreases sizably with respect to calculations performed with NMR data alone. Altogether, the present data thus indicate that SAXS by itself is less restrictive than NMR data by themselves, but still

useful to constrain the allowed conformational space. The simultaneous use of all the data is advantageous. The MO values resulting for 300 randomly selected conformations are shown in figure 2.15 with a color code, as in 2.12. The MO values of the different conformations ranged from 0 to 0.31 when all restraints were used, from 0 to 0.35 when the paramagnetic restraints only were used and from 0.23 to 0.72 when the SAXS restraints only were used. The larger the number of restraints, the smaller the MO values of conformations occurring less frequently (figure 2.15). The latter situation corresponds to an enhanced capability of the MaxOcc portal to distinguish the conformations that can occur only for small or very small time lengths and the conformations that may be populated in solution for relatively long times.

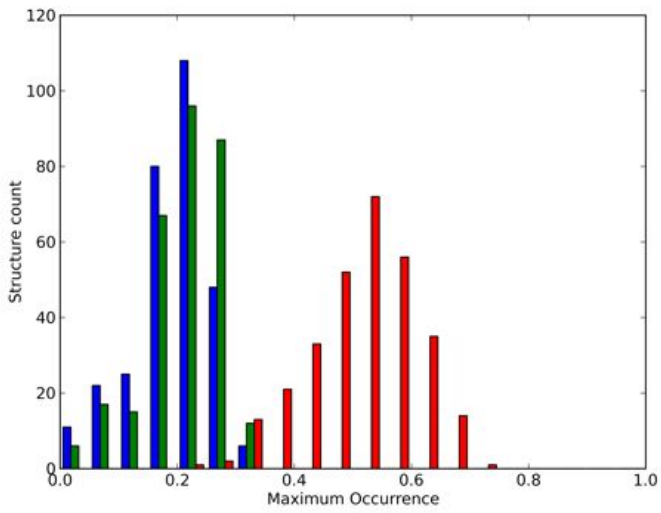

Figure 2.14: Distribution of the MO values calculated for 300 structures from SAXS restraints (red), PCS and RDC restraints (green) and both of them (blue).

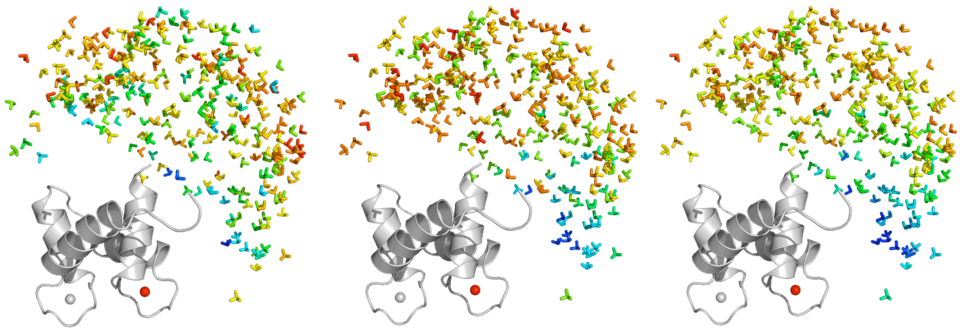

Figure 2.15: Orientation tensors positioned in the centre of mass of the C-terminal domain color coded with respect to the MO of the corresponding conformation. The N-terminal domain is depicted as a cartoon, with the position of the paramagnetic Ln^{3+} ion as a red sphere. MO values obtained from pcs, rdc and SAXS data (range 0.00-0.35, from blue to red) (right), from pcs and rdc data (range 0.00-0.35) (center), and from SAXS data (range 0.23-0.72) alone (left).

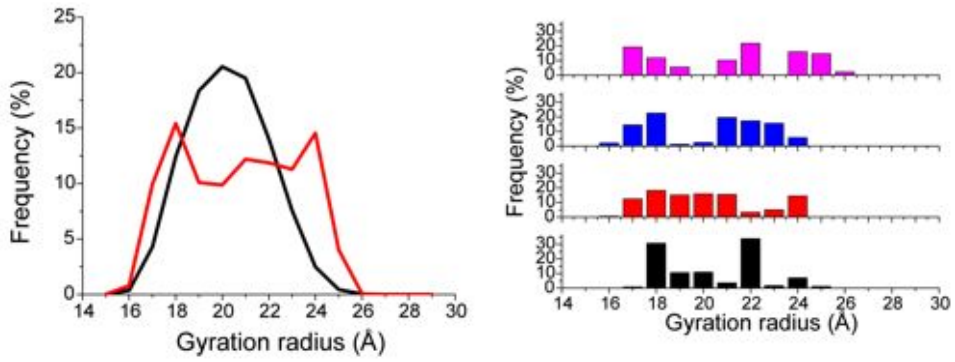

Figure 2.16: Frequency of the conformations as a function of the radius of gyration in the initial pool of structures with randomized interdomain linkers (black) and in the selected ensembles (red) (left panel). The latter distribution is obtained by the averaging of several calculations. The right panel shows the distribution of gyration radius within four different ensembles, all in featuring the same level of agreement with the experimental data, in order to appreciate the distribution variability. All histograms are normalized to the integral value of unity.

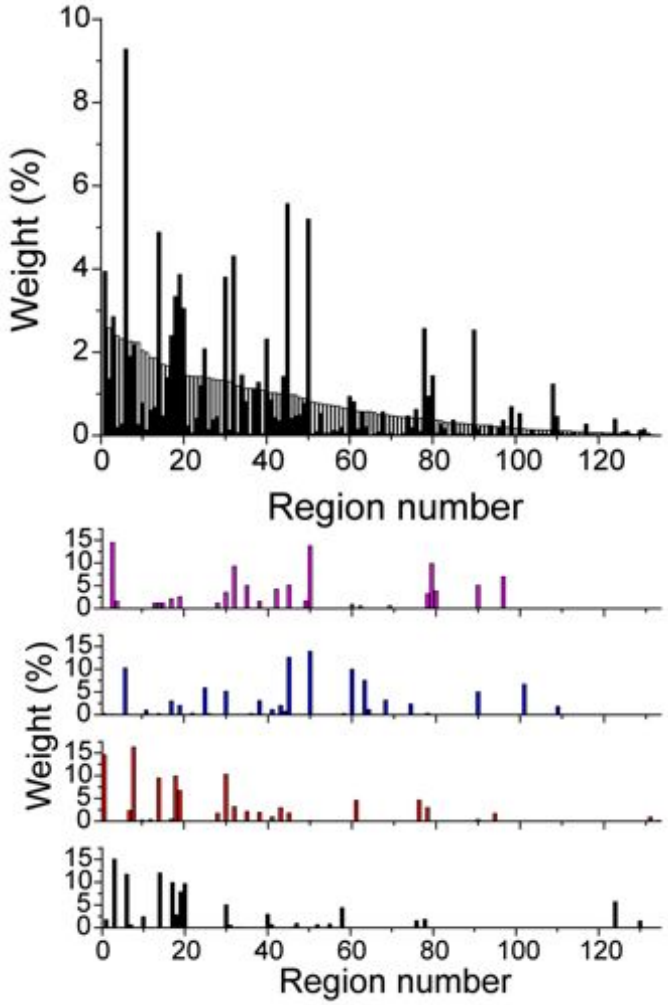

Figure 2.17: Total weight of the conformations within the 2,000 best fit ensembles for the different regions of the conformational space (see text) (black bars in the upper panel). The white bars indicate the percentage of structures within each region of the random pool of 50,000 structures. In the lower panel, the weights of the conformations belonging to the different regions calculated in four best-fit ensembles are shown.

From figures 2.16 and 2.17 it is important to notice that a defined set of substructures are selected preferentially over the random distribution generated by RanCh (left panel in figure 2.16 and top panel 2.17). But it is also important to notice that the preferential selection varies from one ensemble

to the next (right panel in figure 2.16 and lower panel 2.17).
This statistical analysis of the best fit ensembles can provide some information on the structural variability of the investigated system . [11, 111–113, 151, 164, 165]
This analysis is beyond the concept of MO, for which the MaxOcc program and portal have been developed. Nevertheless, MaxOcc does calculate the conformational ensembles in best agreement with the experimental data; this prompted us to analyze whether there are common features in these ensembles that can be used to characterize the system. 2,000 best fit ensembles of calmodulin conformations were calculated allowing the program to freely select 50 structures among the pool of 50,000 conformations, with a negligible weight for the imposed structure. Figure 2.16 shows the distribution of the radius of gyration for the pool of 50,000 conformations generated randomly by RanCh, compared to that calculated from all best fit ensembles of conformations. The former distribution is rather broad, peaked at 20 Å, and covers a range of gyration radii from about 16 to 25 Å, corresponding to compact and completely extended conformations, respectively. Instead, the distribution of gyration radii of the selected ensembles is flatter, with comparable contributions from relatively compact (18 Å) and extended (24 Å) structures. However, individual best-fit ensembles (any of which might actually represent the real system) differ considerably from one another, as shown in the right panel of figure 2.16. A common feature is the absence of any preference for structures with gyration radius around 20 Å. The calculated best fit ensembles are all composed of different conformations. In order to devise some overall similarities we clustered the conformations into regions, using the following relationship:

$$\Delta t + f(1 - \cos\alpha) \leq 10 \tag{2.19}$$

where Δt is the cartesian distance (in Å) between the center of the region and the center of mass of the C-terminal domain of the conformation under consideration, α is the angle between the quaternions corresponding to the average orientation of the region and that of the conformation under con-

sideration, and f was set to 26. In short, two conformations belong to the same region if their centers of mass coincide and their orientation differs for less than 100 °, or if their orientation is the same and they are translated by less than 20 Å. 133 regions were defined. In figure 2.17 the weight of the conformations belonging to all the 2,000 best fit ensembles for the different regions is reported. The percentage of structures within each region of the random pool of 50,000 structures is also shown as white bars. Analogously to what observed in Figure 2.16, the system tends to select structures in specific regions, independently of the distribution in the random pool. Anyway, the presence of most populated regions is not a guarantee that those regions will be populated in solution. The lower panel in figure 2.17 shows the weight of the conformations belonging to 4 different best fit ensembles, again clustered according to the same regions. It is clear that there is no univocal solution, i.e. conformations may or may not belong to any region. Therefore, these results suggest that when interpreting average data with ensemble approaches, only a large number of ensembles can grasp some of the features of the ensemble and that none of the obtained conformations is endowed with physical significance. [108] The use of MO is thus the only numerically accurate approach when structural information on the ensemble is sought.

2.2.3 The use of PRE

Theoretical discussion of the validity of the averaging approaches

The use of PRE as a restraint for ensemble averaging calculations is rather tricky. PREs are averaged over conformations with a rate of interconversion larger than the nuclear relaxation rate (which for proteins is usually of the order of 10^3 s^{-1} for transverse relaxation) and fast on the chemical shift timescale (differently, several NMR signals will be observed, each with its own relaxation rate). Differently from PCSs and RDCs, which amount to the weighted average of the values corresponding to all conformations sampled by the protein, in the presence of mobility PRE cannot be obtained from the weighted average of the PRE of the single conformations in all cases

when the time of interconversion among conformers is shorter or of the order of the time constants modulating the electron-nucleus interaction that gives rise to the PRE values themselves for the rigid conformations. Therefore, the model that should be used to analyze the PRE data depends on the correlation times of the intervening motions.
In the presence of a fast exchange regime, i.e. of a rate of interconversion over the different conformations much larger than the differences in nuclear chemical shifts among the conformations, when this rate is large with respect to the relaxation rates but slower than the inverse of the correlation time for the relaxation process itself, averaged relaxation rates are obtained depending on $\langle r^{-6} \rangle$, where r is the nucleus-electron distance in the different conformations (to be noted that $\langle r^{-6} \rangle$ can be sizably larger than $\langle r \rangle^{-6}$). If, on the contrary, the conformational changes occur on time scales shorter than or of the same order of the overall tumbling time of the protein in those particular conformations and of the electron relaxation time, this time of interconversion among different conformations becomes a correlation time modulating the electron-nucleus interaction. The influence of this time on the nuclear relaxation rate depends on the degree of spatial restriction of the motion. An order parameter is usually defined according to the model free and the N-site discrete jump model. This order parameter has a radial dependence and an angular dependence. The former indicates the fluctuations of the distance r between the dipolarly coupled spins in the sampled conformations; the latter the fluctuations in the orientation of the vector connecting the dipolarly coupled spins. As far as the radial dependence is concerned, the resulting averaged relaxation rates depend on $\langle r^{-3} \rangle^2$. [166]
In summary,

$$\Gamma_2^{PRE} = \langle r^{-6} \rangle \left[S_s^2 f(\tau_c) + (1 - S_s^2) f(\tau_f) \right] \tag{2.20}$$

with $f(\tau) = k'_{SBM} + k'_{Curie}$ and $\tau_f^{-1} = \tau_c^{-1} + \tau_{interdomain}^{-1}$. The S_s^2 order parameter is given by [166–168]:

$$S_s^2 = \frac{1}{\sum_i w_i r_i^{-6}} \sum_{i,j} \frac{w_i w_j}{r_i^3 r_j^3} \left[\frac{3}{2} \left(\frac{\mathbf{r}_i \cdot \mathbf{r}_j}{r_i r_j} \right)^2 - \frac{1}{2} \right] \tag{2.21}$$

where i and j are indices cycling on the elements of the structure ensemble which are sampled by the system during its overall reorientation time, r_i are the nucleus-metal vectors and w_i the weights of the corresponding conformations. Whereas all conformations sampled by the system should be considered for the calculation of $\langle r^{-6} \rangle$ in Equation (2.20), if the interdomain reorientation occurs in a time shorter than the overall reorientation time of the protein, only the subset of the protein conformations sampled during the reorientation of the protein must be considered in the calculation of S_s^2. Since the Lorentzian dispersions in k'_{SBM} and k'_{Curie} provide a negligible contribution with respect to the non dispersive terms for correlation times of the order of nanoseconds, an effective correlation time, $\tilde{\tau}_c = S_s^2 \tau_c + (1 - S_s^2)\tau_f$, can be defined to take into account both the overall protein reorientation time and the fast motions, [109, 166–170] besides the electron relaxation time. If the Curie contribution is not negligible, an effective correlation time, $\tilde{\tau}_{Curie}$, can also be defined, which does not depend in any case on the electron relaxation time but only on the global and local mobility. This results in

$$\Gamma_2^{PRE} = \langle r^{-6} \rangle \left[\left(\frac{\mu_0}{4\pi} \right)^2 \frac{\gamma_H^2 g_e^2 \mu_B^2 S(S+1) 4\tilde{\tau}_c}{15} + \frac{1}{5} \left(\frac{\mu_0}{4\pi} \right)^2 \frac{\omega_H^2 g_e^4 \mu_B^4 S^2 (S+1)^2 4\tilde{\tau}_{Curie}}{(3k_B T)^2} \right] \tag{2.22}$$

If $\tau_e > \tau_r$, the Curie contribution can be neglected and the value of $\tilde{\tau}_c$ can be empirically determined in order to calculate PRE values in best agreement with the experimental data. [109] A value for τ_e can be estimated from the kind of involved paramagnetic center and the applied magnetic field.The value for τ_r can be estimated from the protein molecular weight, calculated through the hydrodynamics properties with HYDRONMR, [171] or obtained experimentally from 15N relaxation measurement.

Results

As noted before (2.1.3) the Maximum Occurrence of a conformation can be calculated analytically applying equation (2.9).

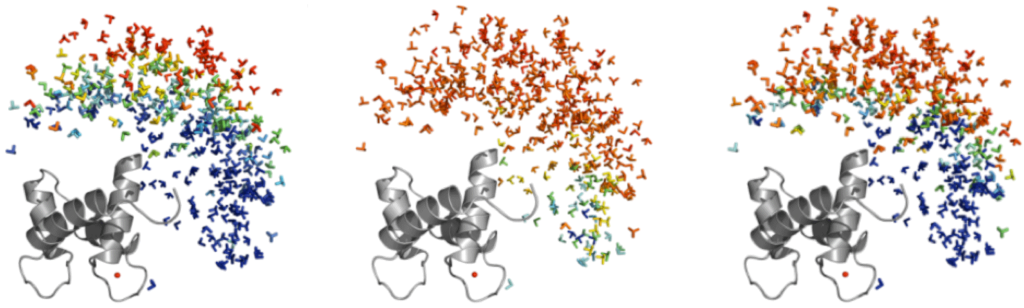

Figure 2.18: MO from PREs alone (left, range 0-1), from PCSs and RDCs (center, ranges: 0.05-0.3) and from the minimum between the two calculations made separately (right, ranges: 0.05-0.3) at 298 K.

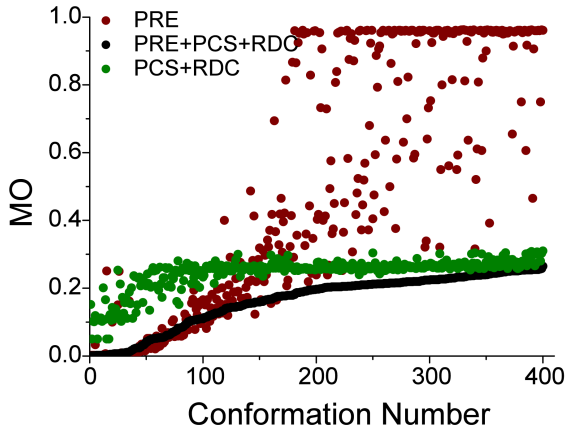

Figure 2.19: MO calculated for the selected 400 conformations from PREs, PCSs and RDCs (black symbols), PCSs and RDCs (green symbols) and PREs only (red symbols) at 298 K. The conformations have been sorted according to the MO values calculated from PRE, PCS and RDC data.

In order to evaluate the effect of PRE data on the MO analysis, the MO

values calculated using the three sets of data simultaneously were compared to the MO values calculated from the minimum of the MO values obtained using PRE data separately from PCSs and RDCs. Figure 2.18 shows the MO calculated from PREs alone (left), and the MO calculated from PCSs and RDCs (center). The minimum of the two is depicted in the right panel. Each map is color coded with respect to the minima and maxima of the MO obtained with that specific set of data. It is immediately evident that although PRE restraints alone are quite effective in reducing the MO of some more compact conformations, they are not very efficient in the case of extended conformations. On the contrary, PCSs and RDCs are equally effective in restricting the MO of all conformations in the overall conformational space. These two sets of independent information can complement each other for providing a more informative MO ranking of the protein conformations. The MO values of the 400 conformations calculated from all the three sets of restraints simultaneously, or from the PRE data, or the PCS and RDC data alone, are better shown in figure 2.19, sorted by ordering the conformations by increasing MO values calculated from all data: the limiting MO is determined by PREs for some conformations and by PCSs and RDCs for others. To be noted that the maximum MO calculated for any of the 400 conformations, using only the PRE data, is not 1 but 0.96. Therefore, even the extended conformation cannot reproduce the PRE data, if taken alone. This is in agreement with the results recently obtained through the analysis of PRE data arising from the presence of spin labels attached to the protein. [172] On the other hand, the calculations performed using PCSs and RDCs indicated that no single conformation can have a weight larger than 0.3.

2.3 MMP1

2.3.1 Implications of mobility in MMP1 biochemisty

Matrix Metalloproteinases (MMPs) are a family of proteases capable of hydrolizing variety of structurally unrelated substrates. The breadth of the

enzymatic activity is the result of evolution in the function of protein domains and the flexibility of the overall shape. In particular, several active MMPs, including MMP-1, are two-domain [catalytic (CAT) and hemopexin-like (HPX)] enzymes capable of catalyzing the hydrolysis of highly structured substrates such as triple-helical, interstitial (types I-III) collagen
Interdomain flexibility has a capital role in the hydrolisis of type I triple helical collagen. In this case flexibility is particularly important to allow for unraveling the collagen and accomodate the otherwise inaccessible peptide chain in the active site.
The active form of MMP1 was observed by NMR and SAXS to experience sizable interdomain flexibility [173],that may lead to an open-closed equilibrium. The compact arrangement of the two domains on MMP1 observed in the crystal structures (1SU3, 2CLT, 4AUO) is not fully representative of the conformations sampled by the protein in solution: for at least one third of the time, the enzyme exists in a extended conformation, [173] consistently with the idea that CAT-HPX should move apart to reveal some exosites involved in substrate recognition. In the crystal structure, some of the residues of HPX domain, essential for collagen binding (291,292,301,311-326) are buried between the two domains and inaccessible to the substrate. [174] More recently, the interaction of MMP-1 with a THP has been investigated utilizing NMR spectroscopy, leading to a plausible multistep mechanism for collagenolysis. [175] In this model, the initial binding of the HPX domain to the THP is followed by the interaction of the CAT domain with the THP in front of the cleavage site, and by a subsequent back rotation of the CAT and HPX domains toward the closed conformation that drives the unwinding of the triple-helix and causes the displacement of one peptide chain into the active site. While there is experimental evidence for the formation of the initial MMP-1/THP complex, the mechanism that leads to the observed two-domain interaction is still unclear.

2.3.2 MO analysis

The MO analysis was performed using as restraints the motionally averaged PCSs and RDCs for the HPX domain and the SAXS data.
Most of the conformations (80%) have a MO smaller than 20%. Only 0.3% have a MO larger than 40%. To visualize the results, the CAT domains of all the structures were superimposed and the position of the HPX domain was schematized as done above (figure 2.12) for calmodulin. Tensors were color-coded with respect to the MO of the corresponding conformations, from blue (MO lower than 5%) to red (highest MO, 47%) Figure 2.20.

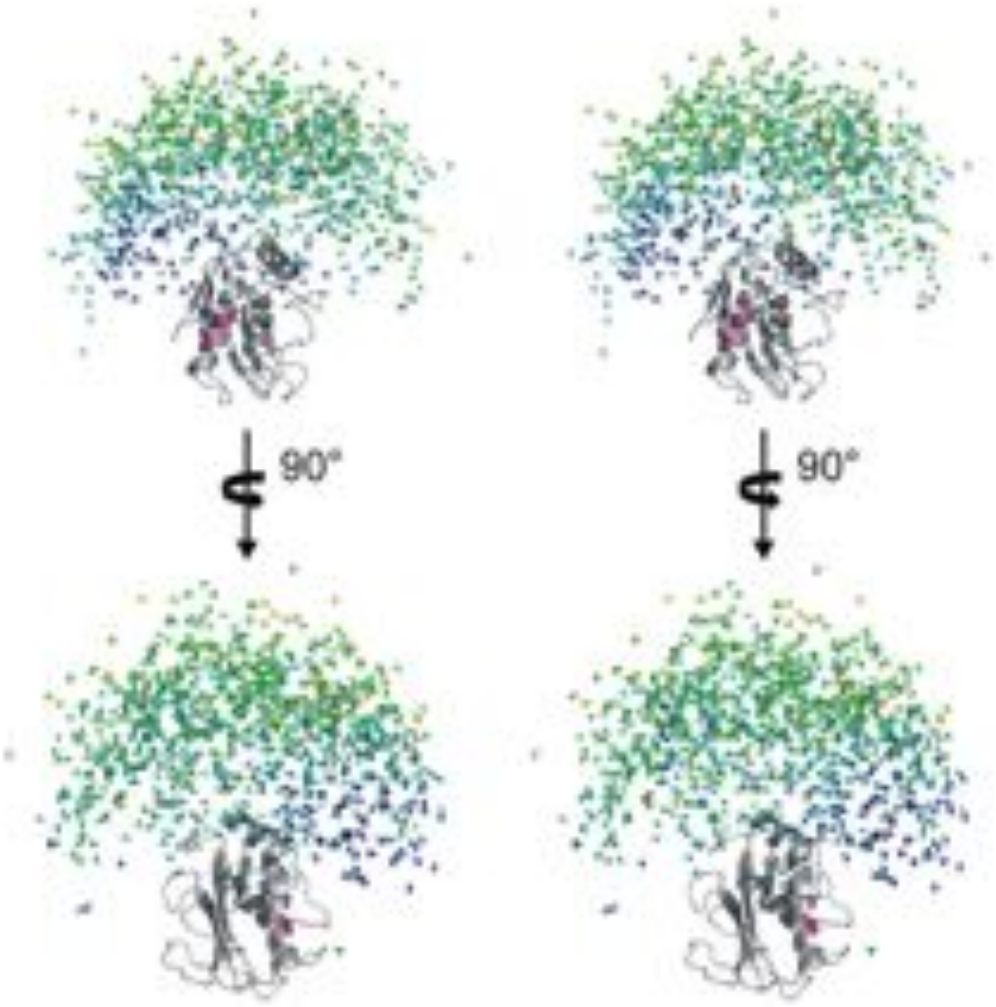

Figure 2.20: Stereo view (cross-eye) of the MO of MMP-1 conformations obtained with PCS, RDC and SAXS. The MO of 1000 conformations is represented.

The conformations having the HPX domain in the region proximal to (Ln^{3+})CLaNP-5 (and distal to the catalytic site cleft) were found to have a negligible weight in solution, with MO values below 5% (blue tensors in Figure 2.20). Thus, these conformations are not sampled significantly by the protein. A striking finding is that most of the conformations with the

highest MO (orange-red tensors in Figure 2.20) are clustered in a well-defined region of the distribution, corresponding to relatively elongated structures. To increase the resolution of the regions populated by the conformations with the highest MO, the MO values of 281 additional conformations near those with high MO, have been evaluated (Figure 2.21).

Figure 2.21: Stereo view (cross-eye) of the high-MO conformations of MMP-1

All available X-ray structures of human full-length MMP-1 (pdb entries: 1SU3 [176] and 2CLT [177, 178]) display relatively closed conformations. It is crucial to understand how much these structures are represented in the ensemble sampled by the protein in solution. The MO values obtained for all the X-ray [178, 179] structures are below 20%. The radii of gyration (R_g) of the two crystallographic structures are 25.5 and 25.7 Å, respectively, whereas the structures with highest MO (>35%) have R_g of 28.9 ± 1.3 Å. This range of R_g is in better agreement with the experimentally determined values from the SAXS data alone, confirming that the X-ray structures are more compact than the average solution conformation. Furthermore, the relative orientations of the HPX and CAT domains in the structures with the highest MO are different from those in the X-ray crystallographic structures.

High MO structures as the prologue to collagenolysis

Analysis of the conformational space experienced by the free protein is useful to investigate the mechanism of binding, and to determine the role of the domains in the identification of substrates, to predict new possible substrates, and to investigate natural and new mechanisms of inhibition. [114, 117, 126, 170, 172, 180, 181]
The MO analysis performed for MMP-1 can provide a description of the conditions that are required for the onset of catalysis. The MMP-1 conformations with large MO values (up to 47%) are restricted into a relatively small conformational region, and all conformations with high MO values largely differ from the closed MMP-1 structures obtained by X-ray crystallography. In all the high MO conformations, the CAT and HPX domains are not in tight contact, and the residues of the HPX domain reported to be responsible for the binding to the THP are solvent exposed.

This indicates that all highest MO structures are characterized by an interdomain orientation and position that can be defined relatively well.
In the highest MO structures, the residues of the HPX domain essential for the binding to collagen are not buried between the CAT and HPX domains, and the open space between the two domains is wider than in the crystallographic structures. Furthermore, and more importantly, the secondary binding sites (exosites) of the HPX domain responsible for collagen interaction, and the active site of the CAT domain, face the same side. If triple helical collagen is modeled in its experimentally determined bound position to the HPX domain, [175] the CAT domain faces closely the collagen cleavage site, and in about half of the highest MO structures even penetrates the triple helical substrate. Actually all the high MO conformations (MO >35%) fall along the boundary between the penetrating and non penetrating conformations. Therefore, the largest MO conformations sampled by MMP-1 in solution appear to be much more poised for interaction with collagen than the compact X-ray crystallographic structures. Therefore, the high MO conformations that are not colliding can be seen as a possible antecedent step for the recently proposed mechanism of collagenolysis. [175]

Chapter 3

Sed-NMR

3.1 Sed-NMR

3.1.1 Contribution of this thesis

During the research work described in this thesis, I have contributed to this field of research through the following papers:

- "Solid-state NMR of proteins sedimented by ultracentrifugation" where the method is first proposed. [182]
- "NMR properties of sedimented solutes", where the theory is further developed, the effect of sample preparation are discussed and the use of ultracentrifugal devices is proposed. [183]
- "On the use of ultracentrifugal devices for sedimented solute NMR", where the use of ultracentrifugal devices is explained and theoretically described and the guidelines for device production are given. [184]

During my Ph.D. I spent a period at MIT under the guidance of Prof. Dr. Robert Guy Griffin: the methodology of SedDNP was experimentally verified. The results are under consideration in a peer-reviewed journal.

3.1.2 Theoretical Background

Magic angle spinning (1.3.1) is required to obtain high resolution solid state NMR spectra. It can also be used to increase the resolution in inhomogeneus liquids/suspensions. In this case, Magic Angle Spinning generates a moderate to strong centrifugal force on the solutes. Even separation of cream from milk was observed under moderate MAS. [185] Based on this observation we developed the technique referred to as "Sedimented Solutes NMR" (SedNMR). [182, 183]

Depending on the molecular weights of the solutes, they can form a concentration gradient. For a protein of molecular weight M, density $\rho_{protein}$ in solution in a solvent of density $\rho_{solvent}$, sealed in a rotor of radius b and spinning at ω_r the equation describing the concentration as a function of the distance from the rotation axis is the following [186]:

$$c(r) = \frac{RT}{M\left(1-\rho_{solvent}/\rho_{protein}\right)\omega_r^2 r}\frac{dc}{dr} \tag{3.1}$$

an empirical limit to the concentration can be imposed, [187] reflecting a phase separation, modifying the previous equation into

$$c(r) = \frac{RT}{M\left(1-\rho_{solvent}/\rho_{protein}\right)\left(1-c/c_{limit}\right)\omega_r^2 r}\frac{dc}{dr} \tag{3.2}$$

Introducing, for the sake of simplicity

$$K^{Sed} = \frac{M\left(1-\rho_{solvent}/\rho_{protein}\right)\omega_r^2}{RT} \tag{3.3}$$

equation (3.2) recasts into:

$$\frac{dc/c_{limit}}{dr} = K^{Sed}\frac{c}{c_{limit}} r\left(1-\frac{c}{c_{limit}}\right) \tag{3.4}$$

Substituting for

$$x = \frac{c}{c_{limit}} \tag{3.5}$$

and separating the variables, the following form is obtained:

$$\frac{dx}{x(1-x)} = K^{Sed} r \, dr \tag{3.6}$$

That integrates to

$$c(r) = \frac{c_{limit}}{A \exp\left[-\dfrac{M\left(1-\rho_{solvent}/\rho_{protein}\right)\omega_r^2 r^2}{2RT}\right] + 1} \tag{3.7}$$

For the cylindric geometry of the problem, for a rotor of radius b and a starting concentration c_0, the integration constant A can be evaluated analytically:

$$A = \frac{\exp\left[\dfrac{M\left(1-\rho_{solvent}/\rho_{protein}\right)\omega_r^2 b^2}{2RT}\left(1-\dfrac{c_o}{c_{limit}}\right)\right] - 1}{1 - \exp\left[-\dfrac{M\left(1-\rho_{solvent}/\rho_{protein}\right)\omega_r^2 b^2}{2RT}\dfrac{c_o}{c_{limit}}\right]} \tag{3.8}$$

To exemplify the meaning of this equation, concentrations profiles for apoferritin of 480 kDa at 60 mg/ml in a 4 mm rotor are reported in figure 3.1. [183]

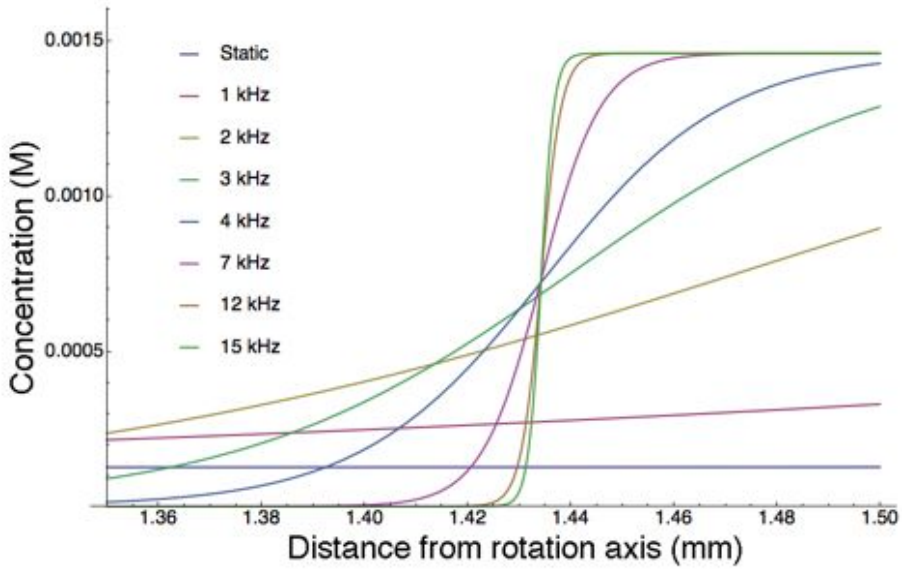

Figure 3.1: Concentration as a function of the distance from the rotor axis at different spinning rates for apoferritin (MW 480 kDa and concentration in the static solution 60 mg/ml), in a 4 mm rotor (1.5 mm internal radius). The limiting concentration is set to 700 mg/ml.

At the MAS rate attainable in commercial probeheads, molecules above 30 kDa molecular weight can sediment. By sedimentation, a protein may reach concentrations of the order of 700 mg/ml. [188, 189] These concentrations, that are not reached even inside the crowded environment of a cell, [190] cause a) the translational diffusion of the molecule to stop, b) the rotational diffusion to slow down. [191] If rotational diffusion is slowed and interconversion between different orientation occurs at a frequency lower than the frequencies of the nuclear interactions, the system appears as a solid, thus it can be polarized by CP and be observed as a solid. [182, 183, 192] If translational diffusion is hampered, the system appears as a glass. It has been experimentally observed that such glass has no long range order. [193]
This approach, called "MAS-induced" or *in-situ* sedimentation, was applied to apoferritin (480 kDa), [182] bovine serum albumin (64 kDa) and carbonic anhydrase II (30 kDa). [183]

High resolution spectra are obtained, [182] with transverse relaxation rates up to 5 times longer than what observed for the microcrystals: [183] sediments are, in principle, more amenable for multidimensional NMR than the corresponding microcrystals.

In situ sedimentation allows for the study of large complexes that exceed the size for satisfactory characterization by solution state NMR. This category comprises a number of biologically relevant targets, such as complexes and oligomeric aggregates.
In particular, the possibility of having a water solution containing the interacting partners allows for detection of small proteins in large complexes, either stable or fluxional. In fact, under sedimentation, the complexed molecules are subtracted from the bulk solution, so that the equilibrium is further shifted towards the formation of the complexes. On the other hand, at the sediment level, the kinetics of dissociation is expected to be hampered by self crowding.
When MAS is stopped, the protein tends to diffuse back from the sediment layer to the bulk solution, in a time that depends on the translational diffusion coefficient of the protein: at some point, the sample will be again in the initial buffer/concentration conditions. This allows for the sample handling that is common for solution state NMR, even allowing for titration. [194]

Anyway, this approach suffers from an intrinsic limitation: at its best, a solution sample contains no more that 2/3 of the corresponding microcrystalline preparation [183]: also in the case of extremely soluble proteins dissolution at concentrations higher than 400 mg/ml is unpractical while the average concentrations in crystals is of the order of 650 mg/ml [195]. This results in lower overall sensitivity of the experiment.
To partially alleviate this problem we have proposed to pre-sediment the sample *via* common ultracentrifugation, [183] allowing one to fill the rotor with the sedimented material ("UC-induced" or *ex-situ* sedimentation). This is summarized in figure 3.2

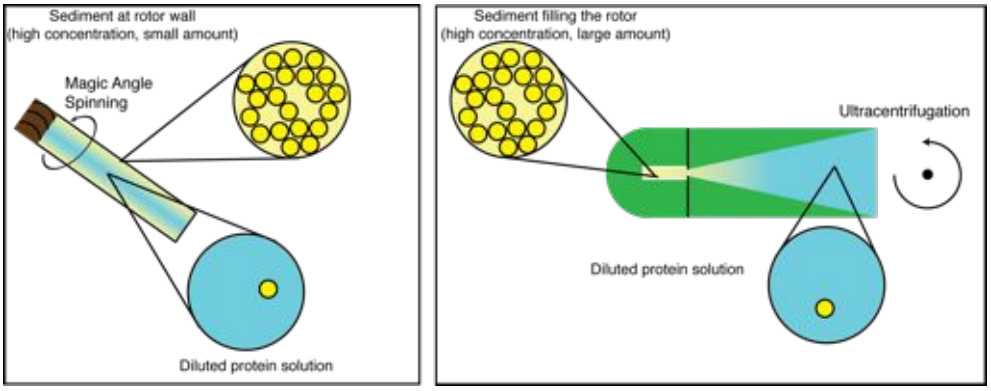

Figure 3.2: Comparison of *in-situ* and *ex-situ* sediment preparations

This is especially important in the case of the 1.3 rotor that can be used to achieve high resolution spectra but has a low internal volume ($\sim 2\ \mu l$). To prove this, two different preparations of apoferritin were compared: the first was 2 μl of a 90 mg/ml solution, the second was the same volume of a sediment obtained by *ex situ* sedimentation (Figure 3.3. The increase in sensitivity is dramatic, since the same signal-to-noise ratio was achieved at 64-fold reduced experimental time (in line with the calculated amount of sedimented material inside the rotor in the two cases).

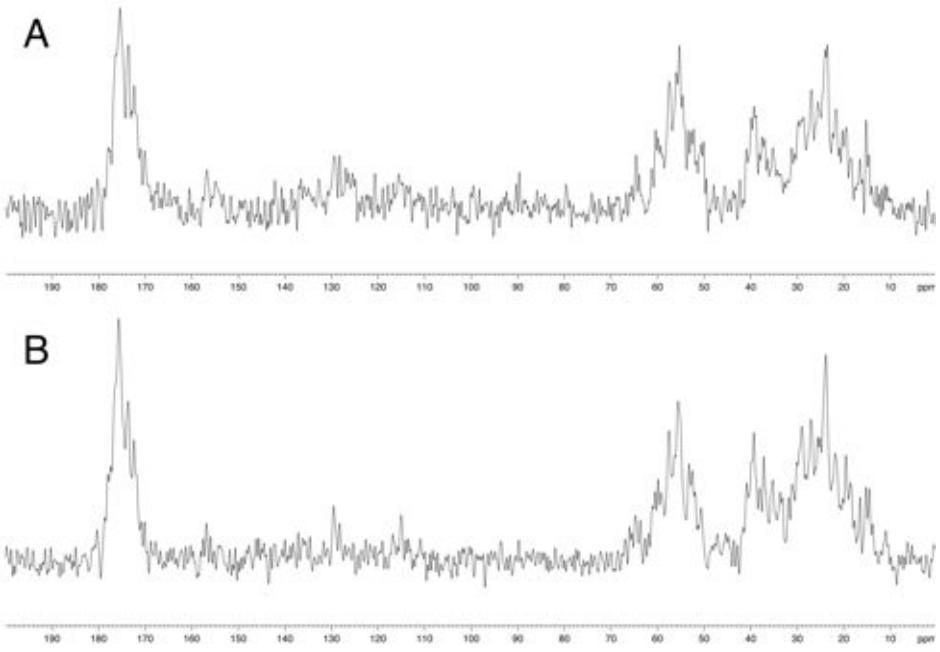

Figure 3.3: $^{1}H - ^{13}C$ CP spectra of *in-situ* (top) and *ex-situ* (bottom) sediment preparations of apoferritin. The top spectrum was acquired with 1024 scans, the bottom spectrum with 16 scans.

On the same preparation a 2D NCO spectrum with S_3E homonuclear decoupling [196] was recorded to prove the feasibility of multidimensional experiments on sedimented samples. [183]
Sample manipulation can be further simplified upon the introduction of the use of especially designed fixtures [197], like the one depicted in figure 3.4, that can withstand ultracentrifugal operations and that can be used to increase the amount of solute that sediments into the NMR rotor from a large solution reservoir. [184, 193].

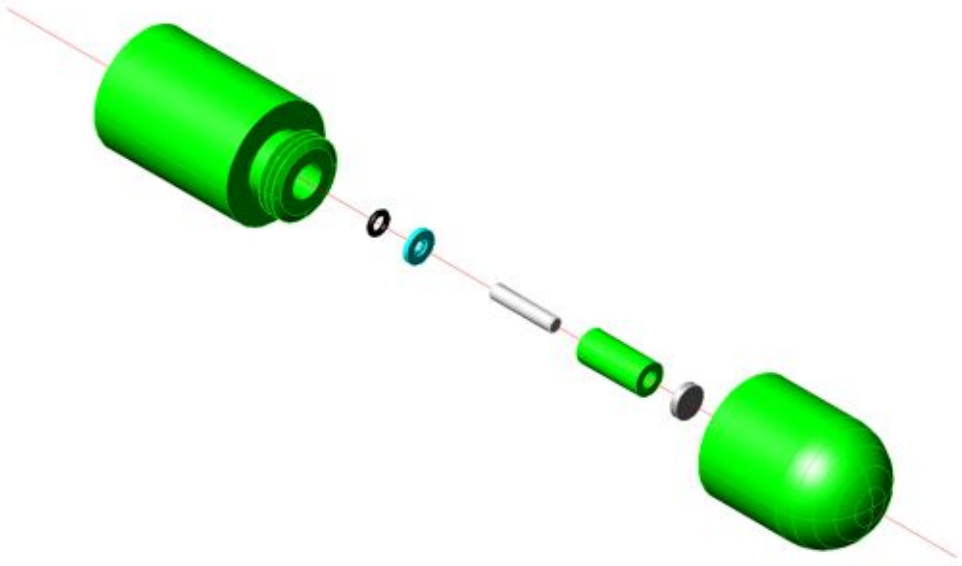

Figure 3.4: Ultracentrifugal device for *ex-situ* sedimentation designed as a part of the present thesis and realized in collaboration with the workshop of the University of Florence. The device is made of Delrin acetal resin and has an inner volume of 1.36 ml.

The use of ultracentrifugal devices, other than alleviating the insensitivity of the method, has a number of features that come from the practice of ultracentrifugation: it allows for sedimentation of even smaller solutes and allows for fractional ultracentrifugation, i.e.: selecting the desired fraction out of a complex mixture.
For *ex-situ* sedimentation using ultracentrifugal devices a further point can be considered: it is not possible to integrate analytically equation (3.7). Numerical integration is anyway possible in the following form:

$$\int c(r)S(r)d\,r = c_0 V_{device}. \tag{3.9}$$

Relation to previous methods and notable applications

Sedimented Solutes NMR was developed by us starting from the observation by Mainz et al. [192] that αB crystallin, a multimer of 600 kDa, in a solution containing 20% glycerol, spun at 12 kHz at 260 K gave rise to solid state NMR spectra. The glycerol and the low temperature can indeed increase viscosity and slow down the molecular tumbling, but not to the extent that would be required for MAS to properly average the anisotropic interactions. Thus, for solid state NMR to work on such preparations a further contribution from sedimentation itself is needed. The preparation involving glycerol has successfully been applied to the study of αB crystallin. [192, 194, 198]. Sedimented solutes NMR was also applied to the study of a dodecameric elicase of 708 kDa. [193]

3.2 Dynamic Nuclear Polarization on Sedimented Solutes

Since the sediment is expected to contain only bound water or water with limited mobility. [182,183], upon freezing this water may not form an ordered ice crystal but an amorphous glassy state.
Formation of ice crystals forces the radicals to segregate in grains and the addition of a glass-forming agent is common to prevent this phenomenon from occurring.
We proposed that the sedimented protein itself may act as the glass forming agent: to some extent the sediment may be regardad as a separate phase that may not be as susceptible to glass formation within the bulk solvent as a homogeneous solution is. In turn, this would allow for ideal dispersion of the biradical throughout the sample. The idea has been proven both as in-situ preparation [199] or ex-situ [200].

3.2.1 *In situ* SedDNP

To demonstrate the difference between frozen sediment and frozen solution, we studied three samples:

1. ApoF sedimented by MAS at room temperature from an aqueous solution containing TOTAPOL and then frozen;
2. aqueous solution of ApoF and TOTAPOL frozen sans sedimentation;
3. same as the first but without the addition of TOTAPOL.

All samples were prepared at a protein concentration of 30 mg/ml in 90/10 (v/v) D_2O/H_2O in 3 mM tris-(hydroxymethyl)-aminomethane (Tris) buffer. Samples (i) and (ii) also contain 2 mM TOTAPOL. The reduced proton concentration of 10% in the matrix is known to yield optimal conditions for ^1H DNP. [33] Figure 3.5 illustrates significant gains in signal from DNP under micro-wave irradiation (on-signal) for sample (i), indicating the incorporation of the radical into the sediment.

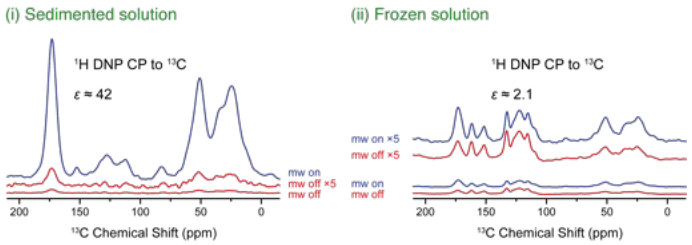

Figure 3.5: Comparison of DNP enhanced signals from a frozen sedimented sample (i) and for a frozen solution (ii) using cross-polarization ($^{13}C -^1 H$) DNP enhanced spectra (on-signals) are given in blue while thermal polarization spectra (off-signals) are given in red. Spectra are also scaled by a factor 5 for better visualization (given in light red or blue color).

^1H polarization yielded a 42-fold increase in signal by comparison with the off-signal Direct polarization of ^{13}C was observed via a Bloch decay; the enhancement factor was determined to be 22 in this case. Due to the absence of ^{13}C in the matrix, polarization has to be transferred directly

Therefore the protein must be in close proximity to TOTAPOL, limiting the distance between the unpaired electron spins and the protein. In contrast, the frozen solution (ii) provides very poor enhancements ($\epsilon \approx 2$) for both ^1H and ^{13}C polarization due to the inability to form a glass with consequent phase separation of water, protein and TOTAPOL, that inhibits effective electron-nuclear spin polarization (Figure 3.5). This shows that sedimentation provides a layer of glassy-like protein on the wall of the sapphire rotor, which enables the biradical to be homogeneously dispersed throughout the sediment providing glass-like properties and efficient $e^- \rightarrow {}^1H(^{13}C)$ polarization transfer. We measured polarization build-up time constants (T_B) and found them to be unusually short for the sedimented sample (i), suggesting direct protein-TOTAPOL interaction is occurring.

Sample	$^1H\ T_B$ **(s)**	$\epsilon(^1H)$	$^1H\ \xi$ $\mathbf{s}^{-1/2}$
Frozen sediment 2 mM TOTAPOL (i)	1.2	42	38
Frozen solution 2 mM TOTAPOL (ii)	2.1	2.1	1.4
Frozen sediment w/o TOTAPOL (iii)	2.1	-	-
Cryoprotected 2 mM TOTAPOL (A)	20.4/1.1	70	16
Cryoprotected 15 mM TOTAPOL (B)	5.0/0.6	100	45

Table 3.1: Summary of DNP enhancements and ^1H, ^{13}C polarization build-up time constants sediment and standard samples.

^1H spin-polarization build-up time constants were found to be of the order of 1.2 s (i) and 2.1 s (ii and iii). As suggested (*vide supra*), the difference in the observed polarization times could indicate an increased TOTAPOL concentration in the sediment with respect to the bulk solution.

The results reported here represent an important step towards DNP of proteins sedimented into an MAS rotor by ultracentrifugation; experiments that are currently underway.

3.2.2 *Ex situ* SedDNP

Cryoprotection is often required when temperature cycling a protein, in order to avoid degeneration and maintain the protein structure at low temperatures (<-75 °C). [201, 202] The addition of a glass-forming solvent, often glycerol, is used to lower the freezing point of water and inhibit bulk ice crystallization to form, enabling the formation of an amorphous solid that protects the protein. However, if the highly concentrated soluble protein forms a glassy state upon freezing of the water matrix due to self-crowding, the addition of a cryoprotectant becomes unnecessary. [199] BSA is also known as a cryoprotectant. [203] The feasibility of cryoprotectant-free DNP by sedimenting BSA was tested from a 100 mg/ml solution in 90/10 (v/v) D_2O/H_2O to which 5 mM TOTAPOL was added. After centrifugation, the sediment ($\sim$50 μl) was packed into a sapphire rotor and inserted into the NMR probe, which had been pre-cooled to cryogenic temperatures (between 85 and 100 K). Irradiation with 8 W of 140 GHz microwaves resulted in a 66-fold enhancement (ε) of the protein CPMAS NMR signal (Figure 3.6).

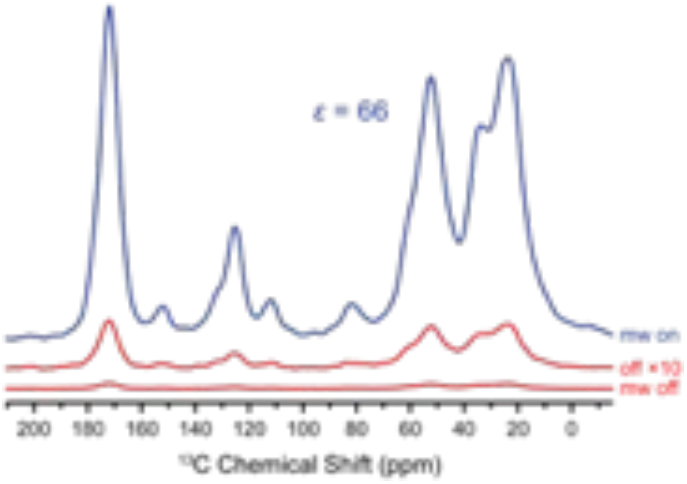

Figure 3.6: DNP-enhanced and thermal equilibrium polarization ^{13}C-CPMAS spectrum of natural abundance BSA sedimented from a 100 mg/ml solution in 90/10 (v/v) D_2O/H_2O solution with 5 mM TOTAPOL.

The magnitude of this enhancement is comparable to typical DNP experiments on proteins, where samples have been prepared in a conventional way (i.e., by dissolving the protein in a glycerol/water mixture). However, the ^1H polarization build-up time constant of T_B= 1.8 s is short compared to a conventional approach. The fast polarization build-up of protons is most likely caused by the high protein proton concentration in the sediment in combination with an increased biradical concentration due to potential protein-TOTAPOL interactions. In the previous section the preferential enrichment of TOTAPOL in the sediment layer in SedDNP is reported. In order to further investigate this situation, TOTAPOL concentration of the BSA solution was varied prior to sedimentation. Three samples were prepared with 200 mg/mL BSA each and 2.5, 5, and 10 mM TOTAPOL concentration, respectively. The results are shown in figure 3.7. DNP enhancements increase, ε = 29, 48, and 64, with increasing TOTAPOL concentration, while ^1H buildup times showed an inverse trend with T_B = 3.6, 2.6, and 1.6 s, respectively. In a control experiment the spin-lattice relaxation time constant, T_1 = 6.3 s was measured for a sample prepared under identical manner sans TOTAPOL.

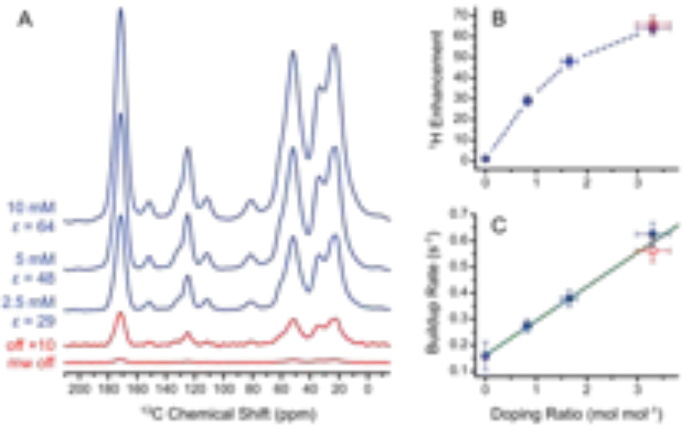

Figure 3.7: TOTAPOL concentration effect (2.5, 5, and 10 mM) on natural abundance BSA sedimented from 200 mg/ml (about 3 mM) protein solutions in 90/10 (v/v) D_2O/H_2O. (A) DNP enhancements with increasing TOTAPOL concentration, (B) ^1H enhancement and (C) polarization buildup rates as doping ratio (i.e., TOTAPOL/protein) increase. Red open circles represent the sediment obtained from 100 mg/ml BSA solutions in 90/10 (v/v) D_2O/H_2O where 5 mM TOTAPOL was added.

Interestingly, when comparing enhancements as well as build-up rates (i.e., T_B^{-1}) in Figure 3.7B and 3.7C, we observe very similar values for the sample sedimented from 100 mg/ml BSA and 5 mM TOTAPOL (red open circle) and that sedimented from 200 mg/ml BSA and 10 mM TOTAPOL (blue filled circles). This suggests that it is not the TOTAPOL concentration in the solution prior to centrifugation that is determining the TOTAPOL amount in the sediment, but rather the TOTAPOL to protein ratio (doping ratio), that is preserved during sedimentation. That would be the case if TOTAPOL were tightly binding to the protein: this is represented in the upper panel of Figure 3.8. Further parameters and details for all samples are listed in Table 1. A similar case has been observed during *in situ* SedDNP of apoferritin, but it is important to notice that the enrichment in the apoferritin layer is most probably aspecific and due to the fact that the hydrophobic patches present on the protein surface are more concentrated in the sediment, thus providing the radical with a more suitable environment to partition in

with respect to the bulk solution, as represented in the middle panel of Figure 3.8. The bottom panel in Figure 3.8 represents a situation that has not been encountered yet where the protein does not interact with the radical. In this case, it is expected that the concentration of the radical will be uniform throughout the sample, regardless of the gradient formed by the protein, and this situation is not different from the radical distribution observed in the usual DNP sample. Direct binding between BSA and TOTAPOL is not unexpected. BSA contains two hydrophobic binding sites which might provide a preferential environment for the partially hydrophobic TOTAPOL. Furthermore, TOTAPOL possesses a relatively flexible structure, allowing it to adopt a conformation suitable for binding. The combination of amphiphilicity and flexibility could further improve TOTAPOLs tendency to interact with the protein, allowing a molar ratio between bound TOTAPOL and BSA larger than 2. Yet, it is important to observe that radical binding to the protein is not intrinsic of the method but rather of the chemistry of the biomolecule under investigation, in this case a protein that is able to bind a number of aspecific partners. [204]

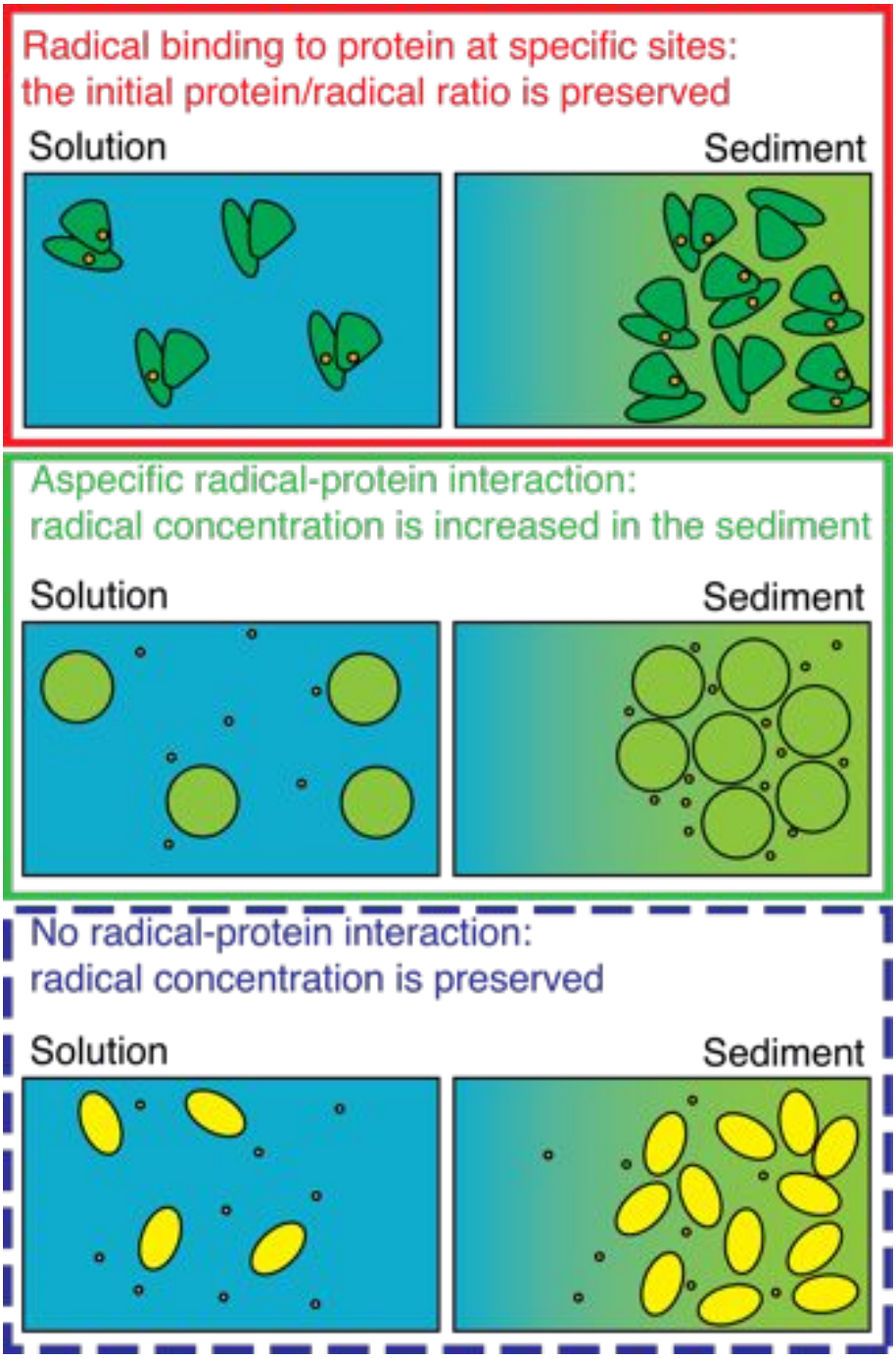

Figure 3.8: Comparison between different SedDNP preparationsTop panel depicts the situation described in the present work: the radical binds to the protein thus the protein/radical ratio is preserved when moving from solution to the sediment. Middle panel depicts the situation described for ApoF where the radical is partitioned in the more hydrophilic sediment. Bottom panel shows a theoretical case in which the protein is sedimented but the radical preserves the same distribution along the sample. Dashed line indicates that this is a theoretical situation, not yet observed in any case.

c_{BSA} **mmol dm^{-3}**	$c_{TOTAPOL}$ **mmol dm^{-3}**	$c_{TOTAPOL}/c_{BSA}$	ϵ	T_B **s**
3.03	0	0	1	6.3
3.03	2.5	0.83	29	3.6
3.03	5	1.7	48	2.6
3.03	10.0	3.29	64	1.6
1.52	5.0	3.3	66	1.8

Table 3.2: Summary of DNP enhancements and ^1H, ^{13}C polarization build-up time for sedimented samples at different protein/radical concentrations.

Using the model globular protein BSA, *ex situ* sedimentation preparation has been demonstrated, which reduces the need for a cryoprotecting matrix in DNP experiments. Proteins in natural isotopic abundance impose a rather delicate sample system in DNP studies. While DNP is particularly well suited to compensate for the intrinsically low sensitivity due to the low ^{13}C content, the signals in the $C_\alpha - C_\alpha$ region are often masked by the unfortunate artifact observed from the equally enhanced cryoprotectant background being up to several factors stronger in intensity than the signals of interest. In order to reduce the glycerol signal one can use ^{13}C-depleted and fully deuterated glycerol. However, for high volume sample screening and/or optimization of DNP conditions this is often not affordable due to high costs associated with either biological samples or cryoprotectant.

Chapter 4

Pseudocontact shift in the solid state

4.1 Previous literature

Pseudocontact shift has the same symmetry with respect to spin operation as the chemical shift operator, as shown in equation (B.23). Thus, Magic Angle Spinning can average the dipolar powder pattern (Curie-spin/ DSA) as it averages the Chemical Shielding Anisotropy (CSA) powder pattern.
This symmetry relationship is evident also in relaxation: since interconversion between two positions in the powder pattern is unlikely at extreme rotational correlation times, [95] Curie-spin relaxation is suppressed in Solid state NMR, [205] as well as CSA relaxation. [206] This makes solid state NMR extremely attractive for paramagnetic substances, expecially for large molecules in which Curie-spin relaxation broadens beyond detection the lines in solution NMR spectra.

4.1.1 PRE for faster acquisition

The development of fast MAS methods allowed, among other things, for the introduction of low power $^{1}H-^{13}C$ decoupling schemes. [207–210] The lower deposition of power into the sample and the probehead, in turn, would in principle allow for faster recycling. Anyway, relaxation in the solid state is

usually slower as compared to what is observed in solution: ^1H relaxation rates range from one to several seconds, whereas in solution they can be as fast as 500 ms for proteins of comparable size.
The Ishii group has introduced a smart approach to solve this problem: doping the sample with paramagnetic ions that shorten ^1H relaxation. [211–213] This is even simpler if the paramagnetic center is bound to the protein. [51]

4.1.2 Paramagnetism for structural determination

Determination of high-quality atomic-resolution structures by solid state NMR is far from being routine. The main culprit in the structure determination process is the difficulty in obtaining unambiguous long-range distance restraints, that are needed to obtain the fold of the protein.
Recovering through-space connectivity under MAS is tricky, because direct $^{13}C - {}^{13}C$ recoupling is susceptible to artifacts due to dipolar truncation, [214, 215] and spin diffusion is inhibited at high field and high spinning frequencies. [216, 217] Several experimental techniques have been proposed recently to overcome such limitations. Another great problem persists anyway: if a cross-peak due to a long-range interaction appears in the spectrum, it must be correctly assigned and tracked through a series of spectra that become more and more crowded with increased mixing times.
In this regard, paramagnetic interactions are far simpler to frame: once the sequential assignment of the spectrum is available, one can easily measure PCS and PRE for any single nucleus. Moreover, while nucleus-nucleus interactions can be detected up to 9-10 Å, paramagnetic interactions can extend well beyond 20 Å.
In the assumption that only dipolar relaxation (SBM mechanism) is active, relaxation is inversely proportional to the metal-nucleus distance to the sixth power ($1/r^6$), although deviations can be expected for heteronuclei. [218] PRE have been used to obtain structural information in several solid-state NMR works, through paramagnetic tagging [219, 220] or in metalloproteins. [53]
Other than PRE, if the paramagnetic center has anisotropic magnetic sus-

ceptibility, pseudocontact shifts are present. With a $1/r^3$ dependence the restraints available have longer range with respect to PRE and can be detected with higher precision. The presence of PCS gives unambiguous long range restraints and it is also useful to assist in the assignment of the spectra [221] and finally to achieve high-resolution structures in the solid state. [62]

4.2 NMR Crystallography

4.2.1 Contribution of this thesis

During this thesis work i contributed to this area of research with the paper "Solid state NMR crystallography through paramagnetic restraints", [61] where we used pseudocontact shifts for refinement of protein structure and crystal arrangement at the same time.

4.2.2 NMR Crystallography made crystal clear

NMR crystallography is rapidly gaining interest. For micro- and nanocrystalline powders of small molecules it permits to access both the atomic details of the molecular structure and the crystal packing. [222–227] A small protein (GB1, 6kDa) has also been recently tackled. [63] But what makes pseudocontact shift a very suitable restraint for for protein NMR crystallography? As mentioned above, PCS are long range structural restraints [12, 24, 228–230] In crystals, molecules sense a magnetic environment that is affected by both intramolecular and intermolecular contributions, so that the presence of paramagnetic ions in neighboring molecules determines PCS values. [221,229,231] As a result, while intramolecular PCS are valuable restraints to obtain the molecular structure, intermolecular PCS can provide information on the relative arrangement of the molecules within the crystal. [231] Indeed, PCS have been used to perform NMR crystallography of small lanthanide chelates. [231] For paramagnetic proteins, the two contributions to the total PCS can be separated using dilution of the paramagnetic molecules in combination with different labelling strategies. [64, 93, 219–221] Intramolecular PCS were mea-

sured with this approach for the high-spin cobalt(II)-substituted protein matrix metalloproteinase 12 (CoMMP-12), a 17 kDa protein, for nuclei up to about 20 Åfrom the metal. [62, 221, 232] The use of intramolecular PCS, coupled to distance restraints measured for the native zinc(II)-containing protein, provided a protein structural family with reasonably high resolution. [62] The strength of using PCS to help solid state NMR (SS-NMR) structure determination resides in the fact that PCS are immediately available, while many distance restraints are initially ambiguous and can be introduced only after the first emerging structural family can be used to resolve ambiguities. A drawback is represented by the need of diluting the paramagnetic protein into a larger amount of the diamagnetic protein to ensure that the measured PCS are all intramolecular. Thus the concentration of the labelled species is low, resulting in a low signal-to-noise ratio. Furthermore, all information on the crystal lattice is lost. We have shown, on the same metalloprotein system CoMMP-12, that total PCS can be used as such on non-diluted systems, and that a refinement scheme based on such PCS data and other intramolecular restraints is feasible to obtain:

1. the structure of the protein and
2. the crystal packing, once the cell parameters are available from powder spectra.

Due to the higher signal-to-noise, the number of measureable total PCS (476) [51] is sizably larger than that of intramolecular PCS (319). [62]
Solving protein structures through solid state NMR can be advantageous for those systems which cannot be solved through solution NMR, like large proteins, membrane proteins, or fibrils.

1. Calculation of the molecular structure using total PCS:

    - The total PCS, together with dihedral angle restraints and the unambiguous distance restraints found initially, are used to perform ab initio structure calculations for the determination of the

protein fold, including from the very beginning n+1 (1 intra and n inter) independently positioned paramagnetic metals. The number n can be empirically found as the minimum number of metals needed to reproduce the experimental data at best (i.e., for which there is no significant further decrease of the target function, TF).

- The obtained structural fold is used to iteratively resolve ambiguities in the NMR spectra and to determine additional distance restraints, similarly to what has been shown to work in the case of purely intramolecular PCS. [62]
- The values of the anisotropy parameters, $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, initially set to literature values typical of the paramagnetic metal,are refined by performing a grid search. The anisotropy parameters are thus fixed to the values providing the minimum of the TF.

In the case of CoMMP-12, the total PCS, the dihedral angle restraints and the 240 unambiguous distance restraints found initially, used as upper distance limits (upl) between protein nuclei, provided a protein structure with lowest target function by including 3+1 paramagnetic metals (line 1 of Table 4.1). As expected, the obtained structural family permitted to determine additional distance restraints. The anisotropy parameters were then fixed to the refined values, which corresponded to $(9\pm1)10^{-32}$ and $(3.0\pm0.3)10^{-32}m^3$ for $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, respectively. By doing so, we could obtain a structure of approximately the same quality as the one previously obtained with intramolecular PCS only: the 20 structures with lowest TF have a backbone (BB) root mean square deviation (RMSD) to the mean of 0.95 and an RSMD to the X-ray structure of 1.4 (line 2 of Table 4.1 and Figure 4.1).

2. Calculation of the crystal structure using total PCS, at this stage, the global information on the position of all crystal mates should be added:

   - For this final step, a crystallographic space group must be imposed. In X-ray crystallography, such information is usually extracted during the structural refinement, [233] and in the present

case we relied on the existing unit cell parameters from the single crystal X-ray structure. [234] Obviously, a crystal structure is not available if one resorts to NMR crystallography, where microcrystalline powders are employed. However, a powder X-ray diffraction pattern is in principle enough to obtain dimension and shape of the unit cell. [235] Microcrystalline powders are frequently obtained in failed attempts to grow the desired crystals. [236] The unit cell dimensions extracted from powder spectra can be very accurate, [237] with a precision even larger than that usually achieved from single-crystal X-ray diffraction. [236]

- Once the cell parameters have been determined, the correct symmetry must be identified. Structure calculations including all paramagnetic metals in the crystal must be then performed for each of the possible space groups.
- The implementation of the symmetry-generated PCS is not straightforward and requires a substantial modification of the PARAMAGNETIC CYANA program. [238, 239] A pseudoresidue is introduced to freely position the crystallographic origin and reference frame. The positions of the metals of neighboring molecules, and the orientations of the corresponding magnetic susceptibility anisotropy tensors, with respect to this crystallographic frame are obtained according to the symmetry rules. In turn, the position of the crystallographic frame is determined during the minimization, by introducing in the TF the intermolecular contributions to PCS due to all the neighboring molecules in the crystal. It should be noted that the introduction of the pseudoresidue mimicking the crystallographic frame reduces the number of degrees of freedom with respect to those present in step 1.

		TF range	RMSD BB	RMSD mean to X-ray structure	Secondary structures RMSD mean to X-ray structure
		$Å^2$	Å	Å	Å
1	240 upl + 476 PCS, 4 independent metals	2.97-7.09	4.3	4.5	2.2
2	727 upl + 476 PCS, 4 independent metals	0.70-0.86	0.9	1.4	1.0
3	Crystal ($P2_12_12$)	0.86-1.22	0.9	1.2	0.9
4	Crystal ($P2_12_12_1$)	1.97-2.72	0.9	1.6	0.9
5	Paramagnetically diluted protein [62]	2.66-4.22	1.0	1.3	0.9

Table 4.1: Parameters for the MMP-12 structures obtained through solid-state NMR restraints.

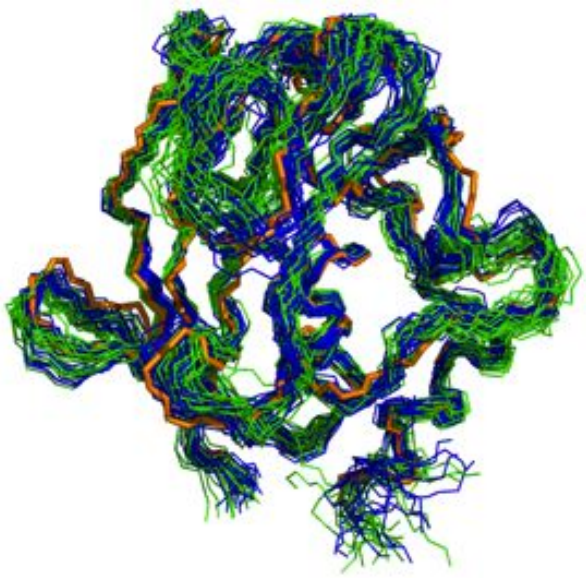

Figure 4.1: Structural family of MMP-12 calculated using four independently positioned metal ions (green) and metal ions positioned according to the imposed crystallographic symmetry (blue), superimposed to the X-ray structure (orange)

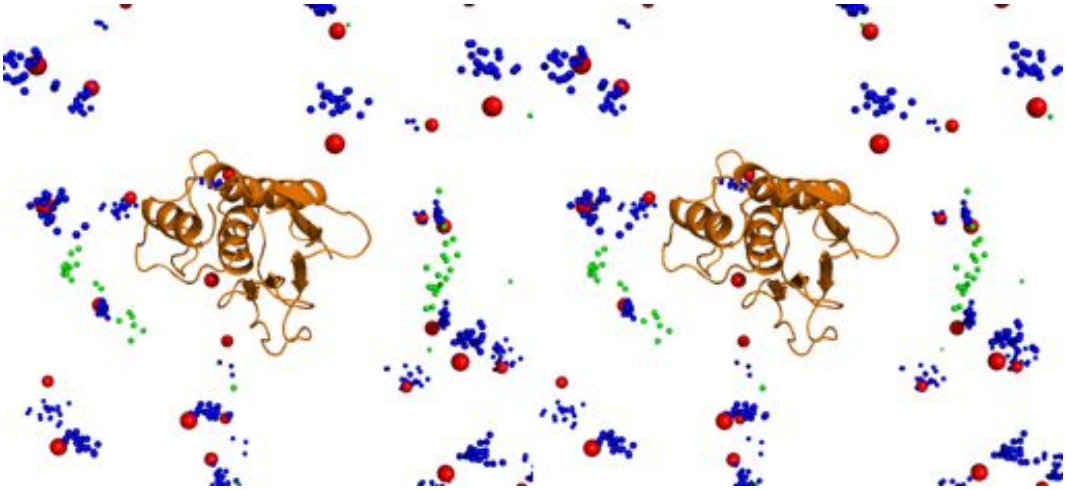

Figure 4.2: The position of the metals if independently positioned (green), positioned according to the crystallographic symmetry (blue) and crystallographically determined (orange)

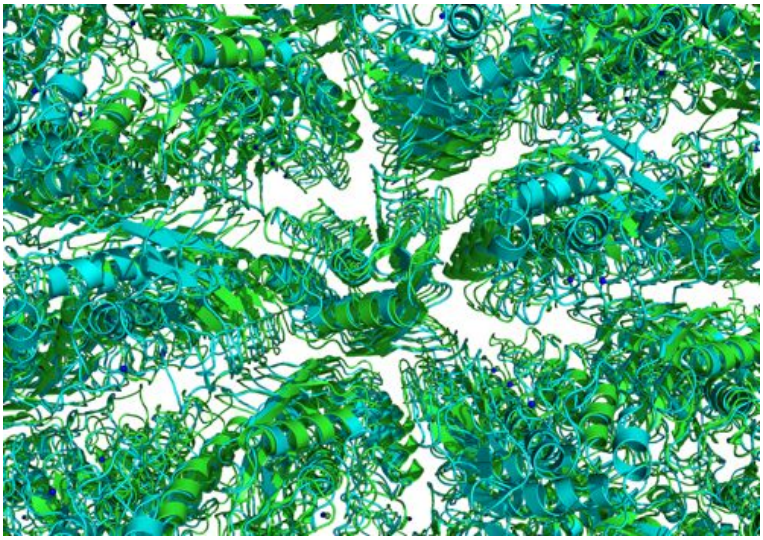

Figure 4.3: Cartoon representation of crystal packing obtained by NMR (cyan) as compared to the packing obtained by X-ray (green).

In the case of CoMMP-12 the good quality of the NMR spectra and the absence of signal doubling ensure lack of crystal heterogeneity. The unit cell parameters are $a = 69.194$, $b = 62.564$, $c = 37.262$, $\alpha = \beta = \gamma = 90°$, indicating an orthorhombic symmetry. [234] Nine different orthorhombic symmetries are possible, differing by the number of molecules per unit cell. The Matthews coefficient [240, 241] indicates that it is reasonable to expect four molecules per unit cell. Therefore, only the $P222$, $P222_1$, $P2_12_12$ and $P2_12_12_1$ space groups are allowed.

The first two symmetries ($P222$ and $P222_1$) did not provide any acceptable solutions because all structures showed severe compenetration among symmetry-related molecules. In the case of the $P2_12_12_1$ symmetry, non compenetrating structures were calculated, but with a TF >2 times larger than in the $P2_12_12$ case (lines 3 and 4 of Table 4.1 for comparison). Therefore, the experimental data strongly point to the $P2_12_12$ symmetry. Figure shows the superposition of the lowest TF structure with the X-ray structure together with all surrounding symmetry-related molecules. The agreement is very good, showing that PCS have indeed been able to reconstruct the crystal with good accuracy, once the cell parameters are available. This indicates that

the intermolecular contributions to PCS have been indeed able to correctly locate the crystallographic origin and reference frame. This achievement is quite striking, considering that no information on van der Waals attractive or repulsive potentials has been included in the calculations. Finally, it is noteworthy that the structural family calculated with the crystallographic restraints shows a further non negligible improvement with respect to the family calculated using independently positioned metal ions, the BB RMSD to the X-ray structure decreasing from 1.4 to 1.2 Å(see lines 2 and 3 in Table 4.1).

Appendix A

Methods

A.1 Solid State Studies

A.1.1 Solid State NMR

Magic Angle Spinning solid state NMR spectroscopy (MAS-NMR) experiments were acquired at CERM on 700 MHz and 850 MHz instruments, equipped with 4 mm and 3.2 mm probeheads and 3.2 mm and 1.3 mm probeheads respectively.
Rotors of different sizes have different inner radii, volumes and are capable of different spinning rates, as summarized in table A.1.

Outer diameter (mm)	**Inner radius (mm)**	**Inner volume (μl)**	**Max. spinning rate (kHz)**
4	1.5	52	15
3.2	1.06	22	23
1.3	0.35	2	60

Table A.1: Characteristics of the rotors for MAS NMR available at CERM

The experimental conditions are reported along with the spectra in the "Results" chapter.

^1H decoupling

At low spinng rates ($< 35\,kHz$), high power ($\omega_h/2\pi \geq 75\,kHz$) H-X decoupling is required and is optimized maximizing the X-nuclei T_2. [242] At higher spinning rates, low power decoupling ($\omega_h \leq \omega_{rot}/2$) can be sufficient. [209,243]

H-X Cross Polarization

CP is used to increase the sensitivity and the time efficiency of the solid state NMR experiments. It is accomplished through spin-locking of H and X nuclei a the same nutation frequency or at frequencies that are related by sum/subtraction of multiples of the sample rotation. [244–247]

A.1.2 DNP

Dynamic nuclear polarization experiments were performed at Francis Bitter Magnet Laboratory (Massachusetts Institute of Technology, Cambridge, MA) using a custom-built 212 MHz (5 T, ^1H) NMR spectrometer (courtesy of David Ruben) and a custom-built 140 GHz gyrotron oscillator high power microwave source generating up to 15 W. [248] Spectra were recorded using a custom-built 4 mm triple resonance (^1H, ^{13}C and ^{15}N) MAS DNP NMR probe utilizing an overmoded circular corrugated waveguide to efficiently couple microwaves to the sample and a sample eject mechanism allowing sample changing during cryogenic operation. [249] Experimental temperatures were maintained between 80 and 90 K by cooling the bearing and drive gas (N_2) using a custom-built external heat exchanger. [250] The magnetic field was set to the value yielding the maximum DNP enhancement for each biradical using a superconducting sweep coil with 50 mT sweep width. One-dimensional experiments involved destruction of nuclear thermal equilibrium polarization by a pre-saturation pulse train on both ^1H and ^{13}C, polarization of the proton matrix by continuous microwave irradiation during a variable polarization period followed by $^1H^{13}C$ ramped cross-polarization (CP). [251]

^1H and ^{13}C r.f. field strengths were adjusted to 100 kHz (γB_1) for each sample; spin-lock field strength of ^1H was set to 100 kHz, while that of ^{13}C was optimized for efficient Hartmann-Hahn matching conditions at a MAS frequency ($\omega_r/2\pi$) of 4.80 kHz. The CP contact time was found to be optimal at 1.2 ms. All spectra were acquired with high power ^1H decoupling using TPPM [252] with 100 kHz field strength. ^1H build-up times (T_{DNP}) were measured by varying the polarization period using an exponential increase from 0.1 to 64 seconds; recycle delays where chosen to be $1.3T_{DNP}$ in order to maximize spectral S/N per unit of time. Depending on the sample concentration and sensitivity between 8 and 90,000 co-added transients were collected.

Appendix B

Basics of NMR

B.1 Boltzmann distribution for a spin 1/2

Due to statistical thermodynamics, the population of the spin levels of an ensemble of like sins is given by the Boltzmann distribution:

$$p_{\alpha/\beta} = \frac{e^{\pm\gamma B_0\hbar/2k_BT}}{2\cosh(\gamma B_0\hbar/2k_BT)}. \tag{B.1}$$

The denominator $2\cosh(\gamma B_0\hbar/2k_BT)$ is related to the partition function of the ensemble and ensures normalization $p_\alpha + p_\beta = 1$.
The z-magnetization at equilibrium is given by $\mu_\alpha p_\alpha + \mu_\beta p_\beta$, where $\mu_{\alpha/\beta} = \pm 1/2\gamma h$. Expressing the p in terms of Boltzmann populations we get:

$$(p_\alpha - p_\beta) = \frac{\sinh(\gamma B_0\hbar/2k_BT)}{\cosh(\gamma B_0\hbar/2k_BT)} \approx 1/2\gamma h(\gamma B_0\hbar/k_BT). \tag{B.2}$$

That evaluates to 10^{-4} for $\gamma B_0/2\pi = 600\ MHz$

B.2 Magnetic Susceptibility

The effect of an external magnetic field $\mathbf{H}_0$ on a particle with spin S is to to split the energies of the S manifolds. In a system made up by many particles, this results in the populations of these levels being different according to the

Boltzmann law, and thus the lower energy state, the one with the particles having magnetic moment aligned with the field, is more populated than the higher energy state with the magnetic moment opposite to the field, as described in appendix B.1. As a consequence, an induced magnetic moment μ_{ind} along $\mathbf{H_0}$ is established. The magnetization per unit volume, M, corresponds to the induced magnetic moment per unit volume and, for many substances, is found to be proportional to the applied magnetic field $\mathbf{H_0}$:

$$M = \frac{\mu_{ind}}{V} = \chi_V H_0 \tag{B.3}$$

where χ_V (dimensionless) is called magnetic susceptibility per unit of volume. A magnetic flux density B_0 is defined as $\mu_0(H_0 + M)$; where μ_0 is the permeability of vacuum. Since $\chi_V \ll 1$ (except for ferromagnetic systems), $B_0 \simeq \mu_0 H_0$ and

$$M = \frac{1}{\mu_0}\chi_V B_0 \tag{B.4}$$

The magnetic susceptibility per molecule χ is defined as: [36]

$$\chi = \frac{\chi_M}{N_A} = \frac{V_M \chi_V}{N_A} = \frac{\mu_0 \langle \mu \rangle}{B_0}. \tag{B.5}$$

B.2.1 Anisotropy

If the electron orbital magnetic moment contribution is non null, the magnetic susceptibility becomes anisotropic.
This can be represented by an anisotropy of the g_e factor, thus a $\mathbf{g}$ tensor is introduced with g_{kk} components for any direction of the magnetic field given, to the first order, by

$$\langle \phi | L_{kk} + g_e S_{kk} | \phi \rangle = g_{kk} M_S \tag{B.6}$$

where the ϕ indicates the Zeeman eigenfunctions.
Then, the values of $\langle \boldsymbol{\mu} \rangle$ for the three main directions of the reference frame

are:

$$\langle \mu_{kk} \rangle = -\mu_B g_{kk} \langle S_{kk} \rangle = -\mu_B \frac{\sum_i \langle \phi_i | L_{kk} + g_e S_{kk} | \phi_i \rangle (1 - E_{ikk}/k_B T)}{\sum_i (1 - E_{ikk}/k_B T)} \quad \text{(B.7)}$$

Where $E_{ikk} = g_{kk} \mu_B M_S B_0$ is the Zeeman energy. This results in $\boldsymbol{\chi}$ being a tensorial quantity.
A consequence of the anisotropy of the magnetic susceptibility is that $\langle \boldsymbol{\mu} \rangle$ depends on the orientation of $\boldsymbol{\chi}$ with respect to $\mathbf{B}_0$:

$$\langle \boldsymbol{\mu} \rangle = \frac{\boldsymbol{\chi} \mathbf{B}_0}{\mu_0}. \quad \text{(B.8)}$$

B.2.2 Anisotropy effects

Equation (B.8) state that $\langle \boldsymbol{\mu} \rangle$ is proportional to the magnetic field. Being χ anisotropic, the proportionality constant is different in different directions. In explicit form, equation (B.8) reads:

$$\langle \boldsymbol{\mu} \rangle_i = \frac{1}{\mu_0} \sum_j \chi_{ij} B_{0j} \quad \text{(B.9)}$$

The energy of a magnetically anisotropic molecule has an orientation-dependent contribution given by

$$E = -\int_0^{\langle \boldsymbol{\mu} \rangle} \mathbf{B}_0 d\langle \boldsymbol{\mu} \rangle = -\frac{\mathbf{B}_0 \boldsymbol{\chi} \mathbf{B}_0}{2\mu_0} \quad \text{(B.10)}$$

and this results in removing the degeneracy among the different orientations. The latter, thus, become not equal , so that the paramagnetic center (being an ion or a cluster) acts as an internal orienting medium. When not all the orientations of the molecule are equally likely, all of the interactions will average to values different from the isotropic one.
Partial orientation at high magnetic field is induced by magnetic anisotropy. The induced partial orientation tensor $\mathbb{S}$ will be proportional to the molecular

$\boldsymbol{\chi}$ tensor, for the derivation please refer to [253].

$$S_{ii} = \frac{B_0^2}{15\mu_0 k_B T} \Delta\chi_{ax} \tag{B.11}$$

Residual Dipolar Couplings

The Hamiltonian for the interaction of two dipoles of two different nuclei A and B is given by

$$\hat{\mathrm{H}}_{Dipole-dipole} = -\frac{\mu_0}{4\pi} \frac{\hbar^2 \gamma_A \gamma_B}{r_{AB}^3} I_z^A I_z^B (3cos^2\vartheta - 1) \tag{B.12}$$

where ϑ is the angle between the interspin vector $\mathbf{r}_{AB}$ and the applied magnetic field.
The average value obtained by integration over the angle ϑ when every orientation of $\mathbf{r}_{AB}$ has equal probability is:

$$\frac{1}{2}\int_{-1}^{1} (3\cos^2\vartheta - 1) d(\cos\vartheta) = \frac{1}{2}(\cos\vartheta^3 - \cos\vartheta)|_{-1}^{1} = 0. \tag{B.13}$$

In the presence of a partial orientation described by the tensor $\mathbb{S}$, the former equation does not hold since different weights must be considered for different directions. Therefore the so-called Residual Dipolar Coupling arises, provided by

$$\Delta\nu^{rdc} = \frac{\Delta E}{h} = -\frac{\mu_0 \hbar \gamma_A \gamma_B}{4\pi^2 r_{AB}^3} T_{zz} \tag{B.14}$$

where

$$T_{zz} = \langle (3\cos^2\vartheta - 1)/2 \rangle = Tr(\mathbb{T}) \frac{2}{3} \sum_{pq} S_{pq} T_{pq} \tag{B.15}$$

where p, q are the principal directions of the molecular frame. Since $Tr(\mathbb{T})$ provides no contribution, equation (B.14) recasts to

$$\Delta\nu^{rdc} = -\frac{\mu_0 \hbar \gamma_A \gamma_B}{4\pi^2 r_{AB}^3} \frac{2}{3} \left[S_{xx} \left(\frac{3l^2 - 1}{2} \right) + S_{yy} \left(\frac{3m^2 - 1}{2} \right) + S_{zz} \left(\frac{3n^2 - 1}{2} \right) \right]$$

(B.16)

Where l, m, n are the direction cosines of the internuclear vector with respect to the principal directions of the traceless tensor $\mathbb{S}$. The previous equations may be rewritten in the form

$$\Delta\nu^{rdc} = -\frac{\mu_0 \hbar \gamma_A \gamma_B S_{LS}}{8\pi^2 r_{AB}^3} \left[S_{zz}(3\cos^2\Theta - 1) + (S_{xx} - S_{yy})\sin^2\Theta\cos 2\Phi \right] \tag{B.17}$$

Where Θ is the angle between the $\mathbf{r}_{AB}$ vector and the z axis of the $\mathbb{S}$ tensor and Φ is the angle between the x axis of that tensor and the projection of $\mathbf{r}_{AB}$ on the xy plane.
S_{LS} (not to be confused with the $\mathbb{S}$ tensor) is the Lipari-Szabo order parameter introduced to take onto account the internal motion of the internuclear vector. Usually, the observed tensor is reduced by about 20%.
Substituting equation (B.11) into equation (B.17), the following equation for the RDC is obtained.

$$\Delta\nu^{rdc} = -\frac{1}{4\pi}\frac{B_0^2}{15 k_B T}\frac{\gamma_A \gamma_B \hbar}{2\pi r_{AB}^3}\left[\Delta\chi_{ax}(3\cos^2\Theta - 1) + \frac{3}{2}\Delta\chi_{rh}\sin^2\Theta\cos 2\Phi \right] \tag{B.18}$$

Pseudocontact Shift

The nuclei sense the sum of the external magnetic field and the field originated by the electron static magnetic moment, resulting in an effect on the chemical shift termed the pseudocontact shift.
The Hamiltonian which describes the dipolar interaction between two magnetic moment $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ is:

$$\mathrm{H}_D = -\frac{\mu_0}{4\pi}\left[\frac{3(\boldsymbol{\mu}_1 \cdot \mathbf{r})(\boldsymbol{\mu}_2 \cdot \mathbf{r})}{r^5} - \frac{\boldsymbol{\mu}_1 \cdot \boldsymbol{\mu}_2}{r^3} \right] \tag{B.19}$$

When the interaction is between a nucleus and the electron, the generic form of the magnetic moment is replaced by:

$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_I = \hbar\gamma_I \mathbf{I}$ for the nucleus, and
$\boldsymbol{\mu}_2 = \boldsymbol{\mu}_S = -g_e\mu_B \mathbf{S}$ for the electron
under the assumption that the electron is localized on the metal nucleus, with $\mathbf{I}$ and $\mathbf{S}$ being the spin moment of the nucleus and of the electron, respectively. For the interaction between the proton and $\boldsymbol{\mu}_I$ and the average induced electron magnetic moment $\langle\bar{\boldsymbol{\mu}}\rangle$, the Hamiltonian reads:

$$\mathrm{H}_D = -\frac{\mu_0}{4\pi}\left[\frac{3(\hbar\gamma_I\mathbf{I}\cdot\mathbf{r})(\langle\bar{\boldsymbol{\mu}}\rangle\cdot\mathbf{r})}{r^5} - \frac{\hbar\gamma_I\mathbf{I}\cdot\langle\bar{\boldsymbol{\mu}}\rangle}{r^3}\right] \tag{B.20}$$

When the zeeman splitting is larger than the hyperfine interaction, the nuclear spin is quantized along the direction of the external magnetic field represented it can be represented as $I_k\mathbf{k}$. By substituting the previous relation and using equation (B.8), one gets:

$$\mathrm{H}_D = -\frac{\hbar\gamma_I B_0}{4\pi r^5} I_k \mathbf{k}^\dagger[3\mathbf{r}\otimes(\mathbf{r}\boldsymbol{\chi}) - r^2\boldsymbol{\chi}]\mathbf{k} \tag{B.21}$$

A dipolar shielding tensor $\boldsymbol{\sigma}$ is defined as:

$$\boldsymbol{\sigma} = 3\mathbf{r}\otimes(\mathbf{r}\boldsymbol{\chi}) - r^2\boldsymbol{\chi} \tag{B.22}$$

so that the dipolar energy is:

$$E = -\hbar\gamma_I B_0 M_I \mathbf{k}^\dagger\boldsymbol{\sigma}\mathbf{k}. \tag{B.23}$$

If a reference frame is chosen on the protein, the external magnetic field continuosly moves in this frame and an orientational average must be performed. Under the saecular approximation (already implicit in the equation (B.21)), this is performed by taking the average of the values calculated along three principal directions e_x, e_y, e_z and the pseudocontact shift is:

$$\delta^{PCS} = \frac{E}{\hbar\gamma_I B_0 M_I} = -\frac{1}{3}Tr(\boldsymbol{\sigma}) \tag{B.24}$$

Saecular Approximation Consider a single spin-1/2, with the following Hamiltonian terms

$\hat{A} = \omega_0 \hat{I}_z$
$\hat{B} = \omega_x \hat{I}_x + \omega_z \hat{I}_z$
The term $\hat{A}$ represent the interaction of the spin with a strong field along the z-axis, while $\hat{B}$ represents the interaction of the spin with a small additional field, with longitudinal component proportional to ω_z and a transverse component proportional to ω_x. Assuming that $\omega_z, \omega_x \ll \omega_0$, $\hat{B}$ corresponds to the full form of the chemical shift interaction. The eigenbasis of $\hat{A}$ is defined by the two kets $|1\rangle = |+1/2\rangle$ and $|2\rangle = |-1/2\rangle$:
$\hat{A}|1\rangle = a_1|1\rangle$
$\hat{A}|2\rangle = a_2|2\rangle$
with eigenvalues:
$a_1 = +1/2\omega_0$
$a_2 = -1/2\omega_0$
The matrix representation of $\hat{B}$ in the basis of $\hat{A}$ is:

$$\hat{B} = \frac{1}{2} \begin{pmatrix} \omega_z & \omega_x \\ \omega_x & -\omega_z \end{pmatrix} \tag{B.25}$$

Therefore the secular approximation of $\hat{B}$ has the form

$$\hat{B} = \frac{1}{2} \begin{pmatrix} \omega_z & 0 \\ 0 & -\omega_z \end{pmatrix} \tag{B.26}$$

I.e.: a matrix element of $\hat{B}$ may be dropped if its magnitude is small compared to the corresponding difference in the eigenvalues of $\hat{A}$.

B.3 Paramagnetic Relaxation

B.3.1 Solomon Relaxation

In the absence of a magnetic field, nucleus and electron magnetic moments are randomly oriented and thus average to zero. When an external magnetic field $\mathbf{B}_0$ is applied, the magnetic moments μ orient along the magnetic field direction. The energy of the interaction between two magnetic moments $\boldsymbol{\mu}_1$

and $\boldsymbol{\mu}_2$ depends on the relative orientation of the two vectors, and is given by equation (B.19). Fluctuations in this value can arise by changes in the r modulus, the $\mathbf{r}$ orientation, and the direction of the electron magnetic moment. Physically, these correspond to changes in the metal-nucleus distance (e.g., detachment of the coordinated molecule), rotation of the complex with respect to the external magnetic field, and electron spin relaxation, respectively. Each of the above processes is a random process and therefore leads to nuclear relaxation. In paramagnetic metalloproteins the electron spin relaxation is usually the fastest process and is considered to be at equilibrium with the lattice.

The equation for R_{1M} , called the Solomon equation, is expressed in terms of the probabilities of the transitions involving nuclear spin flipping. Such nuclear spin flipping involves the nuclear frequency ω_I only (single quantum transition), the sum of the nuclear frequency and the frequency corresponding to the transition frequency between the various electron spin levels, $\omega_S + \omega_I$ (double quantum transition), and $\omega_S - \omega_I$ (zero quantum transition).

$$R_{1M} = \frac{2}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2 g_e^2 \mu_B^2 S(S+1)}{r^6} \left[\frac{\tau_c}{1+(\omega_I-\omega_S)^2\tau_c^2} + \frac{3\tau_c}{1+\omega_I^2\tau_c^2} + \frac{6\tau_c}{1+(\omega_I+\omega_S)^2\tau_c^2}\right] \tag{B.27}$$

where τ_c is the correlation time related to the mechanism responsible for the relaxation:

$$\tau_c^{-1} = \tau_s^{-1} + \tau_s^{-1} + \tau_M^{-1} \tag{B.28}$$

where τ_s is the electron relaxation time, τ_r is the reorientational time and τ_M is the exchange time. An analogous equation may be derived for the transverse relaxation rate (recall the cylindrical symmetry of an NMR ex-

periment)

$$R_{2M} = \frac{1}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2 g_e^2 \mu_B^2 S(S+1)}{r^6} \left[4\tau_c + \frac{\tau_c}{1+(\omega_I-\omega_S)^2\tau_c^2} + \frac{3\tau_c}{1+\omega_I^2\tau_c^2} + \frac{6\tau_c}{1+(\omega_I+\omega_S)^2\tau_c^2} + \frac{6\tau_c}{1+(\omega_S)^2\tau_c^2}\right] \tag{B.29}$$

Both of the previous equations hold in the point dipole-point dipole approximation. Although the unpaired electrons are partially delocalized over the ligand atoms, the deviation from such approximation is neglibible for nuclei $7-8\mathring{A}$ away from the metal ion.

B.3.2 Curie-spin Relaxation

The interaction of the nuclear spins with the static magnetic moment related to $\langle S_z\rangle$ provides a further relaxation contribution. Of course, such an interaction cannot be modulated by electron relaxation because $\langle S_z\rangle$ is already an average over the electron spin states. The correlation time for the coupling is only determined by τ_r (or possibly by τ_M). This relaxation mechanism is usually called magnetic susceptibility relaxation or Curie spin relaxation. The contributions to R_{1M} and R_{2M} provided by this mechanism are

$$R_{1M} = \frac{2}{5}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2 g_e^2 \mu_B^2 S(S+1)}{r^6}\langle S_z\rangle^2 \frac{3\tau_r}{1+\omega_I^2\tau_r^2} \tag{B.30}$$

$$R_{2M} = \frac{1}{5}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2 g_e^2 \mu_B^2 S(S+1)}{r^6}\langle S_z\rangle^2 \left(4\tau_r + \frac{3\tau_r}{1+\omega_I^2\tau_r^2}\right) \tag{B.31}$$

It is to be noted that for the lanthanoids **S** should be replaced by **J**, because of the very strong spin-orbit coupling.
Since the Curie-spin shielding (or Dipolar Shielding) and the Chemical Shielding have the same symmetry with respect to the spin and space operations, they also act in the same way to affect relaxation, thus the Curie-spin relaxation can be seen as the chemical shielding anisotropy relaxation due to spanning the shielding values provided by equation (B.23)

Bibliography

[1] Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Phillips, D. C., and Shore, V. C. (1960) *Nature* **185**, 422–427.

[2] Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., and North, A. C. T. (1960) *Nature* **185**, 416–422.

[3] Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., and et al. (2001) *Science* **291**, 1305–1351.

[4] Fragai, M., Luchinat, C., and Parigi, G. (2006) *Acc.Chem.Res.* **39**, 909–917.

[5] Benvenuti, M. and Mangani, S. (2007) *Nature Protocols* **2**, 1633–1651.

[6] Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., and Bax, A. (1992) *Biochemistry* **31**, 5269–5278.

[7] Fischer, M. W., Losonczi, J. A., Weaver, J. L., and Prestegard, J. H. (1999) *Biochemistry* **38(28)**, 9013–9022.

[8] Skrynnikov, N. R., Goto, N. K., Yang, D., Choy, W.-Y., Tolman, J. R., Mueller, G. A., and Kay, L. E. (2000) *J.Mol.Biol.* **295**, 1265–1273.

[9] Chou, J. J., Li, S., Klee, C. B., and Bax, A. (2001) *Nature Struct. Mol. Biol.* **8**, 990–997.

[10] Poon, D. K. Y., Withers, S. G., and McIntosh, L. P. Jan 2007 *J.Biol.Chem.* **282(3)**, 2091–2100.

[11] Bertini, I., Calderone, V., Fragai, M., Jaiswal, R., Luchinat, C., Melikian, M., Mylonas, E., and Svergun, D. (2008) *Journal of the American Chemical Society* **130**, 7011–7021.

[12] Bertini, I., Kursula, P., Luchinat, C., Parigi, G., Vahokoski, J., Willmans, M., and Yuan, J. (2009) *Journal of the American Chemical Society* **131**, 5134–5144.

[13] Rabi, I. I., Zacharias, J. R., Millman, S., and Kusch, P. (1938) *Phys.Rev.* **53(4)**, 318–318.

[14] Purcell, E. M., Torrey, H. C., and Pound, R. V. (1946) *Phys.Rev.* **69**, 37.

[15] Bloch, F., Hansen, W. W., and Packard, M. (1946) *Physical Review* **69**, 127–127.

[16] Aue, W. P., Bartholdi, E., and Ernst, R. R. (1976) *J.Chem.Phys.* **64**, 2229–2235.

[17] Baumann, R., Kumar, A., Ernst, R. R., and Wüthrich, K. (1981) *J.Magn.Reson.* **44**, 76.

[18] Bartels, C., Xia, T. H., Billeter, M., Güntert, P., and Wüthrich, K. (1995) *J.Biomol.NMR* **5**, 1–10.

[19] Green, D. W., Ingram, V. M., and Perutz, M. F. (1954) *Proc. R. Soc. Lond. A* **225(1162)**, 287–307.

[20] Perutz, M. F. (1956) *Acta Crystallographica* **9(11)**, 867–873.

[21] Karle, J. (1980) *International Journal of Quantum Chemistry* **18(S7)**, 357–367.

[22] Hendrickson, W. A., Smith, J. L., and Sheriff, S. (1985) *Methods in Enzymology* **115**, 41–55.

[23] Hendrickson, W. A. (1995) *Science* **254(5028)**, 51–58.

[24] Bertini, I., Luchinat, C., Parigi, G., and Pierattelli, R. (2008) *Dalton Trans.* **2008**, 3782–3790.

[25] Andreini, C., Bertini, I., Cavallaro, G., Holliday, G. L., and Thornton, J. M. (2008) *J.Biol.Inorg.Chem.* **13(8)**, 1205–1218.

[26] Degtyarenko, K. Oct 2000 *Bioinformatics* **16(10)**, 851–864.

[27] Overhauser, A. W. (1953) *Phys.Rev.* **92**, 411–415.

[28] Carver, T. R. and Slichter, C. P. (1953) *Phys.Rev.* **92**, 212.

[29] Höfer, P., Parigi, G., Luchinat, C., Carl, P., Guthausen, G., Reese, M., Carlomagno, T., Griesinger, C., and Bennati, M. (2008) *Journal of the American Chemical Society* **130**, 3254–3255.

[30] Höfer, P., Carl, P., Guthausen, G., Prisner, T. F., Reese, M., Carlomagno, T., Griesinger, C., and Bennati, M. (2008) *Applied Magnetic Resonance* **34** , 393–398.

[31] Prandolini, M. J., Denysenkov, V. P., Gafurov, M., Lyubenova, S., Endeward, B., Bennati, M., and Prisner, T. F. (2008) *Applied Magnetic Resonance* **34** , 399–407.

[32] Bennati, M., Luchinat, C., Parigi, G., and Türke, M.-T. (2010) *PhysChemChemPhys* **12**, 5902–5910.

[33] Hu, K. N., Yu, H. H., Swager, T. M., and Griffin, R. G. (2004) *Journal of the American Chemical Society* **126 (35)**, 10844–10845.

[34] Matsuki, Y., Maly, T., Ouari, O., Karoui, H., Le Moigne, F., Rizzato, E., Lyubenova, S., Herzfeld, J., Prisner, T. F., Tordo, P., and Griffin, R. G. (2009) *Angew.Chem.Int.Ed.* **121** , 5096–5100.

[35] Kiesewetter, M. K., Corzilius, B., Smith, A. A., Griffin, R. G., and Swager, T. M. (2012) *J. Am. Chem. Soc.* **134(10)**, 4537–4540.

[36] Bertini, I., Luchinat, C., and Parigi, G. (2002) *Progress in Nuclear Magnetic Resonance Spectroscopy* **40**, 249–273.

[37] Bertini, I., Luchinat, C., and Parigi, G. (2002) *Concepts in Magnetic Resonance* **14**, 259–286.

[38] Petoukhov, M. V. and Svergun, D. I. (2007) *Current Opinion in Structural Biology* **17**, 562–571.

[39] Franke, D. and Svergun, D. I. (2009) *Journal Of Applied Crystallography* **42**, 342–346.

[40] Levitt, M. H. (2008) Spin Dynamics - Basics of Nuclear Magnetic Resonance, John Wiley Sons, Ltd., Chichester, 2nd edition.

[41] Duer, M. J. (2002) Solid-State NMR Spectroscopy Principles and Applications, Blackwell Science, .

[42] Pines, A., Gibby, M. G., and Waugh, J. S. (1973) *J. Chem. Phys* **59**, 569–590.

[43] Lowe, I. J. (1959) *Physical Review Letters* **2**, 285–287.

[44] Andrew, E. R., Bradbury, A., and Eades, R. G. (1958) *Nature* **182**, 1659–1659.

[45] Andrew, E. R., Bradbury, A., and Eades, R. G. (1959) *Nature* **183**, 1802–1803.

[46] Schaefer, J. and Stejskal, E. O. (1976) *Journal of the American Chemical Society* **98**, 1031–1032.

[47] Chevelkov, V., vanRossum, B. J., Castellani, F., Rehbein, K., Diehl, A., Hohwy, M., Steuernagel, S., Engelke, F., Oschkinat, H., and Reif, B. Jul 2003 *Journal of the American Chemical Society* **125(26)**, 7788–7789.

[48] Zhou, D. H., Graesser, D. T., Franks, W. T., and Rienstra, C. M. (2006) *J.Magn.Reson.* **178**, 297–307.

[49] Zhou, D. H., Shah, G., Cormos, M., Mullen, C., Sandoz, D., and Rienstra, C. M. (2007) *Journal of the American Chemical Society* **129**, 11791–11801.

[50] Chevelkov, V., Rehbein, K., Diehl, A., and Reif, B. (2006) *Angew.Chem Int.Ed Engl.* **45**, 3878–3881.

[51] Bertini, I., Emsley, L., Lelli, M., Luchinat, C., Mao, J., and Pintacuda, G. (2010) *Journal of the American Chemical Society* **132**, 5558–5559.

[52] Knight, M. J., Webber, A. L., Pell, A. J., Guerry, P., Barbet-Massin, E., Bertini, I., Felli, I. C., Gonnelli, L., Pierattelli, R., Emsley, L., Lesage, A., Hermann, T., and Pintacuda, G. (2011) *Angew.Chem.Int.Ed.* **50**, 11697–11701.

[53] Knight, M. J., Pell, A. J., Bertini, I., Felli, I. C., Gonnelli, L., Pierattelli, R., Hermann, T., Emsley, L., and Pintacuda, G. (2012) *Proc.Natl.Acad.Sci.USA* **109**, 11095–11100.

[54] Lewandowski, J. R., Dumez, J.-. N., Akbey, Ü., Franks, W. T., Emsley, L., and Oschkinat, H. (2011) *J Phys Chem Lett* **2**, 2205–2211.

[55] Bjerring, M., Paaske, B., Oschkinat, H., Akbey, Ü., and Nielsen, N. C. (2012) *Journal of Magnetic Resonance* **214**, 324–328.

[56] Webber, A. L., Pell, A. J., Barbet-Massin, E., Knight, M. J., Bertini, I., Felli, I. C., Pierattelli, R., Emsley, L., Lesage, A., and Pintacuda, G. (2012) *ChemPhysChem* **13**, 2405–2411.

[57] Loening, N. M., Bjerring, M., Nielsen, N. C., and Oschkinat, H. (2012) *Journal of Magnetic Resonance* **214**, 81–90.

[58] Nielsen, A. B., Szekely, K., Gath, J., Ernst, M., Nielsen, N. C., and Meier, B. H. (2012) *J Biomol NMR* **52(4)**, 283–288.

[59] Lamley, J. M. and Lewandowski, J. R. (2012) *Journal of Magnetic Resonance* **218**, 30–34.

[60] Giffard, M., Hediger, S., Lewandowski, J. R., Bardet, M., Simorre, J. P., Griffin, R. G., and De Päepe, G. (2012) *Phys.Chem.Chem.Phys.* **14**, 7246–7255.

[61] Luchinat, C., Parigi, G., Ravera, E., and Rinaldelli, M. (2012) *Journal of the American Chemical Society* **134(11)**, 5006–5009.

[62] Bertini, I., Bhaumik, A., De Paepe, G., Griffin, R. G., Lelli, M., Lewandowski, J. R., and Luchinat, C. Jan 2010 *Journal of the American Chemical Society* **132(3)**, 1032–1040.

[63] Nieuwkoop, A. J. and Rienstra, C. M. (2010) *Journal of the American Chemical Society* **132**, 7570–7571.

[64] Castellani, F., vanRossum, B., Diehl, A., Schubert, M., Rehbein, K., and Oschkinat, H. Nov 2002 *Nature* **420(6911)**, 98–102.

[65] Lewandowski, J. R., Sein, J., Blackledge, M., and Emsley, L. (2010) *Journal of the American Chemical Society* **132(4)**, 1246–1248.

[66] Lewandowski, J. R., Sein, J., Sass, H. J., Grzesiek, S., Blackledge, M., and Emsley, L. (2010) *Journal of the American Chemical Society* **132(24)**, 8252–8254.

[67] Lewandowski, J. R., Sass, H. J., Grzesiek, S., Blackledge, M., and Emsley, L. Sep 2011 *Journal of the American Chemical Society* **133(42)**, 16762–16765.

[68] Harbison, G. S., Herzfeld, J., and Griffin, R. G. (1983) *Biochemistry* **22**, 1–5.

[69] Harbison, G. S., Smith, S. O., Pardoen, J. A., Courtin, J. M. L., Lugtenburg, J., Herzfeld, J., Mathies, R. A., and Griffin, R. G. (1985) *Biochemistry* **24**, 6955–6962.

[70] Lewis, B. A., Harbison, G. S., Herzfeld, J., and Griffin, R. G. (1985) *Biochemistry* **24**, 4671–4679.

[71] Lugtenburg, J., Mathies, R. A., Griffin, R. G., and Herzfeld, J. (1988) *Trends Biochem. Sci.* **13**, 388–393.

[72] Costa, P. R., Kocisko, D. A., Sun, B. Q., Lansbury, P. T., and Griffin, R. G. (1997) *J. Am. Chem. Soc.* **119**, 10487–10493.

[73] Benzinger, T. L. S., Gregory, D. M., Burkoth, T. S., Miller-Auer, H., Lynn, D. G., Botto, R. E., and Meredith, S. C. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 13407–13412.

[74] Balbach, J. J., Ishii, Y., Atzutkin, O. N., Leapman, R. D., Rizzo, N. W., Dyda, F., Reed, J., and Tycko, R. (2000) *Biochemistry* **39**, 13748–13759.

[75] Habenstein, B., Bousset, L., Sourigues, Y., Kabani, M., Loquet, A., Meier, B. H., Melki, R., and Böckmann, A. (2012) *Angew.Chem Int.Ed Engl.* **51(32)**, 7963–7966.

[76] Lv, G., A., K., Giller, K., Orcellet, M. L., Riedel, D., Fernandez, C. O., Becker, S., and Lange, A. (2012) *Journal of Molecular Biology* **420(1-2)**, 99–111.

[77] Lopez delAmo, J. M., Schmidt, M., Fink, U., Dasari, M., Fändrich, M., and Reif, B. (2012) *Angew.Chem Int.Ed Engl.* **51(25)**, 6136–6139.

[78] Habenstein, B., Wasmer, C., Bousset, L., Sourigues, Y., Schutz, A., Loquet, A., Meier, B. H., Melki, R., and Bockmann, A. Nov 2011 *Journal of Biomolecular NMR* **51(3)**, 235–243.

[79] Lewandowski, J. R., Van derWel, P. C. A., Rigney, M., Grigorieff, N., and Griffin, R. G. Sep 2011 *Journal of the American Chemical Society* **133(37)**, 14686–14698.

[80] Bayro, M. J., Debelouchina, G. T., Eddy, M. T., Birkett, N. R., MacPhee, C. E., Rosay, M. M., Maas, W., Dobson, C. M., and Griffin, R. G. (2011) *Journal of the American Chemical Society* **133**, 13967–13974.

[81] Bayro, M. J., Maly, T., Birkett, N. R., MacPhee, C. E., Dobson, C. M., and Griffin, R. G. Sep 2010 *Biochemistry* **49(35)**, 7474–7484.

[82] Bertini, I., Gonnelli, L., Luchinat, C., Mao, J., and Nesi, A. (2011) *Journal of the American Chemical Society* **133**, 16013–16022.

[83] Debelouchina, G. T., Platt, G. W., Bayro, M. J., Radford, S. E., and Griffin, R. G. Aug 2010 *Journal of the American Chemical Society* **132(30)**, 10414–10423.

[84] Baldus, M. (2007) *Eur.Biophys.J.* **36**, 37–48.

[85] Luhrs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Doeli, H., Schubert, D., and Riek, R. Nov 2005 *Proceedings of the National Academy of Sciences of the United States of America* **102(48)**, 17342–17347.

[86] Shewmaker, F., Wickner, R. B., and Tycko, R. (2006) *Proc.Natl.Acad.Sci.USA* **103**, 19754–19759.

[87] Jaroniec, C. P., MacPhee, C. E., Bajaj, V. S., McMahon, M. T., Dobson, C. M., and Griffin, R. G. Jan 2004 *Proc.Natl.Acad.Sci.USA* **101(3)**, 711–716.

[88] Jaroniec, C. P., MacPhee, C. E., Astrof, N. S., Dobson, C. M., and Griffin, R. G. Dec 2002 *Proc.Natl.Acad.Sci.USA* **99(26)**, 16748–16753.

[89] Tycko, R. and Ishii, Y. Jun 2003 *Journal of the American Chemical Society* **125(22)**, 6606–6607.

[90] Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F., and Tycko, R. Dec 2002 *Proceedings of the National Academy of Sciences of the United States of America* **99(26)**, 16742–16747.

[91] Fiorin, G., Biekofsky, R. R., Pastore, A., and Carloni, P. (2005) *Proteins-Structure Function and Bioinformatics* **61(4)**, 829–839.

[92] Paramasivam, S., Suiter, C. L., Hou, G., Sun, S. J., Palmer, M., Hoch, J. C., Rovnyak, D., and Polenova, T. (2012) *J.Phys.Chem.B.* **116(25)**, 7416–7427.

[93] Loquet, A., Giller, K., Becker, S., and Lange, A. Nov 2010 *Journal of the American Chemical Society* **132(43)**, 15164–15166.

[94] Sun, S. J., Siglin, A., Williams, J. C., and Polenova, T. Jul 2009 *Journal of the American Chemical Society* **131(29)**, 10113–10126.

[95] Suwelack, D., Rothwell, W. P., and Waugh, J. S. Sep 1980 *J Chem Phys* **73 (6)**, 2559–2569.

[96] Ravera, E., Parigi, G., Mainz, A., Religa, T. L., Reif, B., and Luchinat, C. (2013).

[97] Huang, T. H., Bachovchin, W. W., Griffin, R. G., and Dobson, C. M. (1984) *Biochemistry* **23**, 5933–5937.

[98] Auger, M., McDermott, A. E., Robinson, V., Castelhano, A. L., Billedeau, R. J., Pliura, D. H., Krantz, A., and Griffin, R. G. (1993) *Biochemistry* **32**, 3930–3934.

[99] Weliki, D. P., Bennet, A. E., Zvi, A., Anglister, J., Steinbach, P. J., and Tycko, R. (1999) *Nat. Struct. Mol. Biol.* **6**, 141–145.

[100] Hall, D. A., Maus, D. C., Gerfen, G. J., Inati, S. J., Becerra, L. R., Dahlquist, F. W., and Griffin, R. G. (1997) *Science* **276**, 930–932.

[101] Martin, R. W. and Zilm, K. W. (2003) *J.Magn.Reson.* **165**, 162–174.

[102] Siemer, A. B., Huang, K.-Y., and McDermott, A. E. (2012) *PLOS ONE* **7(10)**, e47242.

[103] Linden, A. H., Franks, W. T., Akbey, Ü., Lange, S., vanRossum, B.-J., and Oschkinat, H. (2011) *Journal of Biomolecular NMR* **51(3)**, 283–292.

[104] Rothgeb, T. M. and Oldfield, E. (1981) *J. Biol. Chem.* **256**, 1432–1446.

[105] Keniry, M. A., Rothgeb, T. M., Smith, R. L., Gutowsky, H. S., and Oldfield, E. (1983) *Biochemistry* **22**, 1917–1926.

[106] Baianu, I. C., Gutowsky, H. S., and Oldfield, E. (1984) *Biochemistry* **23**, 3105–3110.

[107] Bertini, I., Giachetti, A., Luchinat, C., Parigi, G., Petoukhov, M. V., Pierattelli, R., Ravera, E., and Svergun, D. I. (2010) *Journal of the American Chemical Society* **132**, 13553–13558.

[108] Bertini, I., Ferella, L., Luchinat, C., Parigi, G., Petoukhov, M. V., Ravera, E., and Rosato, A. (2012) *J.Biomol.NMR* **53(4)**, 271–280.

[109] Bertini, I., Luchinat, C., Nagulapalli, M., Parigi, G., and Ravera, E. (2012) *Phys.Chem.Chem.Phys.* **14**, 9149–9156.

[110] Wang, X., Srisailam, S., Ye, A. A., Lemak, A., Arrowsmith, C., Prestegard, J. H., and Tian, F. (2007) *J.Biomol.NMR* **39** , 53–61.

[111] Lindorff-Larsen, K., Best, R. B., DePristo, M. A., Dobson, C. M., and Vendruscolo, M. (2005) *Nature* **433**, 128–132.

[112] Lange, O. F., Lakomek, N.-A., Farés, C., Schröder, G. F., Walter, K. F. A., Becker, S., Meiler, J., Grubmüller, H., Griesinger, C., and deGroot, B. L. (2008) *Science* **320**, 1471–1475.

[113] Clore, G. M. and Schwieters, C. D. (2006) *J.Mol.Biol.* **355** , 879–886.

[114] Zhang, Q., Stelzer, A. C., Fisher, C. K., and Al-Hashimi, H. M. (2007) *Nature* **450**, 1263–1267.

[115] Vögeli, B., Kazemi, S., Güntert, P., and Riek, R. (2012) *Nat. Struct. Mol. Biol.* **19**, 1053–1057.

[116] Iwahara, J., Schwieters, C. D., and Clore, G. M. May 2004 *Journal of the American Chemical Society* **126(18)**, 5879–5896.

[117] Tang, C., Iwahara, J., and Clore, G. M. Nov 2006 *Nature* **444**, 383–386.

[118] Clore, G. M. and Schwieters, C. D. (2004) *Journal of the American Chemical Society* **126** , 2923–2938.

[119] Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M. R., Zweckstetter, M., and Blackledge, M. (2010) *Journal of the American Chemical Society* **132** , 8407–8418.

[120] Huang, J. and Grzesiek, S. (2010) *Journal of the American Chemical Society* **132** , 694–705.

[121] Allison, J. R., Varnai, P., Dobson, C. M., and Vendruscolo, M. (2009) *Journal of the American Chemical Society* **131**, 18314–18326.

[122] Fisher, C. and Stultz, C. (2011) *Current opinion in structural biology* **21(3)**, 426–431.

[123] Fisher, C., Huang, A., and Stultz, C. (2010) *Journal of the American Chemical Society* **132(42)**, 14919.

[124] Volkov, A. N., Ubbink, M., and Van Nuland, N. A. J. (2010) *J.Biomol.NMR* **48** , 225–236.

[125] Volkov, A. N., Worrall, J. A. R., Holtzmann, E., and Ubbink, M. (2006) *Proc.Natl.Acad.Sci.USA* **103** , 18945–18950.

[126] Bashir, Q., Volkov, A. N., Ullmann, G. M., and Ubbink, M. (2010) *Journal of the American Chemical Society* **132**, 241–247.

[127] Tang, C., Schwieters, C. D., and Clore, G. M. (2007) *Nature* **449** , 1078–1082.

[128] Mackereth, C., Madl, T., Bonnal, S., Simon, B., Zanier, K., Gasch, A., Rybin, V., Valcárcel, J., and Sattler, M. (2011) *Nature* **475(7356)**, 408–411.

[129] Iwahara, J. and Clore, G. M. Apr 2006 *Nature* **440(7088)**, 1227–1230.

[130] Garrett, D. S., Seok, Y. J., Liao, D. I., Peterkofsky, A., Gronenborn, A. M., and Clore, G. M. (1997) *Biochemistry* **36**, 2517–2530.

[131] Fawzi, N. L., Doucleff, M., Suh, J. Y., and Clore, G. M. Jan 2010 *Proc.Natl.Acad.Sci.U.S.A.* **107**, 1379–1384.

[132] Xu, X., Reinle, W., Hannemann, F., Konarev, P. V., Svergun, D. I., Bernhardt, R., and Ubbink, M. (2008) *Journal of the American Chemical Society* **130** , 6395–6403.

[133] Xu, X., Keizers, P. H. J., Reinle, W., Hannemann, F., Bernhardt, R., and Ubbink, M. (2009) *J.Biomol.NMR* **43** , 247–254.

[134] Bertini, I., Janik, M. B. L., Lee, Y.-M., Luchinat, C., and Rosato, A. (2001) *Journal of the American Chemical Society* **123(18)**, 4181–4188.

[135] Bertini, I., Del Bianco, C., Gelis, I., Katsaros, N., Luchinat, C., Parigi, G., Peana, M. ., Provenzani, A., and Zoroddu, M. A. (2004) *Proc.Natl.Acad.Sci.USA* **101**, 6841–6846.

[136] Wohnert, J., Franz, K. J., Nitz, M., Imperiali, B., and Schwalbe, H. Nov 2003 *Journal of the American Chemical Society* **125(44)**, 13338–13339.

[137] Prudencio, M., Rohovec, J., Peters, J. A., Tocheva, E., Boulanger, M. J., Murphy, M. E., Hupkes, H. J., Koster, W., Impagliazzo, A., and Ubbink, M. (2004) *Chemistry - A European Journal* **5**, 3252–3260.

[138] Martin, L. J., H?hnke, M. J., Wöhnert, J., Silvaggi, N. R., Allen, K. N., Schwalbe, H., and Imperiali, B. (2007) *Journal of the American Chemical Society* **129**, 7106–7113.

[139] Keizers, P. H., Desreux, J. F., Overhand, M., and Ubbink, M. (2007) *Journal of the American Chemical Society* **129**, 9292–9293.

[140] Su, X. C., Man, B., Beeren, S., Liang, H., Simonsen, S., Schmitz, C., Huber, T., Messerle, B. A., and Otting, G. (2008) *Journal of the American Chemical Society* **130**, 10486–10487.

[141] Keizers, P. H. J., Saragliadis, A., Hiruma, Y., Overhand, M., and Ubbink, M. (2008) *Journal of the American Chemical Society* **130**, 14802–14812.

[142] Häussinger, D., Huang, J., and Grzesiek, S. (2009) *Journal of the American Chemical Society* **131**, 14761–14767.

[143] Hass, M. A. S., Keizers, P. H. J., Blok, A., Hiruma, Y., and Ubbink, M. (2010) *Journal of the American Chemical Society* **132**, 9952–9953.

[144] Man, B., Su, X. C., Liang, H., Simonsen, S., Huber, T., Messerle, B. A., and Otting, G. (2010) *Chem.Eur.J.* **16**, 3827–3832.

[145] Das Gupta, S., Hu, X., Keizers, P. H. J., Liu, W.-M., Luchinat, C., Nagulapalli, M., Overhand, M., Parigi, G., Sgheri, L., and Ubbink, M. (2011) *J.Biomol.NMR* **51**, 253–263.

[146] Keizers, P. H. J. and Ubbink, M. (2011) *Prog.Nucl.Magn Reson.Spectrosc.* **58(1-2)**, 88–96.

[147] Longinetti, M., Luchinat, C., Parigi, G., and Sgheri, L. (2006) *Inv.Probl.* **22**, 1485–1502.

[148] Longinetti, M., Parigi, G., and Sgheri, L. (2002) *J.Phys.A:Math.Gen.* **35**, 8153–8169.

[149] Bertini, I., Gupta, Y. K., Luchinat, C., Parigi, G., Peana, M., Sgheri, L., and Yuan, J. (2007) *Journal of the American Chemical Society* **129**, 12786–12794.

[150] Luchinat, C., Nagulapalli, M., Parigi, G., and Sgheri, L. (2012) *J.Magn.Reson.* **215**, 85–93.

[151] Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. May 2007 *Journal of the American Chemical Society* **129(17)**, 5656–5664.

[152] Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1988) Numerical Recipes in C-The Art of Scientific Computing, Cambridge University Press, New York.

[153] Kleywegt, G. J. (1997) *J.Mol.Biol.* **273** , 371–376.

[154] Babu, Y. S., Bugg, C. E., and Cook, W. J. (1988) *J.Mol.Biol.* **204**, 191–204.

[155] Chattopadhyaya, R., Meador, W. E., Means, A. R., and Quiocho, F. A. (1992) *J.Mol.Biol.* **228**, 1177–1192.

[156] Seaton, B. A., J.F., H., and Richardson, F. M. (1985) *Biochemistry* **24(24)**, 6740–6743.

[157] Majava, V., Petoukhov, M. V., Hayashi, N., Pirila, P., Svergun, D. I., and Kursula, P. (2008) *BMC.Struct.Biol.* **8**, 10.

[158] Wriggers, W., Mehler, E., Pitici, F., Weinstein, H., and Schulten, K. (1998) *Biophys.J.* **74**, 1622–1639.

[159] Barton, N. P., Verma, C. S., and Caves, L. S. D. (2002) *J.Phys.Chem.B* **106**, 11036–11040.

[160] Baber, J. L., Szabo, A., and Tjandra, N. Mar 2001 *Journal of the American Chemical Society* **123**, 3953–3959.

[161] Chang, S.-G., Szabo, A., and Tjandra, N. (2003) *Journal of the American Chemical Society* **125(37)**, 11379–11384.

[162] Wilson, M. A. and Brunger, A. T. (2000) *J.Mol.Biol.* **301(5)**, 1237–1256.

[163] Fallon, J. L. and Quiocho, F. A. (2003) *Structure* **11(10)**, 1303–1307.

[164] Bernado, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R., and Blackledge, M. (2005) *Proc.Natl.Acad.Sci.U.S.A* **102(47)**, 17002–17007.

[165] Gabel, F., Simon, B., Nilges, M., Petoukhov, M. V., Svergun, D., and Sattler, M. (2008) *J.Biomol.NMR* **41(4)**, 199–208.

[166] Brüschweiler, R., Roux, B., Blackledge, M., Griesinger, C., Karplus, M., and Ernst, R. R. (1992) *Journal of the American Chemical Society* **114**, 2289–2302.

[167] Lipari, G. and Szabo, A. (1982) *Journal of the American Chemical Society* **104**, 4546–4559.

[168] Iwahara, J. and Clore, G. M. (2010) *Journal of the American Chemical Society* **132** , 13346–13356.

[169] Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) *Journal of the American Chemical Society* **112**, 4989–4991.

[170] Ryabov, Y. E. and Fushman, D. (2007) *Journal of the American Chemical Society* **129(11)**, 3315–3327.

[171] Bernado, P., Garcia de laTorre, J., and Pons, M. (2002) *J.Biomol.NMR* **23**, 139–150.

[172] Anthis, N. J., Doucleff, M., and Clore, G. M. (2011) *Journal of the American Chemical Society* **133**, 18966–18974.

[173] Bertini, I., Fragai, M., Luchinat, C., Melikian, M., Mylonas, E., Sarti, N., and Svergun, D. (2009) *J.Biol.Chem.* **284(19)**, 12821–12828.

[174] Arnold, L. H., Butt, L. E., Prior, S. H., Read, C. M., Fields, G. B., and Pickford, A. R. Dec 2011 *Journal of Biological Chemistry* **286(52)**, 45073–45082.

[175] Bertini, I., Fragai, M., Luchinat, C., Melikian, M., Toccafondi, M., Lauer, J. L., and Fields, G. B. (2012) *Journal of the American Chemical Society* **134** , 2100–2110.

[176] Jozic, D., Bourenkov, G., Lim, N. H., Visse, R., Nagase, H., Bode, W., and Maskos, K. Mar 2005 *J.Biol.Chem.* **280(10)**, 9578–9585.

[177] Iyer, S., Visse, R., Nagase, H., and Acharya, K. R. Sep 2006 *J.Mol.Biol.* **362(1)**, 78–88.

[178] Manka, S. W., Carafoli, F., Visse, R., Bihan, D., Raynal, N., Farndale, R. W., Murphy, G., Enghild, J. J., Hohenester, E., and Nagase, H. Jul 2012 *Proc.Natl.Acad.Sci.U.S.A* **109(31)**, 12461–12466.

[179] Li, J., Brick, P., Ohare, M. C., Skarzynski, T., Lloyd, L. F., Curry, V. A., Clark, I. M., Bigg, H. F., Hazleman, B. L., Cawston, T. E., and Blow, D. M. Jun 1995 *Structure* **3(6)**, 541–549.

[180] Rosenblum, G., Van denSteen, P. E., Cohen, S. R., Grossmann, J. G., Frenkel, J., Sertchook, R., Slack, N., Strange, R. W., Opdenakker, G., and Sagi, I. (2007) *Structure* **15**, 1227–1236.

[181] Nagulapalli, M., Parigi, G., Yuan, J., Gsponer, J., Deraos, S., Bamm, V. V., Harauz, G., Matsoukas, J., dePlanque, M. R. R., Gerothanassis, I. P., Babu, M. M., Luchinat, C., and Tzakos, A. G. (2012) *Structure* **20**, 522–533.

[182] Bertini, I., Luchinat, C., Parigi, G., Ravera, E., Reif, B., and Turano, P. (2011) *Proc.Natl.Acad.Sci.USA* **108**, 10396–10399.

[183] Bertini, I., Engelke, F., Luchinat, C., Parigi, G., Ravera, E., Rosa, C., and Turano, P. (2012) *Phys.Chem.Chem.Phys.* **14**, 439–447.

[184] Bertini, I., Engelke, F., Gonnelli, L., Knott, B., Luchinat, C., Osen, D., and Ravera, E. (2012) *J.Biomol.NMR* **54(2)**, 123–127.

[185] Doty, F., Entzminger, G., and Yang, Y. (1998) *Concepts in Magnetic Resonance* **10(4)**, 239–260.

[186] Van Holde, K. E. and Baldwin, R. L. (1958) *J.Phys.Chem.* **62(6)**, 734–743.

[187] Minton, A. P. and Lewis, M. S. (1981) *Biophys.Chem.* **14** , 317–324.

[188] Lundh, S. (1980) *Journal of Polymer Science: Polymer Physics Edition* **18**, 1963–1978.

[189] Lundh, S. Aug 1985 *Archives of Biochemistry and Biophysics* **241(1)**, 265–274.

[190] Wang, Y., Li, C., and Pielak, G. J. (2010) *Journal of the American Chemical Society* **132**, 9392–9397.

[191] Koenig, S. H. and Brown, R. D. (1990) *Prog. Nucl. Magn. Res. Spect.* **22(6)**, 487–567.

[192] Mainz, A., Jehle, S., vanRossum, B. J., Oschkinat, H., and Reif, B. (2009) *Journal of the American Chemical Society* **131**, 15968–15969.

[193] Gardiennet, C., Schütz, A. K., Hunkeler, A., Kunert, B., Terradot, L., Böckmann, A., and Meier, B. H. (2012) *Angew.Chem.Int.Ed* **51**, 7855–7858.

[194] Mainz, A., Bardiaux, B., Kuppler, F., Multhaup, G., Felli, I. C., Pierattelli, R., and Reif, B. (2012) *J.Biol.Chem.* **287**, 1128–1138.

[195] Andersson, K. and Hovmoller, S. Jul 2000 *Acta Crystallogr D Biol Crystallogr* **56(7)**, 789–790.

[196] Laage, S., Lesage, A., Emsley, L., Bertini, I., Felli, I. C., Pierattelli, R., and Pintacuda, G. (2009) *Journal of the American Chemical Society* **131**, 10816–10817.

[197] Böckmann, A., Gardiennet, C., Verel, R., Hunkeler, A., Loquet, A., Pintacuda, G., Emsley, L., Meier, B. H., and Lesage, A. (2009) *Journal of Biomolecular NMR* **45**, 319–327.

[198] Baldwin, A. J., Walsh, P., Hansen, D. F., Hilton, G. R., Benesh, J. L. P., Sharpe, S., and Kay, L. E. (2012) *J. Am. Chem. Soc.* **134(37)**, 15343–15350.

[199] Ravera, E., Corzilius, B., Michaelis, V. K., Rosa, C., Griffin, R. G., Luchinat, C., and Bertini, I. Dynamic nuclear polarization of sedimented solutes (2013).

[200] Ravera, E., Corzilius, B., Michaelis, V. K., Griffin, R. G., Luchinat, C., and Bertini, I. Systematic investigation of natural abundance bovine serum albumin sediments and concentrated solutions using high field dnp-enhanced mas nmr (2013).

[201] Keith, S. C. (1913) *Science* **37**, 877–879.

[202] Polge, C., Smith, A. U., and Parkes, A. S. (1949) *Nature* **164**, 666–666.

[203] Tamiya, T., Okahashi, N., Sakuma, R., Aoyama, T., Akahane, T., and Matsumoto, J. J. (1985) *Cryobiology* **22**, 446–456.

[204] S., G. D. (1958) *J. Am. Chem. Soc.* **80(15)**, 3892–3898.

[205] Kervern, G., Steuernagel, S., Engelke, F., Pintacuda, G., and Emsley, L. (2007) *Journal of the American Chemical Society* **129**, 14118–14119.

[206] Alla, M. and Lippmaa, E. (1982) *Chem.Phys.Lett.* **87**, 30–33.

[207] Ernst, M., Samoson, A., and Meier, B. H. (2001) *Chem Phys Lett* **348**, 293–302.

[208] Kotecha, M., Wickramasinghe, N. P., and Ishii, Y. (2007) *Magn. Reson. Chem.* **45**, S221–S230.

[209] Laage, S., Sachleben, J., Steuernagel, S., Pierattelli, R., Pintacuda, G., and Emsley, L. (2008) *J.Magn.Reson.* **196**, 133–141.

[210] Weingarth, M., Bodenhausen, G., and Tekely, P. J. (2009) *J. Magn. Reson.* **199**, 238–241.

[211] Wickramasinghe, N. P. and Ishii, Y. Aug 2006 *Journal of Magnetic Resonance* **181(2)**, 233–243.

[212] Wickramasinghe, N. P., Shaibat, M. A., Jones, C. R., Casabianca, L. B., deDios, A. C., Harwood, J. S., and Ishii, Y. (2008) *J. Chem. Phys* **128**, xx–xx.

[213] Wickramasinghe, N. P., Parthasarathy, S., Jones, C. R., Bhardwaj, C., Long, F., Kotecha, M., Mehboob, S., Fung, L. W. M., Past, J., Samoson, A., and Ishii, Y. (2009) *Nature Meth.* **6**, 215–218.

[214] De Paépe, G., Lewandowski, J. R., Loquet, A., Böckmann, A., and Griffin, R. G. (2008) *J.Chem.Phys.* **129**, 245101.

[215] Reif, B. and Griffin, R. G. Jan 2003 *J.Magn.Reson.* **160(1)**, 78–83.

[216] Grommek, A., Meier, B. H., and Ernst, M. (2006) *Chem.Phys.Lett.* **427**, 404–409.

[217] Hu, B., Lafon, O., Trébosc, J., Chen, Q., and Amoureux, J.-P. (2011) *Journal of Magnetic Resonance* **212 (2)**, 320–329.

[218] Ma, L., Jorgensen, A. M. M., Sorensen, G. O., Ulstrup, J., and Led, J. J. (2000) *Journal of the American Chemical Society* **122**, 9473–9485.

[219] Nadaud, P. S., Helmus, J. J., Höfer, N., and Jaroniec, C. P. (2007) *Journal of the American Chemical Society* **129(24)**, 7502–7503.

[220] Nadaud, P. S., Helmus, J. J., Kall, S. L., and Jaroniec, C. P. (2009) *Journal of the American Chemical Society* **131**, 8108–8120.

[221] Balayssac, S., Bertini, I., Bhaumik, A., Lelli, M., and Luchinat, C. (2008) *Proc.Natl.Acad.Sci.USA* **105**, 17284–17289.

[222] Elena, B. and Emsley, L. (2005) *Journal of the American Chemical Society* **127**, 9140–9146.

[223] Elena, B., Pintacuda, G., Mifsud, N., and Emsley, L. Jul 2006 *Journal of the American Chemical Society* **128(29)**, 9555–9560.

[224] Webber, A. L., Emsley, L., Claramunt, R. M., and Brown, S. P. (2010) *J Phys Chem A* **2010(114)**, 10435–10442.

[225] Bradley, J. P., Velaga, S. P., Antzutkin, O. N., and Brown, S. P. Jun 2011 *Cryst.Growth Des.* **11** , 3463–3471.

[226] Dumez, J.-N. and Emsley, L. (2011) *Phys.Chem.Chem.Phys.* **13**, 7363–7370.

[227] Geppi, M., Mollica, G., Borsacchi, S., and Veracini, C. A. (2008) *Appl.Spectr.Rev.* **43** , 202–302.

[228] Pintacuda, G., John, M., Su, X. C., and Otting, G. (2007) *Acc.Chem.Res.* **40** , 206–212.

[229] Bertini, I., Luchinat, C., and Parigi, G. (2011) *Coord.Chem.Rev.* **255**, 649–663.

[230] Schmitz, C., Stanton-Cook, M. J., Su, X. C., Otting, G., and Huber, T. (2008) *J.Biomol.NMR* **41** , 179–189.

[231] Kervern, G., D'Aléo, A., Toupet, L., Maury, O., Emsley, L., and Pintacuda, G. (2009) *Angew.Chem Int.Ed Engl.* **48**, 3082–3086.

[232] Balayssac, S., Bertini, I., Lelli, M., Luchinat, C., Maletta, M., and Yeo, K. J. (2007) *Journal of the American Chemical Society* **129**, 2218–2219.

[233] Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr D Biol Crystallogr* **1 (53)**, 240–255.

[234] Bertini, I., Calderone, V., Cosenza, M., Fragai, M., Lee, Y.-M., Luchinat, C., Mangani, S., Terni, B., and Turano, P. (2005) *Proc.Natl.Acad.Sci.USA* **102**, 5334–5339.

[235] Cullity, B. D. and Stock, S. R. Feb 2001 Elements of X-Ray Diffraction, volume **3rd**, Prentice Hall, .

[236] Margiolaki, I., Wright, J. P., Wilmanns, M., Fitch, A. N., and Pinotsis, N. (2007) *Journal of the American Chemical Society* **129** , 11865–11871.

[237] Von Dreele, R. B. (2003) *Methods Enzymol.* **368** , 254–267.

[238] Herrmann, T., Güntert, P., and Wüthrich, K. May 2002 *J.Mol.Biol.* **319(1)**, 209–227.

[239] Balayssac, S., Bertini, I., Luchinat, C., Parigi, G., and Piccioli, M. (2006) *Journal of the American Chemical Society* **128**, 15042–15043.

[240] Matthews, B. W. (1968) *Journal of Molecular Biology* **33**, 491–497.

[241] Matthews, B. W. (1976) *Annu.Rev.Phys.Chem.* **27**, 493–523.

[242] Rothwell, W. P. and Waugh, J. S. Mar 1981 *J Chem Phys* **74(5)**, 2721–2732.

[243] Laage, S., Marchetti, A., Sein, J., Pierattelli, R., Sass, H. J., Grzesiek, S., Lesage, A., Pintacuda, G., and Emsley, L. (2008) *Journal of the American Chemical Society* **130**, 17216–17217.

[244] Ray, S., Ladizhansky, V., and Vega, S. (1998) *J. Magn. Reson.* **135**, 427–434.

[245] Marks, D. and Vega, S. (1996) *J. Magn. Reson. Ser. A* **118**, 157–172.

[246] Stejskal, E. O., Schaefer, J., and Waugh, J. S. (1977) *J. Magn. Reson.* **28**, 105–112.

[247] Engelke, F., Kind, T., Michel, D., Pruski, M., and Gerstein, B. C. (1991) *J. Magn. Reson.* **95**, 286–298.

[248] Becerra, L. R., Gerfen, G. J., Temkin, R. J., Singel, D. J., and Griffin, R. G. (1993) *Phys.Rev.Lett.* **71**, 3561.

[249] Barnes, A. B., Mak-Jurkauskas, M. L., Matsuki, Y., Bajaj, V. S., Van derWel, P. C. A., DeRocher, R., Bryant, J., Sirigiri, J. R., Temkin, R. J., Lugtenburg, J., Herzfeld, J., and Griffin, R. G. (2009) *Journal of Magnetic Resonance* **198(2)**, 261–270.

[250] Allen, P. J., Creuzet, F., deGroot, H. J. M., and Griffin, R. G. (1991) *Journal of Magnetic Resonance* **92(3)**, 614–617.

[251] Pines, A., Gibby, M. G., and Waugh, J. S. (1972) *J Chem Phys* **56**, 1776–1777.

[252] Bennett, A. E., Rienstra, C. M., Auger, M., Lakshmi, K. V., and Griffin, R. G. Oct 1995 *Journal of Chemical Physics* **103(16)**, 6951–6958.

[253] Lohman, J. A. B. and Maclean, C. (1978) *Chemical Physics* **35**, 269–274.