DOTTORATO DI RICERCA IN SCIENZE CHIMICHE

CICLO XXXIII

COORDINATORE Prof. PIERO BAGLIONI

Theory and Applications of Magnetic Resonance

to Biomolecules in solution and in the solid state

Teoria e Applicazioni della Risonanza Magnetica

di Biomolecole in soluzione e allo stato solido

Settore Scientifico Disciplinare CHIM/03

Dottorando
Lucia Gigli

Tutore
Dr. Enrico Ravera

Coordinatore
Prof. Piero Baglioni

Anni 2017/2020

Abstract

This thesis aims at expanding the applicability of Magnetic Resonance to biological systems. The presented work relies upon a fine interplay among the development of theoretical models, their numerical implementation and their experimental verification, and can be conceptually divided in two parts.

Part I, "Theoretical Chemistry for NMR", proposes a novel approach for the structural refinement through the combined use of paramagnetic NMR and Quantum Chemical calculations, and the recent advancements of a method aimed at studying the conformational heterogeneity of flexible macromolecules.

Part II, "NMR for Biomaterials", describes the experimental results obtained in the frame of a project for the development of a prototypical protein-based catalyst for water depuration. The atomic-level characterization of the interactions among the components of a hybrid composite, by means of solid-state NMR combined with other techniques, is provided. A denoising method for 2D solid-state NMR spectra is also presented.

Contents

Preface **5**

I Theoretical Chemistry for NMR **9**

1 Magneto-structural correlations via pNMR and *ab initio* methods **12**

1.1 Introduction 12

1.1.1 The magnetic susceptibility 13

1.1.2 The magnetic susceptibility in modern QC 14

1.1.3 Paramagnetic NMR shifts 15

1.2 Structural refinement through PCS and FCS 17

1.2.1 Introduction 17

1.2.2 Methods 18

1.2.3 Results and discussion 22

1.3 PCS-driven structural refinement for metalloproteins 27

1.3.1 Introduction 27

1.3.2 Methods 29

1.3.3 Results and discussion 31

2 Conformational heterogeneity in biomolecules through the MaxOcc method 42

2.1 Introduction 42

2.1.1 Ill-posed inverse problems 43

2.1.2 Ensemble reconstruction methods 44

2.2 The MaxOcc method 46

2.3 The MaxOcc applications 49

2.4 Integration of DEER data in the MaxOcc analysis 50

2.4.1 Introduction 50

2.4.2 Methods 51

2.4.3 Results and discussion 55

2.5 MaxOcc for Intrinsically Disordered Regions 61

2.5.1 Introduction 61

2.5.2 Methods 65

2.5.3 Results and discussion 65

3 Conclusions and outlook 69

II NMR for Biomaterials 72

4 Proteins as catalysts 76

4.1 Introduction 76

4.2 Photoinduced catalysis for water depuration 82

4.2.1 Introduction 82

4.2.2 Methods 82

4.2.3 Results and discussion 84
4.3 Integrated studies on lysozyme-based composites 88
4.3.1 Introduction 88
4.3.2 Methods 89
4.3.3 Results and discussion 93
4.4 Multivariate Curve Resolution for 2D ssNMR 107
4.4.1 Introduction 107
4.4.2 Methods 110
4.4.3 Results and discussion 112

5 Conclusions and outlook **119**

6 Appendices **122**
6.1 Experimental techniques 122
6.2 Supporting Information 126

Bibliography **133**

List of Tables **191**

List of Figures **201**

Preface

Nature is undoubtedly as fascinating as complex, and the attempt to unveil the rules that govern it is probably as old as humanity. This search, initially driven by pure intellectual curiosity, showed however, since almost the beginning, to have also very practical consequences. The knowledge thus gained could indeed be exploited for improving the tools that help the survival, the health, or just the wellness of mankind. In other words, the comprehension of the natural laws is an essential prerequisite for what today we call technology.

This thesis addresses both the theoretical and practical aspects of research, showing how some pure theoretical findings can have important consequences for the comprehension of complex systems like biological macromolecules and can support experiments in the development of technological devices.

Although occurring in very different forms, the leitmotiv of all the presented topics is *Nuclear Magnetic Resonance* (NMR), because of its deep relationship with electronic structure and magnetism. NMR is a physical phenomenon in which nuclear spins interact with the electromagnetic radiation in the presence of an external magnetic field and respond by pro-

ducing an electromagnetic signal whose frequency is characteristic of the magnetic field present at the nucleus. Exploiting this effect, NMR spectroscopy is thus able to provide atomic-level information on molecules, and has currently reached such a level of maturity to allow studying very complex systems, both diamagnetic and paramagnetic, in solution (Part I) and in the solid state (Part II).

The presented work is mainly focused on biological systems, whose importance does not need explanations. In particular, the complexity of the human genome and in general of living systems poses the problem of solving the 3D structures of gene products, namely proteins. The most powerful methods to investigate their structure are NMR, X-ray Diffraction (XRD), and cryo-Electron Microscopy (cryo-EM), highly complementary techniques that look at proteins from different perspectives. Crystallography is fast, does not suffer from protein size limitations, and provides a unique structure, but it cannot solve unfolded proteins and, above all, obtaining crystals is not always possible [1]. Cryo-EM does not need crystallization, but the samples must be highly homogeneous, and high resolution is at the present time a challenge for proteins smaller than 200 kDa [2]. NMR, on the contrary, does not need particular sample handling, can investigate proteins both in solution and in the solid state and, despite the lower effective resolution, provides additional atomic-level information on the structural properties of the system. Furthermore, it allows studying conformational dynamics and operates closer to biologically relevant conditions [3].

A fine interplay among different experimental techniques is therefore needed [4, 5], and integration must rely upon strong computational tools. Solving the structure of proteins is indeed crucial for the purposes of both

basic and applied science, as this thesis tries to demonstrate. Part I, showing some results from theoretical research activities, is aimed at providing new methods to investigate the structure of matter through the use of simulated and experimental paramagnetic NMR (pNMR) observables. Part II, more practically oriented, is aimed at developing a protein-based catalyst for the removal of water pollutants, whose realization is made possible by the integration of solid-state NMR (ssNMR) with other techniques.

Part I

Theoretical Chemistry for NMR

Introduction

The approach to scientific investigation is essentially based on the schematization of reality by means of models. A model is a simplified and possibly idealized representation of a system or a process aimed at transforming the observation of reality into something that can be analyzed with the tools of logics and mathematics. An optimal model is a compromise between simplicity and accuracy: it must include all the features of the system that are significant for the problem under investigation, but nothing more, because any redundant or negligible additional element is just a useless complication that reduces the comprehension of the phenomenon and the applicability of the model. The model is usually built on the basis of previous knowledge and then refined for explaining the results of experimental observations, and is used for making predictions. A reasonably accurate model, however, is in most cases so complicated to make impossible a manual handling, and, in the largest majority of situations, it can only be used and analyzed via a computer implementation.

Computers are fast, never get bored or tired, and are virtually error free. Moreover, they can be considered an additional exploratory tool. When the conditions of the test are extreme, or the investigated proper-

ties experimentally inaccessible, this has no effect in a simulated world: the computer does not collapse in a molten state if asked to reproduce the pressure at the center of the Earth. It is thus natural that the computational techniques are increasingly becoming the bread-and-butter of the modern scientist, and chemists make no exception. It is in fact now hard thinking to work without simulations in the interpretation of experimental results, and modern Chemistry strongly relies on computers.

Theoretical Chemistry, in particular, makes large use of computers to satisfy a request which has profound implications in many fields: getting the highest possible structural resolution, as well as information on the conformational heterogeneity. Even very minor structural and conformational changes can indeed have important consequences, which, in the case of biological systems, can result in the insurgence of a disease.

This part of the thesis is devoted to show the importance of molecular simulations for the investigation of the structural and conformational preferences of macromolecules, with a particular focus on proteins. Due to the profound relationship between electronic structure and magnetic properties, NMR parameters, combined with computational techniques, are an invaluable source of information. The subject is presented in two chapters: Chapter 1 describes two structural refinement approaches that combine Quantum Chemical (QC) methods and paramagnetic NMR, Chapter 2 shows the recent advancements in MaxOcc, an ensemble reconstruction method.

Chapter 1

Magneto-structural correlations via pNMR and *ab initio* methods

1.1 Introduction

The term *magneto-structural correlations* encompasses all the relationships between the magnetic properties of a material and its molecular and electronic structure.

The property addressed in this chapter is the *magnetic susceptibility*, a quantity that depends on the electronic structure and determines NMR observables such as hyperfine interactions.

1.1.1 The magnetic susceptibility

The magnetic susceptibility χ quantifies how a material responds to an external magnetic field [6, 7]. In the simplest case, the system response is isotropic, so χ is a scalar quantity. More often, however, the response is *anisotropic*, therefore the susceptibility is mathematically described by a 2-rank tensor $\boldsymbol{\chi}$ [8], whose components (χ_{ij}) are the partial derivatives, at constant temperature, of the magnetization ($\boldsymbol{M}$) over the magnetic field ($\boldsymbol{H}$):

$$\chi_{ij} = \left(\frac{\partial M_j}{\partial H_i}\right)_T . \tag{1.1}$$

It is convenient to define the *magnetic susceptibility anisotropy*, $\Delta\boldsymbol{\chi}$, as the difference between the total $\boldsymbol{\chi}$ and its isotropic part. The axial and rhombic components of $\Delta\boldsymbol{\chi}$ can be thus defined as:

$$\Delta\chi_{ax} = \tilde{\chi}_{zz} - \frac{\tilde{\chi}_{xx} + \tilde{\chi}_{yy}}{2}; \qquad \Delta\chi_{rh} = \tilde{\chi}_{xx} - \tilde{\chi}_{yy}, \tag{1.2}$$

where $\tilde{\chi}_{xx}, \tilde{\chi}_{yy}, \tilde{\chi}_{zz}$ are the components of the magnetic susceptibility tensor in its principal frame (i.e. the one defined by its eigenvectors), and $\tilde{\chi}_{zz}$ is the component that differs most from the other two.

Being a magnetic quantity, $\boldsymbol{\chi}$ intimately depends on the electronic properties of the system and is thus a reporter of the interactions in which electrons are involved. Each electron interacts with the nucleus, with the other electrons, and with its own motion, through the so-called spin-orbit coupling (SOC), and the relative weights of these contributions can signif-

icantly change depending on the system. As a consequence, the nucleus-electron and the electron-electron interactions determine the relation of χ with NMR restraints, such as hyperfine shifts (Section 1.1.3), and with the effective magnetic moment.

The magnetic susceptibilty can be measured, directly, with several techniques, such as Superconducting Quantum Interference Device (SQUID), torque magnetometry [9–11] and NMR [12–17], or, indirectly, from EPR parameters [18–23], and can be calculated through Quantum Chemical methods [24, 25, 25, 26].

1.1.2 The magnetic susceptibility in modern QC

The common attempt of Quantum Chemistry, except for few special cases, such as in the presence of relativistic effects, is that of solving the Schrödinger equation [27]. The aim of modern QC is thus two-fold: developing new theoretical methods and improving the already existing ones [28]. Both these tasks can be achieved exploiting the always increasing computational power [26, 29–32].

Currently, the available methods range from Hartree-Fock (HF) and simpler models based on effective electrostatics or angular overlap [33–37], to Density Functional Theory (DFT) and post-HF methods. In particular, Multireference Self-Consistent Field methods, such as State-Averaged Complete Active Space Self-Consistent Field (SA-CASSCF) [38, 39], are proven successful in reproducing *ab initio* both the electronic and the magnetic properties of matter [25, 30], and the results can be further improved by reintroducing dynamic correlation through the application of Complete Active Space with Second-order Perturbation Theory (CASPT2) [40] or

N-Electron Valence state Perturbation Theory (NEVPT2) [41, 42].

These calculations are usually performed using Quantum Chemistry packages, of which a wide variety is available. Among the most popular, it can be mentioned MOLCAS [43] and ORCA [24, 44], which has been extensively used in this thesis.

1.1.3 Paramagnetic NMR shifts

The interaction between nuclei and unpaired electrons is named *hyperfine* and affects the chemical shift through two contributions [45].

(i) *Fermi-Contact shifts* (FCS) arise from the presence of unpaired spin density at the resonating nucleus, and are thus transferred through chemical bonds. They are described by the equation [46]:

$$\delta^{FCS} = \frac{A}{\hbar}\frac{g_e\mu_B S(S+1)}{3\gamma_I k_B T}, \tag{1.3}$$

where A is the contact coupling constant, g_e is the free electron g value, γ_I is the nuclear gyromagnetic ratio, and S is the electron spin operator.

(ii) *Pseudocontact shifts* (PCS) arise from the distortion of the field caused by the magnetic susceptibility anisotropy and are transferred through the space. They are described by the Kurland-McGarvey equation [47]:

$$\delta^{PCS} = \frac{1}{3}Tr\left[\frac{1}{4\pi r^3}\Delta\chi \cdot \left(\frac{3\boldsymbol{r}\boldsymbol{r}^T}{r^2} - \mathbf{1}\right)\right], \tag{1.4}$$

where $\Delta\chi$ is the magnetic susceptibility anisotropic tensor (Section

1.1.1) and $\boldsymbol{r}$ is the metal-nucleus distance vector.

Experimentally, the hyperfine shifts are obtained by subtracting the chemical shifts measured in a paramagnetic system to those measured in an analogous diamagnetic system. The separation of PCS and FCS contributions, however, is challenging [48–50], and estimates of the magnetic anisotropy are required.

Whatever is the investigated magnetic property and the selected approach to measure or calculate it, the knowledge of magneto-structural correlations allows to:

(i) design materials with tailored magnetic properties, as in the case of Molecular Magnetism, where single-molecule [51–53] and single-ion [29, 54–58] magnets find widespread applications in the fields of spintronics [59] and quantum computing [60];

(ii) reconstruct the structure of a molecule by measuring its magnetic properties [48], as in the works presented in Sections 1.2 and 1.3.

1.2 Structural refinement through PCS and FCS

1.2.1 Introduction

Thanks to the efforts of computational and experimental groups, Quantum Chemical analysis of paramagnetic NMR shifts is recently gathering momentum, but the used methods are not always in agreement with one another, as far as the calculation of the PCS contribution is concerned [61, 62]. Until very recently, the discrepancy was not completely understood [61], but now the inclusion of further terms in the rigorous QC treatment [63] has allowed to resolve the ambiguity and has provided a profound proof towards the use of the Kurland-McGarvey (eq. 1.4) [47] or the McConnell-Robertson equations [64].

In the presented work [65], QC methods have been tested to accurately predict hyperfine shifts, and the calculated shifts have been verified by ^{1}H NMR experiments. The test cases are two complexes of nickel(II) by pentadentate salicylaldiminates with dipropylenetriamines (SAL-DPT) [66–70], ligands that are sufficiently rigid and bulky to enforce pentacoordination and to sterically discourage the access to the sixth coordination position. In particular, NiSAL-MeDPT and NiSAL-HDPT (Fig. 1.1, panels a and d) are prototypical of pentacoordination for this metal and have very particular spectroscopic features that make them an optimal benchmark for testing the prediction of QC methods:

(i) the hyperfine shifts originate from both contact and pseudocontact contributions, none of the two being negligible for some protons;

(ii) the ^{1}H NMR spectra may extend to almost 1000 ppm, therefore are used as benchmarks for NMR hardware development and testing [71];

(iii) being intrinsically not symmetric, there are no magnetically equivalent protons;

(iv) however, the two enantiomers can interconvert into one another, therefore the signal of each proton (except for those on the apical nitrogen) is always linked by chemical exchange to the signal of the proton that is in the mirror position.

1.2.2 Methods

Experiments

The complexes were prepared as described in reference [66], recrystallized from dichloromethane/toluene, and finally dissolved in deuterated chloroform for performing the NMR experiments. The ^{1}H NMR spectra were recorded on three instruments operating at 80, 400, and 1200 MHz:

(i) a Bruker Benchtop NMR Fourier 80 device (1.9 T, with the nutation frequency of the hard pulse of ca. 27 kHz which allows uniform excitation over the considered chemical shift range);

(ii) a Bruker Avance III spectrometer operating at 400 MHz ^{1}H Larmor frequency (9.4 T, with the nutation frequency of the hard pulse of ca. 90 kHz and a 5 mm, ^{1}H selective probe dedicated to paramagnetic systems to mitigate the effects of relaxation);

(iii) a Bruker Avance NEO spectrometer operating at 1.2 GHz with a 28.2 T HTS/LTS hybrid magnet, using a 3 mm, triple resonance TCI cryo-probehead (the nutation frequency of the hard pulse is ca. 37 kHz, therefore the excitation is achieved with a small flip-angle 200 ns pulse length).

Calculations

The NiSAL-MeDPT structure was obtained from the Cambridge Crystallographic Data Centre (CCDC), accession number 1254189 (SAIMNI10) [72]. Conversely, the NiSAL-HDPT structure is not available, as it does not easily crystallize, therefore it was obtained by substituting in the NiSAL-MeDPT structure the methyl group with a hydrogen atom.

All the calculations were carried out using the ORCA 4.2.1 Quantum Chemistry package [24, 44]. Both the structures were subjected to refinement at the DTF level of theory, with B3LYP functional [73–76], using Ahlrichs polarized basis set def2-TZVP [77, 78] and Grimme's dispersion correction D3 [79, 80]. The resolution of identity approximation [81, 82] was employed with auxiliary basis set def2-TZVP/J in order to speed up the calculations. The Conductor-like Polarizable Continuum Model (CPCM) implicit solvent (chloroform) was used [83].

The shifts were obtained by summing up the three contributions.

(i) The diamagnetic contribution was calculated at the DFT level.

(ii) The PCS were calculated with the Kurland-McGarvey equation (eq. 1.4), using the magnetic susceptibility anisotropy tensor computed with SA-CASSCF [38, 39], accounting for the dynamic correlation by NEVPT2 [41, 42], as described in reference [26]. The active space was chosen to contain eight electrons in the five nickel 3d-based molecular orbitals. All microstates arising from the d^8 configuration were included in the calculation. The SOC was treated using the Spin-Orbit Mean Field (SOMF) approximation as implemented in ORCA [84].

(iii) The FCS were calculated at the DFT level, like the diamagnetic contribution and the geometry optimization, but accounting for scalar

relativistic effects through the Douglas-Kroll-Hess (DKH) Hamiltonian [85]. In fact, the FCS cannot be properly obtained by the approach used to calculate the PCS, because the CASSCF wavefunction lacks spin polarization and the NEVPT2 correction only pertains to energy. However, it must also be noticed that DFT has the tendency to overestimate the electron delocalization [86, 87] and to underestimate the molecular g-matrix due to the lack of configuration interaction and dynamic correlation [86, 88, 89], and that the isotropic g-shift is about 6 to 10% smaller than that calculated at the Multireference level. The FCS would thus be increased by the same amount.

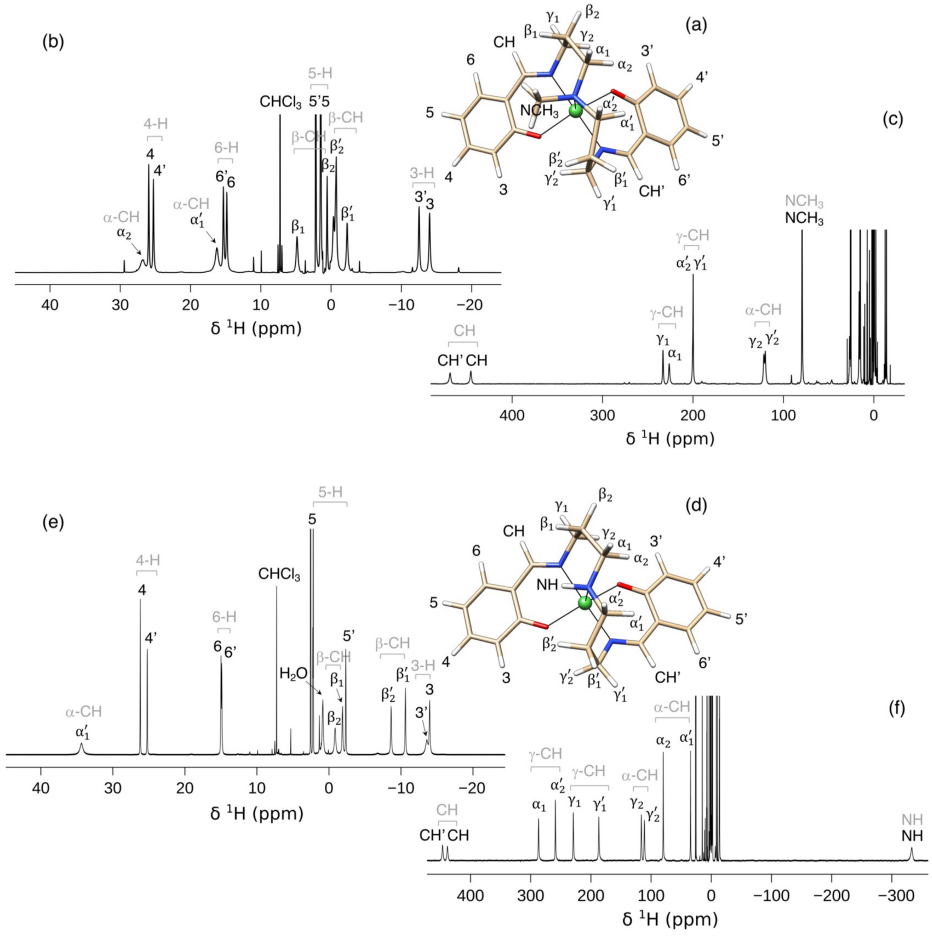


Figure 1.1: *Structures and ^{1}H NMR spectra acquired at 1200 MHz of NiSAL-MeDPT (panels a, b, c) and NiSAL-HDPT (panels d, e, f). The theory-based reassignment (black) is compared with the one proposed by Sacconi and LaMar [70] (grey) [65].*

1.2.3 Results and discussion

In line with the expectations, the metal coordination in the DFT-refined structures of both NiSAL-MeDPT (Fig. 1.2, panel a) and NiSAL-HDPT (Fig. 1.2, panel b) is square pyramidal, with a slight trigonal distortion: the two oxygen donors are bent out of plane by about 12°.

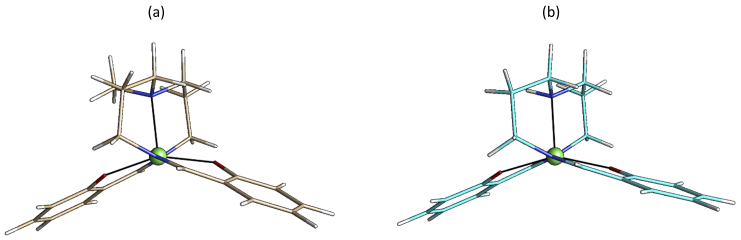


Figure 1.2: *Structures of NiSAL-MeDPT (panel a) and optimized NiSAL-HDPT (panel b) [65].*

The effective magnetic moment obtained with CASSCF/NEVPT2 is in line with the experimental one: 3.33 Bohr Magneton (BM) for NiSAL-MeDPT (exp. 3.34 BM) and 3.32 BM for NiSAL-HDPT (exp. 3.32 BM). Most of the calculated shifts agree rather well with the experimental data of both complexes (Fig. 1.3, panels a and b) using the assignments reported in reference [70]. However, if the assignment of the methylene signals is altered, the agreement becomes nearly perfect for NiSAL-MeDPT and reasonably good also for NiSAL-HDPT (Fig. 1.3, panels d and e, respectively).

NiSAL-HDPT has been used to confirm this theory-based reassignment because of the larger separation of the resonances. A COSY spectrum recorded at 80 MHz (Fig. 1.4, panel a) shows three crosspeaks at: (i) 288 to 79.9 ppm, (ii) 259 to 35.2 ppm, and (iii) 79.9 to -0.82 ppm. (i) and (ii) are

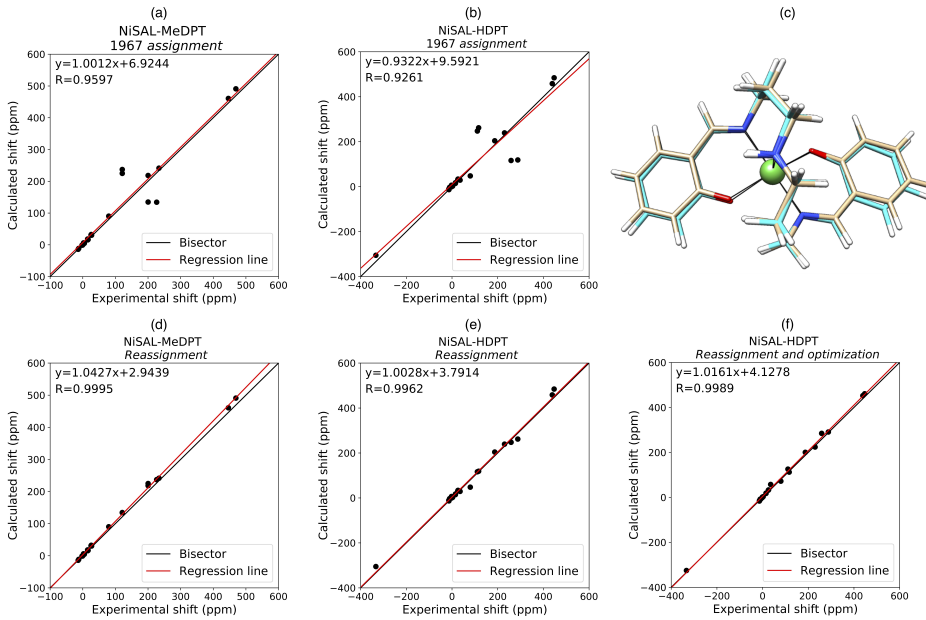


Figure 1.3: *Agreement between experimental and QC-calculated shifts for NiSAL-MeDPT and NiSAL-HDPT, with the assignment proposed by Sacconi and LaMar in 1967 (panels a and b) [70] and the theory-based reassignment (panels d and e). Comparison between initial (beige) and optimized (cyan) NiSAL-HDPT structures (panel c) and agreement between experimental and QC-calculated shifts for the NiSAL-HDPT optimized structure (panel f). The equation and the correlation coefficient are given in the figure, and the linear regression and the bisecting lines are shown in red and black, respectively. The experimental and QC-calculated shifts, together with their FC, PC, and diamagnetic contributions, of NiSAL-MeDPT and optimized NiSAL-HDPT are reported in Tables 6.1 and 6.2, respectively [65].*

unambiguously attributable to two geminal pairs, whereas (iii) links one of the α methylene protons to the vicinal β methylene protons, establishing an unambiguous connectivity. Finally, the presence of a strong NOE response of the resonance at 288 ppm to the irradiation of the one at 229 ppm, and the presence of an exchange response of the resonance at -1.9 ppm to the irradiation of the one at -8.5 (Fig. 1.4, panel b) remove all the ambiguities in the two propylene branches, except the one between the two signals at 117 and 111 ppm, which could not be resolved experimentally. The so obtained connectivity is perfectly compatible with the assignment obtained from the comparison between computed and experimental spectra. The COSY spectrum acquired at 400 MHz allows for tracing the connectivity in the two rings (Fig. 1.4, panel c), resulting in a pattern of the type (from downfield to upfield, Fig. 1.1, panels b and e) 4-4', 6'-6, 5-5', 3-3', which is again confirmed by the calculations.

The agreement between experiment and calculations thus allows to interpret the small discrepancies between observed and calculated shifts obtained for the HDPT derivative (whose structure is not experimentally determined but only modeled on that of NiSAL-MeDPT) in terms of minor structural changes. Most of the deviation relates to the α methylene protons, where the calculation appears to be underestimating the contact contribution, especially for α_1. Bearing in mind that this structure is not experimental but derived from the MeDPT derivative as a template, it is reasonable to think that the structure may need to be altered. Therefore, it has been performed a very coarse scan (9 steps, ca. 7° for each) of the dihedral angle formed by α_1 with its carbon, the apical nitrogen, and the nickel ion, around the position found in the optimized structure. From a

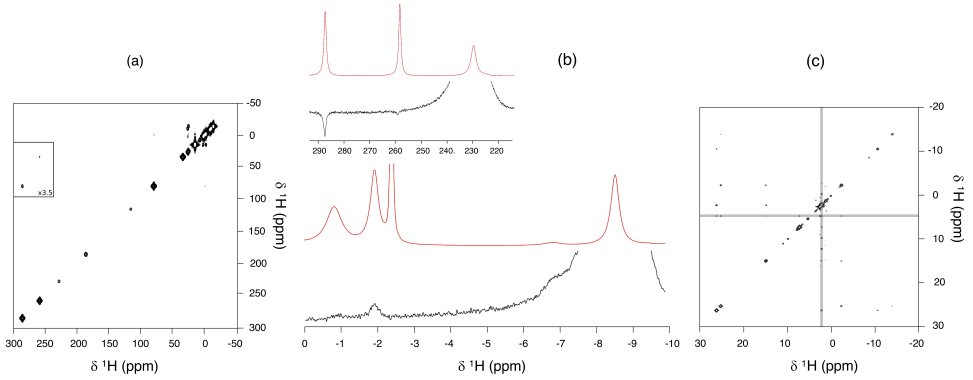


Figure 1.4: *COSY spectrum of NiSAL-HDPT acquired at 80 MHz (panel a). 1D difference traces obtained by irradiating the signals at 229 ppm and -8.5 ppm (panel b, top and bottom, respectively). COSY spectrum of the aromatic region of NiSAL-HDPT acquired at 400 MHz (the areas shaded in grey are T1 noise and artifacts at the transmitter frequency arising from imperfections in 90° pulses) [65].*

qualitative viewpoint, this angle is likely the strongest determinant of the contact shift because it controls the overlap of the hydrogen nuclei with the orbitals of the donor atom overlapped with the metal orbitals bearing the unpaired electron [90–94]. When the angle is equal to 172.9° (Fig. 1.3, panel c), the agreement with the experimental data becomes very good and comparable in quality to that of the MeDPT derivative (Fig. 1.3, panel f).

In line with recent literature on the calculation of magnetic [95, 96], EPR, and NMR properties [48, 97–102], single-point calculations have been performed on the minimum positions achieved by the DFT optimization. The results thus do not include effects due to mobility or dynamics, which influence the relaxation properties [103–105]. Conversely, in the calculation of the shift and susceptibility values, a symmetric averaging around a single minimum position is not expected to yield average values significantly

different from those calculated in the average position.

These results demonstrate the fundamental support that QC methods can provide to experimental observations and the possibility to refine the structure of metal complexes starting from homologous compounds. The combination of QC calculations and pNMR restraints has been thus extended to a more complex and biologically relevant case, a paramagnetic protein, leading to the development of an innovative PCS-driven structural refinement approach (Section 1.3).

1.3 PCS-driven structural refinement for metalloproteins

1.3.1 Introduction

As introduced in the Preface, NMR has a privileged view on the structure of proteins. However, a large share of the proteome consists of metalloproteins [106, 107] that naturally bind paramagnetic ions or can bind them with minimal manipulation [108–110]. Paramagnetism is caused by the presence of unpaired electrons [111], therefore is profoundly related to the electronic structure, but constitutes a problem for all the structural techniques. In XRD it induces density flattening in the vicinity of the metal [112], in cryo-EM the reconstruction is complicated by density deviations from the average distribution found in proteins [113], in NMR several effects that complicate the spectra are observed (Paragraphs 1.1.3 and 6.1). Most paramagnetic metal ions, indeed, affect the neighboring nuclei causing their signals to shift and broaden [114, 115], up to the point that a blind zone can be present in the vicinity of the metal. On the other hand, these drawbacks are more than compensated by the wealth of information encoded in the paramagnetic NMR observables outside of the blind zone [116], which can be so accurate that it is becoming increasingly frequent to substitute a naturally coordinated diamagnetic metal ion with a paramagnetic one.

PCS, in particular, depend on $\Delta\chi$, which is highly sensitive to the coordination geometry. However, while the $\Delta\chi$ tensor is easily back-calculated from PCS through the Kurland-McGarvey equation (eq. 1.4), the reconstruction of the coordination geometry from $\Delta\chi$ had never been attempted. The presented work [117] proposes a PCS-driven structural refinement, at

an unprecedented level of resolution, of the first coordination sphere of paramagnetic metalloproteins. The structure is optimized by minimizing the disagreement between experimental and QC-calculated PCS, using the steepest descent algorithm [129], a method able to find the minimum of a function by moving in the inverse direction of its gradient.

The test protein is the human matrix metalloproteinase catalytic domain, a 17 kDa metalloprotein, coordinated by the inhibitor N-isobutyl-N-[4-methoxyphenylsulphonyl]glycyl hydroxamic acid (NNGH), and in which the native Zn(II) metal center is replaced with high-spin Co(II) (CoMMP-12, Fig. 1.5). Extensive experimental studies of the structure and pNMR shifts are available for this system [118–121].

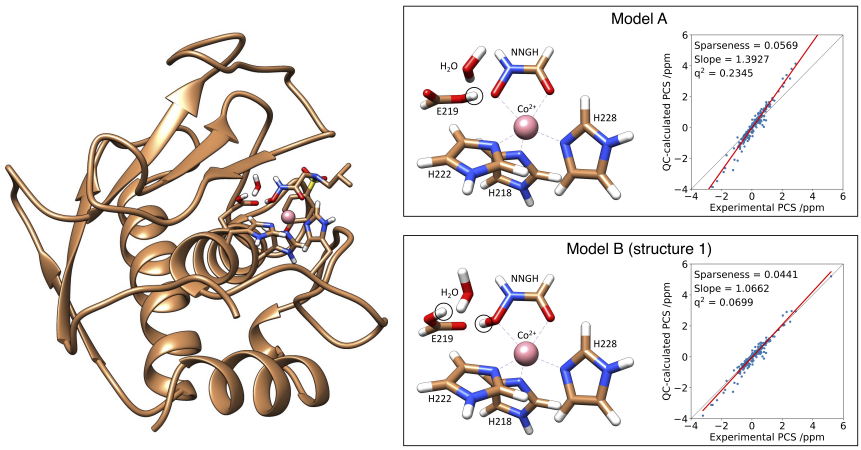


Figure 1.5: *X-ray CoMMP-12 structure with details of the active site (left). Reduced models of the active site (right): with the hydroxamate (model A) or the hydroxamic acid (model B, structure 1) moieties of the NNGH inhibitor (the different protonation state is marked with black circles), and the corresponding agreements between experimental and QC-calculated (at the NEVPT2 level) PCS [117].*

1.3.2 Methods

The best fit of the raw experimental PCS data (246 ^{13}C shifts, measured at 298 K), taken from reference [122], was obtained using FANTEN [123] through a 5-parameter fit to the Kurland-McGarvey equation (eq. 1.4). The best-fit values (called "experimental PCS" throughout) represent a denoised PCS dataset that can be completely reproduced through a single tensor as is obtained by QC calculations.

All the *ab initio* calculations were carried out using the ORCA 4.2.0 QC computational package [24, 44] as described in Table 1.1.

	Geometry optimization	**Final calculations**
Method	SOC-CASSCF(7,5) [38, 39, 124]	SOC-CASSCF(7,5)/NEVPT2 [41, 42]
Quartets	10	10
Doublets	11	40
Basis set for Co	DKH-def2-TZVP [85]	DKH-def2-TZVP
Basis set for other atoms	DKH-def2-SVP [77, 78]	DKH-def2-TZVP

Table 1.1: *Details of the calculations of the magnetic susceptibility tensors [117].*

The quality of the agreements between QC-calculated and experimental PCS were evaluated using three parameters:

1. $q^2 = \frac{\sum_i^N (\delta^i_{calc} - \delta^i_{exp})^2}{\sum_i^N (\delta^i_{exp})^2}$, where N is the number of PCS values, δ^i_{calc} and δ^i_{exp} are QC-calculated and experimental PCS, respectively;
2. *slope* of the linear least-square regression ($bestfit$);
3. $sparseness = \frac{\sum_i^N (\delta^i_{calc} - bestfit)^2}{N}$.

The algorithms for the geometry optimization were implemented using python2.7.

Three structural refinements were carried out using PHENIX (version dev-3900 [125]), using a default protocol with the addition of optimization of stereochemistry/ADP weights. (i) The starting coordinates were obtained performing a new refinement of the PDB entry 5LAB [99], including a manual clean-up of the water positions after the first optimization. The hydrogen positions were calculated using REDUCE [126] and, in the models used in the geometry optimization, they were further optimized at the DFT level, using the B3LYP functional and the def2-SVP basis set for all atoms and def2-TZVP for Co. At the end of the geometry optimization, the resulting structure for the active site (Model B, structure 4) was reinserted into the crystallographic structure and subjected to two parallel refinements: in one case, (ii) all atoms of the structure were left free to refine, whereas, in the second case, (iii) the atoms corresponding to the reduced model were kept fixed (including the relevant hydrogen atoms).

The final energy minimization was carried out using the routine PARAmagnetic restraints for Xplor-NIH version 2.29 [127], similarly to what reported in reference [128]. The refinement protocol was composed of two steps: a rigid minimization and an internal dynamics and minimization in which the structure was subjected to a simulated annealing at 100 K and 200 K, respectively. The backbone atoms were fixed, the refined first coordination sphere and the tensor axes were moved rigidly together, and PCS restraints with a force constant of 10 $kJmol^{-1}ppm^{-2}$ were added.

1.3.3 Results and discussion

The proposed method, schematized in Fig. 1.6, is divided in three steps:

1. *Reduced model construction*
   In order to speed up the calculations, the coordinates to be optimized (initial_structure.xyz) are extracted from the initial pdb file.
2. *Geometry refinement*
   The optimization first requires the selection of the N degrees of freedom v_k ($k = 1, ..., N$) to be adjusted (among coordination bond lengths, pseudobonds, angles, and dihedral angles), and the extent of the variations h_k ($k = 1, ..., N$). The starting values of the degrees of freedom and of the increments are stored in the N-dimensional $\boldsymbol{v}^{init}$ and $\boldsymbol{h}$ vectors, respectively.
   (i) Algorithm 1 takes as input $\boldsymbol{h}$ and, depending on the iteration, initial_structure.xyz and $\boldsymbol{v}^{init}$, or best_structure.xyz and $\boldsymbol{v}^{best}$, and produces as outputs dist_structures.allxyz, a file containing initial_structure.xyz and $2N$ distortions of it, in which each v_k is independently incremented and decremented of h_k.
   (ii) For each of the $2N + 1$ structures, the $\Delta\chi$ tensor is calculated with ORCA at the SOC-CASSCF level.
   (iii) Algorithm 2 computes the $2N + 1$ PCS datasets from the $2N + 1$ QC-calculated $\Delta\chi$ tensors and evaluates their disagreements with the experimental PCS in terms of q^2. The numerical gra-

dient $\boldsymbol{G}$ of the disagreement is then calculated as:

$$\boldsymbol{G} = \left(\frac{q_2^2 - q_1^2}{h_1}, \cdots, \frac{q_{2N+1}^2 - q_1^2}{h_N} \right). \tag{1.5}$$

The steepest descent method is thus applied: an output file (opt_structures.allxyz), composed of M sets of coordinates, is obtained by simultaneously changing all the degrees of freedom as a function of the gradient $\boldsymbol{v}_j^i = \boldsymbol{v}_j^{i-1} - \alpha_j \boldsymbol{G}^{i-1}$ ($j = 1, ..., M$, i = iteration index, and $\boldsymbol{\alpha}$ = vector containing M arbitrarily selected weights of the gradient).

(iv) For each of the M structures, the $\Delta\chi$ tensor is calculated with ORCA at the SOC-CASSCF/NEVPT2 level.

(v) Algorithm 3 computes the M PCS datasets from the M QC-calculated $\Delta\chi$ tensors, evaluates their disagreements with the experimental PCS in terms of q^2, and finds the minimum q^2 by interpolation. If the minimum q^2 is lower than a prefixed threshold, the corresponding structure (best_structure.xyz) is used to reconstruct the full model, otherwise it constitutes the starting point for the next iteration.

3. *Full model reconstruction*

The final pdb is reconstructed by replacing initial_structure.xyz with best_structure.xyz in initial.pdb.

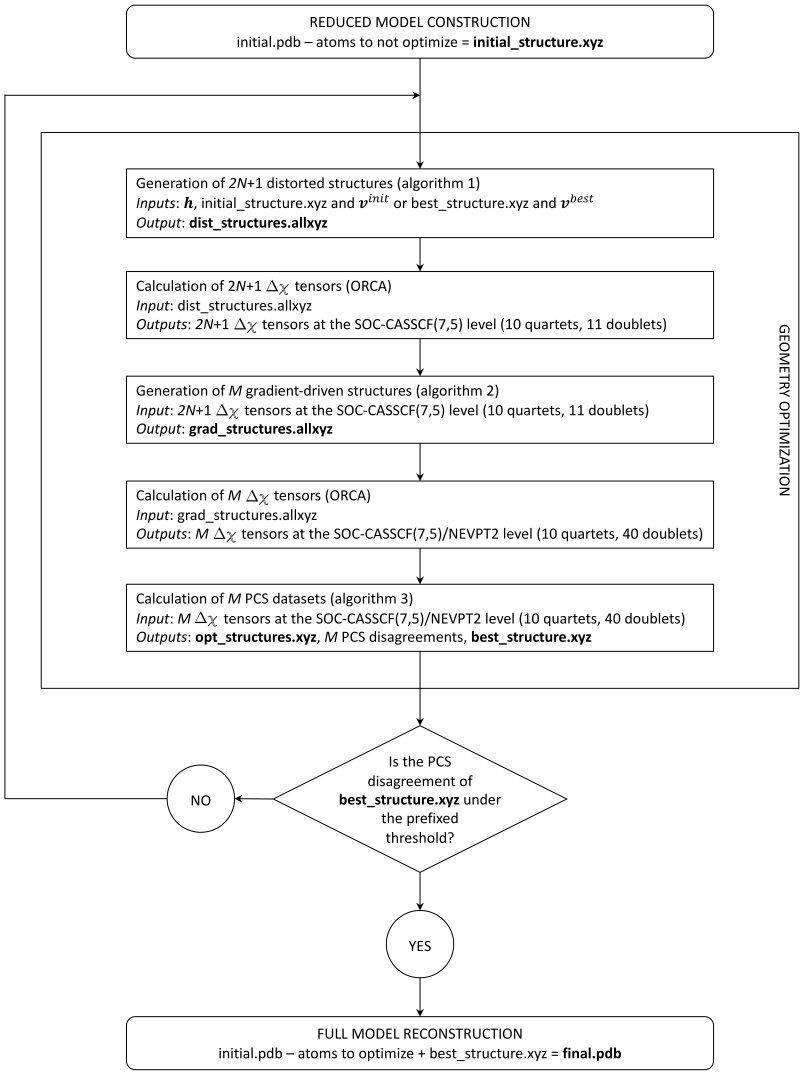


Figure 1.6: *Flowchart of the structural refinement procedure.*

This procedure has been applied to CoMMP-12, in order to refine its first coordination sphere.

1. Reduced model construction

The crystal structure of ZnMMP-12 gives a very good agreement to the cobalt-bound analogue PCS (Fig. 1.7), confirming that a robust estimate of the experimental $\Delta\chi$ tensor is available at the beginning of the procedure.

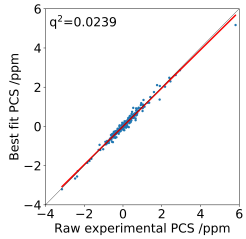


Figure 1.7: *Agreement between raw experimental and best fit PCS for the X-ray structure [117].*

The hydrogen positions in X-ray structures are notoriously difficult to detect, because the electron density depends on the atomic number. In this case, the only possible ambiguity lies in the protonation state of NNGH O_N and, as a consequence, in that of E219. The two initial reduced models from 5LAB [99] (Fig. 1.5, left) are thus composed of: the metal, the imidazole rings of H218, H222, and H228, the carboxylate of E219, one water molecule [99], and either the hydroxamate (Fig. 1.5, model A) or the hydroxamic acid (Fig. 1.5, model B) moiety of the NNGH inhibitor. The *ab initio* calculations show sizable differences in the $\Delta\chi$ tensors. In model A, the optimization of the hydrogen positions binds the proton to E219, confirming the picture adopted by most MMP researchers [130–134], but

the PCS agreement is not so good (Fig. 1.5, top right). Conversely, if the hydroxamate is protonated, the agreement improves significantly (Fig. 1.5, bottom right). To force the protonation of the hydroxamate, a second proton bound to E219 is needed, and the initial structure of model B hereafter will be called structure 1. This proton distribution may have interesting implications for the action mode of MMPs, as it will be discussed later.

2. Geometry optimization

The initial parameters were fixed: $\boldsymbol{\alpha} = 0.25, 0.5, ..., 3.5$, thus $M = 14$, q^2 threshold $= 0.05$, $N = 19$. The 19 degrees of freedom (3 coordination bond lengths (B1-B3), 1 pseudobond (P), 11 planar (A1-A11) and 4 dihedral (D1-D4) angles) do not alter the structure of the ligands and have been incremented at each step of 0.02 Å (B1-B3, P), 1.2° (A1-A11), and 5.2° (D1-D4) (Fig. 1.8).

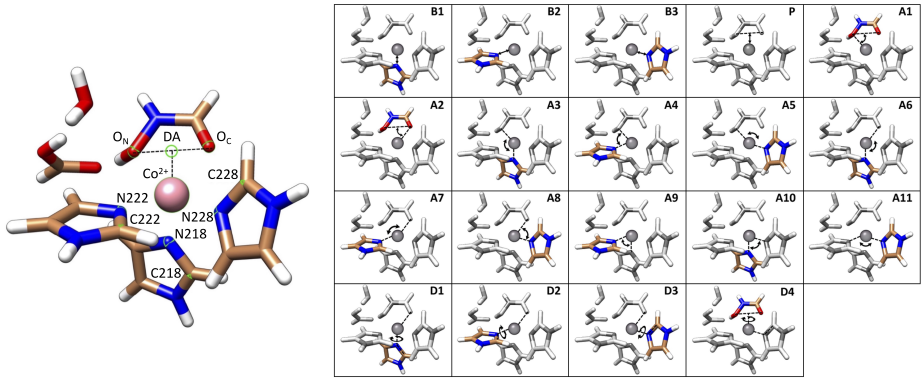


Figure 1.8: *Graphical representation of the degrees of freedom of the geometry optimization. The involved atoms are labeled and highlighted in green (DA stands for "Dummy Atom" and represents a fictitious atom centered in the intersection of the two dotted lines) (left). Only the moving molecules are colored (right) [117].*

In four steps of geometry refinement (Figs. 6.1 to 6.4), relatively modest structural changes (Fig. 1.9, left, and Fig. 6.5) lead to a very good PCS agreement (Fig. 1.9, center). The hydroxamic acid fragment (NNGH) gets closer to the metal, the residues H222 and H218 tend to decrease the angle formed with the metal and, respectively, the NNGH O_C and H228, according to the variations of the experimental and QC-calculated magnetic susceptibility tensor axes (Fig. 1.9, right). On the contrary, no improvement is achieved for model A (data not shown).

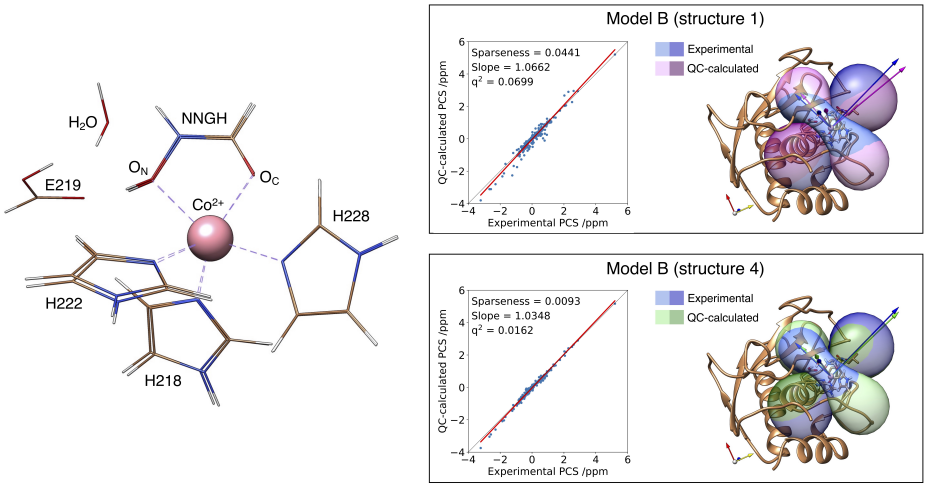


Figure 1.9: *Comparison between structures 1 and 4 in terms of: geometry (left), agreement between experimental and QC-calculated (at the NEVPT2 level) PCS (top and bottom center), and isoPCS surfaces (calculated at ±1.5 ppm) for QC-calculated and experimental PCS (top and bottom right) [117].*

The magnetic anisotropy in structure 4 is smaller than in structure 1 (Table 1.2) due to a larger ligand field associated with the shorter Co-N218 distance that defines the main axis of the ligand field (Fig. 1.10, panel a)

of a distorted square pyramidal coordination (Fig. 1.11). The energy of the first excited state that largely determines the magnetic anisotropy is connected with the splitting between d_{xy} and d_{xz} orbitals: the d_{xz} orbital is destabilized due to the π-antibonding interaction with the H218 π-system, and this effect becomes more pronounced as the Co-N218 bond shortens (Fig. 1.10, panel b).

	$\Delta\chi_{ax}$ /10^{-32} m^3	$\Delta\chi_{rh}$ /10^{-32} m^3	$\alpha/°$	$\beta/°$	$\gamma/°$
Structure 1	10.43	-4.95	-102.42	-77.97	-105.08
Structure 4	9.71	-4.67	-70.33	-98.93	70.58
Experimental	10.0 ±0.1	-2.8 ±0.1	-74.9 ±1	-98.2 ±0.5	68.6 ±0.2

Table 1.2: *Comparison of the experimental and QC-calculated (at the NEVPT2 level) magnetic susceptibility anisotropy tensors at 298 K for structures 1 and 4 [117].*

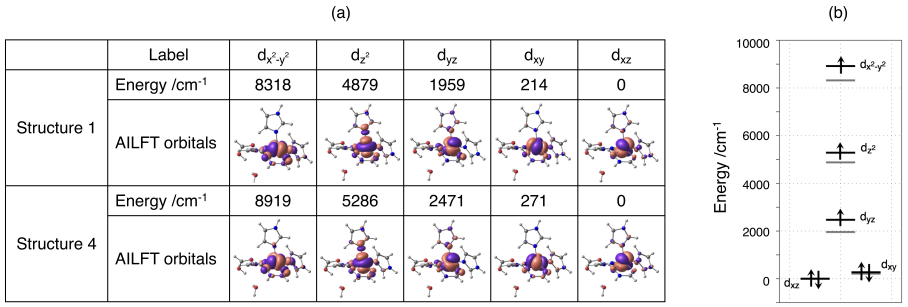


Figure 1.10: *Ab Initio Ligand Field Theory (AILFT) 3d orbitals with their energies (panel a) and splitting (panel b) based on SOC-CASSCF(7,5)/NEVPT2 calculations of 10 quartet and 40 doublet states (panel a). Grey and black bars correspond to structures 1 and 4, respectively (panel b).*

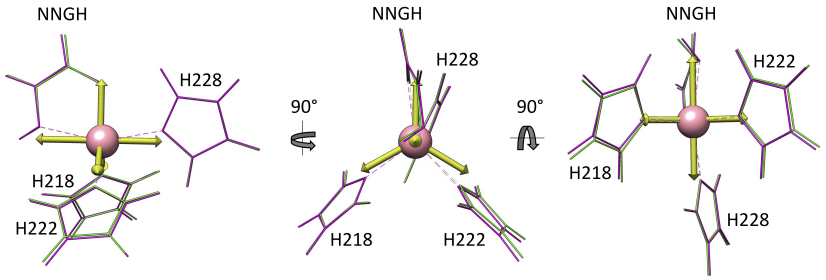


Figure 1.11: *Different orientations of structures 1 (magenta) and 4 (green) with the idealized trigonal bipyramidal geometry represented in yellow [117].*

Coordination chemists are inclined to think that metal coordination geometries are rather stiff restraints to the molecular shape. However, it has long been known that metalloproteins are not canonical coordination systems, because the protein itself poses significant strains to the idealized geometries [135]. Even for such a low-symmetry system, it is instructive to relate the starting and the minimized structures to ideality. In ranging from octahedral to tetrahedral coordination, high spin cobalt(II) can pass through different flavors of pentacoordination [136]. In the present case, the structure of the first coordination sphere is intermediate between square pyramidal and trigonal bipyramidal (Fig. 1.11), and the refinement brings it somewhat closer to the latter.

The refined first coordination sphere structure still fits into the ZnMMP-12 electron density map (Fig. 1.12). Therefore, the obtained structure is not only more precise but also has higher accuracy, because it simultaneously satisfies the X-ray experimental data and the $\Delta\chi$ tensor determined from the experimental PCS.

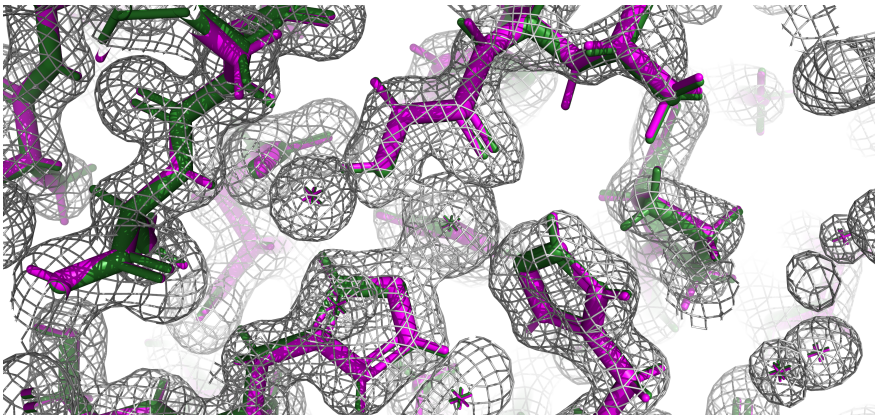


Figure 1.12: *$2F_O$-F_C map of ZnMMP-12 with the original structure in magenta and the constrained one in green. The density is shown at 1.0σ [117].*

It is possible that different geometrical modifications of the CoMMP-12 coordination sphere could yield the same magnetic susceptibility tensor, and therefore the same agreement to the experimental PCS. However, the steepest descent minimization ensures that the movement is towards the local minimum that is geometrically the closest to the starting coordinates.

3. Full model reconstruction

The full model has been finally reconstructed, and the side chains have been refined through Xplor-NIH [137, 138], applying as restraints: the experimental PCS, the coordinates of the reduced model atoms, and the orientation of the magnetic susceptibility tensor obtained from the previous optimization [107]. The family of the 20 structures with the lowest energy, out of the 35000 generated structures, satisfies the data and does not show geometrical strains with respect to the crystallographic positions.

Recently, it has been shown that by comparing experimental and calculated pNMR observables it is possible to select, among two structures, the one that best represents the first coordination sphere in a metalloprotein [48]. The present results demonstrate that the information encoded by the magnetic susceptibility in the PCS allows for an extremely detailed refinement of the active site coordination geometry using relativistic multiconfigurational QC methods, to the point to suggest the presence of a further proton in the active site, with implications on the mechanism of inhibition and catalysis.

The protonation state of the hydroxamate ligand in inhibited MMPs has been a subject of debate in the early years of the discovery of this class of enzymes, which are of potential pharmacological interest due to their implications in inflammation and in the spreading of metastatic cells (references [131], [133], and references therein). Carboxylic acids have pKas around 5, while hydroxamic acids around 9, so at a physiological pH of around 7 it would be expected that E219 is deprotonated and anionic, and the hydroxamic inhibitor protonated and neutral. However, multiple experimental evidences point to a situation where the proton forms an H-bond between the two groups and is actually residing mostly on E219, leaving a negative hydroxamate moiety. This has been rationalized by considering that the bulky hydroxamic inhibitor completely occludes the active site cavity, so that the carboxylate moiety of E219 is not accessible to solvent molecules. Under these conditions a low dielectric environment is created, and the pKa of E219 can increase by several units, while the negative charge of the hydroxamate can be stabilized by the dipositive zinc (or in this case cobalt) cation. The proposed arrangement of protons in the

active site of MMPs would provide a plausible explanation for the proton uptake observed by calorimetry when hydroxamic acid inhibitors bind the enzyme: E219 could be deprotonated in the free enzyme but could regain a proton from the hydroxamic acid moiety, which in turn could regain a proton from the solvent. In the free, active enzyme, the protonation of E219 by a metal coordinated water molecule, which then becomes a hydroxide ion, has also been invoked to explain the efficient nucleophilic attack on the peptide bond, which constitutes the first step of the hydrolytic action of the enzyme.

Chapter 2

Conformational heterogeneity in biomolecules through the MaxOcc method

2.1 Introduction

The high level of complexity reached by biological systems can be attributed, at least in part, to the considerable flexibility of biological macromolecules, such as nucleic acids and proteins. These molecules often perform their tasks by reorganizing their 3D structures according to the cell's conditions (i.e. pressure, temperature, pH, and salinity), creating ensembles of conformations which are strongly function-related [139–141]. Patholog-

ical alterations of these conditions, including mutations, may significantly alter the conformational preferences of the system [142, 143]. As a consequence, individual conformations must be replaced by more representative *structural ensembles*, i.e. sets of conformations complemented by their statistical weights [144–146].

The comprehension of the conformational heterogeneity of such systems thus starts with the evaluation of the chemically accessible conformations. The number of degrees of freedom of the molecule under investigation is calculated on the basis of its chemical properties (i.e. bond rotation, steric clashes, etc..), and determines the dimensionality of the space that represents the system. The ensemble of all the possible conformations is indeed described by the so-called *conformational space*: a set of points, each of which univocally represents a specific conformation. Once the pool of possible structures is generated, the conformations have to be properly weighed with a procedure, the *ensemble reconstruction*, which falls into the category of the *inverse ill-posed problem*.

2.1.1 Ill-posed inverse problems

Mathematical and, more generally, scientific problems can be divided in two categories: *direct* and *inverse* problems. The first consist in finding the effects from known causes, the second in finding the causes from the effects. Solving inverse problems is usually much more difficult, because they are usually affected by issues of posedness and conditioning. A problem is said to be well-posed, i.e. - at least in principle - solvable, if it satisfies the three *Hadamard conditions* [147]:

1. *existence*: the problem must admit at least one solution;
2. *uniqueness*: the solution must be unique;
3. *continuity*: the solution must depend with continuity on all the parameters of the problem.

If at least one of these conditions is not fulfilled, the problem is termed *ill-posed*. Most numerical treatments overcome the issue of ill-posedness by including additional information, controlled by the so-called *regularization parameters* [148].

2.1.2 Ensemble reconstruction methods

There exists a variety of ensemble reconstruction methods which make use of the two available sources of information: experimental data, theoretical simulations, or both [149, 150].

Purely experimental methods use the most direct investigation tool, but suffer from three principal limitations. They are always affected by systematic and random errors (i), and provide sparse information, because in many cases it is not possible to measure all the properties of interest (ii). Moreover, a sample comprises a number of molecules of the order of the Avogadro's constant, each occurring in a different state and undergoing several conformational changes during the timeframe of the experiment. What is actually measured is a spatial and temporal average instead of instantaneous conformations of every single molecule (iii).

Purely theoretical methods are limited by the computational resources, the complexity of the model, and the level of theory used, which inevitably lead to a loss of information.

In the light of all this, the most promising strategy seems to combine both experimental data and theoretical simulations in the so-called *hybrid methods*, which can be divided in two categories: those determining a large ensemble, following the *Maximum Entropy* principle, and those determining a minimal set that fit the data, in the spirit of the *Maximum Parsimony* principle. The *Maximum Occurrence* (MaxOcc) method here presented [151–155], however, does not follow in its setting none of the two approaches [156], as it is not aimed at finding an ensemble, but rather at studying individual conformations.

2.2 The MaxOcc method

The Maximum Occurrence (MaxOcc) w_c of a conformation c is the maximum probability that c occurs in an ensemble that fits the experimental data. Mathematically, it corresponds to an upper bound of its statistical weight.

The procedure

The procedure to find the MaxOcc value for each conformation starts with the:

1. collection of experimental data and estimation of the uncertainty;
2. creation of a pool of allowed conformations and calculation of the simulated data.

More specifically, in point 2 the molecule is modeled with a proper software, e.g. Xplor-NIH [137, 138], and the flexible parts are moved in order to generate a number of possible conformations. These structures, which are sampled from the conformational space, populate the pool of conformations that will describe the system. The simulated data are then calculated as follows: for each conformation is predicted and stored the value, corresponding to the experimental observable, that would have been measured if the system assumed solely this conformation. In this way, the contribution coming from each conformation is calculated and can be used to generate simulated ensembles with the corresponding averaged simulated data.

The method

The basic idea is fixing, for a selected conformation, an initial arbitrarily low weight w_c, and building around it a number of possible statistical ensembles with the remaining conformations. The agreements between the experimental data and the averaged data calculated on each ensemble are evaluated in terms of residual norm, and considered good when the norm equals the experimental uncertainty. Once one of these ensembles is found in agreement with the experimental data, the weight w_c is increased and the procedure repeated. When it is no longer possible to find a completing ensemble that fits the data, the process is stopped, and the MaxOcc value (the last w_c in agreement with the data) is found. This method is then applied for each conformation.

The mathematical formulation

Mathematically, this procedure consists in varying the statistical weights vector $\mathbf{w}$ with the aim of minimizing the discrepancy between theory and experiment, and simultaneously maximizing the weight w_c of the selected conformation. In order to circumvent the problem of ill-posedness, the initial problem:

$$\underset{\mathbf{w}}{\operatorname{argmin}} \|\mathbb{P}\mathbf{w} - \mathbf{d}\|_2^2 , \qquad s.t. \mathbf{w} \geq 0 \tag{2.1}$$

has indeed been replaced by an approximate form which includes additional information about the system:

$$\underset{\mathbf{w}}{\operatorname{argmin}}[\|\mathbb{P}\mathbf{w} - \mathbf{d}\|_2^2 + \xi_1(\|\mathbf{w}\|_1 - 1)^2 - \xi_2 w_c], \qquad s.t. \mathbf{w} \geq 0. \tag{2.2}$$

Argmin means finding the values of $\mathbf{w}$ (which is imposed to be positive in order to be physically meaningful) for which the function inside the square brackets reaches its minimum. The first term is the square norm of the discrepancy between the simulated ($\mathbb{P}$ matrix, with the corresponding statistical weights vector $\mathbf{w}$) and the experimental ($\mathbf{d}$ vector) data. The second and third terms, controlled by the parameters ξ_1 and ξ_2 (found through an L-curve approach [157]), introduce two information on the system: the normalization restraint (the sum over the statistical weights must be unitary) and the maximization of the single conformation weight w_c.

The interpretation of the output

The algorithm gives as output the vector $\mathbf{w}$ containing the MaxOcc values w_c associated to each conformation of the pool. Differently from the majority of the ensemble reconstruction methods, indeed, the MaxOcc aim is not to find the most probable ensembles, but to associate to each conformation a number, the MaxOcc value, which gives an information on the probability of that conformation to occur. This number is roughly defined as "information" because knowing the MaxOcc value of a conformation does not mean knowing its statistical weight, but being sure that this conformation will occur with a probability lower than this value. As a consequence, it makes no sense creating ensembles composed of the conformations with the highest MaxOcc values, nor the ensembles containing the conformations with their maximum statistical weight found in agreement with the experimental data are by themselves meaningful. The ensembles made of the conformations out of the considered are only ancillary.

2.3 The MaxOcc applications

Although of general validity, the Maximum Occurrence method has been tested on two particular systems that experience a huge conformational variability, with two different objectives:

1. extending, through DEER integration, the biological validity of MaxOcc on highly flexible multidomain systems (Section 2.4);
2. demonstrating the effectiveness of residual dipolar couplings (RDC) as restraints in MaxOcc for the assessment of unstructured protein domains (Section 2.5).

2.4 Integration of DEER data in the MaxOcc analysis

2.4.1 Introduction

The MaxOcc method is expected to converge towards the true statistical weights by increasing the number of independent experimental datasets. According to this property, in addition to the already tested techniques, i.e. NMR (PCS, Section 1.1.3 and RDC, Paragraph 6.1) and SAXS (Paragraph 6.1) [158], it has been selected Double Electron Electron Resonance (DEER) (Paragraph 6.1), because its time-dependence partially solves the issue of time averaging.

The proposed analysis has been tested on calmodulin (*Cal*cium-*modul*ated prote*in*, CaM), a multifunctional intermediate calcium-binding messenger protein expressed in eukaryotic cells which naturally binds four Ca^{2+} ions. Its sizable mobility is due to a 4-residues (78-81) linker which connects the two negatively charged N-terminal and C-terminal domains (NTD and CTD, respectively) of the protein (Fig. 2.1, panel a) [159, 160]. Calmodulin was subjected to structural modifications necessary for the NMR and DEER experiments. In order to collect the experimental pNMR data, one Ca^{2+} ion has been substituted by a lanthanoid ion (Tb^{3+}, Dy^{3+}, and Tm^{3+}, Fig. 2.1, panel b) [161], but it has been demonstrated that the presence of large lanthanoid binding tags does not affect the fold and the conformational variability of CaM [162]. For the DEER measurements, the protein has been labeled with two Gd^{3+} 1,4,7,10-Tetraazacyclododecane-1,4,7,10-tetraacetic acid-maleimide (DOTA-M) tags, one in each domain (Fig. 2.1, panel c). Even if the presence of the tag could in principle reduce the ac-

cessibility of compact conformations, those are already virtually excluded by the other available experimental data.

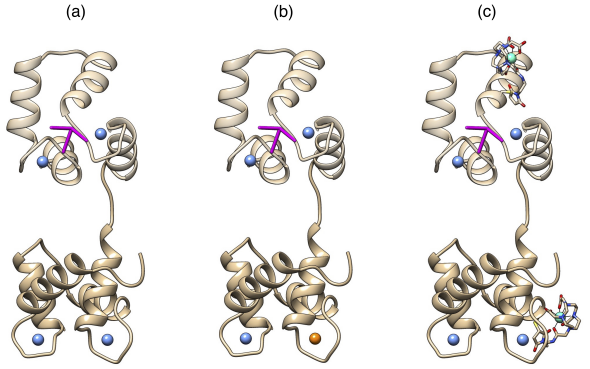


Figure 2.1: *Generic CaM conformation: the NTD is depicted as a solid ribbon, while the CTD is represented both explicitly as a transparent ribbon and implicitly as a triad of axes placed in its center of mass. The triad schematization, useful for simplifying the visualization of several conformations in the same picture, is oriented with respect to an arbitrary reference position, which corresponds to the axes (1,0,0; 0,1,0; 0,0,1) of an extended 1CLL conformation [163]. The Ca^{2+} ions are shown as light blue spheres in the native CaM used for SAXS experiments (panel a), the lanthanoid ion replacing one Ca^{2+} ion to obtain PCS and RDC is represented as an orange sphere (panel b), and the Gd^{3+} ion, coordinated to the DOTA-M tag for the DEER measurements, as a sea-green sphere (panel c) [164].*

2.4.2 Methods

Experiments

The experimental SAXS and NMR (PCS and RDC) data have been taken from the literature (references [158] and [154], respectively), where CaM was dissolved in 20 mmol/dm^3 MES and 400 mmol/dm^3 KCl, pH 6.5. For SAXS data, the uncertainty was estimated using mean and standard

deviation. The NMR data, instead, are not the result of multiple measurements, and the accuracy of the measured couplings was estimated (the authors report errors of ±30% for PCS and ±0.5 Hz for RDC).

The DEER datasets were acquired on samples of the mutants of human calmodulin N53C-T110C, N53C-A103C and T34C-T117C, with final CaM concentration of 50 μL in a 20 mmol/dm^3 phosphate buffer, pH 6.8, in a mixture of D_2O/glycerol-d_8 (8/2 v/v). The mutants were labeled with two Gd^{3+} DOTA-M tags situated in the CTD and NTD [165]. The DEER traces were recorded on a home-built W-band spectrometer operating at 94.9 GHz at 10 K [166, 167]. The dead-time free 4-pulse DEER sequence $(\pi/2)_{mw1} - \tau_1 - (\pi)_{mw1} - (\tau_1 + t) - (\pi)_{mw2} - (\tau_2 - t) - (\pi)_{mw1} - \tau_2$−echo was used. The lengths of the observed $\pi/2$ and π microwave pulses were 15 ns and 30 ns, respectively, the pump pulse length was 15 ns, τ_1 was 350 ns, and the repetition time was 0.8 ms. The value of τ_2 varied depending on the experiment. In order to maximize the modulation depth, the DEER experiments were carried out by setting the pump pulse frequency on the maximum of the Gd^{3+} spectrum. The detection pulses frequency was 100 MHz higher. Exponential background subtraction was applied. As an estimate of the uncertainty in each DEER datapoint, for the N53C-T110C and N53C-A103C datasets, the standard deviation of the consecutive differences between 60 and 160 ns was used. The data for T34C-T117C were obtained as average over three different measurements (after background subtraction) and the uncertainty was in this case taken as the largest difference between the three curves.

Simulations

The simulated data have been calculated on a pool of 33000 conformations obtained through Xplor-NIH [137, 138] by rigidly moving the CTD with respect to the NTD, around the backbone bonds of the residues of the linker [158, 168], and excluding the conformers with significant clashes.

In order to correctly combine the different datasets, literature approaches have been used to select their relative weight. For quantifying the agreement for paramagnetic NMR data, it has been used the q-factor [169], defined as:

$$q = \frac{\|\mathbb{P}\mathbf{w} - \mathbf{d}\|_2}{\|\mathbf{d}\|_2}, \tag{2.3}$$

where $\mathbb{P}$ is the matrix of the predicted data, $\mathbf{w}$ is the weights vector, and $\mathbf{d}$ is the vector of the experimental data (for either PCS or RDC). For SAXS and DEER it has been used the reduced χ^2 [170], defined as:

$$\chi^2 = \frac{1}{N}\|(c\mathbb{P}\mathbf{w} - \mathbf{d}) \oslash \sigma\|_2^2, \tag{2.4}$$

where N is the number of data points, $\oslash$ is the Hadamard division, c is a scaling factor, $\mathbf{w}$ is the weights vector, $\mathbb{P}$ and $\mathbf{d}$ are, respectively, the predicted data matrix and the experimental data vector (for either SAXS or DEER), and σ is the standard deviation of the points of the curve. The scaling factor c is needed because, while paramagnetic NMR data magnitudes are fixed by the magnitude of the $\Delta\chi$ tensor and thus $\mathbb{P}\mathbf{w}$ is directly comparable to $\mathbf{d}$, SAXS and DEER have arbitrary scales. For

SAXS it is calculated as:

$$c = \frac{\|\mathbb{P}\mathbf{w} \circ \mathbf{d}\|_1}{\|\mathbb{P}\mathbf{w}_2 \circ \mathbb{P}\mathbf{w}\|_1}, \tag{2.5}$$

where $\mathbb{P}$ and $\mathbf{d}$ are, respectively, the predicted data matrix and the experimental data vector of SAXS, and $\circ$ is the Hadamard product. For DEER the modulation depth has been calculated as:

$$\lambda = \frac{\|\mathbb{M}\mathbf{w} \circ \mathbf{m}\|_1}{\|\mathbb{M}\mathbf{w} \circ \mathbb{M}\mathbf{w}\|_1}, \tag{2.6}$$

where $\mathbb{M} = \mathbb{P} - \mathbb{1}$ and $\mathbf{m} = \mathbf{d} - \mathbf{1}$ ($\mathbb{1}$ and $\mathbf{1}$ are, respectively, a matrix and a vector of ones).

2.4.3 Results and discussion

The MaxOcc values have been calculated for all the conformations using different combinations of experimental datasets, and a selection of the 1000 with the highest MaxOcc values has been reported in Fig. 2.2. It can be observed that:

- single experimental datasets are quite inefficient in narrowing the set of highly probable conformations, especially in the case of SAXS only (panel a) and DEER only (panel b), for which the highest MaxOcc value is around 65%, in contrast with ca. 35% for NMR only;
- the fact that NMR is rather selective in the MaxOcc analysis is confirmed also when compared to the SAXS+DEER combination (panel c), which causes a general decrease in the MaxOcc values, but does not significantly affect the highest ones, that are still higher than 35%;
- when combining NMR with SAXS (panel d) or with different combinations of DEER datasets (panels e, f, g, h), all the values drop below the NMR-only threshold, although with different extents. Among the DEER datasets, the mutant N53C-A103C (panel e) yields the smallest subset of highly probable conformations, followed by the N53C-T110C (panel f), then by the T34C-T117C (panel g), but the highest decrease is obtained when all the mutants are simultaneously used (panel h).

The high selectivity of NMR demonstrates that PCS and RDC are a precious source of information and highly complementary restraints [171]. The improved discrimination of DEER over SAXS is likely linked to the

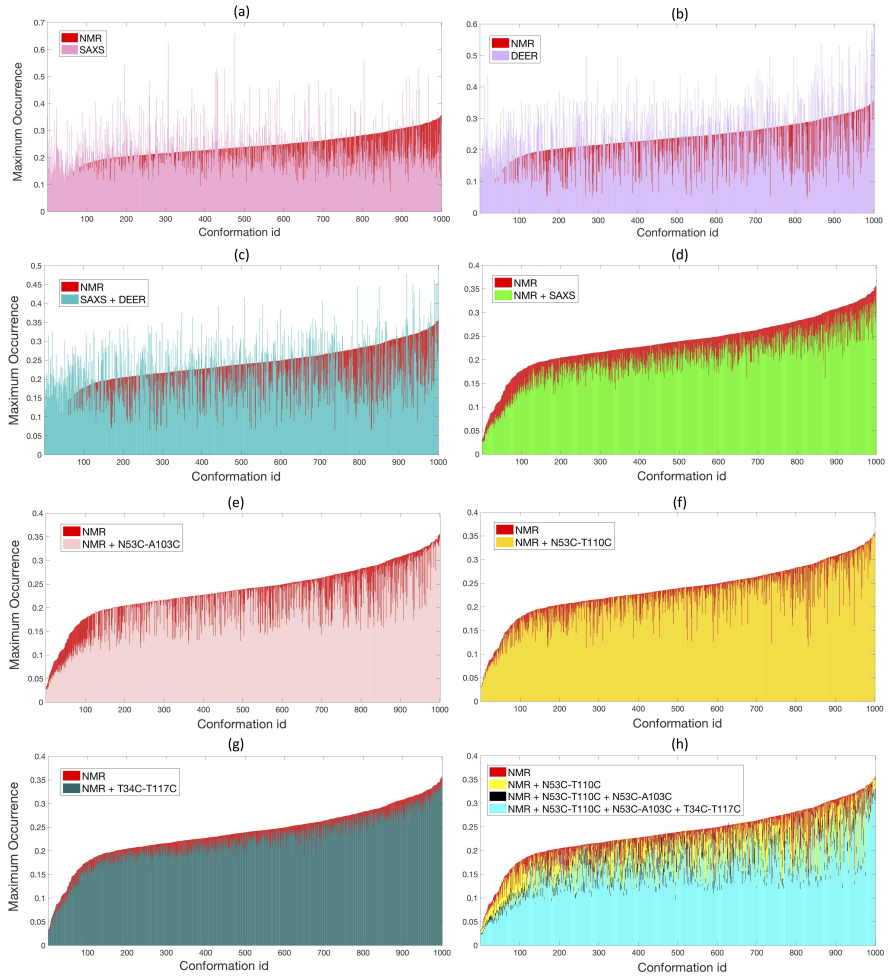


Figure 2.2: *Comparison between the MaxOcc values obtained for 1000 conformations with NMR only (red, sorted in ascending order) and with different combinations of experimental datasets: SAXS (a), DEER (b), SAXS+DEER (c), NMR+SAXS (d), single DEER datasets (N53C-A103C (e), N53C-T110C (f), and T34C-T117C (g), respectively), and comparison between different combinations of DEER datasets (h) [164].*

fact that the tags used for measuring DEER are not positioned in the center of the domain, whereas the electron density which scatters the X-rays has its center of mass close to the geometric centers of the domains, therefore the DEER information also depends on the orientational degrees of freedom. This is not completely unexpected, as other spectroscopic distance/conformation measurements based on (e.g.) FRET have built on the same behavior [172].

According to the results reported in Fig. 2.2, the combination of NMR, SAXS, and DEER selects the smallest subset of highly probable conformations (Fig. 2.3).

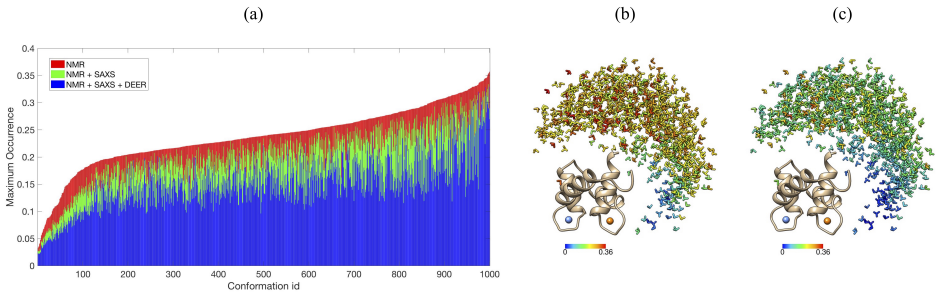


Figure 2.3: *Comparison between the MaxOcc values obtained for 1000 conformations with different combinations of experimental datasets: NMR only (red), NMR+SAXS (green), NMR+SAXS+DEER (blue). The conformations are sorted by ascending NMR only order (panel a). Triad of axes representation of the 1000 conformations, color-coded according to the MaxOcc values (<5% blue, ≥34% red) obtained with NMR only (panel b) and NMR+SAXS+DEER (panel c) [164].*

As expected, most of the conformations undergo a general decrease in the MaxOcc value by increasing the number of the experimental restraints, while the conformations with the highest MaxOcc values (conformation ids

from ca. 980 to 1000) are relatively less affected. This heterogeneous decrease confirms the high complementarity of the information, which indeed lead to the selection of the same subset of conformations. In Fig. 2.3 (panels b and c) also the chemical significance of the highest MaxOcc value conformations can be appreciated: passing from the NMR only (panel b) to the NMR, SAXS, and DEER analysis (panel c), the number of low MaxOcc values (blue triads) increases, and corresponds to conformations with severe steric clashes. On the contrary, the number of high MaxOcc value conformations (red triads) significantly decreases, individuating a subset of nine conformations that share some common features (Fig. 2.4).

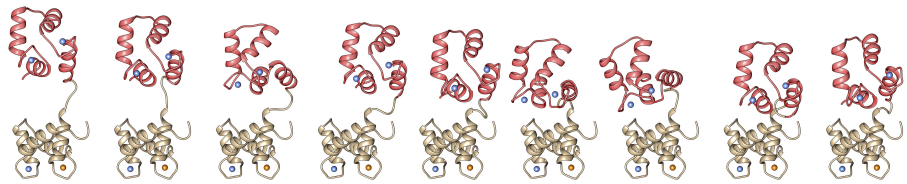


Figure 2.4: *Explicit representation of the nine conformations with MaxOcc ≥30% obtained with the NMR+SAXS+DEER combination.*

They are significantly extended, with the CTD bent towards the first helix of the NTD, as the conformational landscape of CaM is most likely dominated by the repulsive electrostatic interactions between the two highly charged domains. Although open, they do not correspond to the classical elongated, dumbbell-shaped conformation (1CLL [163]), nor to the closed canonical conformations (1YR5 [128]), and neither to averages between the two X-ray extremes. In order to highlight the variation of these structural similarities, the mean linear ($\langle r \rangle$) and angular ($\langle \rho \rangle$) variations among all the conformations have been calculated (Fig. 2.5).

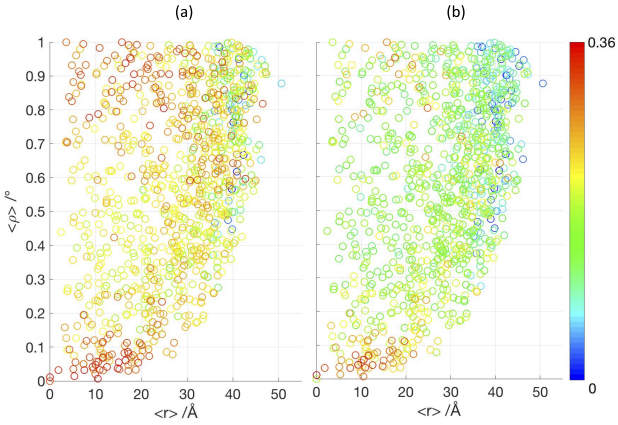


Figure 2.5: *Representation of the MaxOcc values for NMR only (panel a) and NMR+SAXS+DEER (panel b), color-coded according to the MaxOcc values (<5% blue, ≥34% red). Each conformation is located in the plot according to its distance in terms of angular (ρ) and cartesian (r) coordinates from the conformation which, in the NMR+SAXS+DEER analysis, has the highest MaxOcc value [164].*

More specifically, r_{ij} represents the distance between the CTD centers of mass in the i-th (cm_i) and j-th (cm_j) conformation, $r_{ij} = \|cm_i - cm_j\|_2$, and r_{ij} is the angular distance between two CTD orientations expressed in terms of quaternions, $\boldsymbol{q}$, as defined in [173], $\rho_{ij} = 1 - |q_i - q_j|$. In Fig. 2.5, each conformation, represented as a circle, is color-coded according to its MaxOcc value and is located in the "$r\rho$" plane according to its distance from the conformation that in the analysis with NMR, SAXS, and DEER resulted to have the highest MaxOcc value. Getting away from the axis origin means finding conformations increasingly different from the reference, and again, passing from NMR only (panel a) to NMR, SAXS and DEER (panel b), the red circles decrease in number and concentrate in a particular region, confirming similar features also in terms of angles and distances.

Of course, if more MaxOcc values were calculated, other conformations would emerge above the same threshold. However, given that the MaxOcc function is rather smooth, it can be expected that the other conformations would appear in the same region, i.e. they would share with these nine conformations the relative arrangements of the two domains.

The presented results demonstrate that DEER is highly complementary to NMR and SAXS in better defining the conformations that can be sampled for longer times, compatibly with the experimental data. In the specific case of CaM, the set of highly probable conformations has both a chemical and biological significance. The conformations are significantly extended, with the CTD bent towards the first helix of the NTD, according to the steric hindrance, as well as to the repulsive electrostatic interaction between the charged domains, and can give a hint in the comprehension of the mechanisms in which calmodulin is involved.

2.5 MaxOcc for Intrinsically Disordered Regions

2.5.1 Introduction

Intrinsically Disordered Proteins (IDPs) are proteins that lack a fixed or ordered three-dimensional structure, thus representing an extreme case of molecular flexibility. IDPs can exhibit disorder in their entirety or contain only some mobile residues in an overall folded structure, as in the case of Intrinsically Disordered Regions (IDRs) [174–176]. The behavior of both IDPs and IDRs may have profound implications in physiology and pathology [177], therefore it is of capital importance to understand their conformational preferences.

The proposed analysis is based on the hypothesis that the conformational preferences of IDRs are encoded in the observables from the structured part and can be seen through RDC (Paragraph 6.1)[178–181]. This method has been tested on the CTD of the HIV-1 viral capsid protein (Fig. 2.6, panel a). The native, capsid protein comprises two helical domains, a NTD and a CTD, connected by a highly flexible linker [182, 183]. The isolated CTD is dimeric, with residues 150-221 forming a four-helix bundle that possesses an antiparallel helical interface with W184 and M185 at its center. The last 10 residues (221-231), a significant fraction of which is positively charged, form an unstructured tail that is the IDR investigated in this work (Fig. 2.6, panel b).

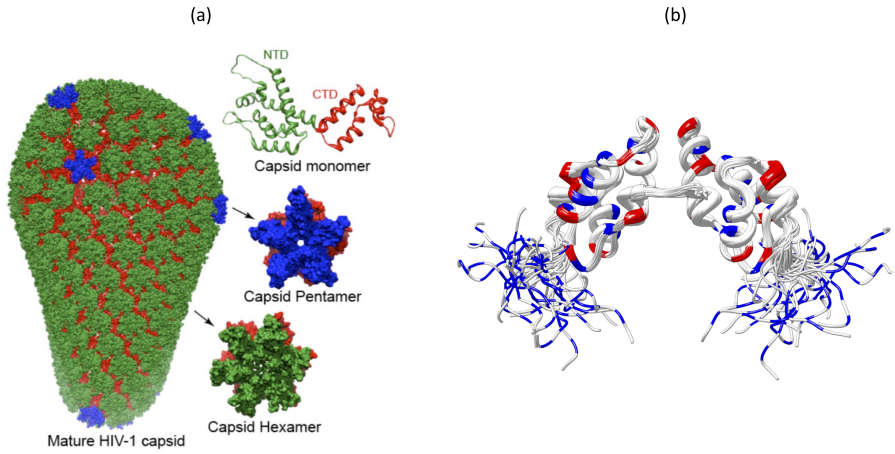


Figure 2.6: *The capsid protein (top right) comprises NTD (green) and CTD (red) [183]. During capsid assembly, the NTD forms pentameric [184] (middle right) and hexameric [185] (bottom right) rings (with the NTDs shown in blue and green, respectively, and the CTDs shown in red in both oligomers). A model of the fully assembled capsid (left) comprises adjacent hexamers connected to each other via CTD dimers and exactly 12 pentamers are required to close the cone [182] (panel a). Distribution of positively (blue) and negatively (red) charged residues in the NMR structural ensemble [186] (panel b).*

This system had been previously characterized by XRD and NMR:

(i) very different interdomain arrangements are observed in a number of crystal structures (1A8O ([187]), 1A43 ([188]), 1BAJ ([188]), and 2BUO ([189])), indicating that crystal packing forces can influence the dimer formation [190];

(ii) the solution structure was determined by NMR, using NOE distance restraints and RDC (PDB: 2KOD [190]) measured on the structured part of the protein, in two different alignment media (Pf1 phage [191, 192] and C12E5 [193], Tables S1 and S2, respectively, in [186]).

The detailed analysis described in Paragraph 6.2 indicates that the protein structure in solution is different from that in the solid state outside the structural inaccuracy of the crystallographic models. All the subsequent calculations were thus performed using the solution structure 2KOD. Since very few experimental distance restraints are available for residues in the tail, large variability is present in the NMR ensemble for this region. Moreover, the conformation of the tail, as well as its motional properties, could influence the predicted RDC, and the positively charged residues could impact the calculation of the electrostatic alignment. Some tests were thus performed, showing that:

- each member of the structural family obtained from the NMR minimization reproduces the RDC [190] extremely well (Fig. 2.7, panels a, b, and c),
- however none of the structures yields a good agreement in the prediction of the RDC with PALES (Fig. 2.7, panel c).
- Removal of the tail residues also results in poor agreement between experimental and predicted RDC, both for the NMR structure (Fig. 2.7, panels c, d, and e) and for the *de novo* dimer structures reconstructed using HADDOCK (Fig. 2.7, panel f). The latter have been generated in order to exclude the possibility that a different dimer conformation could fit the data. The dimer conformations (without C-terminal tail) were indeed used to predict the RDC for the two alignment media, with the result that, for the electrostatic medium, the best-fit dimer has the same conformation of the NMR structures, while, for the steric medium, the best-fit conformation is more elongated.

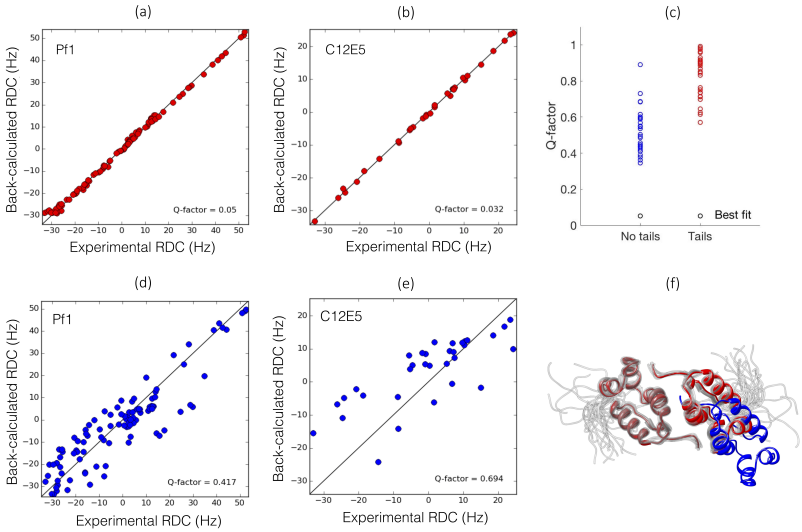


Figure 2.7: *Agreement between experimental and back-calculated RDC for dimers of the NMR ensemble (PDB ID: 2KOD) in Pf1 (panel a) and in C12E5 (panel b). Distribution of the RDC q factors for the structures in the NMR family calculated using PALES with (red) and without (blue) tails. For comparison, the best fit value obtained from back-calculated RDC is shown in black (panel c). Agreement between experimental and PALES-predicted RDC for the capsid CTD dimer, without the C-terminal tail, for Pf1 (panel d) and C12E5 (panel e). Comparison between dimers obtained by HADDOCK and those belonging to the NMR family. The best agreement between experimental and predicted RDC is obtained for the dimers colored in red (Pf1) and in blue (C12E5) (panel f) [186].*

These results indicate the need to model the C-terminal tail to accurately predict the RDC, especially for the steric medium, and suggests that the RDC must contain information on the average orientations of the tail.

2.5.2 Methods

X-ray structures were downloaded from the PDB. Simulated RDC were predicted using PALES [194, 195], whereas experimental RDC were taken from [190] (the authors report an error bar of 0.035 Hz). The RDC fittings were performed with FANTEN [123], the joint structure refinement with REFMAC-NMR [5, 196]. The dimers were reconstructed with HADDOCK [197, 198], by including some restraints derived from interdomain NOEs and by fixing the C2 symmetry for the system. The conformational pool of 9999 tails, with bonded and non-bonded energy terms, and without electrostatic contributions, was generated using Xplor-NIH [127, 137, 138]. The MaxOcc analysis was performed using the MATLAB (The MathWorks, Natick, MA) script described in Section 2.2.

2.5.3 Results and discussion

A simple preliminary check excluded the possibility to describe the system as a unique structure: no single conformer was compatible with both the RDC datasets. It was then evaluated the simplest possible ensemble, i.e. the one obtained by assigning all the members the same probability, and a striking disagreement with the experimental data was again noted. Conversely, when the weights of the conformations in the ensembles were allowed to change by a non-negative least-squares fit, the prediction of the RDC showed a remarkable agreement with the experimental data (Fig. 2.8), suggesting the presence of a distribution of conformational preferences. Being the system heavily underdetermined, however, the reconstruction was not trivial and far from being unique [149, 150, 199–202]. The confor-

mational preferences were thus extracted through the MaxOcc approach [158, 164, 200, 203], using the PALES-predicted RDC and both the experimental RDC datasets.

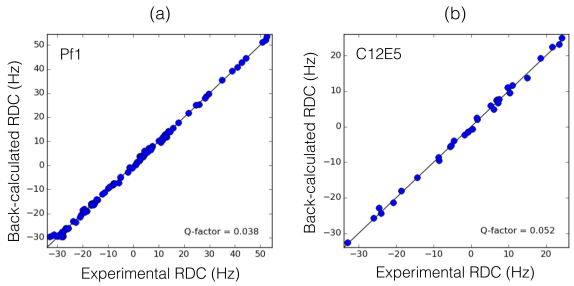


Figure 2.8: *Agreement between predicted and experimental RDC for an ensemble of dimers possessing C-terminal tails generated by Xplor-NIH. The data for alignment in Pf1 and C12E5 are shown in panels a and b, respectively [186].*

Out of the 9999 structures, only 117 conformations have the MaxOcc value $> 10\%$ (Figs. 2.9) and account for a Maximum Occurrence Region (MaxOR) of 98%.

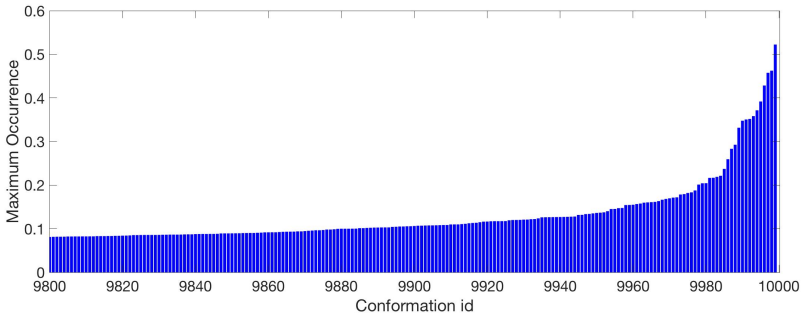


Figure 2.9: *MaxOcc values, sorted by ascending value, for the highest 200 conformations out of the whole set of 9999 [186].*

Strikingly, the 25 most probable tail conformations (Fig. 2.10, panels a and b) share a common orientation, pointing towards the interior of the capsid, and are almost equally distributed in two groups (A and B), containing 5648 and 4351 conformations, respectively.

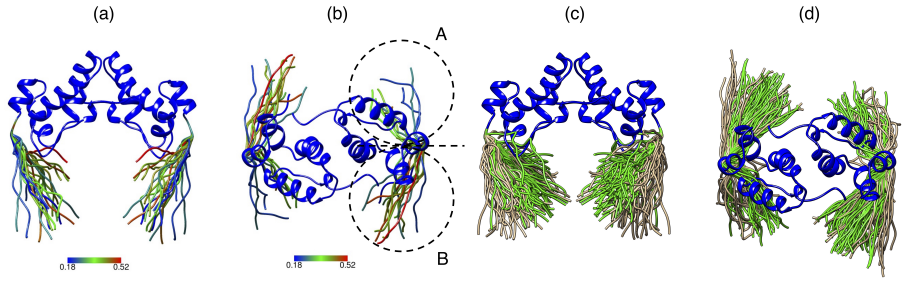


Figure 2.10: *Two different views of the capsid CTD, containing 25 (panels a and b) and 100 (panels c and d) tails with the highest MaxOcc values. The conformations in group A or B (encircled) point in different directions with respect to the plane indicated by the black dotted line, positioned at the C_α of the last residue of the folded part (panel b). The tails in panels c and d have been calculated with two different RDC datasets: C12E5 only (green) and PF1+C12E5 (beige) [186].*

Although each group accounts for a large fraction of the data, the combination of just one A and one B tail yields much better agreement with the experimental data. However, a MaxOR of 100% cannot be achieved even when combining the 11 conformations with MaxOcc $>$ 30% (Table 2.1). The calculations were also repeated using only the steric dataset (C12E5, green, Fig. 2.10, panels c and d), and the structural preferences revealed to be similar to those obtained with both steric (C12E5) and electrostatic (PF1) media (beige), indicating that possible inaccuracies in the PALES prediction of the electrostatic contributions to the alignment do not significantly affect the calculated structural preferences of the protein tail.

A		B	MaxOR (%)
The entire ensemble of A conformations		-	49
-		The entire ensemble of B conformations	75
The A conformation with the highest MaxOcc	+	The B conformation with the highest MaxOcc	87
The 3 A conformations with MaxOcc 30%	+	The 8 B conformations with MaxOcc 30%	94

Table 2.1: *MaxOR values for different combinations of conformations [186].*

In conclusion, RDC have proved to be a precious source of information, not only for the validation of the folded part structure, but also for the description of the disordered part. Moreover, when used as experimental restraints in the MaxOcc method, they lead to biologically relevant results, with potential implications on the mechanisms in which the virus is involved. The selected subset of tail conformations reflects the charge distribution: the positively charged tails point towards the interior of the capsid, which assembles into a cone that encloses the negatively charged viral RNA [182].

Chapter 3

Conclusions and outlook

The results presented in part I of this thesis demonstrate that paramagnetic NMR observables are an invaluable source of information: on the one hand, they reflect the profound relationship between the electronic structure and the magnetic properties, hence provide very localized information, on the other hand, they encode long-range structural information. In the light of the presented results, it is possible to draw two main conclusions.

1. Quantum Chemical methods are now mature to lead the structural refinement to a next level of accuracy. Increasing as much as possible the resolution is indeed required to improve structure-function relationships, which may have a huge impact in fields such as drug discovery.

(i) The comparison between experimental and QC-calculated hyperfine shifts allows for revealing details not easily accessible by experiments only. Moreover, the presented calculations give a strong hint towards

reassigning the signals, and the minor but significant structure modification of the HDPT derivative suggests the possibility to refine the structure of metal complexes starting from homologous compounds. The present results do not include effects of mobility or dynamics, because calculating the magnetic properties from structural ensembles is currently too demanding for routine applications. However, it can be reasonably expected that, in the near future, the improvements in the computational tools will provide access to thermal and solvent effects [204, 205].

(ii) A novel approach that extracts local information from long-range pNMR observables has been presented. Such method demonstrates to be successful in the reconstruction of the first coordination sphere of a metalloprotein, and its sensitivity is expected to be highly relevant also for the design of paramagnetic tags [206–208]. The comparison between the susceptibility tensor calculated *a priori* from the tag structure and the experimental one could be used to verify the absence of relative motion between the tag and the protein, a necessary condition for the quantitative structural analysis through tag-induced PCS [116, 209, 210]. Not least, in the presented example the information is so detailed to even suggest the protonation state of the active site, with implications on the mechanism of inhibition and catalysis.

2. Experimental data contain information about the conformational heterogeneity which are difficult to extract, thus mathematical models are required. MaxOcc demonstrates to be successful in defining the conformations that can be sampled for longer times in highly flexible systems. The conformations with the highest MaxOcc value share common features, with

potential biological implications.

(i) The accuracy of the MaxOcc method significantly improves by increasing the number of complementary experimental restraints. In particular, the addition of DEER is particularly efficient, also because it partially solves the problem of time averaging. The conformations that are most likely to be sampled are all extended and could represent a useful step in the comprehension of the mechanisms in which calmodulin is involved.

(ii) For proteins contaning an IDR, the experimentally determined RDC of residues for the folded domain can be used not only to validate the structure of the folded region but also to describe conformational features of the disordered part, even in the absence of experimental data for the amino acids in the disordered region. The obtained subset of tail conformations, which share a common orientation towards the interior of the capsid, reflects the charge distribution of the positively charged tails and the negatively charged RNA contained in the capsid.

Although the encouraging results, there is ample room for improvements. First, when it will be computationally feasible, the refinement methods can be further complicated by removing the hypothesis of working with static systems. In fact, conformational heterogeneity and dynamics can be crucial for the comprehension of metalloproteins, as demonstrated by the studies on the ensemble reconstruction method. Then, both the refinement and the ensemble reconstruction methods can be improved, in terms of velocity and user-friendliness, and could take advantages from the possibilities offered by Machine Learning [103, 104, 211].

Part II

NMR for Biomaterials

Introduction

As stated in the Introduction to Part I, a model is a somewhat abstract representation of reality, typically expressed in mathematical terms and often transformed into computer programs. Its development, however, is strongly tied to the results obtained by a series of physical observations.

This part of the thesis is again focused on biological systems, even if in a rather broader sense. The aim is no more to investigate the behavior of biomolecules in living systems but to design bioinspired materials and biomaterials: the former are synthetic materials whose structure, properties, or function mimic those of natural materials or living organisms, the latter are substances that have been engineered to interact with biological systems for medical purposes.

The most used tools for the characterization of such systems are spectroscopic techniques, such as NMR. As already seen, indeed, NMR has a privileged view of atomic level and often, if not always, bioinspired materials lack long-range order, being thus inaccessible by other techniques [212–218]. NMR has the advantage of ensuring a very high resolution, but suffers of low sensitivity, mainly because of the low natural abundance of the observed nuclei and the small difference in the population of the nuclear

spin levels. Moreover, the spins of the nuclei which find many applications in Materials Science (Si, Ti, etc.) have low or very low gyromagnetic ratios and therefore are poorly coupled with the environment, so that the relaxation mechanisms can be quite inefficient. In order to mitigate the effects of low sensitivity, the acquisition times are often very long.

The situation is further complicated by the fact that materials are often in the solid state, and ssNMR has to deal with the additional complication of inhomogeneous line broadening, because the nuclear interactions, such as chemical shift anisotropy and dipolar coupling, are not averaged out. In solution the anisotropic interactions are averaged out by molecular motions and have effects only on the relaxation (homogeneous broadening), in a powder solid state spectrum the signals are distributed on a range of thousands of Hz by anisotropic interactions. Furthermore, a sample can contain slightly different environments which cause an additional line broadening (heterogeneous broadening). Inhomogeneous broadening can be decreased by rotating the sample at a speed of the order of the frequency at which occurs the interaction that is intended to be eliminated (Magic Angle Spinning (MAS)), but this imposes a reduction in the amount of sample (Fig. 3.1, panel a). However, the signal intensity can be still not enough for the characterization of materials. Increasing the magnetic field is currently possible up to 28 T (Fig. 3.1, panel b), while decreasing the temperature of the sample requires large amounts of liquefied cryogenic gases; unfortunately, both these options are very expensive.

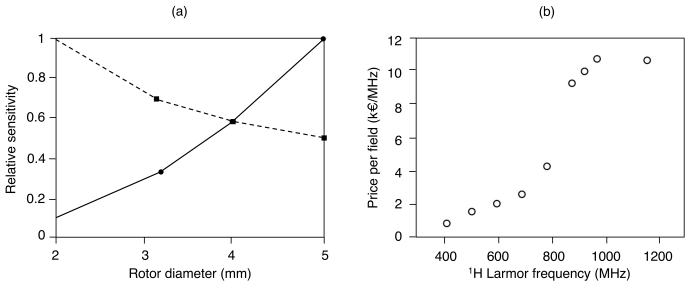


Figure 3.1: *Comparison of probe sensitivity for various sample (rotor) sizes. Shown are relative sensitivity per unit volume (dashed line) and relative sensitivity for a full rotor (solid line) (panel a) [219]. Trend of the price of commercial NMR spectrometers using superconducting magnet systems normalized by the field, at the time of market introduction of the 1.2 GHz instrument (source: Bruker Biospin) (panel b) [220].*

Consistent efforts are thus devoted to the solution of these problems, through a two-fold aim:

1. refining the integrated approaches that combine complementary experimental methods;
2. improving the NMR technique from both the experimental and theoretical point of view: on one side through the use of Dynamic Nuclear Polarization (DNP) [221] and the design of pulse sequences that increase the quality of the signals [222], on the other side through the development of post-processing methods [223].

This part of the thesis presents the characterization of a protein-based catalyst (Chapter 4), at two different levels of complexity (Sections 4.2 and 4.3). The materials are analyzed by means of ssNMR and other techniques, such as X-ray diffraction. Finally, the low quality of some acquired spectra led to the development of a denoising method (Section 4.4).

Chapter 4

Proteins as catalysts

4.1 Introduction

Among the biomolecules involved in biological processes, proteins are the most stable and those that can experience the widest range of 3D arrangements and conformational changes. For these reasons, with the exception of ribozymes, i.e catalytic RNA molecules, they constitute the largest majority of enzymes. Enzymes are potent catalysts able to accomplish complex chemical tasks under physiological, hence environmental friendly, conditions. Due to both their efficiency and inherent Green nature, from key players in biological sciences, they are fairly gaining their role also in biomedical and industrial applications. However, their use has still not reached its full potential for a number of reasons, first and foremost their limited lifetime if compared to the typical industrial catalysts. This problem can be overcome through the immobilization of the protein on a solid support [224], which results in both the protein stabilization and interest-

ing changes in the properties (catalytic, photochemical, biological) of the composite.

In spite of the broad interest that protein-material composites have attracted, their characterization is mainly oriented to bulk properties, such as matrix morphology or enzyme activity, whereas an atomic-level characterization is far from being routine [225]. In the literature are present few reports of experiments probing the conformational state of small peptides [226–228] and their interactions with the matrix [215, 229], but some studies on biosilica-entrapped enzymes prove the feasibility of ssNMR characterization [230, 231], as well as of high-field DNP [232].

Solid-state NMR has thus been exploited to investigate the interactions that occur on the molecular scale, as they affect the composite stability and reactivity and can enable the rational design of catalysts for a variety of applications. The proposed device is aimed at providing a Green solution to the depuration of water from organic pollutants through a photoinduced reaction cascade. The system is composed of an inorganic support, *titania* (TiO_2), and two protein components, *superoxide dismutase 1* (SOD1) and *cytochrome-c* (cyt-c). Upon light irradiation, titania produces superoxide, SOD1 transforms it in peroxide, which, catalyzed by cytochrome-c, oxidizes organic pollutants.

The first tests, performed on the fully assembled material with different mutants of cytochrome-c trapped in anatase titania, showed a significant catalytic activity. However, it has also been observed that the matrix itself has a strong photoreactivity towards the selected substrates. Therefore, in order to reduce the number of variables, the problem has been simplified by substituting the catalytic protein component cyt-c with *lysozyme*, having

similar superficial charge but no redox-active centers, and the matrix with its photo-inactive analogue, *silica* (SiO_2).

Titania

Titanium(IV) oxide (TiO_2), also known as titania, occurs in nature as the minerals rutile and anatase, and contains six oxygen atoms surrounding a central titanium atom. In bioinspired titania synthesis, titanium(IV) bis(ammonium lactato)dihydroxide (TiBALDH) (Fig. 4.1, panel a) is routinely employed as a precursor, because water-soluble and stable at neutral pH and ambient conditions [233–236]. TiBALDH solutions contain several species that are in equilibrium among themselves and with TiO_2 and, at neutral or slightly basic pH, mononuclear hydrated titanium species also include $Ti(OH)_4$ [234, 237–240]. Beyond the applications of titania in the fields of paints, plastics and cosmetics, nanosized anatase is known to exhibit photocatalytic activity under UV irradiation [241].

Silica

Silicon(IV) oxide (SiO_2), also known as silica, is a polymer solid very abundant in nature, containing four oxygen atoms surrounding a central silicon atom, with three different Si coordinations: disilanols ($SiO_2(OH)_2^{2-}$), monosilanols ($SiO_3(OH)^{3-}$), and fully condensed sites (SiO_4^{4-}). In water the species $[H_xSiO_4]^{(4-x)-}$ are the precursors for the formation of silica. The prevailing species at neutral pH is the undissociated and monoprotonated tetraoxosilicic(IV) acid H_4SiO_4 (Fig. 4.1, panel b). Silica is a chemically inert material, whit a negatively charged surface over a wide range of pH values (from 2 to ca. 10), that finds widespread applications

in several fields, e.g. as a support for heterogeneous catalysts [242].

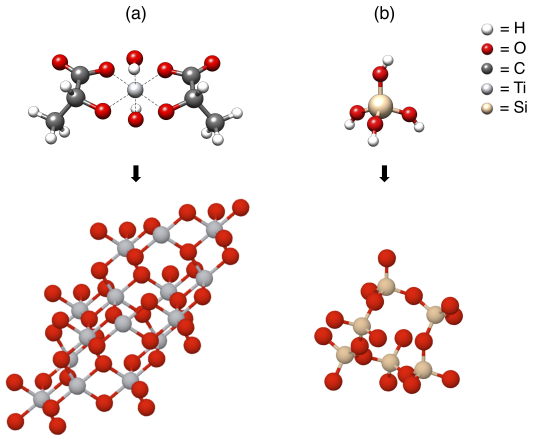


Figure 4.1: *Structures of TiBALDH, anatase titania [243] (panel a, top and bottom, respectively), H_4SiO_4 and silica [244] (panel b, top and bottom, respectively).*

Superoxide dismutase 1

Superoxide dismutase 1 (SOD1) is a 32 kDa homodimer (Fig. 4.2, panel a) which forms a β-barrel and contains an intramolecular disulfide bond and a binuclear Cu/Zn site in each subunit. The copper and zinc ions are responsible for the catalytic activity of this enzyme, which favors the disproportionation of superoxide to hydrogen peroxide and molecular oxygen [245].

Cytochrome-c

Cytochrome-c is a small hemeprotein (12 kDa) belonging to the c-type cytochrome family. It contains a characteristic amino acid motif (CXXCH, i.e. cysteine-any-any-cysteine-histidine) that binds heme, and its backbone

is folded into five α-helices. It has a high surface charge of positive sign (Fig. 4.2, panel b, transparent blue surface) and, unlike the other c-type cytochromes, is highly water-soluble. This enzyme is involved in several cellular processes, that include electron transport, cell apoptosis, and peroxidase activity [246].

Lysozyme

Lysozyme is a small globular protein (14 kDa) found in relevant concentrations in human secretions, such as tears, saliva, and blood serum, in cow-milk and in hen-egg white. Structurally, it consists of a single polypeptide chain of 129 amino acids, which is found to be cross-linked by four disulfide bridges [247] (Fig. 4.2, panel c, highlighted in yellow). The active site of lysozyme consists of a deep crevice, which divides the protein in two domains: one (residues 40-88, Fig. 4.2, panel c, cyan) consists almost entirely of β-sheets, the other (residues 1-39 and 89-129, Fig. 4.2, panel c, magenta) is mainly helical [248]. Although the 80% of its amino acids presents neutral side chains, the remaining charged residues are mostly concentrated close to the surface. The resulting high superficial positive charge (Fig. 4.2, panel c, transparent blue surface) makes lysozyme an antimicrobial enzyme, in particular a glycosidase. It catalyzes, through electrostatic interactions, the hydrolysis of glycosidic bonds in peptidoglycans (the major component of bacterial cell walls, such as Gram-positive), which in turn compromise the integrity of the walls, causing lysis of the bacteria. Lysozyme is not only the prototypical templating molecule in bioinspired mineralization of silica and titania [236], but also particularly suitable for the XRD characterization: it easily crystallizes in a variety of conditions, and the quality of the

crystals is usually high enough to allow for high resolution in-house data collections.

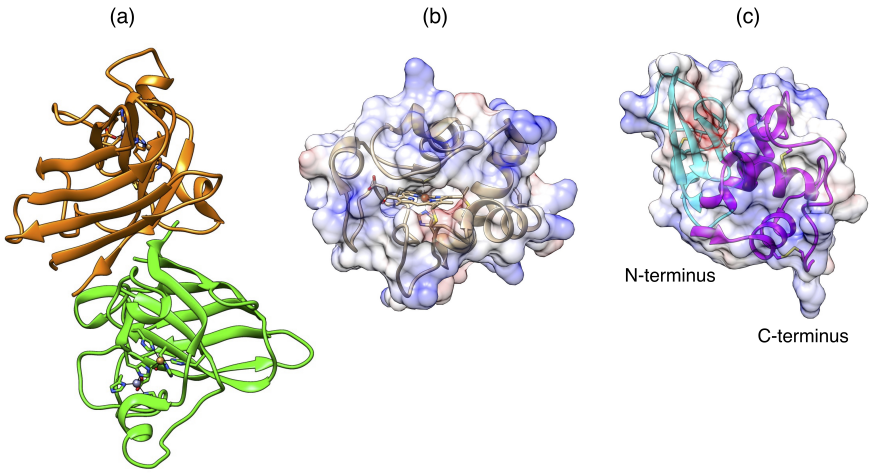


Figure 4.2: *Structures of SOD1 (1UXM [249], panel a), cyt-c (1HRC [250], panel b), and lysozyme (1HEW [251], panel c).*

4.2 Photoinduced catalysis for water depuration

4.2.1 Introduction

The composite has been prepared according to the protocol given in [252] and analyzed using ssNMR and Cyclic Voltammetry (CV), in order to investigate the atomic-level interactions between the components and the redox activity of the material.

Photoactivity tests have been then performed in the presence of pyrogallol, testing different cytochrome-c mutants. Pyrogallol is a typical aromatic substrate for oxidoreductases reactivity tests [253], so its oxidation to purpurogallin with hydrogen peroxide is expected to be catalyzed by cyt-c. The selected mutants are M86H, M86A, and M86A-G47T double mutant. The M86H mutant is inactive: the substitution of methionine with histidine causes a stronger interaction with iron, abolishing the peroxidase activity [254]. The M86A mutant is more active than the wild-type since alanine does not coordinate the metal center and thus the access to the axial position is always free [255]. The double mutant shows a further increase of the activity because the threonine interacts with the histidine bound to the heme and attracts charge [256].

4.2.2 Methods

TiBALDH, SOD1, and recombinant horse heart cyt-c were purchased from Sigma-Aldrich and used without further purification.

100 μL of a 10 mg/mL recombinant horse heart cyt-c solution were mixed with 100 μL of a 1 mg/mL SOD1 solution and 20 μL of Ti-BALDH solution at 25% concentration (w/v).

The photoactivity tests were performed using SOD1 and yeast cyt-c mutants (wild type, M86A, M86H, and G47T-M86A double mutants). All the mutants were purchased from Giottobiotech. The precipitation mixture consisted of buffer PBS, 100 mM Ti-BALDH, lysozyme at a final concentration of 10 mg/mL, each of the cyt-c mutant at a final concentration of 1 mg/mL, and SOD1 at a concentration of 0.1 mg/mL. A blank sample containing only lysozyme and SOD1 was also prepared. Each precipitate was washed thoroughly with deionized water and pyrogallol 4.8 mmol/dm^3 in citrate buffer at pH 3.5 was added. The suspensions were let for a whole day under sunlight.

^{1}H and ^{13}C NMR spectra have been acquired on a Bruker AVANCE-NEO spectrometer operating at 500 MHz ^{1}H Larmor frequency (11.7 T), equipped with cryogenically cooled triple resonance inverse detection probe heads. The ssNMR spectra on the composite containing ^{15}N-labelled cyt-c have been acquired on a Bruker Avance III operating at 850 MHz ^{1}H Larmor frequency and 60 kHz MAS rate.

The CV characterization was performed using carbon-based screen-printed electrochemical cells (Drop-sens SPCE). The precipitation mixture consisted of phosphate-buffered saline (PBS), 100 mmol/dm^3 TiBALDH, and cyt-c (final concentration 10 mg/mL). The mixture was vortexed at room temperature and 10 μL were dropped onto the surface of the working electrode of a SPCE. As blank, a sensor was modified using the same precipitation mixture but in the presence of lysozyme, which has similar structural characteristics of cyt-c but does not contain iron. During the electrochemical characterization, PBS buffer was used as supporting electrolyte.

4.2.3 Results and discussion

The synthesis described in Section 4.2.2 leads to the formation of a brownish precipitate and is highly reproducible.

The absence of signals from ^{13}C and $^{47/49}Ti$ NMR spectra indicates the necessity of isotopic enrichment, while the ^{1}H NMR spectrum shows two water peaks: a sharp one, attributed to free water, and a broader one, likely due to immobilized water either around the protein or interacting with the matrix (Fig. 4.3, panel a, red and black spectra). Spectra recorded with fast recycling display signals around 8, 2, and -5 ppm, which are consistent with the chemical shift of the protein [255] (Fig. 4.3, panel a, cyan spectrum). The sharper signal in the area around 2 ppm appears to correlate with water signals and is consistent with the resonance frequency associated to basic H atoms bound to terminal oxygens on the titania surface [257]. However, the most shifted resonances from the heme protons (from 50 to -20 ppm [255]) are not observed. Two broad resonances, corresponding to the most extreme values, appear upon homonuclear decoupling during the acquisition [258] (Fig. 4.3, panel b). These peaks shift with temperature as expected for paramagnetic systems [17], therefore may correspond to the protons of the heme system.

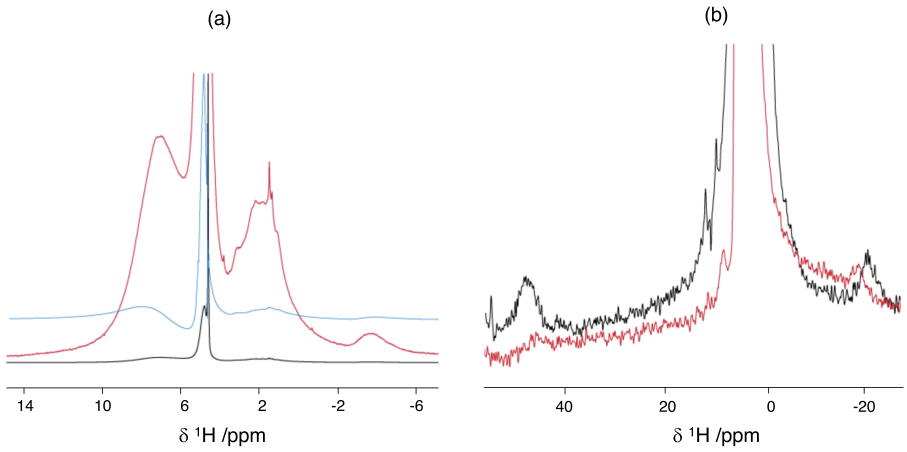


Figure 4.3: *^{1}H spectra of the titania-based composite: direct excitation spectrum (black), multiplied 30 times (red), and relaxation-filtered experiment (cyan) that shows the broader resonances associated with the protein (panel a). ^{1}H windowed-PMLG spectra of the titania-based composite acquired at 294 (red) and 269 (black) K (panel b).*

The ^{1}H-^{15}N heterocorrelation (HETCOR) spectrum maps almost completely with the solution HSCQ spectrum of the free protein (Fig. 4.4). A few missing peaks were identified to be close to charged residues on the surface of the protein as previously observed for other systems [259], and this proves a viable solution for the characterization of the titania based strategy as well.

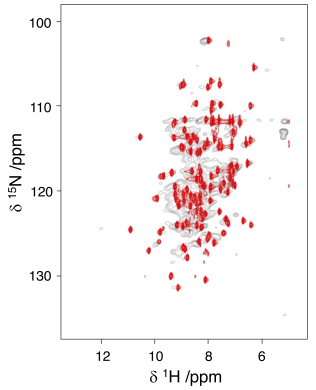


Figure 4.4: *^{1}H-^{15}N solution HSCQ spectrum of the free protein (red) superimposed to the ^{1}H-^{15}N HETCOR spectrum acquired in the solid state (grey).*

The system exhibits an electrochemical activity. From the CV performed on both the composite and a Ti-BALDH/lysozyme blank (Fig. 4.5, panels a and b, respectively), an oxidation peak is observed around 0.2 V.

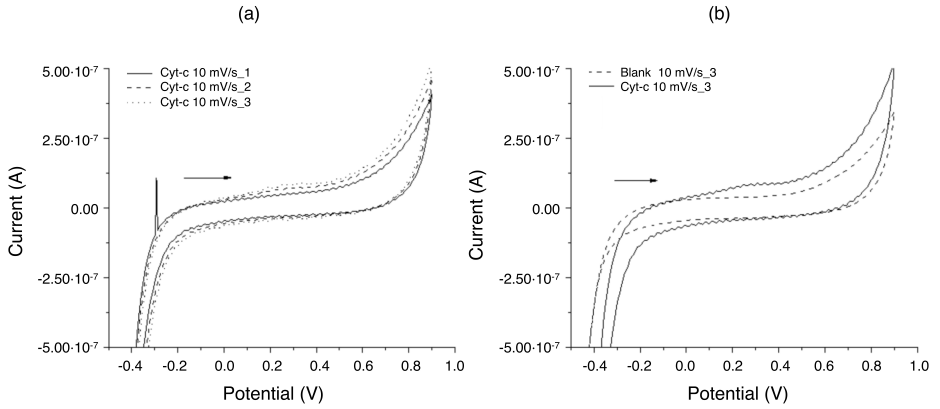


Figure 4.5: *Voltammetric scans obtained for Ti-BALDH/ cyt-c SPCE with PBS buffer, scan rate 10 mV/s (panel a). Comparison between Ti-BALDH/cyt-c SPCE and Ti-BALDH/lysozyme SPCE (blank) voltammetric scans (panel b).*

From the photoactivity tests emerged that all the samples are active in oxidizing pyrogallol. After one day of sunlight exposure, all the suspensions, even the blank and those with the inactive cyt-c, turn brownish (Fig. 4.6), sign of the presence of the oxidation products.

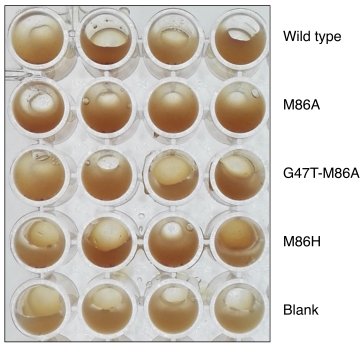


Figure 4.6: *Effect of the composite containing different cyt-c mutants and of a lysozyme+SOD1 solution (blank) on a pyrogallol solution upon sunlight irradiation. Each sample has been split in 4 wells to check whether the amount of material that is pipetted is reproducibly yielding the same amount of oxidation.*

These preliminary results on the titania composite showed both a photochemical and redox activity, but the complexity of the system required a simplification of the problem. The protein components have been substituted with lysozyme, which is extremely well characterized and has a high positive surface charge, and, in most of the experiments, the matrix has been substituted with silica, the photo-inactive analogue of titania. Several studies on bioinspired silica- and titania-based materials are indeed present in the literature [260–263], because their synthesis is cheap and fast, and the tunability of their structure and reactivity offers a wide range of opportunities in several fields [264, 265].

4.3 Integrated studies on lysozyme-based composites

4.3.1 Introduction

The study of the lysozyme-based composites containing SiO_2 and TiO_2 as inorganic matrices started with the analysis of the interactions between the protein and the matrix precursors through NMR titrations and XRD (Paragraph 6.1) [266].

Then, the whole composite has been investigated using different experimental techniques. Among the synthesis present in the literature [267], the one proposed by Luckarift *et al.* [236] has been selected, because particularly convenient. It foresees the use of lysozyme as a catalyst for the condensation of tetraoxosilicic(IV) acid and TiBALDH, thus it is based on cheap and readily available reagents, is fast, can be performed at ambient conditions, and yields the formation of a condensed network in the matrix without thermal processing. The work which followed was focused on the silica-lysozyme composite, on which a number of experiments have been performed, as schematized in Fig. 4.7.

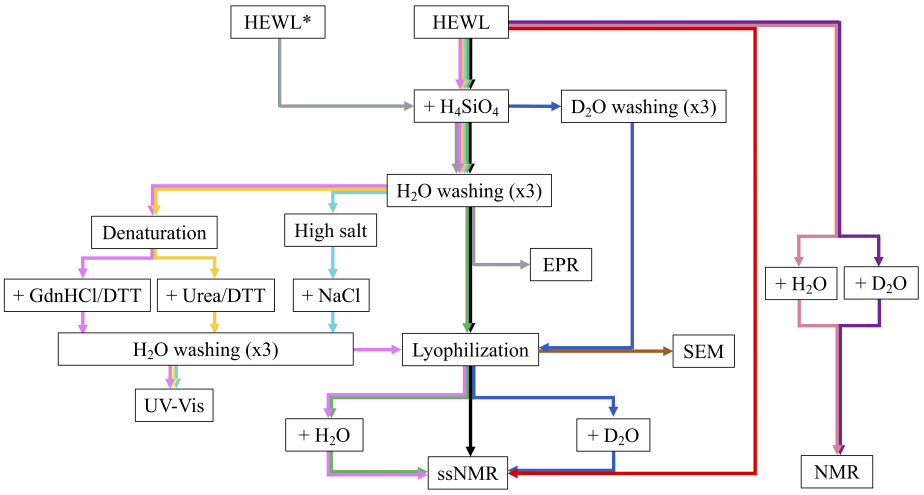


Figure 4.7: *Scheme of the performed experiments on the silica-lysozyme composite.*

4.3.2 Methods

Hen-egg white lysozyme (HEWL), titanium(IV) bis(ammonium lactato)dihydroxide (TiBALDH), and tetraoxosilicic(IV) acid were purchased from Sigma-Aldrich and used without further purification.

Crystal preparation

Crystals of HEWL were obtained in hanging drop by adding 2 μL of protein solution (0.5 mol/dm^3 Tris-HCl, 8 mg/mL HEWL, pH 8.5) to 2 μL of reservoir buffer (0.5 mol/dm^3 Tris-HCl, 0.7 mol/dm^3 NaCl, pH 8.5) and stored at 4 °C. The crystallization conditions were chosen to respect the pH at which the polymerization occurs and to minimize the number of additives. The crystals were afterwards soaked in tetraoxosilicic(IV) acid

and TiBALDH solution, in different concentrations (10 to 100 mmol/dm^3 for both the precursors and 1.7 mol/dm^3 for TiBALDH only), for 5 days.

XRD

The crystals used for the data collection were cryo-cooled using 25% ethylene glycol in the mother liquor, and the dataset was collected in-house. The data were processed keeping Friedel mates separated using the program XDS [268], reduced and scaled using XSCALE [268], and amplitudes were calculated using XDSCONV [268]. The structure was solved using the molecular replacement technique: the successful model used was 2W1X. The orientation and translation of the molecule within the crystallographic unit cell was determined with MOLREP [269]. The refinement was carried out using PHENIX [232], applying TLS restraints and using anisotropic B-factors for Na, Cl, and Ti only. In between the refinement cycles, the model was subjected to manual rebuilding using COOT [270]. Water molecules have been added by using the standard procedures within the ARP/WARP [259] suite. The quality of the refined structure was assessed using the program MOLPROBITY [271]. Data processing and refinement statistics are shown in Table 4.1. Coordinates and structure factors have been deposited at the PDB under the accession code 7A70.

Diffraction source	Bruker D8 Venture
Wavelength (Å)	1.541
Temperature (K)	100
Detector	PHOTON II
Crystal-detector distance (mm)	50
Oscillation range (°)	0.5
Total rotation range (°)	360
Exposure time/image (s)	30
Space group	$P4_32_12$
a, b, c (Å)	78.1, 78.1, 37.3
Mosaicity (°)	0.3
Resolution range (Å)	50.00-1.80 (1.91-1.80)
Total reflections, unique reflections	274135 (24437), 20052 (2909)
Completeness (%)	97.8 (89)
CC1/2	99.9 (45.1)
I/(σI)	16.1 (1.7)
Wilson B-factor (Å^2)	29.9
$R^{\dagger}_{merge}$, $R_{cryst}R^{\ddagger}_{free}$ (%)	0.12 (0.91), 19.625.6
Protein atoms, ligand atoms, water molecules	1001, 13, 82
RMSD bond lengths (Å^2), RMSD bond angles (°)	0.010, 1.950

Table 4.1: *Data collection and refinement statistics parameters.*
$^{\dagger}R_{merge} = \frac{\sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|}{\sum_{hkl}\sum_i I_i(hkl)}$, *where* $I_i(hkl)$ *is the mean intensity of the i-th observation of symmetry-related reflections hkl.*
$^{\ddagger}R = \frac{\sum_{hkl}||F_O| - |F_C||}{\sum_{hkl}|F_O|}$, *where* F_C *is the calculated protein structure factor from the atomic model (*R_{free} *was calculated with a randomly selected 5% of the reflections).*

Calculations

The $Ti(OH)_4$ structure was subjected to refinement at the DTF level of theory, with B3LYP functional [73–76], using Ahlrichs polarized basis set def2-TZVP [77, 78] and Grimme's dispersion correction D3 [79, 80]. The

resolution of identity approximation [81, 82] was employed with auxiliary basis set def2-TZVP/J in order to speed up the calculations. CPCM implicit solvent (water) was used [83]. All calculations were carried out using the ORCA 4.2.1 quantum chemistry package [24, 44].

Composite preparation

HEWL was dissolved in buffer Tris-HCl at pH 7.5 (A solution, 22 mg/dm^3 concentration). TMOS was hydrolyzed with a HCl 1 mmol/dm^3 solution and diluted 1:10 with the same buffer after 5 minutes of agitation at room temperature (B solution, 0.1 mol/dm^3 concentration). The A and B solutions were mixed and the obtained precipitate was centrifuged at 10000 rpm (r = 98 mm) and washed three times with deionized water.

In order to investigate the relative arrangements of the components, the precipitate was washed with three different solutions, then washed with deionized water, by centrifuging at each step for 20 minutes. The solutions were NaCl 6 mol/dm^3, urea 8 mol/dm^3 + dithiothreitol (DTT) 50 mmol/dm^3, and guanidine hydrochloride (GdnHCl) 6 mol/dm^3 + dithiothreitol (DTT) 50 mmol/dm^3, in buffer Tris 20 mmol/dm^3, pH 7.4.

NMR

The samples have been placed in a ZrO_2 3.2 mm rotor. The spectra have been recorded at 265 K and 64 kHz MAS. The 1D ^{1}H NMR spectrum has been acquired on a Bruker Avance III operating at 950 MHz ^{1}H Larmor frequency (22 T). The ssNMR spectra have been recorded on a Bruker Avance II operating at 700, 176, and 139 MHz ^{1}H, ^{13}C, and ^{29}Si Larmor frequency, respectively. The instrument is equipped with a 3.2 BVT MAS

double-resonance probe. The pulses at 90 obtained from the optimization of the acquisition parameters are 2.4, 3.5, and 4.7 μs for ^{1}H, ^{13}C, and ^{29}Si, respectively.

SEM

The SEM images have been obtained using a Carl Zeiss Merlin II field emission Scanning Electron Microscope, using a detector for secondary electrons with acceleration potential of 5 kV, a working distance of 2.8 mm, and magnification factors of 28.82 and 37.06. The sample was not subjected to preliminary treatments.

EPR

The EPR spectrum was acquired in solution at room temperature on a BRUKER ELEXSYS EPR spectrometer operating at X-band equipped with a ER 4102ST rectangular cavity. The modulation and the microwave frequencies were set to 20 kHz and 9.843 GHz, respectively. Centerfield was set to 130 mT and the total field sweep was 65 mT. The fit of the EPR spectrum was performed using EasySpin [272].

4.3.3 Results and discussion

The protein-precursor interaction

The NMR titrations did not provide useful results. For tetraoxosilicic(IV) acid, the only apparent perturbation at the highest precursor concentration is a 5% decrease in the signal intensity, with no detectable shift alteration (Fig. 4.8, panel a). For TiBALDH, shifts in the signals of residues 101-110 can be observed (Fig. 4.8, panels b, c, and d), but this area is known

to be susceptible to variations in response to minor changes in the experimental conditions (pH, ionic strength etc. [273, 274]) and, given that the pH has decreased from 8.5 to 5.1 during the addition, these results cannot be interpreted reliably.

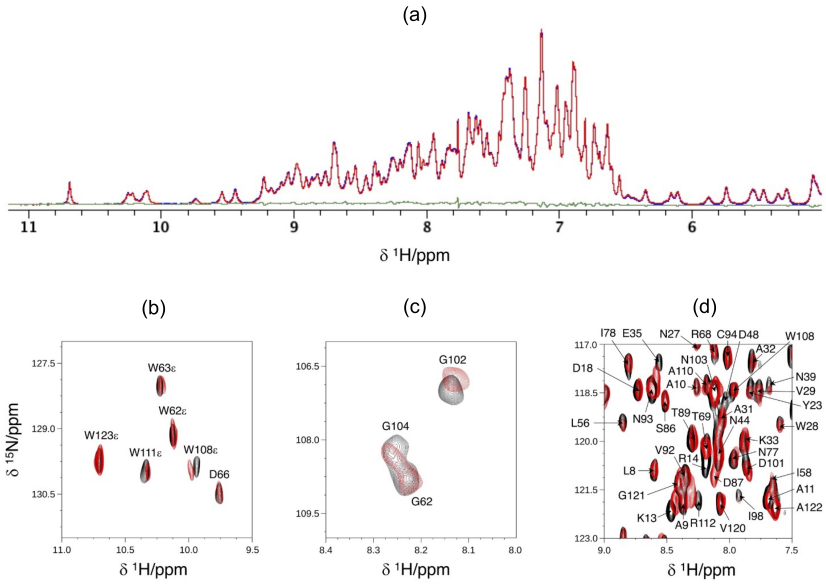


Figure 4.8: *^{1}H NMR spectra of lysozyme 10 mg/mL (red), lysozyme 10 mg/mL with silicic acid 100 mmol/dm^{3} (blue), and their difference (green) (panel a). Enlargements of regions of the ^{1}H-^{15}N HSQC spectra of lysozyme 100 mg/mL (black), lysozyme 100 mg/mL with TiBALDH 100 mmol/dm^{3} (red) [266]. The assignments are taken from references [275, 276].*

Soaking with tetraoxosilicic(IV) acid (buffered at pH 8.5) results in a fast disruption of the crystals, proportional to the concentration of acid added to the drop and, after 3 days, a glassy film is found instead of the liquid drop (Fig. 4.9, panel a). At the concentrations compatible with

the crystals, no conclusive evidence of the presence of silicon species in the crystals can be found.

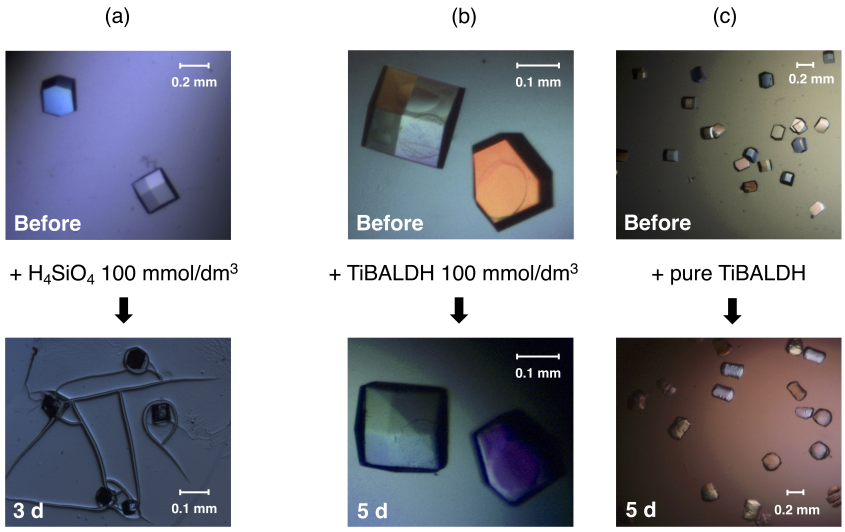


Figure 4.9: *HEWL crystals before and after the soaking with 100 mmol/dm^3 tetraoxosilicic(IV) acid (panel a), 100 mmol/dm^3 and pure TiBALDH (panels b and c, respectively).*

The situation is markedly different for TiBALDH. Soaking at the same concentrations tested for tetraoxosilicic(IV) acid does not cause alterations in the resulted crystals (Fig. 4.9, panel b), that indeed do not reveal a clear presence of titanium species. This is not fully expected due to a report of binding of a titanium species to the same crystal form of lysozyme (PDB entry: 6G5C) [4]. A careful inspection of this entry casts doubts in terms of the presence of titanium (a putative TiO_2 species), because (i) the supposed position of titanium perfectly corresponds to the position of a chloride ion

in all the lysozyme structures belonging to the same space group, (ii) the shape of the $2F_O$-F_C density of the putative titanium is perfectly spherical and fits way better with the chloride ion, and (iii) mononuclear TiO_2 species does not exist as such in solution [230, 277]. The last attempt was thus to perform a soaking with 2 μL of 50% w/w TiBALDH solution (1.69 mol/dm^3, pH 8, subject to change upon the hydrolysis of the components), which does not cause a complete disruption of the crystals, but rather a slight worsening that results in a decrease in resolution. The data had a maximum resolution achievable of 1.8 Å instead of the typical 1.4-1.5 Å in the same diffractometer, however the quality of the data remained very good.

The analysis of the data revealed the presence of a quite intense anomalous peak in an unexpected position. Considering the composition of the crystallization buffer, of the soaking solution, and of the cryo-protectant, the only atoms which could have a similar anomalous effect at the diffractometer wavelength are sulfur, chlorine, and titanium, with the last being the one with the largest expected value. The unexpected peak has been attributed to a labile hydrolyzed titanium compound in the form of $Ti(OH)_4$ for several reasons.

1. The anomalous signal:
   - no density is present in that position in several other examined lysozyme structures with the same space group;
   - the signal is slightly higher than that attributed to chlorides.
2. The $2F_O$-F_C density:
   - the shape is tetrahedral (Fig. 4.10). It is not attributable to sul-

fur, as it is not present free in solution, nor to a chloride ion, as the shape should be spherical, neither to a tetrahedral tetraoxochlorate(VII), because its formation would require a strongly oxidizing environment, incompatible with the protein;
- the Ti-O distances refine well at about 1.9 Å, in agreement with the theoretical value expected for such bond.

3. The PDB parameters:
   - the occupancy is about 0.7;
   - the B-factor has values coherent with those of the other labile or loosely bound atoms in the structure.

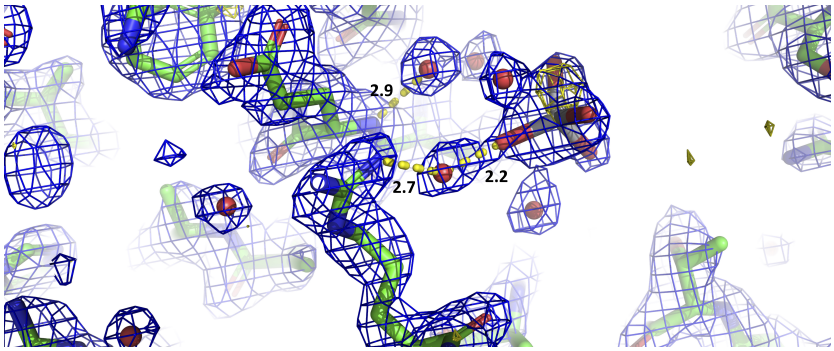


Figure 4.10: *Detail of the hydrogen bonds stabilizing the position of the titanium species through two water molecules represented as red spheres (distances are expressed in Å). The $2F_O$-F_C map is contoured at 1.0σ (blue) and 3.0σ (yellow) level [266].*

With these results at hand, the data collection on the crystal soaked with 100 mmol/dm^3 TiBALDH were re-examined and showed, ex post, a weaker but clear anomalous signal in the same position.

The presence of $Ti(OH)_4$ is also supported by the DFT calculation and by an electrochemical analysis of the mononuclear titanium(IV) species present in aqueous solution as a function of pH [230]. The first indicates that the oxygen atoms can have up to 0.87 e^- partial charge, which can easily account for a preference for forming hydrogen bonds as acceptor, the second shows that the species $Ti(OH)_4$ is prevailing at the working pH.

The position and the binding mode of the ligand suggest that its presence is likely to be an artifact due to the crystal packing and to its very high concentration. This could cause the diffusion of $Ti(OH)_4$ at a higher rate in the solvent channels, until the complex reaches a narrower channel where it is forcibly stopped because of steric hindrance and charge accumulation (Fig. 4.11).

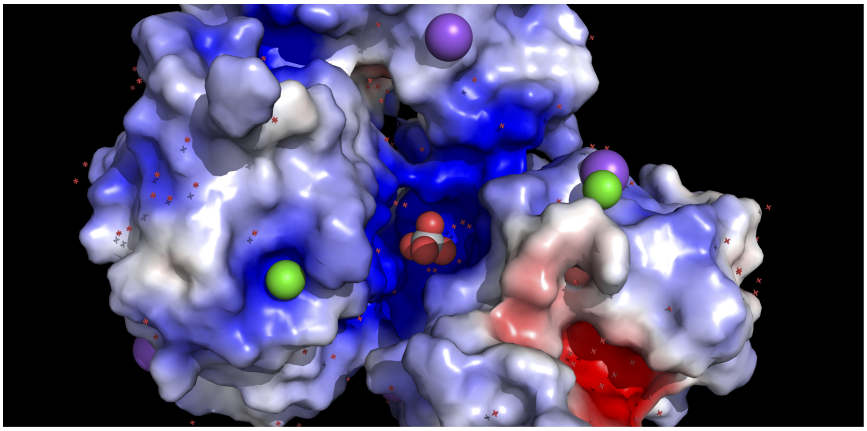


Figure 4.11: *Surface representation (calculated with Pymol APBS plugin) of the environment in which the titanium species (the red and white tetrahedron) is found, colored by electrostatic potential, showing chloride ions in green and sodium in purple [266].*

$Ti(OH)_4$ is placed in the region between two symmetry-related molecules and has no direct chemical interaction with any of the protein atoms. It is held in place by hydrogen bond interactions with two water molecules that are in turn in close interaction with the protein (one with the backbone amide of E7, the other with the sidechain of R14 of a symmetry mate) and with two more labile water molecules, one of which interacts with the backbone amide of C6 (Fig. 4.10).

The protein-matrix interaction

A morphological inspection of the composite at SEM (Fig. 4.12 and Fig. 4.7, brown path) reveals the presence of spheroidal aggregates with diameter of around 0.2 μm.

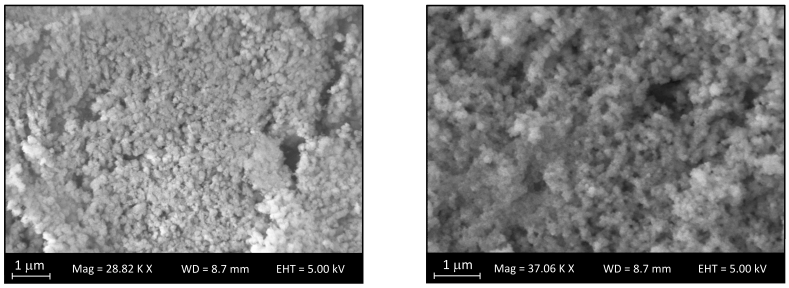


Figure 4.12: *SEM images of the silica-lysozyme composite with magnification factor 28.82 KX (left) and 37.06 KX (right), respectively.*

The analysis of the silica-lysozyme interactions started with the acquisition of the ^{1}H NMR spectra of HEWL (green spectrum in Fig. 4.13, top panel, Fig. 4.7, pink path). As in a typical folded protein, the signals are distributed in the range between 12 and -1 ppm, and the exchange with D_2O causes the attenuation of the amide signals (red spectrum in Fig. 4.13,

top panel, Fig. 4.7, purple path). The solid-state ^{1}H NMR spectrum of the dry composite presents, around the water signal, two shoulders in the same ppm regions, which confirm the presence of lysozyme in the composite (Fig. 4.13, bottom panel, Fig. 4.7, black path).

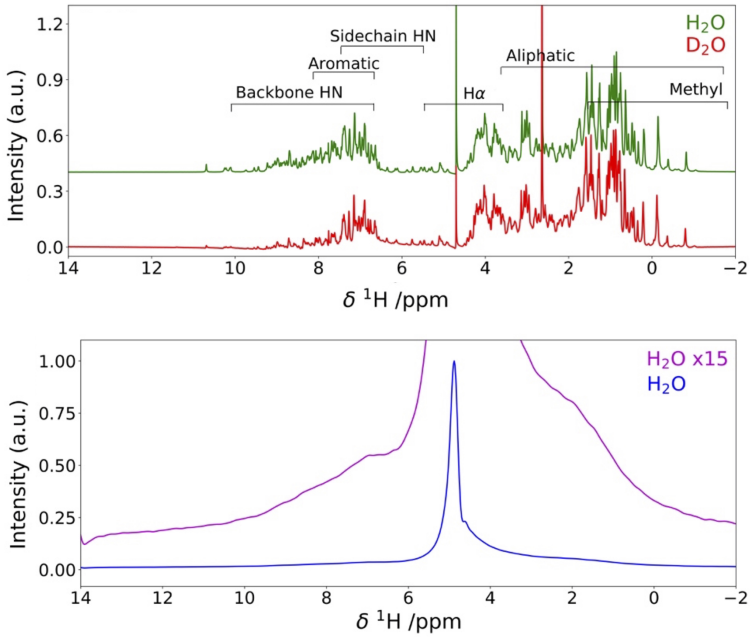


Figure 4.13: *^{1}NMR spectra of lysozyme solution 10 mg/mL in a mixture H_2O/D_2O 90:10 (green) and 98% D_2O (red). The assignments are taken from [278] (top panel). ssNMR spectrum of the silica-lysozyme composite (blue) and its 15 fold magnification (magenta) (bottom panel).*

Since the spread of the ^{13}C NMR signals is a reporter of the folding of the protein [279, 280], ^{1}H-^{13}C cross-polarization ssNMR spectra of the composite under different conditions (dry and rehydrated with H_2O), and of the pure lysozyme have been acquired (Fig. 4.14, red, black, and ma-

genta spectrum, respectively, and Fig. 4.7, black, green, and red path, respectively). As typical for dry protein samples, the resolution of the protein signals is low due to structural heterogeneity. Comparison with the spectrum of the sedimented lysozyme shows that the broader envelope of the composite spectrum encompasses the frequency range that is observed for the sedimented protein. The distinctive features of the folded protein are observed, i.e. signals from the methyl-bearing sidechains, which have shifts lower than 20 ppm when the hydrophobic core of the protein is intact, as well as signals in the overall distribution of the aliphatic region (65 to 5 ppm). This result confirms previous reports [232, 281] and is in contrast with the idea that lysozyme loses its tertiary structure when the silica matrix is formed [282].

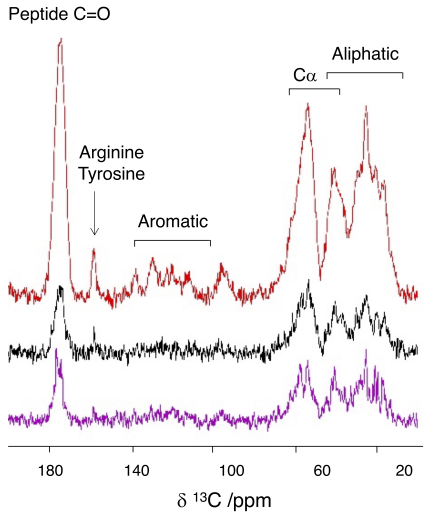


Figure 4.14: *^{1}H-^{13}C CP ssNMR spectra of the composite (freeze dried and rehydrated in red and black, respectively) and of the pure lysozyme sediment (magenta).*

The silica matrix was then characterized. The ^{29}Si NMR spectra show signals originating from siloxanes (Q^4, around -110 ppm), single silanols (Q^3, around -100 ppm), and geminal silanols (Q^2, around -90 ppm), the latter being barely visible (Fig. 4.15). The ratio Q^4/Q^3 of 0.51 is indicative of a high degree of condensation of the silica matrix [231, 283].

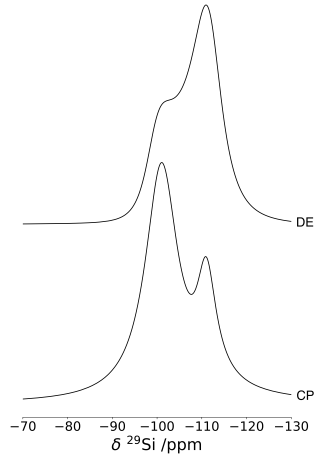


Figure 4.15: *^{29}Si direct excitation (DE) and cross-polarization (CP) ssNMR spectra of the composite.*

Only monodimensional spectra, however, do not provide information about the protein-silica interactions. In order to evaluate if the carbon nuclei, only present in the protein, and the silicon nuclei, only present in the silica matrix, are getting polarized by the same proton sources, a direct correlation of ^{13}C and ^{29}Si signals would be needed. However, being both systems in natural abundance, the sensitivity is limited, thus the ^{1}H traces

have been used to discriminate among the protons that act as polarization sources for the heteronuclear sites. 2D ^{1}H-^{29}Si and ^{1}H-^{13}C HETCOR spectra have been acquired, with wPMLG ^{1}H homonuclear decoupling [284] during the evolution of the indirect ^{1}H dimension and CP for polarization transfer (Fig. 4.16).

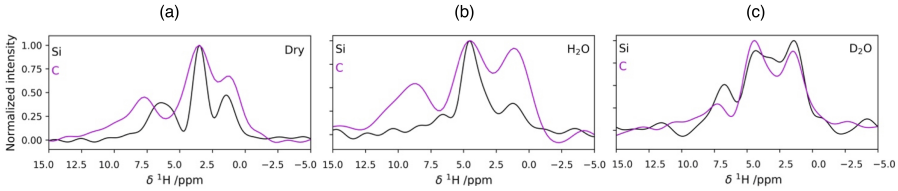


Figure 4.16: *^{1}H traces extracted from the 2D HETCOR ^{1}H-^{13}C (magenta) and ^{1}H-^{29}Si (black) spectra of the dry (panel a), rehydrated with H_2O (panel b), and with D_2O (panel c) sample.*

All the spectra are rather broad and show cross-polarization peaks around 1, 4, and 8 ppm, coming from at least three proton sources. While the first two are consistent with the signals commonly observed for silica, the third could report for the spin diffusion of lysozyme towards silica. At low contact times, the ^{1}H traces of the HETCOR experiments are markedly different: as expected, ^{29}Si is receiving magnetization mainly from silanols, whereas ^{13}C receives magnetization from the protein protons. However, features at similar chemical shifts tend to appear at longer mixing times. In particular, a signal around 7 ppm appears in the traces of the ^{29}Si spectra. This signal disappears when the protein is quantitatively removed by the GnHCl/DTT treatment (Fig. 4.7, magenta path), but all the other conditions remain the same, making it unlikely that this signal corresponds to protonated silanols. This implies that magnetization in transferred (or, equivalently, relayed)

from either aromatic sidechains, from backbone amides, or from the amine and guanidino sidechains of lysine and arginine residues. However, other peaks at lower frequency in the ^{1}H dimension are not observed, unlike other silica-peptide preparations, where clear indication of covalent interaction has been provided [269]. If D_2O is used for rehydration (Fig. 4.7, blue path), the ^{1}H peak corresponding to amides is removed in the traces of both the ^{13}C and ^{29}Si spectra. Collectively, these observations seem to suggest a non-covalent interaction in this sample.

The EPR spectrum of the composite containing labeled lysozyme (Fig. 4.7, grey path) shows the presence of both free and immobilized protein (Fig. 4.17).

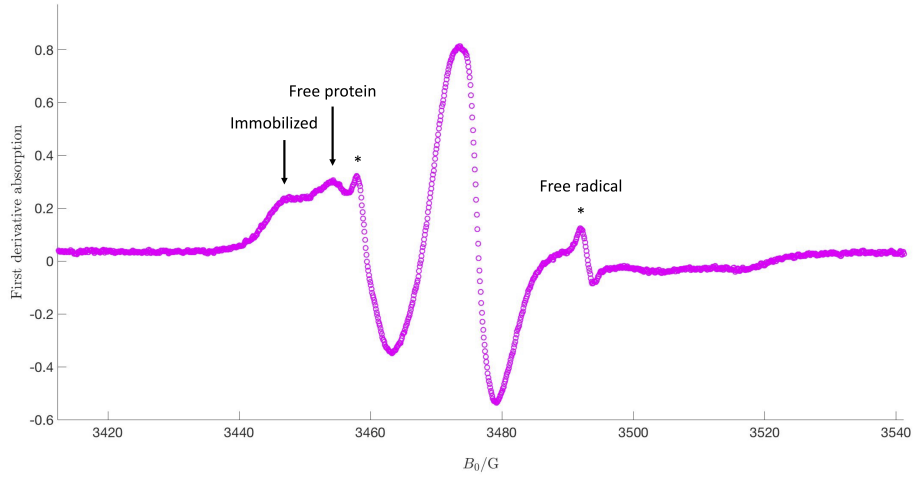
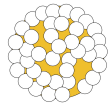
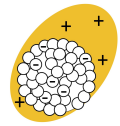
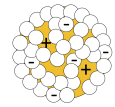


Figure 4.17: *EPR spectrum of the composite containing labeled lysozyme (Fig. 4.7, orange path).*

The possible silica-lysozyme relative arrangements are depicted in Fig. 4.18. The hypothesized models represent the three limiting cases: lysozyme is encapsulated due to steric hindrance, it electrostatically interacts with the surface or from the inside.

Conditions	Expected behavior for the proposed models			Observed behavior
	Steric inside 	Electrostatic outside 	Electrostatic inside 	
Water	✓	✓	✓	✓
NaCl: high ionic strenght	✓	✗	✓	✓
Urea + DTT: denaturing, riducing, low ionic strenght	✗	✓	✓	✗
GdnHCl + DTT: denaturing, riducing, high ionic strenght	✗	✗	✗	✗

Matrix — The protein remains in the composite
Protein — The protein is quantitatively released

Figure 4.18: *Expected and observed behavior of the composite with respect to ionic strength and denaturation.*

Being impossible to explain the system with any of these models, nor their combinations, in the presence of only water, the composite was tested in three different conditions through the addition of different additives. The concentration of the protein in the supernatant was measured by UV-Vis spectroscopy before and after the additions (Fig. 4.7, magenta, yellow, and cyan paths). In a high ionic strength environment, the protein is expected to be quantitatively released only in the presence of electrostatic superficial

interactions, because the 3D arrangement due to the folding prevents the protein passing through the matrix pores. Since after the addition of NaCl 6 mol/dm^3 lysozyme remained in the composite, the possibility of mainly superficial interactions is excluded. The composite was then tested with two denaturing and reducing agents which differ in ionic strength. They are both expected to cause the loss of the 3D protein structure, that can thus leak out of the composite through the matrix pores, but their effect in the electrostatic-based models is opposite. The electrostatic interaction is expected to be significantly perturbed only in case of high ionic strength, thus causing the quantitative release of protein. Since HEWL was quantitatively released by adding both urea 8 mol/dm^3 + DTT 50 mmol/dm^3, and GdnHCl 6 mol/dm^3 + DTT 50 mmol/dm^3, it is possible to conclude that the interaction occurring between silica and lysozyme is mainly due to steric hindrance.

4.4 Multivariate Curve Resolution for 2D ssNMR

4.4.1 Introduction

Denoising methods are aimed at increasing the amount of information that can be extracted from noise-contaminated signals. Several techniques have been proposed and are currently under investigation [223, 285–289]. The approach here presented makes use of the Multivariate Curve Resolution (MCR) [290] and Cadzow [291] methods.

Multivariate Curve Resolution

MCR is a chemometric method commonly used to recover pure components information from data on complex mixtures [292–294]. Initially developed for UV-Vis spectroscopy [295], it has been then extended to a plethora of different applications [296–300], including the deconvolution of 2D solution NMR data of reaction mixtures [301]. The method is based on the assumptions that:

(i) the outcome of the experiment is a continuous curve resulting from the linear combination of nonnegative and linearly independent pure component functions;

(ii) the contribution in the signal coming from each pure component is proportional to its concentration in the sample [302, 303].

Mathematically, it corresponds to decomposing the experimental data matrix $\mathbb{D}$ into the product of a concentration matrix $\mathbb{C}$ and a pure signal (usually a spectrum) matrix $\mathbb{S}$, leaving behind a residual matrix $\mathbb{E}$:

$$\mathbb{D} = \mathbb{C}\mathbb{S}^{\mathrm{T}} + \mathbb{E} \tag{4.1}$$

Such factorization is obtained through the following steps:

(i) Determination of the number of components. It can be known beforehand or determined with methods which typically make use of the Singular Value Decomposition (SVD), such as principal component analysis or purest variables algorithm.

(ii) Generation of the initial guess for $\mathbb{C}$ or $\mathbb{S}$ matrix. Initial estimates can be based on previous knowledge, e.g. spectra of components known to be present in the mixture, or can be obtained with purest variables selection methods.

(iii) Iterative optimization of the $\mathbb{C}$ and $\mathbb{S}$ matrices, through the minimization of the residual matrix $\mathbb{E}$. At each step $\mathbb{C}_k$, $\mathbb{S}_k$, and $\mathbb{E}_k$ are calculated as follows:

$$\mathbb{S}_k^T = \mathbb{C}_{k-1}^+ \mathbb{D} \tag{4.2}$$

$$\mathbb{C}_k = \mathbb{D}\mathbb{S}_k^{T+} \tag{4.3}$$

$$\mathbb{E}_k = \mathbb{D} - \mathbb{C}_k \mathbb{S}_k^T \tag{4.4}$$

where $^+$ denotes the Moore-Penrose pseudoinverse. A very common procedure is MCR-ALS, which makes use of the Alternating Least-Square (ALS) algorithm [303].

Cadzow filtering

The Cadzow algorithm is a 1D signal denoising method based on the relationship between low singular values and noise components. The complex vector $\boldsymbol{y} = \boldsymbol{y'} + \boldsymbol{e}$, sum of the signal $\boldsymbol{y'}$ and the noise $\boldsymbol{e}$, is divided in two segments of length l and k (where $l + k = N$), in order to construct a

Hankel-type matrix:

$$\mathbb{H} = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_k \\ y_2 & y_3 & \cdots & \cdots & y_{k+1} \\ y_3 & \cdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_l & y_{l+1} & \cdots & \cdots & y_N \end{bmatrix}. \tag{4.5}$$

$\mathbb{H}$ is decomposed through the SVD, $\mathbb{H} = \mathbb{U}\mathbb{\Sigma}\mathbb{V}^*$, and the lowest singular values are put to zero. The so modified $\mathbb{\Sigma}'$ matrix is then used to reconstruct a new matrix $\mathbb{J} = \mathbb{U}\mathbb{\Sigma}'\mathbb{V}^*$, which is, however, no longer a Hankel-type. In order to have again a matrix of this kind, each element of an antidiagonal is replaced by the average value along the antidiagonal itself. The first and the last row and column of the final $\mathbb{H}'_k$ are then put back together to reconstruct the denoised $\boldsymbol{y}'$ vector.

The MCR-Cadzow method

The proposed procedure consists in the application of the MCR method on the FIDs of 2D ssNMR data, instead of the signals, thus allowing for further denoising. In the presented case, the MCR-processed spectra have been subjected to the Cadzow filter. In addition, it is possible to coprocess multiple 2D spectra containing signals at the same frequencies but with different intensities. This is particularly relevant in the case of spectra that are acquired by varying some experimental parameters, e.g. the mixing time, which impact the overall sensitivity, causing some of the spectra of the series to have a significantly lower signal-to-noise (S/N) ratio [304].

4.4.2 Methods

The algorithm has been implemented using python2.7, exploiting the class NMR PIPE, contained in the NMRGlue library [305]. The number of components is set to 4, with 10% of the most intense signal as threshold for the noise, and the optimization stops when $\|\mathbb{X}_{\mathrm{k}} - \mathbb{X}_{\mathrm{k\text{-}1}}\|_2 \leq 10^{-5}$ (where $\mathbb{X} = \mathbb{C}, \mathbb{S}, \mathbb{E}$). The initial guess of the $\mathbb{C}$ matrix was generated through the purest variables algorithm [306]. The processing parameters are reported in Table 4.2. The S/N ratio was estimated dividing the maximum signal intensity by the standard deviation of the noise (σ_n), calculated as follows:

$$\sigma_n = \frac{1}{\sqrt{N-1}} \sqrt{\sum_{i=-n}^{n} y(i)^2 - \frac{1}{N}\left[\left(\sum_{i=-n}^{n} y(i)\right)^2 + \frac{3}{N^2-1}\left(\sum_{i=1}^{n} i(y(i) - y(-i))\right)^2\right]} \quad (4.6)$$

where N is the total number of points in the noise region, $n = (N-1)/2$, and $y(i)$ is the i-th point in the noise region. As a representative noise region, a slice of the spectrum where no signal is present was selected.

Dimension	F2	F1
Nucleus	^{29}Si	^{1}H
TD	57	60
Apodization function	Sine bell	Bruker-style GM
	Power 2	LB /Hz -20
	Offset 0.35π	GB 0.1
Zero-filling	2048	512

Table 4.2: *Processing parameters of the 2D spectra [304].*

Solid-state NMR experiments were recorded on a Bruker Avance II spec-

trometer operating at 700 MHz ^{1}H Larmor frequency (16.4 T), corresponding to 139 MHz ^{29}Si Larmor frequency, and equipped with a 3.2 BVT MAS probehead in double resonance mode. The pulse lengths are 2.4 and 4.7 μs for ^{1}H and ^{29}Si, respectively. Cross-polarization was achieved by matching the k=1 Hartmann-Hahn condition [307]. The spectral windows for the different nuclei were 60 and 249 ppm for ^{1}H and ^{29}Si, respectively. During the ^{1}H magnetization evolution under the chemical shift in the indirect dimension of HETCOR experiments, the PMLG decoupling sequence was used to suppress ^{1}H-^{1}H dipolar couplings [258, 284]. Spectra were acquired with CPMG echo train acquisition [222], and then the echoes were coadded.

Synthetic data were generated over the same spectral window as the experimental data. The spectra in the series of three comprise of up to nine cross-peaks of variable intensities and linewidths in the indirect dimension (Table 4.3).

F2		F1		Intensity		
PP /ppm	LW /Hz	PP /ppm	LW /Hz	0.5 ms	6 ms	10 ms
-89.8	1570	7.50	4650	2	1	1
-89.8	1570	4.75	3142	1	1	1
-89.8	1570	0.20	3142	0	0	0.5
-99.8	1570	7.50	4650	3	3	3
-99.8	1570	4.75	3142	2	3	3
-99.8	1570	0.20	3142	2	3	1
-109.8	1570	7.50	4650	0	0	1
-109.8	1570	4.75	3142	0.5	2	2
-109.8	1570	0.20	3142	1	2	2

Table 4.3: *Position (PP), linewidth (LW), and intensity of the cross-peaks in the synthetic spectra [304].*

4.4.3 Results and discussion

The MCR algorithm has been adapted to the 2D ssNMR problem: in the $\mathbb{C}$ matrix the concentrations are replaced by the time evolutions in the indirect dimension, while in the $\mathbb{S}$ matrix the pure spectra components are substituted by the FIDs in the direct dimension. This method is applicable regardless of the line shape of the peaks, therefore no forward model (e.g.: using Gaussians for modeling the peaks), nor any regularization (e.g.: non-negativity of the spectra in the frequency domain, or smoothness), for either the FIDs or evolutions, needs to be applied.

Denoising on synthetic spectra

The method was first tested on a single synthetic spectrum, to which gaussian noise was added, yielding a S/N ratio of 46. The denoising increases the S/N ratio to 221 and 237 after, respectively, MCR and MCR+Cadzow denoising, without altering the relative intensities of the peaks (Fig. 4.19).

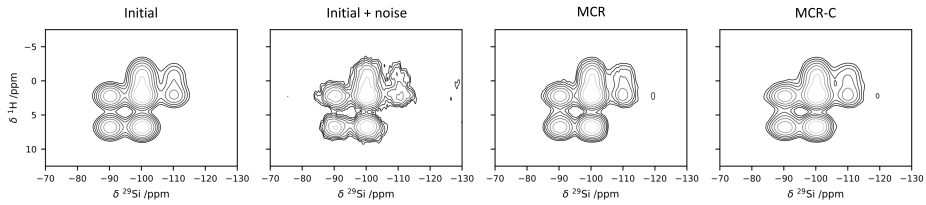


Figure 4.19: *Synthetic spectrum without and with noise, result of the processing of the spectrum with MCR and with MCR+Cadzow denoising [304].*

The test on multiple synthetic spectra shows no alterations in the shape and the relative intensities of the peaks and a quite large improvement in the S/N ratio (Table 4.4, Fig. 4.20). Interestingly, in contrast to the spectra

with 6 and 10 ms of mixing time, the S/N ratio for the spectrum with 0.5 ms improves more when it is coprocessed rather than individually processed.

	Spectrum					
	0.5 ms		6 ms		10 ms	
Step	Single	Coprocessed	Single	Coprocessed	Single	Coprocessed
Initial S/N	45	-	45	-	47	-
MCR S/N	106	206	221	150	250	153
MCR-C S/N	130	253	237	191	299	195

Table 4.4: *S/N ratio in the different spectra across the different processing steps [304].*

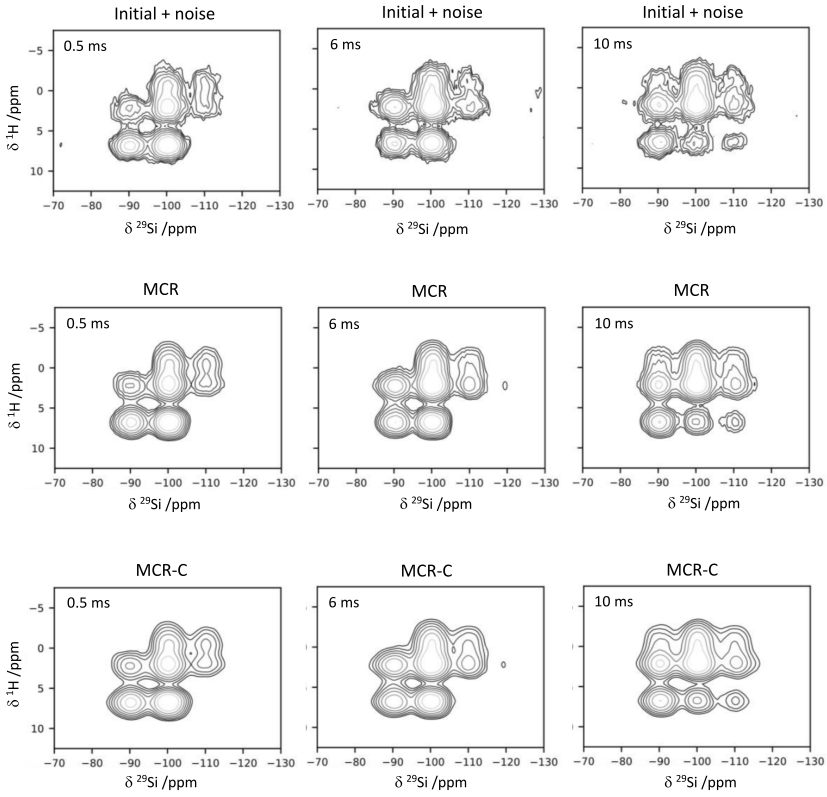


Figure 4.20: *Synthetic spectra at different mixing times (0.5, 6, and 10 ms) with gaussian noise, coprocessed with MCR and with MCR+Cadzow [304].*

Denoising on experimental spectra

The quantitative information on the synthetic spectra appears to be preserved across the denoising steps, therefore the same procedure has been applied to experimental datasets (Fig. 4.21), acquired on the silica-lysozyme composite (Section 4.3). The S/N greatly increases from 9 to 32 and 99 after processing with MCR and MCR+Cadzow, respectively.

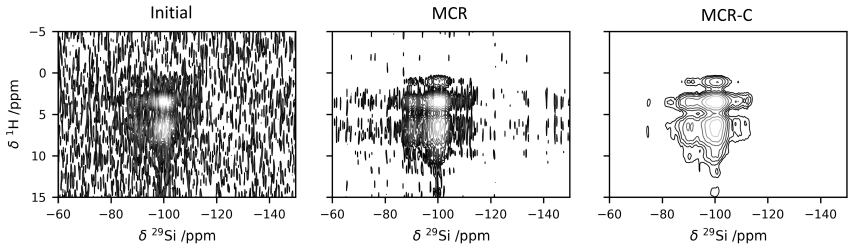


Figure 4.21: *Effect of denoising on a low S/N experimental spectrum: initial, reconstructed with MCR, and with MCR+Cadzow denoising [304].*

The noise properties have been also evaluated during the denoising phases and it remains gaussian (Fig. 4.22). From initial to processed spectra, the noise distribution width is greatly decreased (i.e. FT of the $\mathbb{C}\mathbb{S}^T$ matrix) in the "signal" spectra, while remains the same in the "error" spectra (i.e.: the FT of the $\mathbb{E}$ matrix), indicating that no part of the signal is discarded into the noise.

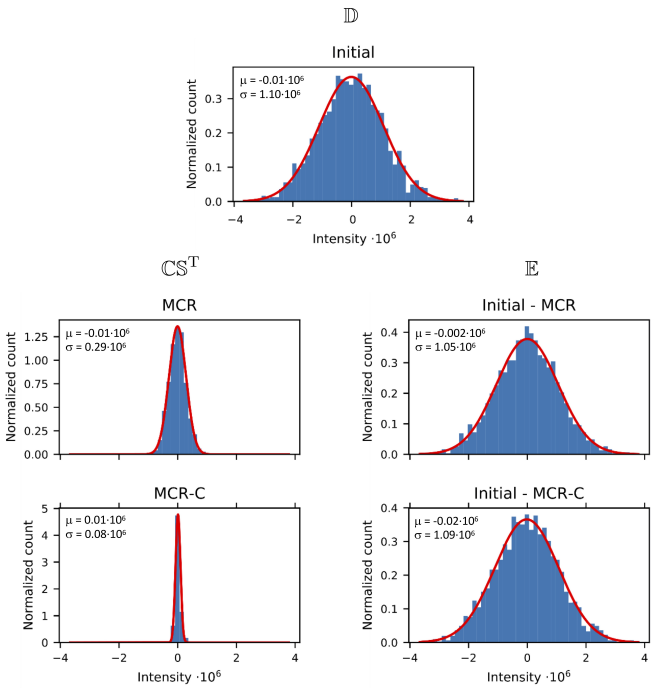


Figure 4.22: *Distribution of the intensities along the points of the row of the spectrum extracted at 50 ppm, which contains only noise, representing the noise in the initial spectrum (initial) and after application of combined denoising. Intensity in the "signal" spectrum* $\mathbb{C}\mathbb{S}^T$ *(left) and in the "noise" spectrum* $\mathbb{E}$ *(right). The red curves are the best fitting gaussian distributions that approximate the intensities [304].*

The combined denoising on a set of experimental spectra has been finally performed (Fig. 4.23). The MCR processing increases the S/N ratio from 10, 43, and 60 to 26, 119, and 160, and the subsequent application of Cadzow denoising further increases it to 28, 123, and 177.

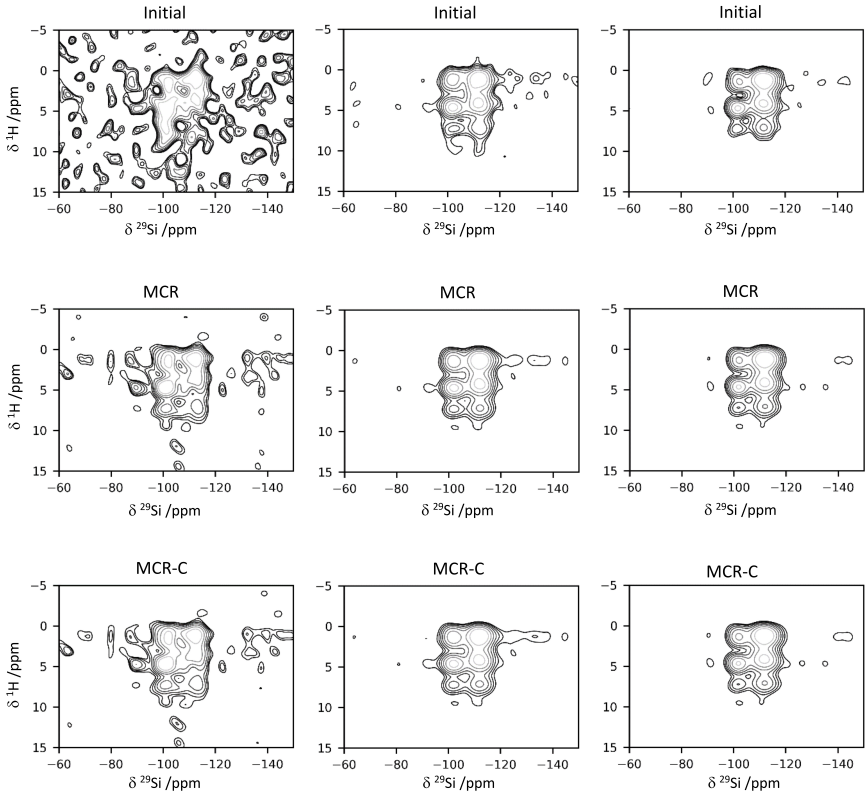


Figure 4.23: *Experimental spectra at different mixing times (0.5, 6, and 10 ms), processed with MCR and with MCR+Cadzow [304].*

MCR is ideally versed toward processing of 2D spectra, handling several transients at the same time, whereas other denoising methods, like Cadzow denoising, work on single transients. The order of application of the different denoising schemes on the reconstruction has been inverted (Fig. 4.24), showing that, while the impact of MCR and Cadzow yields similar improvements in S/N ratio (from 9 to 32 and 33, respectively), MCR is able

to improve the Cadzow procedure when applied first (99 for MCR+Cadzow and 80 for Cadzow+MCR).

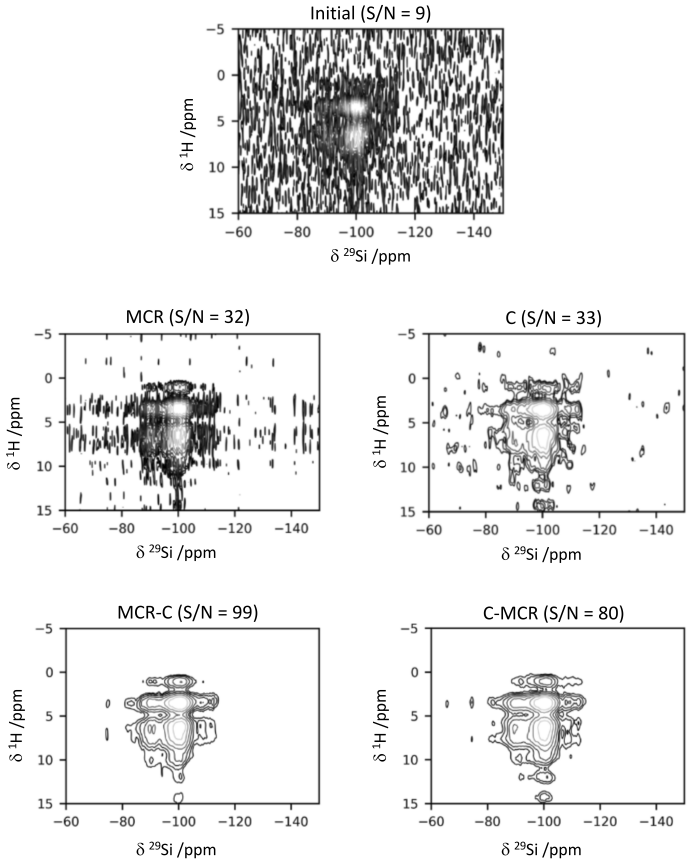


Figure 4.24: *Effect of changing the order of the denoising steps on the same low S/N experimental spectrum [304].*

Chapter 5

Conclusions and outlook

Part II demonstrates that ssNMR, supported by denoising methods and combined with other experimental techniques, is a powerful tool for the fine characterization of biomaterials, allowing for their rational design.

(i) The proposed catalyst, composed of SOD1 and cyt-c as protein components and titania as inorganic matrix, is easy to produce and the synthesis is highly reproducible. Preliminary tests on a series of composites containing different mutants of cyt-c demonstrate that all the samples are redox-active and able to efficiently oxidize aromatic alcohols. However, the complexity of the problem required the simplification of the system, thus SOD1, cyt-c, and titania have been substituted with easily treatable analogues: lysozyme and silica.

(ii) The study of the different stages of the composite formation started with the analysis of the interactions occurring among lysozyme and the silica and titania precursors. While NMR titrations give not useful

hints about the lysozyme-precursor interactions, X-ray crystallography shows that, unlike tetraoxosilicic(IV) acid, $Ti(OH)_4$, among the possible precursors of titania suggested in the literature, interacts with lysozyme before precipitation starts. The interaction is electrostatic in nature and can be mediated by intervening water molecules. This first structure-based evidence can be a proxy of those interactions that drive the initial steps of the oxide formation and can be a starting point for detailed (e.g. computational [308–310]) studies of the structural basis of the bioinspired titania precipitation.

(iii) The analysis of the protein-matrix interactions in the whole composite has been performed only on the silica-lysozyme system, which is composed of spheroidal silica aggregates. The protein interacts with the matrix and the presence of these interactions does not alter the lysozyme shape enough to affect its folding. The fact that the protein can only be detached from the matrix by denaturation in a high ionic strength environment suggests that it is mainly trapped inside silica due to steric hindrance.

(iv) Denoising methods allow the acquisition of data even on nuclei with low gyromagnetic ratio and/or low natural abundance. In particular, the MCR method demonstrates to be successful in reducing the noise of 2D ssNMR spectra still preserving the quantitative information and the gaussian nature of the noise. The presented algorithm has also the intrinsic ability to coprocess, allowing the reconstruction of the spectra with the lowest S/N ratio across a series. Furthermore, as an innovation, the method is applied on time-domain-data, preserving the possibility to apply further processing, such as the Cadzow

filtering, which leads to a further increase in the S/N ratio.

The presented experiments show encouraging results, but the technological goal of the proposed device is still to achieve. The silica-lysozyme composite requires additional measurements, such as matrix porosity through Brunauer-Emmett-Teller (BET) or relaxometry, and investigations about its possible role also as a contrast agent for Magnetic Resonance Imaging are planned. Then, the photoactive complex will be characterized at the atomic-level and tests on samples of polluted water will be performed. Regarding the denoising algorithm, it can be clearly optimized and made more user-friendly, but, more interestingly, it can be extended to other NMR cases, such as other low receptive nuclei, solution NMR, multidimensional experiments acquired on static samples [311–316], and, beyond NMR, to datasets of higher dimensionality and non-uniformly sampled data [317–322].

Chapter 6

Appendices

6.1 Experimental techniques

Residual Dipolar Couplings

Solution NMR of proteins is influenced by their fast reorientations that, in the presence of a magnetic field, result in either an isotropic or anisotropic distribution of all molecular orientations. In the first case, the dipolar coupling, which is traceless, averages out, resulting in a simplification of the NMR spectrum but also in a loss of orientational information. In the second case, which originates from partial alignment, the dipolar coupling does not average to zero anymore and produces the so-called *Residual Dipolar Coupling* ($\Delta\nu^{RDC}$):

$$\Delta\nu^{RDC} = -\frac{B_0^2\gamma_A\gamma_B\hbar}{4\pi 15 k_B T r_{AB}^3}\left[\Delta\chi_{ax}(3cos^2\alpha - 1) + \frac{3}{2}\Delta\chi_{rh}sin^2\alpha cos\beta\right], \tag{6.1}$$

where $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are defined as in eq. 1.2, α is the angle between the AB vector and the z-axis of the $\Delta\chi$ tensor, β is the angle which describes the position of the projection of the AB vector on the xy plane of the $\Delta\chi$ tensor, relative to the x-axis.

This coupling has been extensively used in structure determinations of biomolecules as orientational restraints [178–181, 323].

Small Angle X-ray Scattering

Small Angle Scattering (SAS) is a scattering technique which provides information about size, shape, and orientation of materials in solution and in the solid state. The sample is irradiated with visible light (Small Angle Light Scattering, SALS), X-rays (Small Angle X-ray Scattering, SAXS), or neutrons (Small Angle Neutron Scattering, SANS), and the intensity I of the scattering part of its response is revealed by a detector. $I(\boldsymbol{q})$ contains information about dimensions, shape, and the interaction between the scattering centers present in the sample and has the form:

$$I(\boldsymbol{q}) = kNpV_p^2(\Delta\rho)^2P(\boldsymbol{q})S(\boldsymbol{q}) + B, \tag{6.2}$$

where k depends on the instrument, N_p and V_p are, respectively, the numerical density and the volume of the scattered particles, $\Delta\rho$ is the contrast, $P(\boldsymbol{q})$ and $S(\boldsymbol{q})$ are said, respectively, shape and structure factors, $\boldsymbol{q}$ is the scattering vector, and B is a linear offset depending on the noise and the incoherent scattering. SAXS, in particular, is sensitive to inhomogeneities in the electron density and analyzes only the scattering intensity at small angles.

Double Electron Electron Resonance

Double Electron Electron Resonance (DEER) is an electron paramagnetic technique technique able to measure distance distributions between paramagnetic centers in the range from 1.8 to 6 nm [324]. The interactions between couples of electron spins are observed in the time domain as a decay of the signal $V(t)$ due to transverse electron spin relaxation. This time evolution signal is given by the sum of single couples contributions $V_i(t)$:

$$V_i(t) = V_0 \prod_{k \neq i} [1 - \lambda_{ik}(1 - cos(\omega_{ik} t))], \tag{6.3}$$

where the modulation depth λ_{ik} quantifies the fraction of signal due to excited spin pairs (i, k) and ω_{ik} is the dipolar evolution frequency. The latter is defined as:

$$\omega_{ik}^{d} = \frac{\mu_0}{4\pi\hbar} \frac{g_i g_k \mu_B^2}{r_{ik}^3} (3cos^2\theta_{ik} - 1), \tag{6.4}$$

where g_i and g_k are the electron g values of the two spins, θ_{ik} is the angle between the static field vector $\boldsymbol{B}_0$ and the vector connecting the positions of the two spins, and r the distance between the two spins [325].

The time evolution signal is then separated into a background factor and a form factor, which contains the distance distribution [326].

X-ray Diffraction

X-ray diffraction is a scattering technique used to analyze the long-range structure in crystals. The result of an XRD experiment is a diffraction pattern reporting the intensity as a function of the angle between the ra-

diation and the detector. The angles are then converted in distances using the Bragg's law:

$$n\lambda = 2dsen\theta, \tag{6.5}$$

where n is the diffraction order, λ the wavelength of the incident radiation, d the interplanar distance, and θ the angle between the radiation and the atomic plane.

6.2 Supporting Information

Shifts' contributions of the PCS-FCS refinement

FC contributions	PC contributions	Diamagnetic contributions	Total calculated shifts	Experimental shifts
-12.55509348	-5.983914992	7.258	-11.28100848	-12.43
27.41882505	-3.687380359	7.478	31.20944469	25.33
1.176805777	-1.855071039	6.73	6.051734738	2.4
5.134471067	3.090915801	7.434	15.65938687	15.31
453.7298351	28.18105983	9.083	490.993895	469.2
58.71286404	68.07320269	7.374	134.1600667	121
196.2192461	40.84759994	4.265	241.3318461	233.5
56.37424364	70.84376017	7.437	134.6550038	121
172.5482102	41.43222923	4.327	218.3074394	200.2
-15.75570538	-5.57329558	7.251	-14.07800096	-13.95
28.55065736	-3.444520741	7.488	32.59413662	26.08
-3.208107469	-1.626056186	6.729	1.894836345	1.43
7.503073778	3.340124196	7.423	18.26619797	14.85
423.4476992	28.15100918	9.102	460.7007084	446.3
56.38173922	-40.88945058	2.143	90.32376928	79.6
-9.99160574	9.435379851	2.129	1.572774111	0.64
-7.135790446	9.851274293	2.144	4.859483847	4.84
63.86232626	-36.51603006	3.006	30.3522962	27
257.1882792	-21.94372252	1.86	237.1045567	226.5
244.6556724	-21.36715463	1.598	224.8865178	200.2
40.51360017	-26.12254601	2.259	16.65005416	16.37
-6.401223782	3.149034825	2.381	-0.871188958	-0.64
-13.80685504	10.69012261	2.06	-1.05673243	-2.29

Table 6.1: *FC, PC, and diamagnetic contributions to the total QC-calculated shifts and experimental shifts for NiSAL-MeDPT.*

FC contributions	**PC contributions**	**Diamagnetic contributions**	**Total calculated shifts**	**Experimental shifts**
-17.98026364	-5.559917616	7.288	-16.25218125	-14.07
27.57023744	-2.890169993	7.527	32.20706745	25.3
-7.1456177	-1.306469719	6.774	-1.678087419	-2.4
8.765191072	2.548127062	7.463	18.77631813	14.9
430.2591611	21.73736125	9.034	461.0305223	446.9
61.52879207	44.07344722	6.905	112.5072393	116.5
190.9447013	28.84425236	4.184	223.9729537	229.8
72.49840246	46.26590629	7.09	125.8543087	110.9
166.733579	29.64801499	4.112	200.493594	187
-15.43093518	-5.207556208	7.218	-13.42049139	-13.52
28.49999253	-2.59952338	7.505	33.40546915	26.2
-2.421862033	-0.946998437	6.728	3.359139531	2.4
7.070637451	3.145814868	7.409	17.62545232	15
422.2212784	21.28269543	9.063	452.5669738	438.9
-8.712704898	5.779128446	1.566	-1.367576452	-1.92
-7.850432038	6.870583581	2.665	1.685151543	-0.82
97.58679367	-27.83979219	2.393	72.14000148	79.9
304.4423038	-15.48263005	2.057	291.0166737	287.9
296.2994487	-13.61643104	2.175	284.8580177	258.5
69.65665103	-14.05416096	2.028	57.63049007	35.2
-12.57418771	2.564421572	2.247	-7.762766133	-8.54
-15.10852011	8.940819514	-1.781	-7.948700597	-10.64
-265.9474441	-61.55143751	2.298	-325.2008816	-333.3

Table 6.2: *FC, PC, and diamagnetic contributions to the total QC-calculated shifts and experimental shifts for NiSAL-HDPT.*

Details of the calculations of the PCS-driven refinement method

For the initial and final structures and for each of the intermediate steps, the QC-calculated PCS at the CASSCF and NEVPT2 level are compared, and the reciprocal of the slope of the regression line is used as a scaling factor (b in Figures 6.1 to 6.4) for the subsequent step.

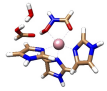
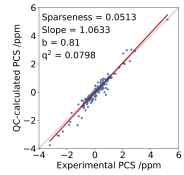
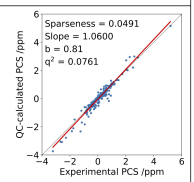
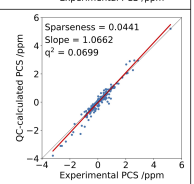
Geometry	Calculation type	ZFS tensor from effective H [cm^{-1}]	Electronic g'$_{ab}$-matrix	Electronic g-matrix from effective H	Magnetic susceptibility tensor [m^3] at T=298K
Energy = -2580.323 E_h  C -1.915460 -2.631390 2.622600 C -0.828151 -2.163410 1.971020 N -2.898620 -1.672870 2.461370 C -2.449020 -0.702922 1.690880 N -1.193710 -0.964697 1.396830 H -3.835040 -1.735830 2.853340 H 0.167550 -2.582730 1.895640 H -3.011130 0.183887 1.406180 H -2.082730 -3.524180 3.220280 C 2.933990 -3.509120 2.015270 O 2.437100 -2.457260 2.474510 O 3.384440 -3.565640 0.881239 H 3.374080 -2.722430 0.376959 H 2.995660 -4.439090 2.608220 C 0.413710 -3.346300 -2.534510 C 0.704579 -2.660130 -1.413260 N -0.481858 -2.576360 -3.240260 C -0.710117 -1.450500 -2.574580 N -0.0152210 -1.482870 -1.446680 H -0.884924 -2.828190 -4.139340 H 1.350820 -2.932250 -0.589094 H -1.368090 -0.651422 -2.907330 H 0.741075 -4.318210 -2.896140 C -2.985200 2.119070 -2.084350 C -2.575270 1.066150 -1.337860 N -1.900400 2.951630 -2.194140 C -0.889375 2.452310 -1.515640 N -1.275800 1.307480 -0.986889 H -1.903210 3.842820 -2.684770 H -3.118480 0.177256 -1.029460 H 0.084993 2.918580 -1.394320 H -3.936820 2.351770 -2.556470 Co 0.000000 0.000000 0.000000 N 2.703960 0.440128 1.074400 O 1.981160 -0.658199 0.670629 O 1.014090 1.693940 0.581911 C 2.108170 1.608150 1.106030 H 3.638980 0.250820 1.430790 H 2.083820 -1.333110 1.442280 H 2.677790 2.455720 1.527960 O 3.479870 -1.019350 -1.707650 H 4.355050 -0.822075 -2.074680 H 3.031980 -1.477780 -2.433530	CASSCF(7,5) 10 quartets 11 doublets	D_{xx} = 34.386801 D_{xy} = -13.988609 D_{xz} = 4.303519 D_{yy} = -9.836008 D_{yz} = -7.941487 D_{zz} = 13.823749 D = -42.203527 E/D = 0.311898	-2.040834 0.828914 0.825504 -6.753479 -1.472281 -0.577230 -1.236063 -1.905706 1.376383 g-factors 1.478968 2.196173 7.327163 Isotropic average = 3.667435 g-shifts -0.523351 0.193854 5.324844 Isotropic average = 1.665116	2.133286 0.137132 -0.018043 0.135998 2.615757 0.070690 -0.025374 0.073924 2.342299 g-factors 2.090950 2.335698 2.664702 Isotropic average = 2.363783 g-shifts 0.088630 0.333379 0.662382 Isotropic average = 0.361464	χ_{xx} = 1.167885 ·10^{-31} χ_{xy} = 3.273446 ·10^{-32} χ_{xz} = -4.498660·10^{-33} χ_{yy} = 2.259749 ·10^{-31} χ_{yz} = 1.993645·10^{-32} χ_{zz} = 1.574805 ·10^{-31} $\Delta\chi_{ax}$ = 10.8·10^{-32} $\Delta\chi_{rh}$ = -4.95·10^{-32} α = -101.39° β = -77.44° γ = -104.54° 
	CASSCF(7,5) 10 quartets 40 doublets	D_{xx} = 33.602567 D_{xy} = -13.654272 D_{xz} = 3.706446 D_{yy} = -9.868109 D_{yz} = -7.736443 D_{zz} = 13.798176 D = -41.634867 E/D = 0.299870	1.991235 -0.942064 -0.806851 6.786878 -0.164484 1.632586 1.082082 -1.572407 1.777045 g-factors 1.446134 2.126813 7.378633 Isotropic average = 3.650527 g-shifts -0.556186 0.124494 5.376314 Isotropic average = 1.648207	2.139358 0.139072 -0.018427 0.136836 2.619150 0.068931 -0.025785 0.071902 2.348783 g-factors 2.096166 2.342878 2.668255 Isotropic average = 2.369100 g-shifts 0.093847 0.340558 0.665936 Isotropic average = 0.366780	χ_{xx} = 1.152788·10^{-31} χ_{xy} = 3.208419·10^{-32} χ_{xz} = -4.069902·10^{-33} χ_{yy} = 2.210742·10^{-31} χ_{yz} = 1.907518·10^{-32} χ_{zz} = 1.546262·10^{-31} $\Delta\chi_{ax}$ = 10.5·10^{-32} $\Delta\chi_{rh}$ = -4.80·10^{-32} α = -101.02° β = -77.60° γ = -104.72° 
	CASSCF(7,5)/ NEVPT2 10 quartets 40 doublets	D_{xx} = 27.394224 D_{xy} = -12.099522 D_{xz} = 3.661280 D_{yy} = -7.628004 D_{yz} = -6.931176 D_{zz} = 11.696675 D = -34.768384 E/D = 0.305245	-1.144102 1.902026 -0.776754 -5.407750 3.532763 1.533117 -0.106951 1.779951 1.837297 g-factors 1.445075 2.108590 7.044543 Isotropic average = 3.532736 g-shifts -0.557244 0.106271 5.042223 Isotropic average = 1.530417	2.118791 0.116270 -0.019305 0.113674 2.509703 0.056095 -0.023732 0.058495 2.285265 g-factors 2.081035 2.282328 2.550399 Isotropic average = 2.304587 g-shifts 0.078716 0.280009 0.548080 Isotropic average = 0.302268	χ_{xx} = 1.400960·10^{-31} χ_{xy} = 3.282665·10^{-32} χ_{xz} = -5.469369·10^{-33} χ_{yy} = 2.444699·10^{-31} χ_{yz} = 1.876312·10^{-32} χ_{zz} = 1.791225·10^{-31} $\Delta\chi_{ax}$ = 10.4·10^{-32} $\Delta\chi_{rh}$ = -4.95·10^{-32} α = -102.42° β = -77.97° γ = -105.08° 

Figure 6.1: *ORCA output parameters for structure 1.*

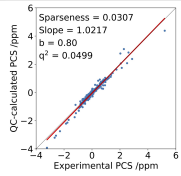
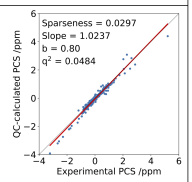
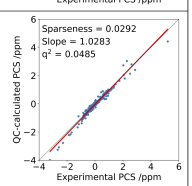
Geometry	Calculation type	ZFS tensor from effective H [cm^{-1}]	Electronic g'$_{eff}$-matrix	Electronic g-matrix from effective H	Magnetic susceptibility tensor [m^3] at T=298 K
Energy = -2580.322 E_h  C -2.021238 -2.523067 2.616517 C -0.907107 -2.086611 1.988806 N -2.986357 -1.558161 2.395470 C -2.499267 -0.615948 1.613440 N -1.238292 -0.900631 1.369120 H -3.935981 -1.598973 2.757458 H 0.085185 -2.519486 1.958688 H -3.040409 0.268845 1.285174 H -2.219473 -3.396391 3.233188 C 2.933990 -3.509120 2.015270 O 2.437100 -2.457260 2.474510 O 3.384440 -3.565640 0.881239 H 3.374080 -2.722430 0.376959 H 2.995660 -4.439090 2.608220 C 0.542575 -3.269720 -2.618970 C 0.794688 -2.608170 -1.473794 N -0.356065 -2.501125 -3.322277 C -0.623765 -1.399387 -2.631407 N 0.048682 -1.446985 -1.490508 H -0.734876 -2.737236 -4.236050 H 1.430200 -2.888135 -0.643932 H -1.293164 -0.606540 -2.956179 H 0.899606 -4.223707 -2.999745 C -2.991745 2.131282 -2.050638 C -2.577983 1.072259 -1.314974 N -1.907942 2.965471 -2.157882 C -0.893725 2.461008 -1.488004 N -1.277181 1.311423 -0.967472 H -1.913397 3.860840 -2.640820 H -3.119321 0.180391 -1.011916 H 0.080859 2.926862 -1.366827 H -3.945512 2.367401 -2.516689 Co 0.000000 0.000000 0.000000 N 2.674688 0.467327 1.066545 O 1.985549 -0.655907 0.672556 O 0.950878 1.665996 0.554762 C 2.045108 1.617799 1.083355 H 3.613533 0.309202 1.428043 H 2.105013 -1.319075 1.451931 H 2.588352 2.486140 1.497768 O 3.479870 -1.019350 -1.707650 H 4.355050 -0.822075 -2.074680 H 3.031980 -1.477780 -2.433530	CASSCF(7,5) 10 quartets 11 doublets	D_{xx} = 30.768906 D_{xy} = -16.394735 D_{xz} = 5.769324 D_{yy} = -6.199839 D_{yz} = -9.878461 D_{zz} = 13.164825 D = 41.207869 E/D = 0.321536	-2.072433 -1.279087 1.108293 -3.490174 -5.469537 -0.842185 0.809833 -2.694135 0.760311 g-factors 1.543297 2.385143 7.098407 Isotropic average = 3.675615 g-shifts -0.459023 0.382823 5.096088 Isotropic average = 1.673296	2.140825 0.161847 -0.038160 0.160823 2.545086 0.091783 -0.041928 0.094458 2.318345 g-factors 2.064984 2.318704 2.620570 Isotropic average = 2.334752 g-shifts 0.062664 0.316385 0.618250 Isotropic average = 0.332433	χ_{xx} = 1.167620 ·10^{-31} χ_{xy} = 3.676621·10^{-32} χ_{xz} = -7.573397·10^{-33} χ_{yy} = 2.057751·10^{-31} χ_{yz} = 2.335154·10^{-32} χ_{zz} = 1.506431·10^{-31} $\Delta\chi_{ax}$ = 9.98·10^{-32} $\Delta\chi_{rh}$ = -5.11·10^{-32} α = -72.21° β = -105.13° γ = -72.15° 
	CASSCF(7,5) 10 quartets 40 doublets	D_{xx} = 30.108963 D_{xy} = -16.101328 D_{xz} = 5.162598 D_{yy} = -6.336381 D_{yz} = -9.716360 D_{zz} = 13.226283 D = -40.150268 E/D = 0.330073	-2.598501 -0.009003 0.634951 -5.637872 3.084729 -1.495338 -1.651164 -0.398151 -2.309745 g-factors 1.504893 2.301209 7.166043 Isotropic average = 3.657382 g-shifts -0.497426 0.298890 5.163724 Isotropic average = 1.655062	2.146414 0.164589 -0.039045 0.162791 2.549342 0.091164 -0.042821 0.093680 2.324518 g-factors 2.068655 2.325736 2.625886 Isotropic average = 2.340092 g-shifts 0.066335 0.323417 0.623567 Isotropic average = 0.337773	χ_{xx} = 1.179314·10^{-31} χ_{xy} = 3.710387·10^{-32} χ_{xz} = -7.315717·10^{-33} χ_{yy} = 2.066093·10^{-31} χ_{yz} = 2.322364·10^{-32} χ_{zz} = 1.514635·10^{-31} $\Delta\chi_{ax}$ = 9.99·10^{-32} $\Delta\chi_{rh}$ = -5.09·10^{-32} α = -72.40° β = -105.02° γ = -71.92° 
	CASSCF(7,5)/ NEVPT2 10 quartets 40 doublets	D_{xx} = 24.644331 D_{xy} = -14.000468 D_{xz} = 4.987222 D_{yy} = -4.546860 D_{yz} = 4.987222 D_{zz} = 11.108135 D = 34.126373 E/D = 0.319628	-0.932401 -2.419286 0.632249 -4.832664 -3.695167 -1.386050 -0.289744 -1.530419 -2.373732 g-factors 1.540623. 2.346400 6.799538 Isotropic average = 3.562187 g-shifts -0.461697 0.344081 4.797219 Isotropic average = 1.559868	2.123716 0.134961 -0.034586 0.133053 2.447204 0.073627 -0.036943 0.075565 2.264224 g-factors 2.057923 2.267526 2.509695 Isotropic average = 2.278381 g-shifts 0.055604 0.265207 0.507376 Isotropic average = 0.276062	χ_{xx} = 1.429178·10^{-31} χ_{xy} = 3.731399·10^{-32} χ_{xz} = -8.676722·10^{-33} χ_{yy} = 2.289848·10^{-31} χ_{yz} = 2.276917·10^{-32} χ_{zz} = 1.759915·10^{-31} $\Delta\chi_{ax}$ = 9.76·10^{-32} $\Delta\chi_{rh}$ = -5.27·10^{-32} α = -71.36° β = -104.73° γ = 71.58° 

Figure 6.2: *ORCA output parameters for structure 2.*

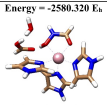
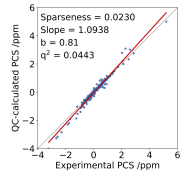
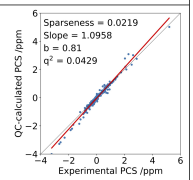
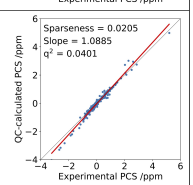
Geometry	Calculation type	ZFS tensor from effective H [cm^{-1}]	Electronic g'$_{eff}$-matrix	Electronic g-matrix from effective H	Magnetic susceptibility tensor [m^3] at T=298K
Energy = -2580.320 E_h  C -2.010981 -2.554813 2.578318 C -0.897624 -2.108726 1.956027 N -2.977620 -1.588674 2.369644 C -2.492106 -0.635883 1.599541 N -1.230734 -0.915505 1.351457 H -3.927122 -1.635531 2.731220 H 0.095328 -2.539655 1.920290 H -3.034663 0.252161 1.282582 H -2.207773 -3.436181 3.183907 C 2.933990 -3.509120 2.015270 O 2.437100 -2.457260 2.474510 O 3.384440 -3.565640 0.881239 H 3.374080 -2.722430 0.376959 H 2.995660 -4.439090 2.608220 C 0.549130 -3.305678 -2.590107 C 0.798339 -2.631894 -1.451446 N -0.347453 -2.544514 -3.304041 C -0.616741 -1.435351 -2.625783 N 0.052648 -1.470806 -1.482646 H -0.723878 -2.790367 -4.216229 H 1.431569 -2.903041 -0.616925 H -1.285089 -0.645910 -2.960854 H 0.906956 -4.263779 -2.959640 C -3.007329 2.104680 -2.040573 C -2.583584 1.047446 -1.308025 N -1.931367 2.949239 -2.145579 C -0.912372 2.452324 -1.477311 N -1.285005 1.297721 -0.960034 H -1.945252 3.845904 -2.625936 H -3.116532 0.149703 -1.007455 H 0.057850 2.926890 -1.354942 H -3.963339 2.333240 -2.505797 Co 0.000000 0.000000 0.000000 N 2.666640 0.466213 1.042418 O 1.987410 -0.656660 0.630594 O 0.922637 1.653941 0.575508 C 2.022963 1.607927 1.091491 H 3.611191 0.312044 1.390514 H 2.123455 -1.335415 1.393662 H 2.559717 2.473680 1.519550 O 3.479870 -1.019350 -1.707650 H 4.355050 -0.822075 -2.074680 H 3.031980 -1.477780 -2.433530	CASSCF(7,5) 10 quartets 11 doublets	D_{xx} = 31.988588 D_{xy} = -16.877618 D_{xz} = 5.367063 D_{yy} = -7.683876 D_{yz} = -9.371536 D_{zz} = 14.821828 D = -42.886501 E/D = 0.316509	-2.433765 0.397763 1.000405 -6.483917 -1.473503 -1.046468 -1.191851 -2.311328 0.607873 g-factors 1.465228 2.200179 7.248055 Isotropic average = 3.637821 g-shifts -0.537092 0.197860 5.245735 Isotropic average = 1.635501	2.134487 0.167066 -0.033484 0.165377 2.569005 0.085067 -0.037665 0.088739 2.305575 g-factors 2.062037 2.306423 2.640611 Isotropic average = 2.336357 g-shifts 0.059718 0.304104 0.638292 Isotropic average = 0.334038	χ_{xx} = 1.157487·10^{-31} χ_{xy} = 3.814571·10^{-32} χ_{xz} = -6.619363·10^{-33} χ_{yy} = 2.116846·10^{-31} χ_{yz} = 2.209139·10^{-32} χ_{zz} = 1.476783·10^{-31} $\Delta\chi_{ax}$ = 10.6·10^{-32} $\Delta\chi_{rh}$ = -4.88·10^{-32} α = -73.32° β = -103.10° γ = 72.18° 
	CASSCF(7,5) 10 quartets 40 doublets	D_{xx} = 31.324081 D_{xy} = -16.586088 D_{xz} = 4.739578 D_{yy} = -7.821932 D_{yz} = -9.223019 D_{zz} = 14.925004 D = -42.588692 E/D = 0.301143	1.242739 -2.266770 0.626947 0.310421 -6.630209 -1.415761 1.203442 -1.130593 -2.031002 g-factors 1.422014 2.111233 7.314928 Isotropic average = 3.616058 g-shifts -0.580306 0.108914 5.312609 Isotropic average = 1.613739	2.139849 0.169810 -0.034202 0.167250 2.573142 0.084376 -0.038373 0.087836 2.311544 g-factors 2.065678 2.313062 2.645799 Isotropic average = 2.341513 g-shifts 0.063359 0.310742 0.643480 Isotropic average = 0.339194	χ_{xx} = 1.168946·10^{-31} χ_{xy} = 3.848126·10^{-32} χ_{xz} = -6.320277·10^{-33} χ_{yy} = 2.125201·10^{-31} χ_{yz} = 2.195679·10^{-32} χ_{zz} = 1.484358·10^{-31} $\Delta\chi_{ax}$ = 10.7·10^{-32} $\Delta\chi_{rh}$ = -4.84·10^{-32} α = -73.56° β = -103.00° γ = 71.97° 
	CASSCF(7,5)/ NEVPT2 10 quartets 40 doublets	D_{xx} = 25.609124 D_{xy} = -14.411354 D_{xz} = 4.668132 D_{yy} = -5.776613 D_{yz} = -8.178779 D_{zz} = 12.512134 D = -35.439746 E/D = 0.317016	0.577733. 2.511475 -0.619554 -1.708499 6.044631 1.319881 0.710323 1.347090 2.114318 g-factors 1.458815 2.158725 6.944426 Isotropic average = 3.520655 g-shifts -0.543504 0.156406 4.942107 Isotropic average = 1.518336	2.118992 0.139491 -0.030930 0.137063 2.467652 0.068351 -0.033508 0.070896 2.254341 g-factors 2.055977 2.257772 2.527238 Isotropic average = 2.280329 g-shifts 0.053658 0.255452 0.524919 Isotropic average = 0.278010	χ_{xx} = 1.418921·10^{-31} χ_{xy} = 3.874341·10^{-32} χ_{xz} = -7.747841·10^{-33} χ_{yy} = 2.349304·10^{-31} χ_{yz} = 2.159927·10^{-32} χ_{zz} = 1.731248·10^{-31} $\Delta\chi_{ax}$ = 10.4·10^{-32} $\Delta\chi_{rh}$ = -5.03·10^{-32} α = -72.37° β = -102.74° γ = 71.62° 

Figure 6.3: *ORCA output parameters for structure 3.*

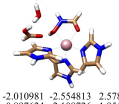
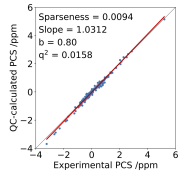
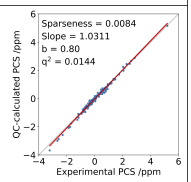
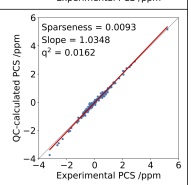
Geometry	Calculation type	ZFS tensor from effective H [cm^{-1}]	Electronic g'$_{eff}$-matrix	Electronic g-matrix from effective H	Magnetic susceptibility tensor [m^3] at T=298K
Energy = -2580.321 E_h  C -2.010981 -2.554813 2.578318 C -0.897624 -2.108726 1.956027 N -2.977620 -1.588674 2.369644 C -2.492106 -0.635883 1.599541 N -1.230734 -0.915505 1.351457 H -3.927122 -1.635531 2.731220 H 0.095328 -2.539655 1.920290 H -3.034663 0.252161 1.282582 H -2.207773 -3.436181 3.183907 C 2.933990 -3.509120 2.015270 O 2.437100 -2.457260 2.474510 O 3.384440 -3.565640 0.881239 H 3.374080 -2.722430 0.376959 H 2.995660 -4.439090 2.608220 C 0.549130 -3.305678 -2.590107 C 0.798339 -2.631894 -1.451446 N -0.347453 -2.544514 -3.304041 C -0.616741 -1.435351 -2.625783 N 0.052648 -1.470806 -1.482646 H -0.723878 -2.790367 -4.216229 H 1.431569 -2.903041 -0.616925 H -1.285089 -0.645910 -2.960854 H 0.906956 -4.263779 -2.959640 C -3.007329 2.104680 -2.040573 C -2.583584 1.047446 -1.308025 N -1.931367 2.949239 -2.145579 C -0.912372 2.452324 -1.477311 N -1.285005 1.297721 -0.960034 H -1.945252 3.845904 -2.625936 H -3.116532 0.149703 -1.007455 H 0.057850 2.926890 -1.354942 H -3.963339 2.333240 -2.505797 Co 0.000000 0.000000 0.000000 N 2.643533 0.460344 1.102168 O 1.940951 -0.668772 0.751068 O 0.958195 1.663211 0.483301 C 2.037894 1.622138 1.041545 H 3.569728 0.303572 1.495489 H 2.027399 -1.291273 1.567468 H 2.588833 2.500727 1.422879 O 3.479870 -1.019350 -1.707650 H 4.355050 -0.822075 -2.074680 H 3.031980 -1.477780 -2.433530	CASSCF(7,5) 10 quartets 11 doublets	D_{xx} = 29.077086 D_{xy} = -16.375222 D_{xz} = 6.099210 D_{yy} = -7.458469 D_{yz} = -7.570825 D_{zz} = 15.108390 D = -40.218605 E/D = 0.317793	-0.753408 2.130227 1.607666 0.227606 2.307524 6.275638 2.074309 0.959389 0.790688 g-factors 1.471437 2.197921 7.162581 Isotropic average = 3.610646 g-shifts -0.530883 0.195601 5.160262 Isotropic average = 1.608327	2.146786 0.160980 -0.042086 0.159076 2.546912 0.064968 -0.045814 0.068639 2.280107 g-factors 2.070916 2.292558 2.610334 Isotropic average = 2.324603 g-shifts 0.068597 0.290239 0.608015 Isotropic average = 0.322284	χ_{xx} = 1.183457 ·10^{-31} χ_{xy} = 3.696127 ·10^{-32} χ_{xz} = -8.728731·10^{-33} χ_{yy} = 2.066328·10^{-31} χ_{yz} = 1.680694·10^{-32} χ_{zz} = 1.430345·10^{-31} $\Delta\chi_{ax}$ = 9.93·10^{-32} $\Delta\chi_{rh}$ = -4.54·10^{-32} α = -71.15° β = -99.39° γ = 71.16° 
	CASSCF(7,5) 10 quartets 40 doublets	D_{xx} = 28.310874 D_{xy} = -16.075090 D_{xz} = 5.450471 D_{yy} = -7.570848 D_{yz} = -7.330963 D_{zz} = 15.167529 D = -39.799662 E/D = 0.301021	-2.515868 -0.964008 0.669162 -6.632841 0.794628 -0.957404 -0.983978 -0.798730 -1.972084 g-factors 1.424006 2.100919 7.232078 Isotropic average = 3.585668 g-shifts -0.578314 0.098600 5.229759 Isotropic average = 1.583348	2.152422 0.163435 -0.042725 0.160846 2.550278 0.064369 -0.046453 0.067864 2.286068 g-factors 2.074823 2.298981 2.614968 Isotropic average = 2.329590 g-shifts 0.072504 0.296661 0.612648 Isotropic average = 0.327271	χ_{xx} = 1.195556·10^{-31} χ_{xy} = 3.724936·10^{-32} χ_{xz} = -8.429872·10^{-33} χ_{yy} = 2.072700·10^{-31} χ_{yz} = 1.660331·10^{-32} χ_{zz} = 1.437780·10^{-31} $\Delta\chi_{ax}$ = 9.92·10^{-32} $\Delta\chi_{rh}$ = -4.49·10^{-32} α = -71.39° β = -99.27° γ = 70.92° 
	CASSCF(7,5)/ NEVPT2 10 quartets 40 doublets	D_{xx} = 23.231291 D_{xy} = -13.925065 D_{xz} = 5.235850 D_{yy} = -5.629409 D_{yz} = -6.552123 D_{zz} = 12.676629 D = -33.106017 E/D = 0.319895	2.566243 -0.787798 0.664808 4.271506 -4.657547 -0.888628 1.162441 -0.046796 -2.058623 g-factors 1.469746 2.165985 6.864118 Isotropic average = 3.499950 g-shifts -0.532573 0.163666 4.861798 Isotropic average = 1.497630	2.127707 0.133734 -0.037823 0.131464 2.448196 0.051512 -0.040181 0.054041 2.232971 g-factors 2.062200 2.245643 2.501032 Isotropic average = 2.269625 g-shifts 0.059881 0.243323 0.498713 Isotropic average = 0.267306	χ_{xx} = 1.444146·10^{-31} χ_{xy} = 3.730804·10^{-32} χ_{xz} = -9.760219·10^{-33} χ_{yy} = 2.297444·10^{-31} χ_{yz} = 1.628830·10^{-32} χ_{zz} = 1.684723·10^{-31} $\Delta\chi_{ax}$ = 9.71·10^{-32} $\Delta\chi_{rh}$ = -4.76·10^{-32} α = -70.33° β = -98.93° γ = 70.58° 

Figure 6.4: *ORCA output parameters for structure 4.*

Movement type	**Initial value** (v_i)	**Final value** (v_f)	$\lvert(v_i - v_f)/h_k\rvert$
Pseudobond (P)	1.7002 Å	1.6434 Å	2.8
Angle (A7)	142.196°	138.822°	2.8
Angle (A10)	105.023°	102.376°	2.2
Angle (A1)	50.84°	48.745°	1.7
Angle (A9)	97.647°	99.733°	1.7
Angle (A5)	152.819°	154.861°	1.7
Angle (A6)	118.404°	120.418°	1.7
Bond (B1)	2.075 Å	2.044 Å	1.6
Angle (A8)	85.346°	83.489°	1.5
Bond (B2)	2.072 Å	2.089 Å	$8.5 \cdot 10^{-1}$
Angle (A11)	96.564°	97.371°	$6.7 \cdot 10^{-1}$
Bond (B3)	2.076 Å	2.063 Å	$6.5 \cdot 10^{-1}$
Dihedral (D2)	-92.304°	-89.533°	$5.3 \cdot 10^{-1}$
Dihedral (D3)	32.228°	30.154°	$4.0 \cdot 10^{-1}$
Dihedral (D1)	86.271°	84.335°	$3.7 \cdot 10^{-1}$
Angle (A2)	55.694°	56.126°	$3.6 \cdot 10^{-1}$
Angle (A3)	100.039°	99.811°	$1.9 \cdot 10^{-1}$
Dihedral (D4)	21.333°	21.902°	$1.1 \cdot 10^{-1}$
Angle (A4)	90.306°	90.348°	$3.5 \cdot 10^{-2}$

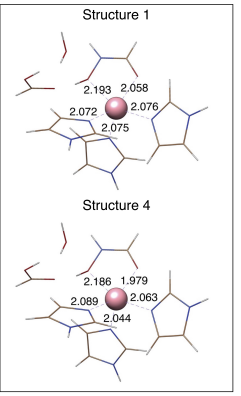


Figure 6.5: *Variations of the degrees of freedom between structures 1 and 4. The differences, sorted in descending order, are expressed relatively to the corresponding increment element h_k, with k = angle, dihedral, bond. The nomenclature and the graphical representation of the movement types are reported in Fig. 1.8 (left). Comparison between structures 1 and 4 with shown the coordination distances (right).*

NMR and XRD analysis of HIV-1 viral capsid protein

Differences between atomic structures determined using NMR and XRD are not uncommon, due to crystal packing forces or differences in the environment (ionic strength, pH, temperature, etc.) or in the procedural practices [5, 327–330]. It was thus performed a joint structure refinement through REFMAC-NMR [5, 196, 329] (Figs. S2-S11 and Tables S3-S12 in reference [186], which showed that the helix 2 position systematically violates the solution data of the X-ray structures 1A8O, 1A43, and 1BAJ, whereas the monomer structure of 2BUO exhibits reasonable agreement. However, even for the 2BUO structure, the differences after the joint refinement are significantly larger than the experimental uncertainty of the RDC values.

Bibliography

[1] E.S. Ameh. A review of basic crystallography and X-ray diffraction applications. *Int. J. Adv. Manuf. Technol.*, 105:3289–3302, 2019. doi: https://doi.org/10.1007/s00170-019-04508-1.

[2] D. Radostin, Y. Haruaki, and K. Masahide. Cryo-electron microscopy methodology: Current aspects and future directions. *Trends Biochem. Sci.*, 44(10):837–848, 2019. doi: https://doi.org/10.1016/j.tibs.2019.04.008.

[3] E. Luchinat and L. Banci. In-cell NMR in human cells: Direct protein expression allows structural studies of protein folding and maturation. *Acc. Chem. Res.*, 51(6):1550–1557, 2018. doi: https://doi.org/10.1021/acs.accounts.8b00147.

[4] A. Schiró, A. Carlon, G. Parigi, G. Murshudov, V. Calderone, E. Ravera, and C. Luchinat. On the complementarity of X-ray and NMR data. *J. Struct. Biol.: X*, 4:100019, 2020. doi: https://doi.org/10.1016/j.yjsbx.2020.100019.

[5] A. Carlon, E. Ravera, G. Parigi, G.N. Murshudov, and C. Luchinat.

Joint X-ray/NMR structure refinement of multidomain/multisubunit systems. *J. Biomol. NMR*, 73(6):265–278, 2019. doi: https://doi.org/10.1007/s10858-018-0212-3.

[6] D. Gatteschi, R. Sessoli, and J. Villain. *Molecular Nanomagnets*. Oxford University Press, USA, 2 edition, 2006.

[7] A. Abragam and B. Bleaney. *Electron Paramagnetic Resonance of Transition Ions*. Oxford University Press, USA, 1986.

[8] C. Malgrange, C. Ricolleau, and Schlenker M. Second-rank tensors. In *Symmetry and Physical Properties of Crystals*, pages 205–223. Springer, 2014.

[9] M. Perfetti. Cantilever torque magnetometry on coordination compounds: from theory to experiments. *Coord. Chem. Rev.*, 348:171–186, 10 2017. doi: https://doi.org/10.1016/j.ccr.2017.08.013.

[10] M. Perfetti, E. Lucaccini, L. Sorace, J.P. Costes, and R. Sessoli. Determination of magnetic anisotropy in the LnTRENSAL complexes (Ln = Tb, Dy, Er) by torque magnetometry. *Inorg. Chem.*, 54 (7):3090–3092, 2015. doi: https://doi.org/10.1021/acs.inorgchem.5b00288.

[11] G. Cucinotta, M. Perfetti, J. Luzon, M. Etienne, P.E. Car, A. Caneschi, G. Calvez, K. Bernot, and R. Sessoli. Magnetic anisotropy in a dysprosium/DOTA Single-Molecule Magnet: beyond simple magneto-structural correlations. *Angew. Chem. Int. Ed.*, 51 (7):1606–1610, 2012. doi: https://doi.org/10.1002/anie.201107453.

[12] O.A. Blackburn, N.F. Chilton, K. Keller, C.E. Tait, W.K. Myers, E.J. McInnes, A.M. Kenwright, P.D. Beer, C.R. Timmel, and S. Faulkner. Spectroscopic and crystal field consequences of fluoride binding by $[Yb{\cdot}DTMA]^{3+}$ in aqueous solution. *Angew. Chem. Int. Ed.*, 54(37): 10783–10786, 2015. doi: https://doi.org/10.1002/anie.201503421.

[13] O.A. Blackburn, R.M. Edkins, S. Faulkner, A.M. Kenwright, D. Parker, N.J. Rogers, and S. Shuvaev. Electromagnetic susceptibility anisotropy and its importance for paramagnetic NMR and optical spectroscopy in lanthanide coordination chemistry. *Dalton Trans.*, 45 (16):6782–6800, 2016. doi: https://doi.org/10.1039/c6dt00227g.

[14] Y. Takano, R. Tashita, M. Suzuki, S. Nagase, H. Imahori, and T. Akasaka. Molecular location sensing approach by anisotropic magnetism of an endohedral metallofullerene. *JACS*, 138(25):8000–8006, 2016. doi: https://doi.org/10.1021/jacs.6b04037.

[15] Y. Zhang, D. Krylov, M. Rosenkranz, S. Schiemenz, and A.A. Popov. Magnetic anisotropy of endohedral lanthanide ions: paramagnetic NMR study of $MSc_2N@C_{80}$-I_h with M running through the whole 4f row. *Chem. Sci.*, 6:2328–2341, 2015. doi: https://doi.org/10.1039/C5SC00154D.

[16] S. Donald Emerson and Gerd N. La Mar. NMR determination of the orientation of the magnetic susceptibility tensor in cyanometmyoglobin: a new probe of steric tilt of bound ligand. *Biochemistry*, 29(6):1556–1566, 1990. doi: https://doi.org/10.1021/bi00458a029.

[17] I. Bertini, C. Luchinat, and G. Parigi. Magnetic susceptibility in

paramagnetic NMR. *Prog. Nucl. Magn. Reson. Spectrosc.*, 40(3):249–273, 2002. doi: https://doi.org/10.1016/S0079-6565(02)00002-X.

[18] J.J. Henderson, C. Koo, P.L. Feng, E. del Barco, S. Hill, I.S. Tupitsyn, P.C.E. Stamp, and D.N. Hendrickson. Manifestation of spin selection rules on the quantum tunneling of magnetization in a Single-Molecule Magnet. *Phys. Rev. Lett.*, 103:017202, 2009. doi: https://doi.org/10.1103/PhysRevLett.103.017202.

[19] L. Sorace, M.E. Boulon, P. Totaro, A. Cornia, J. Fernandes-Soares, and R. Sessoli. Origin and spectroscopic determination of trigonal anisotropy in a heteronuclear Single-Molecule Magnet. *Phys. Rev. B*, 88:104407, 2013. doi: https://doi.org/10.1103/PhysRevB.88.104407.

[20] S. Hill, S. Maccagnano, Kyungwha Park, R.M. Achey, J. M. North, and N.S. Dalal. Detailed single-crystal EPR line shape measurements for the Single-Molecule Magnets Fe_8Br and Mn_{12}-acetate. *Phys. Rev. B*, 65:224410, 2002. doi: https://doi.org/10.1103/PhysRevB.65.224410.

[21] A.L. Barra. High-frequency EPR spectroscopy of Single-Molecule Magnets: A case study. *Appl. Magn. Reson.*, 21:619–628, 12 2001. doi: https://doi.org/10.1007/BF03162434.

[22] M.A. Sørensen, H Weihe, M.G. Vinum, J.S. Mortensen, L.H. Doerrer, and J. Bendix. Imposing high-symmetry and tuneable geometry on lanthanide centres with chelating Pt and Pd metalloligands. *Chem. Sci.*, 8(5):3566–3575, 2017. doi: https://doi.org/10.1039/c7sc00135e.

[23] S. Ghosh, S. Datta, L. Friend, S. Cardona-Serra, A. Gaita-Ariño, E. Coronado, and S. Hill. Multi-frequency EPR studies of a mononuclear holmium Single-Molecule Magnet based on the polyoxometalate $[Ho^{III}(W_5O_{18})_2]^{9-}$. *Dalton Trans.*, 41:13697–13704, 2012. doi: https://doi.org/10.1039/C2DT31674A.

[24] F. Neese. The ORCA program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2:73–78, 2012. doi: https://doi.org/10.1002/wcms.81.

[25] E. A. Suturina, J. Nehrkorn, J.M. Zadrozny, J. Liu, M. Atanasov, T. Weyhermüller, D. Maganas, S. Hill, A. Schnegg, E. Bill, J.R. Long, and F. Neese. Magneto-structural correlations in pseudotetrahedral forms of the $[Co(SPh)4]^{2-}$ complex probed by magnetometry, MCD spectroscopy, advanced EPR techniques, and ab initio electronic structure calculations. *Inorg. Chem.*, 56(5):3102–3118, 2017. doi: https://doi.org/10.1021/acs.inorgchem.7b00097.

[26] M. Atanasov, D. Aravena, E. Suturina, E. Bill, D. Maganas, and F. Neese. First principles approach to the electronic structure, magnetic anisotropy and spin relaxation in mononuclear 3d-transition metal Single Molecule Magnets. *Coord. Chem. Rev.*, 289-290:177–214, 2015. doi: https://doi.org/10.1016/j.ccr.2014.10.015.

[27] E. Cancés, M. Defranceschi, W. Kutzelnigg, C. Le Bris, and Y. Maday. Computational quantum chemistry: A primer. In *Computational Chemistry, Special Volume*, volume 10 of *Handbook of Numerical Analysis*, pages 3–270. Elsevier, 2003. doi: https://doi.org/10.1016/S1570-8659(03)10003-8.

[28] A. Szabo and N.S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Publications, Inc., 1 edition, 1996.

[29] J. Dreiser. Molecular lanthanide Single-Ion Magnets: from bulk to submonolayers. *J. Condens. Matter Phys.*, 27(18):183203, 2015. doi: https://doi.org/10.1088/0953-8984/27/18/183203.

[30] L.F. Chibotaru and L. Ungur. Ab initio calculation of anisotropic magnetic properties of complexes. I. Unique definition of pseudospin Hamiltonians and their derivation. *J. Chem. Phys.*, 136(6):064112, 2012. doi: https://doi.org/10.1063/1.4739763.

[31] D. Aravena, M. Atanasov, and F. Neese. Periodic trends in lanthanide compounds through the eyes of multireference ab initio theory. *Inorg. Chem.*, 55(9):4457–4469, 2016. doi: https://doi.org/10.1021/acs.inorgchem.6b00244.

[32] D. Aravena, F. Neese, and D.A. Pantazis. Improved segmented all-electron relativistically contracted basis sets for the lanthanides. *J. Chem. Theory Comput*, 12(3):1148–1156, 2016. doi: https://doi.org/10.1021/acs.jctc.5b01048.

[33] H. Schilder and H. Lueken. Computerized magnetic studies on d, f, d-d, f-f, and d-S, f-S systems under varying ligand and magnetic fields. *J. Magn. Magn. Mater.*, 281:17–26, 2004. doi: https://doi.org/10.1016/j.jmmm.2004.03.041.

[34] J. Leusen, M. Speldrich, H. Schilder, and P. Kögerler. Comprehensive insight into molecular magnetism via CONDON: Full vs. effective

models. *Coord. Chem. Rev.*, 289-290:137 – 148, 2015. ISSN 0010-8545. doi: https://doi.org/10.1016/j.ccr.2014.10.011.

[35] M. Speldrich, H. Schilder, H. Lueken, and P. Kögerler. A computational framework for magnetic polyoxometalates and molecular spin structures: CONDON 2.0. *Isr. J. Chem.*, 51:215 – 227, 02 2011. doi: https://doi.org/10.1002/ijch.201100013.

[36] A. Bronova, T. Bredow, R. Glaum, M.J. Riley, and W. Urland. BonnMag: Computer program for ligand-field analysis of f^n systems within the angular overlap model. *J. Comput. Chem.*, 39(3):176–186, 2018. doi: https://doi.org/10.1002/jcc.25096.

[37] J.J. Baldoví, S. Cardona-Serra, J. Clemente-Juan, E. Coronado, A. Gaita-Ariño, and A. Palii. SIMPRE: A software package to calculate crystal field parameters, energy levels, and magnetic properties on mononuclear lanthanoid complexes based on charge distributions. *J. Comput. Chem.*, 34:1961–1967, 2013. doi: https://doi.org/10.1002/jcc.23341.

[38] P.E.M. Siegbahn, J. Almlöf, A. Heiberg, and B.O. Roos. The complete active space SCF (CASSCF) method in a Newton-Raphson formulation with application to the HNO molecule. *J. Chem. Phys.*, 74 (4):2384–2396, 1981. doi: https://doi.org/10.1063/1.441359.

[39] B.O. Roos, P.R. Taylor, and P. E. M. Sigbahn. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *J. Chem. Phys.*, 48(2):157–173, 1980. doi: https://doi.org/10.1016/0301-0104(80)80045-0.

[40] J. Finley, P. Malmqvist, B.O. Roos, and L. Serrano-Andrés. The multi-state CASPT2 method. *Chem. Phys. Lett.*, 288(2):299–306, 1998. doi: https://doi.org/10.1016/S0009-2614(98)00252-8.

[41] C. Angeli and R. Cimiraglia. Multireference perturbation configuration interaction V. Third-order energy contributions in the Møller-Plesset and Epstein-Nesbet partitions. *Theor. Chem. Acc.*, 107:313–317, 2002. doi: https://doi.org/10.1007/s00214-002-0336-z.

[42] C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger, and J.P. Malrieu. Introduction of n-electron valence states for multireference perturbation theory. *J. Chem. Phys.*, 114:10252–10264, 2001. doi: https://doi.org/10.1063/1.1361246.

[43] G. Karlström, R. Lindh, M. Per-Åke, O.R. Björn, U. Ryde, V. Veryazov, P. Widmark, M. Cossi, B Schimmelpfennig, P. Neogrady, and L. Seijo. MOLCAS: a program package for computational chemistry. *Comput. Mater. Sci.*, 28(2):222–239, 2003. doi: https://doi.org/10.1016/S0927-0256(03)00109-5.

[44] F. Neese. Software update: the ORCA program system, version 4.0: Software update. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 8:e1327, 2017. doi: https://doi.org/10.1002/wcms.1327.

[45] I. Bertini, C. Luchinat, G. Parigi, and E. Ravera. The hyperfine shift. In *NMR of Paramagnetic Molecules: Applications to Metallobiomolecules and Models*, pages 25–60. Elsevier, 2 edition, 2017.

[46] H. Mcconnell and D.B. Chesnut. Theory of isotropic hyperfine inter-

actions in π-electron radicals. *J. Chem. Phys.*, 28:107–117, 1958. doi: https://doi.org/10.1063/1.1744052.

[47] R.J. Kurland and B.R. McGarvey. Isotropic NMR shifts in transition metal complexes: The calculation of the fermi contact and pseudocontact terms. *J. Magn. Reson.*, 2(3):286–301, 1970. doi: https://doi.org/10.1016/0022-2364(70)90100-9.

[48] A. Bertarello, L. Benda, K.J. Sanders, A.J. Pell, M.J. Knight, V. Pelmenschikov, L. Gonnelli, I.C. Felli, M. Kaupp, L. Emsley, R. Pierattelli, and G. Pintacuda. Picometer resolution structure of the coordination sphere in the metal-binding site in a metalloprotein by NMR. *JACS*, 142(39):16757–16765, 2020. doi: https://doi.org/10.1021/jacs.0c07339.

[49] Harold M. Goff. Iron(iii) porphyrin-imidazole complexes. analysis of carbon-13 Nuclear Magnetic Resonance isotropic shifts and unpaired spin delocalization. *JACS*, 103(13):3714–3722, 1981. doi: https://doi.org/10.1021/ja00403a017.

[50] L. Banci, I. Bertini, R. Pierattelli, M. Tien, and A.J. Vila. Factoring of the hyperfine shifts in the cyanide adduct of lignin peroxidase from p. chrysosporium. *JACS*, 117:8659–8667, 1995. doi: https://doi.org/10.1021/ja00138a023.

[51] N. Ishikawa, M. Sugita, T. Ishikawa, S. Koshihara, and Y. Kaizu. Lanthanide double-decker complexes functioning as magnets at the single-molecular level. *JACS*, 125(29):8694–8695, 2003. doi: https://doi.org/10.1021/ja029629n.

[52] S.T. Liddle and J. van Slageren. Improving f-element Single Molecule Magnets. *Chem. Soc. Rev.*, 44:6655–6669, 2015. doi: https://doi.org/10.1039/C5CS00222B.

[53] R. Sessoli, H.L. Tsai, A.R. Schake, S. Wang, J.B. Vincent, K. Folting, D. Gatteschi, G. Christou, and D.N. Hendrickson. High-spin molecules: $[Mn_{12}O_{12}(O_2CR)_{16}](H_2O)_4]$. *JACS*, 115(5):1804–1816, 1993. doi: https://doi.org/10.1021/ja00058a027.

[54] S.G. McAdams, A.M. Ariciu, A.K. Kostopoulos, J.P.S. Walsh, and F.a Tuna. Molecular Single-Ion Magnets based on lanthanides and actinides: Design considerations and new advances in the context of quantum technologies. *Coord. Chem. Rev.*, 346:216–239, 2017. ISSN 0010-8545. doi: 10.1016/j.ccr.2017.03.015.

[55] G.A. Craig and M. Murrie. 3d Single-Ion Magnets. *Chem. Soc. Rev.*, 44:2135–2147, 2015. doi: https://doi.org/10.1039/C4CS00439F.

[56] R. Sessoli, D. Gatteschi, A. Caneschi, and M.A. Novak. Magnetic bistability in a metal-ion cluster. *Nature*, 365(6442):141–143, 1993. doi: https://doi.org/10.1038/365141a0.

[57] N. Ishikawa, M. Sugita, T. Okubo, N. Tanaka, T. Iino, and Y. Kaizu. Determination of ligand-field parameters and f-electronic structures of double-decker bis(phthalocyaninato)lanthanide complexes. *Inorg. Chem.*, 42(7):2440–2446, 2003. doi: https://doi.org/10.1021/ic026295u.

[58] D.E. Freedman, W.H. Harman, T.D. Harris, G.J. Long, C.J. Chang,

and J.R. Long. Slow magnetic relaxation in a high-spin iron(II) complex. *JACS*, 132(4):1224–1225, 2010. doi: https://doi.org/10.1021/ja909560d.

[59] S. Sanvito. Molecular spintronics. *Chem. Soc. Rev.*, 40:3336–3355, 2011. doi: https://doi.org/10.1039/C1CS15047B.

[60] J. Lehmann, A. Gaita-Ariño, E. Coronado, and D. Loss. Quantum computing with molecular spin systems. *J. Mater. Chem.*, 19:1672–1677, 2009. doi: https://doi.org/10.1039/B810634G.

[61] G. Parigi, L. Benda, E. Ravera, M. Romanelli, and C. Luchinat. Pseudocontact shifts and paramagnetic susceptibility in classical and quantum chemistry theories. *J. Chem. Phys.*, 150:144101–144111, 2018. doi: https://doi.org/10.1063/1.5037428.

[62] L. Cerofolini, J.M. Silva, E. Ravera, M. Romanelli, C. Geraldes, A.L. Macedo, M. Fragai, G. Parigi, and C. Luchinat. How do nuclei couple to the magnetic moment of a paramagnetic center? A new theory at the gauntlet of the experiments. *J. Phys. Chem. Lett.*, pages 3610–3614, 2019. doi: https://doi.org/10.1021/acs.jpclett.9b01128.

[63] L. Lang, E. Ravera, G. Parigi, C. Luchinat, and F. Neese. Solution of a puzzle: High-level quantum-chemical treatment of pseudocontact chemical shifts confirms classic semiempirical theory. *J. Phys. Chem. Lett.*, 11(0):8735–8744, 2020. doi: https://doi.org/10.1021/acs.jpclett.0c02462.

[64] H.M. McConnell and R.E. Robertson. Isotropic Nuclear Resonance

Shifts. *J. Chem. Phys.*, 29(6):1361–1365, 1958. doi: https://doi.org/10.1063/1.1744723.

[65] E. Ravera, L. Gigli, B. Czarniecki, L. Lang, R. Kümmerle, G. Parigi, M. Piccioli, F. Neese, and C. Luchinat. A quantum chemistry view on two archetypical paramagnetic pentacoordinate nickel(II) complexes offers a fresh look on their NMR spectra. *Inorg. Chem.*, 60(3):2068–2075, 2021. doi: https://doi.org/10.1021/acs.inorgchem.0c03635.

[66] L. Sacconi and I. Bertini. High-spin five-coordinated 3d metal complexes with pentadentate Schiff bases. *JACS*, 88(22):5180–5185, 1966. doi: https://doi.org/10.1021/ja00974a027.

[67] L. Sacconi, M. Ciampolini, and G.P. Speroni. Structure mimicry in solid solutions of 3d metal complexes with N-Methylsalicylaldimine (Msal-Me). *JACS*, 87(14):3102–3106, 1965. doi: https://doi.org/10.1021/ja01092a016.

[68] L. Sacconi, P. L. Orioli, and M. Di Vaira. The structure of five-coordinated high-spin complexes of nickel(II) and cobalt(II) with N-β-Diethylamineethyl-5-chlorosalicylaldimine. *JACS*, 87(9):2059–2059, 1965. doi: https://doi.org/10.1021/ja01087a043.

[69] L. Sacconi, P. Nannelli, N. Nardi, and U. Campigli. Five-coordination in some complexes of nickel(II) with Schiff bases, formed from salicylaldehydes and N,N-substituted ethylenediamines. *Inorg. Chem.*, 4 (7):943–949, 1965. doi: https://doi.org/10.1021/ic50029a004.

[70] Gerd N. La Mar and Luigi Sacconi. Proton Magnetic Resonance Studies of high-spin nickel(II) complexes with pentadentate Schiff

bases. *JACS*, 89(10):2282–2291, 1967. doi: https://doi.org/10.1021/ja00986a009.

[71] C. Luchinat, S. Steuernagel, and P. Turano. Application of 2D NMR techniques to paramagnetic systems. *Inorg. Chem.*, 29(21):4351–4353, 1990. doi: https://doi.org/10.1021/ic00346a039.

[72] P. Orioli, M. Di Vaira, and L. Sacconi. Crystal and molecular structure of the nickel(II) complex with the pentadentate ligand bis(salicylidene-γ-iminopropyl)methylamine. *Inorg. Chem.*, 10(3): 553–558, 1971. doi: https://doi.org/10.1021/ic50097a023.

[73] P.J. Stephens, F.J. Devlin, C.F. Chabalowski, and M.J. Frisch. Ab initio calculation of vibrational absorption and Circular Dichroism spectra using Density Functional Force Fields. *J. Phys. Chem.*, 98 (45):11623–11627, 1994. doi: https://doi.org/10.1021/j100096a001.

[74] S.H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.*, 58(8):1200–1211, 1980. doi: https://doi.org/10.1139/p80-159.

[75] C. Lee, W. Yang, and R.G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, 1988. doi: https://doi.org/10.1103/PhysRevB.37.785.

[76] A.D. Becke. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.*, 98(2):1372–1377, 1993. doi: https://doi.org/10.1063/1.464304.

[77] A. Schäfer, C. Huber, and R. Ahlrichs. Fully optimized contracted Gaussian basis sets of Triple Zeta Valence quality for atoms Li to Kr. *J. Chem. Phys.*, 100(8):5829–5835, 1994. doi: https://doi.org/10.1063/1.467146.

[78] F. Weigend and R. Ahlrichs. Balanced basis sets of split valence, triple zeta valence and Quadruple Zeta Valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7(18): 3297, 2005. doi: https://doi.org/10.1039/B508541A.

[79] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. A consistent and accurate ab initio parametrization of Density Functional Dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, 132 (15):154104, 2010. doi: https://doi.org/10.1063/1.3382344.

[80] S. Grimme, S. Ehrlich, and L. Goerigk. Effect of the damping function in dispersion corrected Density Functional Theory. *J. Comput. Chem.*, 32(7):1456–1465, 2011. doi: https://doi.org/10.1002/jcc.21759.

[81] R.A. Kendall and H.A. Fruchtl. The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. *Theor. Chim. Acta*, 97(1-4):158–163, 1997. doi: https://doi.org/10.1007/s002140050249.

[82] O. Vahtras, J. Almlöf, and M.W. Feyereisen. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.*, 213(5):514–518, 1993. doi: https://doi.org/10.1016/0009-2614(93)89151-7.

[83] V. Barone and M. Cossi. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A*, 102(11):1995–2001, 1998. doi: https://doi.org/10.1021/jp9716997.

[84] F. Neese. Efficient and accurate approximations to the molecular spin-orbit coupling operator and their use in molecular g-tensor calculations. *J. Chem. Phys.*, 122:34107, 2005. doi: https://doi.org/10.1063/1.1829047.

[85] B.A. Hess. Relativistic electronic-structure calculations employing a two-component no-pair formalism with external-field projection operators. *Phys. Rev. A*, 33:3742–3748, 1986. doi: https://doi.org/10.1103/PhysRevA.33.3742.

[86] F. Neese. Theoretical study of ligand superhyperfine structure. Application to Cu(II) complexes. *J. Phys. Chem. A*, 105(17):4290–4299, 2001. doi: https://doi.org/10.1021/jp003254f.

[87] F. Neese. Metal and ligand hyperfine couplings in transition metal complexes: The effect of spin-orbit coupling as studied by coupled perturbed Kohn-Sham theory. *J. Chem. Phys.*, 118(9):3939–3948, 2003. doi: https://doi.org/10.1063/1.1540619.

[88] S.K. Singh, M. Atanasov, and F. Neese. Challenges in Multireference Perturbation Theory for the calculations of the g-tensor of first-row transition-metal complexes. *J. Chem. Theory Comput.*, 14(9):4662–4677, 2018. doi: https://doi.org/10.1021/acs.jctc.8b00513.

[89] F. Neese. Prediction of Electron Paramagnetic Resonance g values using coupled perturbed Hartree-Fock and Kohn-Sham theory. *J. Chem. Phys.*, 115(24):11080–11096, 2001. doi: https://doi.org/10.1063/1.1419058.

[90] Bertini I., Dikiy A., Luchinat C., Macinai R., and Viezzoli M.S. ^{1}h NMR study of the reduced cytochrome c' from rhodopseudomonas palustris containing a high-spin iron(II) heme moiety. *Inorg Chem.*, 22:4814–4821, 1998. doi: https://doi.org/10.1021/ic980531c.

[91] I. Bertini, F. Capozzi, C. Luchinat, M. Piccioli, and A.J. Vila. The iron-sulfur cluster (Fe_4S_4) centers in ferredoxins studied through proton and carbon hyperfine coupling. Sequence-specific assignments of cysteines in ferredoxins from clostridium acidi urici and clostridium pasteurianum. *JACS*, 116(2):651–660, 1994. doi: https://doi.org/10.1021/ja00081a028.

[92] R.H. Holm, P. Kennepohl, and E.I. Solomon. Structural and functional aspects of metal sites in biology. *Chem. Rev.*, 96(7):2239–2314, 1996. doi: https://doi.org/10.1021/cr9500390.

[93] I. Bertini, C. Luchinat, G. Parigi, and F. Walker. Heme methyl ^{1}H chemical shifts as structural parameters in some low-spin ferriheme proteins. *JBIC*, 4:515–519, 1999. doi: https://doi.org/10.1007/s007750050337.

[94] N. Shokhirev and F. Walker. The effect of axial ligand plane orientation on the contact and pseudocontact shifts of low-spin ferri-

heme proteins. *JBIC*, 3:581–594, 1998. doi: https://doi.org/10.1007/s007750050271.

[95] M. Briganti, G. Fernandez Garcia, J. Jung, R. Sessoli, B. Le Guennic, and F. Totti. Covalency and magnetic anisotropy in lanthanide Single Molecule Magnets: the DyDOTA archetype. *Chem. Sci.*, 10:7233–7245, 2019. doi: https://doi.org/10.1039/C9SC01743G.

[96] L. Rigamonti, N. Bridonneau, G. Poneti, L. Tesi, L. Sorace, D. Pinkowicz, J. Jover, E. Ruiz, R. Sessoli, and A. Cornia. A pseudo-octahedral cobalt(II) complex with bispyrazolylpyridine ligands acting as a Zero-Field Single-Molecule Magnet with easy axis anisotropy. *Chem. Eur. J.*, 24(35):8857–8868, 2018. doi: https://doi.org/10.1002/chem.201801026.

[97] M. Vonci, K. Mason, E.A. Suturina, A.T. Frawley, S.G. Worswick, I. Kuprov, D. Parker, E.J.L. McInnes, and N.F. Chilton. Rationalization of anomalous pseudocontact shifts and their solvent dependence in a series of C3-symmetric lanthanide complexes. *JACS*, 139(40):14166–14172, 2017. doi: https://doi.org/10.1021/jacs.7b07094.

[98] M. Hiller, S. Krieg, N. Ishikawa, and M. Enders. Ligand-field energy splitting in lanthanide-based Single-Molecule Magnets by NMR spectroscopy. *Inorg. Chem.*, 56(24):15285–15294, 2017. doi: https://doi.org/10.1021/acs.inorgchem.7b02704.

[99] L. Benda, J. Mareš, E. Ravera, G. Parigi, C. Luchinat, M. Kaupp, and J. Vaara. Pseudo-contact NMR shifts over the paramagnetic met-

alloprotein CoMMP-12 from first principles. *Angew. Chem. Int. Ed.*, 55:14713–14717, 2016. doi: https://doi.org/10.1002/anie.201608829.

[100] J. Krzystek, G. Kohl, H.B. Hansen, M. Enders, and J. Telser. Combining HFEPR and NMR spectroscopies to characterize organochromium(III) complexes with large Zero-Field Splitting. *Organometallics*, 38(9):2179–2188, 2019. doi: https://doi.org/10.1021/acs.organomet.9b00158.

[101] E.A. Suturina, K. Mason, M. Botta, F. Carniato, I. Kuprov, N.F. Chilton, E.J.L. McInnes, M. Vonci, and D. Parker. Periodic trends and hidden dynamics of magnetic properties in three series of triazacyclononane lanthanide complexes. *Dalton Trans.*, 48:8400–8409, 2019. doi: https://doi.org/10.1039/C9DT01069F.

[102] E.A. Suturina, D. Maganas, E. Bill, M. Atanasov, and F. Neese. Magneto-structural correlations in a series of pseudotetrahedral $[Co^{II}(XR)_4]^{2-}$ Single Molecule Magnets: An ab initio ligand field study. *Inorg. Chem.*, 54(20):9948–9961, 2015. doi: https://doi.org/10.1021/acs.inorgchem.5b01706.

[103] A. Lunghi and S. Sanvito. Spin-phonon relaxation in molecular qubits from first principles spin dynamics, 2019.

[104] A. Lunghi, F. Totti, R. Sessoli, and S. Sanvito. The role of anharmonic phonons in under-barrier spin relaxation of Single Molecule Magnets. *Nat. Comm.*, 8:14620, 2017. doi: https://doi.org/10.1038/ncomms14620.

[105] J. Rantaharju, J. Mareš, and J. Vaara. Spin dynamics simulation of electron spin relaxation in Ni^{2+}(aq). *J. Chem. Phys.*, 141(1):014109, 2014. doi: https://doi.org/10.1063/1.4885050.

[106] C. Andreini, I. Bertini, G. Cavallaro, G.L. Holliday, and J.M. Thornton. Metal ions in biological catalysis: from enzyme databases to general principles. *JBIC*, 13(8):1205–1218, 2008. doi: https://doi.org/10.1007/s00775-008-0404-5.

[107] C. Andreini, I. Bertini, and A. Rosato. Metalloproteomes: A bioinformatic approach. *Acc. Chem. Res.*, 42(10):1471–1479, 2009. doi: https://doi.org/10.1021/ar900015x.

[108] I. Bertini and C. Luchinat. Cobalt(II) as a probe of the structure and function of carbonic anhydrase. *Acc. Chem. Res.*, 16(8):272–279, 1983. doi: https://doi.org/10.1021/ar00092a002.

[109] I. Bertini, I. Gelis, N. Katsaros, C. Luchinat, and A. Provenzani. Tuning the affinity for lanthanides of calcium binding proteins. *Biochemistry*, 42(26):8011–8021, 2003. doi: https://doi.org/10.1021/bi034494z.

[110] I. Bertini, C. Luchinat, A. Scozzafava, P. Hemmerich, J. Livorness, V. Massey, H. Michel, C. Schug, and T. Smith. *Carbonic anhydrase: An insight into the zinc binding site and into the active cavity through metal substitution*, volume 48, pages 45–92. Springer, 12 2007. doi: https://doi.org/10.1007/BFb0111580.

[111] A. Abragam. *The principles of Nuclear Magnetism*. Oxford University Press, 1 edition, 1961.

[112] B.D. Belviso, R. Caliandro, D. Siliqi, V. Calderone, F. Arnesano, and G. Natile. Structure of matrix metalloproteinase-3 with a platinum-based inhibitor. *Chem. Commun.*, 49:5492–5494, 2013. doi: https://doi.org/10.1039/C3CC41278D.

[113] M. Spiegel, A.K. Duraisamy, and G.F. Schröder. Improving the visualization of cryo-EM density reconstructions. *J. Struct. Biol.*, 191 (2):207–213, 2015. doi: https://doi.org/10.1016/j.jsb.2015.06.007.

[114] E. Ravera, G. Parigi, and C. Luchinat. What are the methodological and theoretical prospects for paramagnetic NMR in structural biology? A glimpse into the crystal ball. *J. Magn. Reson.*, 306:173–179, 2019. doi: https://doi.org/10.1016/j.jmr.2019.07.027.

[115] E. Ravera, A. Carlon, M. Fragai, G. Parigi, and C. Luchinat. Paramagnetic NMR as a new tool in structural biology. *Emerg. Top. Life Sci.*, 2:ETLS20170084, 2018. doi: https://doi.org/10.1042/ETLS20170084.

[116] I. Bertini, C. Luchinat, G. Parigi, and E. Ravera. *NMR of Paramagnetic Molecules: Applications to Metallobiomolecules and Models.* Elsevier, 2 edition, 2017.

[117] E. Ravera, L. Gigli, E. A. Suturina, V. Calderone, M. Fragai, G. Parigi, and C. Luchinat. A high-resolution view of the coordination environment in a paramagnetic metalloprotein from its magnetic properties. *Angew. Chem. Int. Ed.*, 2021. doi: https://doi.org/10.1002/anie.202101149.

[118] S. Balayssac, I. Bertini, A. Bhaumik, M. Lelli, and C. Luchinat. Paramagnetic shifts in solid-state NMR of proteins to elicit structural information. *PNAS*, 105(45):17284–17289, 2008. doi: https://doi.org/10.1073/pnas.0708460105.

[119] I. Bertini, L. Emsley, M. Lelli, C. Luchinat, J. Mao, and G. Pintacuda. Ultrafast MAS solid-state NMR permits extensive ^{13}C and ^{1}H detection in paramagnetic metalloproteins. *JACS*, 132(16):5558–5559, 2010. doi: https://doi.org/10.1021/ja100398q.

[120] C. Luchinat, G. Parigi, E. Ravera, and M. Rinaldelli. Solid-state NMR crystallography through paramagnetic restraints. *JACS*, 134 (11):5006–5009, 2012. doi: https://doi.org/10.1021/ja210079n.

[121] I. Bertini, A. Bhaumik, G. De Paëpe, R.G. Griffin, M. Lelli, J.R. Lewandowski, and C. Luchinat. High-resolution solid-state NMR structure of a 17.6 kDa protein. *JACS*, 132(3):1032–1040, 2010. doi: https://doi.org/10.1021/ja906426p.

[122] S. Balayssac, I. Bertini, M. Lelli, C. Luchinat, and M. Maletta. Paramagnetic ions provide structural restraints in solid-state NMR of proteins. *JACS*, 129(8):2218–2219, 2007. doi: https://doi.org/10.1021/ja068105a.

[123] M. Rinaldelli, A. Carlon, E. Ravera, G. Parigi, and C. Luchinat. FANTEN: A new web-based interface for the analysis of magnetic anisotropy-induced NMR data. *J. Biomol. NMR*, 61:21–34, 11 2014. doi: https://doi.org/10.1007/s10858-014-9877-4.

[124] D. Ganyushin and F. Neese. A fully variational spin-orbit coupled complete active space self-consistent field approach: application to Electron Paramagnetic Resonance g-tensors. *J. Chem. Phys.*, 138 (10):104113, 2013. doi: https://doi.org/10.1063/1.4793736.

[125] D. Liebschner, P.V. Afonine, M.L. Baker, G. Bunkóczi, V.B. Chen, T.I. Croll, B. Hintze, L.W. Hung, S. Jain, A.J. McCoy, N.W. Moriarty, R.D. Oeffner, B.K. Poon, M.G. Prisant, R.J. Read, J.S. Richardson, D.C. Richardson, M.D. Sammito, O.V. Sobolev, D.H. Stockwell, T.C. Terwilliger, A.G. Urzhumtsev, L.L. Videau, C.J. Williams, and P.D. Adams. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D*, 75(10):861–877, 2019. doi: https://doi.org/10.1107/s2059798319011471.

[126] J.M. Word, S.C. Lovell, J.S. Richardson, and D.C. Richardson. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.*, 285(4):1735–1747, 1999. doi: https://doi.org/10.1006/jmbi.1998.2401.

[127] L. Banci, I. Bertini, G. Cavallaro, A. Giachetti, C. Luchinat, and G. Parigi. Paramagnetism-based restraints for Xplor-NIH. *J. Biomol. NMR*, 28:249–61, 2004. doi: https://doi.org/10.1023/B:JNMR.0000013703.30623.f7.

[128] I. Bertini, P. Kursula, C. Luchinat, G. Parigi, J. Vahokoski, M. Wilmanns, and J. Yuan. Accurate solution structures of proteins from X-ray data and a minimal set of NMR data: Calmodulin-

peptide complexes as examples. *JACS*, 131(14):5134–5144, 2009. doi: https://doi.org/10.1021/ja8080764.

[129] S. Ruder. An overview of gradient descent optimization algorithms, 2017.

[130] V. Borsi, V. Calderone, M. Fragai, C. Luchinat, and N. Sarti. Entropic contribution to the linking coefficient in fragment based drug design: A case study. *J. Med. Chem.*, 53(10):4285–4289, 2010. doi: https://doi.org/10.1021/jm901723z.

[131] J. Cha and D.S. Auld. Site-directed mutagenesis of the active site glutamate in human matrilysin: a investigation of its role in catalysis. *Biochemistry*, 36(50):16019–16024, 1997. doi: https://doi.org/10.1021/bi972223g.

[132] C.M. Holman, C.C. Kan, M.R. Gehring, and H.E. Van Wart. Role of His-224 in the anomalous pH dependence of human stromelysin-1. *Biochemistry*, 38(2):677–681, 1999. doi: https://doi.org/10.1021/bi9822170.

[133] I. Bertini, V. Calderone, M. Fragai, C. Luchinat, M. Maletta, and K. Yeo. Snapshots of the reaction mechanism of matrix metalloproteinases. *Angew. Chem. Int. Ed.*, 45:7952–5, 2006. doi: https://doi.org/10.1002/anie.200603100.

[134] M.H. Parker, E.A. Lunney, D.F. Ortwine, A.G. Pavlovsky, C. Humblet, and C.G. Brouillette. Analysis of the binding of hydroxamic acid and carboxylic acid inhibitors to the stromelysin-1 (matrix metalloproteinase-3) catalytic domain by isothermal titration

calorimetry. *Biochemistry*, 38(41):13592–13601, 1999. doi: https://doi.org/10.1021/bi991222g.

[135] B.L. Vallee and R.J. Williams. Metalloenzymes: the entatic nature of their active sites. *PNAS*, 59(2):498–505, 1968. doi: https://doi.org/10.1073/pnas.59.2.498.

[136] I. Bertini, C. Luchinat, G. Parigi, and E. Ravera. Transition metal ions: shift and relaxation. In *NMR of Paramagnetic Molecules: Applications to Metallobiomolecules and Models*, pages 175–253. Elsevier, 2 edition, 2017.

[137] C.D. Schwieters, J.J. Kuszewski, N. Tjandra, and G.M. Clore. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, 160(1):65–73, 2003. doi: https://doi.org/10.1016/S1090-7807(02)00014-9.

[138] C.D. Schwieters, J.J. Kuszewski, and G.M. Clore. Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.*, 48(1):47–62, 2006. doi: https://doi.org/10.1016/j.pnmrs.2005.10.001.

[139] Yuanpeng J. Huang and Gaetano T. Montelione. Structural biology: Proteins flex to function. *Nature*, 438(7064):36–37, 2005. doi: https://doi.org/10.1038/438036a.

[140] E. Eisenmesser, D. Bosco, M. Akke, and D. Kern. Enzyme dynamics during catalysis. *Science*, 295:1520–3, 2002. doi: https://doi.org/10.1126/science.1066176.

[141] E. Ravera, L. Salmon, M. Fragai, G. Parigi, H. Al-Hashimi, and C. Luchinat. Insights into domain-domain motions in proteins and RNA from solution NMR. *Acc. Chem. Res.*, 47(10):3118–3126, 2014. doi: https://doi.org/10.1021/ar5002318.

[142] D.A. Ragland, E.A. Nalivaika, M.N.L. Nalam, K.L. Prachanronarong, H. Cao, R.M. Bandaranayake, Y. Cai, N. Kurt-Yilmaz, and C.A. Schiffer. Drug resistance conferred by mutations outside the active site through alterations in the dynamic and structural ensemble of HIV-1 protease. *JACS*, 136(34):11956–11963, 2014. doi: https://doi.org/10.1021/ja504096m.

[143] A. Castellanos-Rubio and S. Ghosh. Disease-associated SNPs in inflammation-related lncRNAs. *Front. Immunol.*, 10:420, 2019. doi: https://doi.org/10.3389/fimmu.2019.00420.

[144] J. Schilder and M. Ubbink. Formation of transient protein complexes. *Curr. Opin. Struct. Biol.*, 23(6):911–918, 2013. doi: https://doi.org/10.1016/j.sbi.2013.07.009.

[145] N.J. Anthis and G.M. Clore. Visualizing transient dark states by NMR spectroscopy. *Q. Rev. Biophys.*, 48(1):35–116, 2015. doi: https://doi.org/10.1017/S0033583514000122.

[146] G. Schreiber and A.E. Keating. Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol.*, 21(1):50–61, 2011. doi: https://doi.org/10.1016/j.sbi.2010.10.002.

[147] A.N. Tikhonov and V.Y. Arsenin. *Solutions of ill-Posed Problems.* Winston, 1 edition, 1977.

[148] H. Engl and P. Kügler. Nonlinear inverse problems: Theoretical aspects and some industrial applications. In *Multidisciplinary Methods for Analysis, Optimization and Control of Complex Systems*, volume 6, pages 3–47. Springer, 2005. doi: https://doi.org/10.1007/3-540-27167-8_1.

[149] B. Massimiliano, T.H. Gabriella, C. Carlo, and V. Michele. Principles of protein structural ensemble determination. *Curr. Opin. Struct. Biol.*, 42:106–116, 2017. doi: https://doi.org/10.1016/j.sbi.2016.12.004.

[150] E. Ravera, L. Sgheri, G. Parigi, and C. Luchinat. A critical assessment of methods to recover information from averaged data. *Phys. Chem. Chem. Phys.*, 18:5686–5701, 2016. doi: https://doi.org/10.1039/C5CP04077A.

[151] M. Longinetti, L. Sgheri, and F. Sottile. Convex hulls of orbits and orientations of a moving protein domain. *Discrete Comput. Geom.*, 43:54–77, 2008. doi: https://doi.org/10.1007/s00454-008-9076-8.

[152] I. Bertini, M. Longinetti, C. Luchinat, G. Parigi, and L. Sgheri. Efficiency of paramagnetism-based constraints to determine the spatial arrangement of α-helical secondary structure elements. *J. Biomol. NMR*, 37(19):123–136, 2002. doi: https://doi.org/10.1023/A:1014214015858.

[153] L. Sgheri. Conformational freedom of proteins and the maximal probability of sets of orientations. *Inverse Probl.*, 26(3):035003, 2010. doi: https://doi.org/10.1088/0266-5611/26/3/035003.

[154] I. Bertini, Y.K. Gupta, C. Luchinat, G. Parigi, M. Peana, L. Sgheri, and J. Yuan. Paramagnetism-based NMR restraints provide Maximum Allowed Probabilities for the different conformations of partially independent protein domains. *JACS*, 129(42):12786–12794, 2007. doi: https://doi.org/10.1021/ja0726613.

[155] R.J. Gardner, M. Longinetti, and L. Sgheri. Reconstruction of orientations of a moving protein domain from paramagnetic data. *Inverse Probl.*, 21(3):879–898, 2005. doi: https://doi.org/10.1088/0266-5611/21/3/006.

[156] D.M. Selegato, C. Bracco, C. Giannelli, G. Parigi, C. Luchinat, L. Sgheri, and E. Ravera. Comparison of different reweighting approaches for the calculation of conformational variability of macromolecules from molecular simulations. *Chem. Phys. Chem.*, 2020. doi: https://doi.org/10.1002/cphc.202000714.

[157] Per Hansen and Dianne O'leary. The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.*, 14: 1487–1503, 11 1993. doi: https://doi.org/10.1137/0914086.

[158] I. Bertini, A. Giachetti, C. Luchinat, G. Parigi, M.V. Petoukhov, R. Pierattelli, E. Ravera, and D.I. Svergun. Conformational space of flexible biological macromolecules from average data. *JACS*, 132(38): 13553–13558, 2010. doi: https://doi.org/10.1021/ja1063923.

[159] J.J. Chou, S. Li, C.B. Klee, and A. Bax. Solution structure of Ca^{2+}-calmodulin reveals flexible hand-like properties of its domains. *Nat.*

Struct. Biol., 8(11):990–997, 2001. doi: https://doi.org/10.1038/nsb1101-990.

[160] G. Barbato, M. Ikura, L.E. Kay, R.W. Pastor, and A. Bax. Backbone dynamics of calmodulin studied by nitrogen-15 relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry*, 31(23):5269–5278, 1992. doi: https://doi.org/10.1021/bi00138a005.

[161] I. Bertini, C. Del Bianco, I. Gelis, N. Katsaros, C. Luchinat, G. Parigi, M. Peana, A. Provenzani, and M.A. Zoroddu. Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. *Proc. Natl. Acad. Sci.*, 101(18):6841–6846, 2004. doi: https://doi.org/10.1073/pnas.0308641101.

[162] S. Dasgupta, X. Hu, P.H.J. Keizers, W. Liu, C. Luchinat, M. Nagulapalli, M. Overhand, G. Parigi, L. Sgheri, and M. Ubbink. Narrowing the conformational space sampled by two-domain proteins with paramagnetic probes in both domains. *J. Biomol. NMR*, 51(3):253–263, 2011. doi: https://doi.org/10.1007/s10858-011-9532-2.

[163] R. Chattopadhyaya, W.E. Meador, A.R. Means, and F.A. Quiocho. Calmodulin structure refined at 1.7 å resolution. *J. Mol. Biol.*, 228(4):1177–1192, 1992. doi: https://doi.org/10.1016/0022-2836(92)90324-D.

[164] L. Gigli, W. Andrałojć, A. Dalaloyan, G. Parigi, E. Ravera, D. Goldfarb, and C. Luchinat. Assessing protein conformational landscapes: integration of DEER data in Maximum Occurrence analysis. *Phys.*

Chem. Chem. Phys., 20:27429–27438, 2018. doi: https://doi.org/10.1039/C8CP06195E.

[165] A. Martorana, G. Bellapadrona, A. Feintuch, E. Di Gregorio, S. Aime, and D. Goldfarb. Probing protein conformation in cells by EPR distance measurements using Gd^{3+} spin labeling. *JACS*, 136(38):13458–13465, 2014. doi: https://doi.org/10.1021/ja5079392.

[166] Mentink-Vigier F., Collauto A., Feintuch A., Kaminker I., Le V.T., and Goldfarb D. Increasing sensitivity of pulse EPR experiments using echo train detection schemes. *J. Magn. Reson.*, 236:117–125, 2013. doi: https://doi.org/10.1016/j.jmr.2013.08.012.

[167] D. Goldfarb, Y. Lipkin, A. Potapov, Y. Gorodetsky, B. Epel, A. Raitsimring, M. Radoul, and I. Kaminker. HYSCORE and DEER with an upgraded 95 GHz pulse EPR spectrometer. *J. Magn. Reson.*, 194: 8–15, 2008. doi: https://doi.org/10.1016/j.jmr.2008.05.019.

[168] I. Bertini, L. Ferella, C. Luchinat, G. Parigi, M. Petoukhov, E. Ravera, A. Rosato, and D. Svergun. MaxOcc: A web portal for Maximum Occurrence analysis. *J. Biomol. NMR*, 53:271–280, 2012. doi: https://doi.org/10.1007/s10858-012-9638-1.

[169] G. Cornilescu, J.L. Marquardt, M. Ottiger, and A. Bax. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *JACS*, 120(27):6836–6837, 1998. doi: https://doi.org/10.1021/ja9812610.

[170] D. Svergun, C. Barberato, and M.H.J. Koch. CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from

atomic coordinates. *J. Appl. Cryst.*, 28:768–773, 1995. doi: https://doi.org/10.1107/s0021889895007047.

[171] W. Andrałojć, K. Berlin, D. Fushman, C. Luchinat, G. Parigi, E. Ravera, and L. Sgheri. Information content of long-range NMR data for the characterization of conformational heterogeneity. *J. Biomol. NMR*, 62:353–371, 2015. doi: https://doi.org/10.1007/s10858-015-9951-6.

[172] S. Kilic, S. Felekyan, O. Doroshenko, I. Boichenko, M. Dimura, H. Vardanyan, L. Bryan, G. Arya, C. Seidel, and B. Fierz. Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1α. *Nat. Commun.*, 9, 2018. doi: https://doi.org/10.1038/s41467-017-02619-5.

[173] J. Kuffner. Effective sampling and distance metrics for 3D rigid body path planning. *Proceedings - IEEE Int. Conf. Robot. Autom.*, 4:3993–3998, 2004. doi: https://doi.org/10.1109/ROBOT.2004.1308895.

[174] I.C. Felli and R. Pierattelli. *Intrinsically Disordered Proteins Studied by NMR Spectroscopy*. Springer International Publishing, 1 edition, 2015. doi: https://doi.org/10.1007/978-3-319-20164-1.

[175] V.N. Uversky. Intrinsically disordered proteins from A to Z. *Int. J. Biochem. Cell Biol.*, 43(8):1090–1103, 2011. doi: https://doi.org/10.1016/j.biocel.2011.04.001.

[176] C.J. Oldfield and A.K. Dunker. Intrinsically Disordered Proteins and Intrinsically Disordered Protein regions. *Annu. Rev.*

Biochem., 83(1):553–584, 2014. doi: https://doi.org/10.1146/annurev-biochem-072711-164947.

[177] V.N. Uversky, C.J. Oldfield, and A. Dunker. Intrinsically Disordered Proteins in human diseases: Introducing the D2 concept. *Ann. Rev. Biophys.*, 37:215–246, 2008. doi: https://doi.org/10.1146/annurev.biophys.37.032807.125924.

[178] I. Bertini, C. Luchinat, G. Parigi, and E. Ravera. The effect of partial orientation: Residual Dipolar Couplings. In *NMR of Paramagnetic Molecules: Applications to Metallobiomolecules and Models*, pages 25–60. Elsevier, 2 edition, 2017.

[179] J.H. Prestegard, C.M. Bougault, and A.I. Kishore. Residual Dipolar Couplings in structure determination of biomolecules. *Chem. Rev.*, 104(8):3519–3540, 2004. doi: https://doi.org/10.1021/cr030419i.

[180] G. Marius Clore, A.M. Gronenborn, and Nico Tjandra. Direct structure refinement against Residual Dipolar Couplings in the presence of rhombicity of unknown magnitude. *J. Magn. Reson.*, 131(1):159–162, 1998. doi: https://doi.org/10.1006/jmre.1997.1345.

[181] G.M. Clore and A.M. Gronenborn. New methods of structure refinement for macromolecular structure determination by NMR. *PNAS*, 95 (11):5891–5898, 1998. doi: https://doi.org/10.1073/pnas.95.11.5891.

[182] L. Deshmukh, C.D. Schwieters, A. Grishaev, R. Ghirlando, J.L. Baber, and G.M. Clore. Structure and dynamics of full-length HIV-1 capsid protein in solution. *JACS*, 135(43):16133–16147, 2013. doi: https://doi.org/10.1021/ja406246z.

[183] S. Du, L. Betts, R. Yang, H. Shi, J. Concel, J. Ahn, C. Aiken, P. Zhang, and J.I. Yeh. Structure of the HIV-1 full-length capsid protein in a conformationally trapped unassembled state induced by small-molecule binding. *J. Mol. Biol.*, 406(3):371–386, 2011. doi: https://doi.org/10.1016/j.jmb.2010.11.027.

[184] O. Pornillos, B. Ganser-Pornillos, and M. Yeager. Atomic-level modelling of the HIV capsid. *Nature*, 469:424–7, 2011. doi: https://doi.org/10.1038/nature09640.

[185] O. Pornillos, B.K. Ganser-Pornillos, B.N. Kelly, Y. Hua, F.G. Whitby, C.D. Stout, W.I. Sundquist, C.P. Hill, and M. Yeager. X-ray structures of the hexameric building block of the HIV capsid. *Cell*, 137 (7):1282–1292, 2009. doi: https://doi.org/10.1016/j.cell.2009.04.063.

[186] A. Carlon, L. Gigli, E. Ravera, G. Parigi, A.M. Gronenborn, and C. Luchinat. Assessing structural preferences of unstructured protein regions by NMR. *Biophys. J*, 117(10):1948–1953, 2019. doi: https://doi.org/10.1016/j.bpj.2019.10.008.

[187] T. Gamble, SH.. Yoo, F. Vajdos, U. Schwedler, D. Worthylake, H. Wang, J. McCutcheon, W. Sundquist, and C. Hill. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science*, 278:849–53, 11 1997. doi: https://doi.org/10.1126/science.278.5339.849.

[188] D.K. Worthylake, H. Wang, S. Yoo, W.I. Sundquist, and C.P. Hill. Structures of the HIV-1 capsid protein dimerization domain at 2.6 Å

resolution. *Acta Cryst. D*, 55(1):85–92, 1999. doi: https://doi.org/10.1107/S0907444998007689.

[189] F. Ternois, J. Sticht, S. Duquerroy, H.G. Kräusslich, and F.A. Rey. The HIV-1 capsid protein C-terminal domain in complex with a virus assembly inhibitor. *Nat. Struct. Mol. Biol.*, 12:678–82, 09 2005. doi: https://doi.org/10.1038/nsmb967.

[190] I.L. Byeon, X. Meng, J. Jung, G. Zhao, R. Yang, J. Ahn, J. Shi, J. Concel, C. Aiken, P. Zhang, and A.M. Gronenborn. Structural convergence between cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell*, 139(4):780–790, 2009. doi: https://doi.org/10.1016/j.cell.2009.10.010.

[191] G.M. Clore, M.R. Starich, and A.M. Gronenborn. Measurement of Residual Dipolar Couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *JACS*, 120(40): 10571–10572, 1998. doi: https://doi.org/10.1021/ja982592f.

[192] M.R. Hansen, L. Mueller, and A. Pardi. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.*, 5(12):1065–1074, 1998. doi: https://doi.org/10.1038/4176.

[193] M. Rückert and G. Otting. Alignment of biological macromolecules in novel nonionic liquid crystalline media for nmr experiments. *JACS*, 122:7793–7797, 2000. doi: https://doi.org/10.1021/ja001068h.

[194] M. Zweckstetter and A. Bax. Characterization of molecular alignment

in aqueous suspensions of Pf1 bacteriophage. *J. Biomol. NMR*, 20(4): 365–377, 2001. doi: https://doi.org/10.1023/a:1011263920003.

[195] M. Zweckstetter. Determination of molecular alignment tensors without backbone resonance assignment: Aid to rapid analysis of protein-protein interactions. *J. Biomol. NMR*, 27:41–56, 10 2003. doi: https://doi.org/10.1023/A:1024768328860.

[196] M. Rinaldelli, E. Ravera, V. Calderone, G. Parigi, G.N. Murshudov, and C. Luchinat. Simultaneous use of solution NMR and X-ray data in REFMAC5 for joint refinement/detection of structural differences. *Acta Crystallogr. D*, 70(4):958–967, 2014. doi: https://doi.org/101107/S1399004713034160.

[197] C. Dominguez, R. Boelens, and A.M.J.J. Bonvin. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *JACS*, 125(7):1731–1737, 2003. doi: https://doi.org/10.1021/ja026939x.

[198] G.C.P. van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastritis, E. Karaca, A.S.J. Melquiond, M. van Dijk, S.J. de Vries, and A.M.J.J. Bonvin. The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.*, 428(4):720–725, 2016. doi: https://doi.org/10.1016/j.jmb.2015.09.014. Comput. Mol. Biol.

[199] W. Andrałojć and E. Ravera. Treating biomacromolecular conformational variability. In *Paramagnetism in Experimental Biomolecular*

NMR, pages 107–133. The Royal Society of Chemistry, 1 edition, 2018. doi: https://doi.org/10.1039/9781788013291-00107.

[200] M. Longinetti, C. Luchinat, G. Parigi, and L. Sgheri. Efficient determination of the most favoured orientations of protein domains from paramagnetic NMR data. *Inverse Probl.*, 22:1485, 2006. doi: https://doi.org/10.1088/0266-5611/22/4/019.

[201] S. Bottaro and K. Lindorff-Larsen. Biophysical experiments and biomolecular simulations: A perfect match? *Science*, 361:355–360, 2018. doi: https://doi.org/10.1126/science.aat4010.

[202] S. Orioli, A.H. Larsen, S. Bottaro, and K. Lindorff-Larsen. How to learn from inconsistencies: Integrating molecular simulations with experimental data, 2019.

[203] W. Andrałojć, Y. Hiruma, W. Liu, E. Ravera, M. Nojiri, G. Parigi, C. Luchinat, and M. Ubbink. Identification of productive and futile encounters in an electron transfer protein complex. *PNAS*, 114(10): E1840–E1847, 2017. doi: https://doi.org/10.1073/pnas.1616813114.

[204] A.A. Auer, V.A. Tran, B. Sharma, G.L. Stoychev, D. Marx, and F. Neese. A case study of Density Functional Theory and domain-based local pair natural orbital coupled cluster for vibrational effects on EPR hyperfine coupling constants: vibrational perturbation theory versus ab initio molecular dynamics. *Mol. Phys.*, 118 (19-20):e1797916, 2020. doi: https://doi.org/10.1080/00268976.2020.1797916.

[205] S. Vogler, J.C.B. Dietschreit, L.D.M. Peters, and C. Ochsenfeld. Important components for accurate hyperfine coupling constants: electron correlation, dynamic contributions, and solvation effects. *Mol. Phys.*, 118(19-20):e1772515, 2020. doi: https://doi.org/10.1080/00268976.2020.1772515.

[206] P. Valverde, J.I. Quintana, J.I. Santos, A. Ardá, and J. Jiménez-Barbero. Novel NMR avenues to explore the conformation and interactions of glycans. *ACS Omega*, 4(9):13618–13630, 2019. doi: https://doi.org/10.1021/acsomega.9b01901.

[207] C. Nitsche and G. Otting. Intrinsic and extrinsic paramagnetic probes. In *New Developments in NMR*, pages 42–48. Royal Society of Chemistry, 2018. doi: https://doi.org/10.1039/9781788013291-00042.

[208] Q. Miao, W.M. Liu, T. Kock, A. Blok, M. Timmer, M. Overhand, and M. Ubbink. A double-armed, hydrophilic transition metal complex as a paramagnetic NMR probe. *Angew. Chem. Int. Ed.*, 58:13093–13100, 2019. doi: https://doi.org/10.1002/anie.201906049.

[209] D. Joss, F. Winter, and D. Häussinger. A novel, rationally designed lanthanoid chelating tag delivers large paramagnetic structural restraints for biomolecular NMR. *Chem. Commun.*, 2020. doi: https://doi.org/10.1039/D0CC04337K.

[210] D. Shishmarev and G. Otting. How reliable are pseudocontact shifts induced in proteins and ligands by mobile paramagnetic metal tags?

A modelling study. *J. Biomol. NMR*, 56(3):203–216, 2013. doi: https://doi.org/10.1007/s10858-013-9738-6.

[211] A. Lunghi, F. Totti, S. Sanvito, and R. Sessoli. Intra-molecular origin of the spin-phonon coupling in slow-relaxing molecular magnets. *Chem. Sci.*, 8:6051–6059, 2017. doi: https://doi.org/10.1039/C7SC02832F.

[212] C. Göger, K. Lutz, and E. Brunner. NMR studies of biomineralisation. *Prog. Nucl. Magn. Reson. Spectrosc.*, 54:54–68, 2009. doi: https://doi.org/10.1016/j.pnmrs.2008.02.003.

[213] M. Sumper and E. Brunner. Silica biomineralisation in diatoms: The model organism thalassiosira pseudonana. *Chem. Bio. Chem.*, 9(8): 1187–1194, 2008. doi: https://doi.org/10.1002/cbic.200700764.

[214] X. Song, Z. Jiang, L. Li, and H. Wu. Immobilization of β-glucuronidase in lysozyme-induced biosilica particles to improve its stability. *Front. Chem. Sci. Eng.*, 8:353–361, 2014. doi: https://doi.org/10.1007/s11705-014-1421-2.

[215] D. Wisser, S.I. Brückner, F. M Wisser, G. Althoff-Ospelt, J. Getzschmann, S. Kaskel, and E. Brunner. ^{1}H-^{13}C-^{29}Si triple resonance and REDOR solid-state NMR. A tool to study interactions between biosilica and organic molecules in diatom cell walls. *Solid State Nucl. Magn. Reson.*, 66-67:33–39, 2015. doi: https://doi.org/10.1016/j.ssnmr.2014.12.007.

[216] S. Kababya, A. Gal, K. Kahil, S. Weiner, L. Addadi, and A. Schmidt. Phosphate-water interplay tunes amorphous calcium

carbonate metastability: Spontaneous phase separation and crystallization vs stabilization viewed by solid state NMR. *JACS*, 137(2): 990–998, 2015. doi: https://doi.org/10.1021/ja511869g.

[217] A. Akiva-Tal, S. Kababya, Y.S. Balazs, L. Glazer, A. Berman, A. Sagi, and A. Schmidt. In situ molecular NMR picture of bioavailable calcium stabilized as amorphous $CaCO_3$ biomineral in crayfish gastroliths. *PNAS*, 108:14763–14768, 2011. doi: https://doi.org/10.1073/pnas.1102608108.

[218] Y. Wang, S. Von Euw, F.M. Fernandes, S. Cassaignon, M. Selmane, G. Laurent, G. Pehau-Arnaudet, C. Coelho, L. Bonhomme-Coury, M. Giraud-Guille, F. Babonneau, T. Azaïs, and N. Nassif. Water-mediated structuring of bone apatite. *Nat. Mater.*, 12(12):1144–1153, 2013. doi: https://doi.org/10.1038/nmat3787.

[219] S. McNeill, P. Gor'kov, J. Struppe, W. Brey, and J. Long. Optimizing ssNMR experiments for dilute proteins in heterogeneous mixtures at high magnetic fields. *MRC*, 45:S209–S220, 2007. doi: https://doi.org/10.1002/mrc.2146.

[220] L. Banci, L. Barbieri, V. Calderone, F. Cantini, L. Cerofolini, S. Ciofi-Baffoni, I.C. Felli, M. Fragai, M. Lelli, C. Luchinat, E. Luchinat, G. Parigi, M. Piccioli, R. Pierattelli, E. Ravera, A. Rosato, L. Tenori, and P. Turano. Biomolecular NMR at 1.2 GHz, 2019.

[221] A. Lesage, M. Lelli, D. Gajan, M.A. Caporini, V. Vitzthum, P. MiÉville, J. Alauzun, A. Roussey, C. Thieuleux, A. Mehdi, G. Bodenhausen, C. Coperet, and L. Emsley. Surface enhanced NMR spec-

troscopy by Dynamic Nuclear Polarization. *JACS*, 132(44):15459–15461, 2010. doi: https://doi.org/10.1021/ja104771z.

[222] J.W. Wiench, V.S.-Y. Lin, and M. Pruski. ^{29}Si NMR in solid state with CPMG acquisition under MAS. *J. Magn. Reson.*, 193(2):233–242, 2008. doi: https://doi.org/10.1016/j.jmr.2008.05.007.

[223] P.P. Man, Bonhomme, and F. C. Babonneau. Denoising NMR time-domain signal by Singular-Value Decomposition accelerated by graphics processing units. *Solid State Nucl. Magn. Reson.*, 61-62: 28–34, 2014. doi: https://doi.org/10.1016/j.ssnmr.2014.05.001.

[224] Wesley D.M.II, A.S. Shaikh, S.J. Muller, and J.D. Keasling. Enzyme immobilization via silaffin-mediated autoencapsulation in a biosilica support. *Biotechnol. Prog.*, 25(2):417–423, 2009. doi: https://doi.org/10.1002/btpr.136.

[225] J.M. Bolivar, I. Eisl, and B. Nidetzky. Advanced characterization of immobilized enzymes as heterogeneous biocatalysts. *Catal. Today*, 259:66–80, 2016. doi: https://doi.org/10.1016/j.cattod.2015.05.004.

[226] G. Goobes, R. Goobes, O. Schueler-Furman, D. Baker, P.S. Stayton, and G.P Drobny. Folding of the C-terminal bacterial binding domain in statherin upon adsorption onto hydroxyapatite crystals. *PNAS*, 103(44):16083–16088, 2006. doi: https://doi.org/10.1073/pnas.0607193103.

[227] J. Koehler and J. Meiler. Expanding the utility of NMR restraints with paramagnetic compounds: Background and practical aspects.

Prog. Nucl. Magn. Reson. Spectrosc., 59:360–89, 2011. doi: https://doi.org/10.1016/j.pnmrs.2011.05.001.

[228] A. Roehrich and G. Drobny. Solid-state NMR studies of biomineralization peptides and proteins. *Acc. Chem. Res.*, 46(9):2136–2144, 2013. doi: https://doi.org/10.1021/ar300321e.

[229] i. Matlahov, T. Iline-Vul, M. Abayev, E.M.Y. Lee, M. Nadav-Tsubery, K. Keinan-Adamsky, J.J. Gray, and G. Goobes. Interfacial mineral-peptide properties of a mineral binding peptide from osteonectin and bone-like apatite. *Chem. Mater.*, 27(16):5562–5569, 2015. doi: https://doi.org/10.1021/acs.chemmater.5b01696.

[230] M. Fragai, C. Luchinat, T. Martelli, E. Ravera, I. Sagi, I. Solomonov, and Y. Udi. ssNMR of biosilica-entrapped enzymes permits an easy assessment of preservation of native conformation in atomic detail. *Chem. Commun.*, 50:421–423, 2014. doi: https://doi.org/10.1039/C3CC46896H.

[231] T. Martelli, E. Ravera, A. Louka, L. Cerofolini, M. Hafner, M. Fragai, C.F. W. Becker, and C. Luchinat. Atomic-level quality assessment of enzymes encapsulated in bioinspired silica. *Chem. A Eur. J.*, 22(1):425–432, 2016. doi: https://doi.org/10.1002/chem.201503613.

[232] E. Ravera, V. Michaelis, T.C. Ong, E. Keeler, T. Martelli, M. Fragai, R. Griffin, and C. Luchinat. Biosilica-entrapped enzymes studied by using Dynamic Nuclear-Polarization-enhanced high-field NMR spectroscopy. *Chem. Phys. Chem.*, 16:2751–2754, 2015. doi: https://doi.org/10.1002/cphc.201500549.

[233] Y. Yan, B. Hao, X. Wang, and G. Chen. Bio-inspired synthesis of titania with polyamine induced morphology and phase transformation at room-temperature: Insight into the role of the protonated amino group. *Dalton Trans.*, 42:12179–12184, 2013. doi: https://doi.org/10.1039/C3DT50359C.

[234] J.L. Sumerel, W. Yang, D. Kisailus, J.C. Weaver, J.H. Choi, and D.E. Morse. Biocatalytically templated synthesis of titanium dioxide. *Chem. Mater.*, 15(25):4804–4809, 2003. doi: https://doi.org/10.1021/cm030254u.

[235] S.L. Sewell and David W. Wright. Biomimetic synthesis of titanium dioxide utilizing the R5 peptide derived from cylindrotheca fusiformis. *Chem. Mater.*, 18(13):3108–3113, 2006. doi: https://doi.org/10.1021/cm060342p.

[236] H.R. Luckarift, Dickerson, M.B., K.H Sandhage, and J.C. Spain. Rapid, room-temperature synthesis of antibacterial bionanocomposites of lysozyme with amorphous silica or titania. *Small*, 2(5):640–643, 2006. doi: https://doi.org/10.1002/smll.200500376.

[237] J. Schmidt and W. Vogelsberger. Aqueous long-term solubility of titania nanoparticles and titanium(IV) hydrolysis in a sodium chloride system studied by adsorptive stripping voltammetry. *J. Solution Chem.*, 38:1267–1282, 2009. doi: https://doi.org/10.1007/s10953-009-9445-9.

[238] A. Hernández-Gordillo, A. Hernandez-Arana, A. Campero-Celis, and L. Vera-Robles. TiBALDH as a precursor for biomimetic TiO_2 syn-

thesis: stability aspects in aqueous media. *RSC Adv.*, 9:34559–34566, 2019. doi: https://doi.org/10.1039/C9RA05923G.

[239] M. Kakihana, K. Tomita, V. Petrykin, M. Tada, S. Sasaki, and Y. Nakamura. Chelating of titanium by lactic acid in the water-soluble diammonium tris(2-hydroxypropionato)titanate(IV). *Inorg. Chem.*, 43(15):4546–4548, 2004. doi: https://doi.org/10.1021/ic0400311.

[240] A. Forgács, K. Moldov án, P. Herman, E. Baranyai, I. Fábián, G. Lente, and J. Kalmár. Kinetic model for hydrolytic nucleation and growth of TiO_2 nanoparticles. *J. Phys. Chem. C*, 122(33):19161–19170, 2018. doi: https://doi.org/10.1021/acs.jpcc.8b04227.

[241] J. Bogdan, A. Jackowska-Tracz, J. Zarzynska, and J. Pławińska-Czarnak. Chances and limitations of nanosized titanium dioxide practical application in view of its physicochemical properties. *Nanoscale Research Letters*, 10, 2015. doi: https://doi.org/10.1186/s11671-015-0753-2.

[242] J. Liang, Z. Liang, R. Zou, and Y. Zhao. Heterogeneous catalysis in zeolites, mesoporous silica, and metal-organic frameworks. *Adv. Mater.*, 29:1701139, 2017. doi: https://doi.org/10.1002/adma.201701139.

[243] K. Persson. Materials Data on TiO_2 (SG:141) by Materials Project, 2014.

[244] K. Persson. Materials Data on SiO_2 (SG:154) by Materials Project, 2014.

[245] R. Rakhit and A. Chakrabartty. Structure, folding, and misfolding of Cu, Zn superoxide dismutase in amyotrophic lateral sclerosis. *Biochim. Biophys. Acta Mol. Basis Dis.*, 1762(11):1025–1037, 2006. doi: https://doi.org/10.1016/j.bbadis.2006.05.004.

[246] Volkov A.N., P. Nicholls, and J.A.R. Worrall. The complex of cytochrome c and cytochrome c peroxidase: The end of the road? *BBA*, 1807(11):1482–1503, 2011. doi: https://doi.org/10.1016/j.bbabio.2011.07.010.

[247] C.C. Blake, D.F. Koenig, G.A. Mair, A.C. North, D.C. Phillips, and V.R. Sarma. Structure of hen egg-white lysozyme. A three-dimensional fourier synthesis at 2 angstrom resolution. *Nature*, 206 (4986):757–761, 1965. doi: https://doi.org/10.1038/206757a0.

[248] A.C. Young, R.F. Tilton, and J.C. Dewan. Thermal expansion of hen egg-white lysozyme: Comparison of the 1.9 å resolution structures of the tetragonal form of the enzyme at 100 K and 298 K. *J. Mol. Biol.*, 235(1):302–317, 1994. doi: https://doi.org/10.1016/S0022-2836(05)80034-8.

[249] M.A. Hough, J.G. Grossmann, S.V. Antonyuk, R.W. Strange, P.A. Doucette, J.A. Rodriguez, L.J. Whitson, P.J. Hart, L.J. Hayward, J.S. Valentine, and S.S. Hasnain. Dimer destabilization in superoxide dismutase may result in disease-causing properties: Structures of motor neuron disease mutants. *PNAS*, 101(16):5976–5981, 2004. doi: https://doi.org/10.1073/pnas.0305143101.

[250] G.W. Bushnell, G.V. Louie, and G.D. Brayer. High-resolution three-

dimensional structure of horse heart cytochrome c. *J. Mol. Biol.*, 214(2):585–595, 1990. doi: https://doi.org/10.1016/0022-2836(90)90200-6.

[251] J.C. Cheetham, PJ.. Artymiuk, and D.C. Phillips. Refinement of an enzyme complex with inhibitor bound at partial occupancy. Hen egg-white lysozyme and tri-N-acetylchitotriose at 1.75 å resolution. *J. Mol. Biol.*, 224(3):613–628, 1992. doi: https://doi.org/10.1016/0022-2836(92)90548-x.

[252] S. Giuntini, L. Cerofolini, E. Ravera, M. Fragai, and C. Luchinat. Atomic structural details of a protein grafted onto gold nanoparticles. *Sci. Rep.*, 7, 12 2017. doi: https://doi.org/10.1038/s41598-017-18109-z.

[253] Tauber H. Oxidation of pyrogallol to purpurogallin by crystallin catalase. *J. Biol. Chem.*, 1(205):395–400, 953.

[254] A.L. Raphael and H.B. Gray. Axial ligand replacement in horse heart cytochrome c by semisynthesis. *Protein Struct. Funct. Bioinform.*, 6 (3):338–340, 1989. doi: https://doi.org/10.1002/prot.340060316.

[255] K.L. Bren, H.B. Gray, L. Banci, I. Bertini, and P. Turano. Paramagnetic ^{1}H NMR spectroscopy of the cyanide derivative of Met80Ala-iso-1-cytochrome c. *JACS*, 117(31):8067–8073, 1995. doi: https://doi.org/10.1021/ja00136a003.

[256] T. Josephs, M. Liptak, G. Hughes, A. Lo, R. Smith, S. Wilbanks, K. Bren, and E. Ledgerwood. Conformational change and human

cytochrome c function: Mutation of residue 41 modulates caspase activation and destabilizes Met-80 coordination. *JBIC*, pages 289–297, 2013. doi: https://doi.org/10.1007/s00775-012-0973-1.

[257] V.M. Mastikhin, I.L. Mudrakovsky, and A.V. Nosov. ^{1}H NMR Magic Angle Spinning (MAS) studies of heterogeneous catalysis. *Prog. Nucl. Magn. Reson. Spectrosc.*, 23(3):259–299, 1991. doi: https://doi.org/10.1016/0079-6565(91)80006-N.

[258] M. Leskes, R.S. Thakur, P.K. Madhu, N.D. Kurur, and S. Vega. Bimodal Floquet description of heteronuclear dipolar decoupling in solid-state Nuclear Magnetic Resonance. *J. Chem. Phys.*, 127(2): 024501, 2007. doi: https://doi.org/10.1063/1.2746039.

[259] E. Ravera, L. Cerofolini, T. Martelli, A. Louka, M. Fragai, and C. Luchinat. ^{1}H-detected solid-state NMR of proteins entrapped in bioinspired silica: A new tool for biomaterials characterization. *Sci. Rep.*, 6:27851, 06 2016. doi: https://doi.org/10.1038/srep27851.

[260] T.M. Stawski, D.B. van den Heuvel, R. Besselink, D.J. Tobler, and L.G. Benning. Mechanism of silica-lysozyme composite formation unravelled by in situ fast SAXS. *Beilstein J. Nanotechnol.*, 10:182–197, 2019. doi: https://doi.org/10.3762/bjnano.10.17.

[261] K. Sugam, Vinod A.K., and P. Callow. pH-dependent interaction and resultant structures of silica nanoparticles and lysozyme protein. *Langmuir*, 30(6):1588–1598, 2014. doi: https://doi.org/10.1021/la403896h.

[262] B. Bharti, J. Meissner, S.H.L. Klapp, and G.H. Findenegg. Bridging interactions of proteins with silica nanoparticles: The influence of pH, ionic strength and protein concentration. *Soft Matter*, 10:718–728, 2014. doi: https://doi.org/10.1039/C3SM52401A.

[263] M.B. Cardoso, H.R. Luckarift, V.S. Urban, H. O'Neill, and G.R. Johnson. Protein localization in silica nanospheres derived via biomimetic mineralization. *Adv. Funct. Mater.*, 20(18):3031–3038, 2010. doi: https://doi.org/10.1002/adfm.201000144.

[264] M. Sumper and E. Brunner. Learning from diatoms: Nature's tools for the production of nanostructured silica. *Adv. Funct. Mater.*, 16: 17–26, 2006. doi: https://doi.org/10.1002/adfm.200500616.

[265] Y. Wang, J. Cai, Y. Jiang, X. Jiang, and D. Zhang. Preparation of biosilica structures from frustules of diatoms and their applications: Current state and perspectives. *Appl. Microbiol. Biotechnol.*, 97:453–460, 11 2012. doi: https://doi.org/10.1007/s00253-012-4568-0.

[266] L. Gigli, E. Ravera, V. Calderone, and C. Luchinat. On the mechanism of bioinspired formation of inorganic oxides: Structural evidence of the electrostatic nature of the interaction between a mononuclear inorganic precursor and lysozyme. *Biomolecules*, 11:43, 2021. doi: https://doi.org/10.3390/biom11010043.

[267] T. Coradin and J. Livage. Aqueous silicates in biological sol-gel applications: new perspectives for old precursors. *Acc. Chem. Res.*, 40 (9):819–826, 2007. doi: https://doi.org/10.1021/ar068129m.

[268] A. Jantschke, E. Koers, D. Mance, M. Weingarth, E. Brunner, and M. Baldus. Insight into the supramolecular architecture of intact diatom biosilica from DNP-supported solid-state NMR spectroscopy. *Angew. Chem. Int. Ed.*, 54(50):15069–15073, 2015. doi: https://doi.org/10.1002/anie.201507327.

[269] Y. Geiger, H.E. Gottlieb, Ü. Akbey, H. Oschkinat, and G. Goobes. Studying the conformation of a silaffin-derived pentalysine peptide embedded in bioinspired silica using solution and Dynamic Nuclear Polarization Magic-Angle Spinning NMR. *JACS*, 138(17):5561–5567, 2016. doi: https://doi.org/10.1021/jacs.5b07809.

[270] E.L. Buckle, J.S. Lum, A.M. Roehrich, R.E. Stote, B. Vandermoon, M. Dracinsky, S.F. Filocamo, and G.P. Drobny. Serine-lysine peptides as mediators for the production of titanium dioxide: Investigating the effects of primary and secondary structures using solid-state NMR spectroscopy and DFT calculations. *J. Phys. Chem. B*, 122(17):4708–4718, 2018. doi: https://doi.org/10.1021/acs.jpcb.8b00745.

[271] S. Patwardhan and S. Clarson. Silicification and Biosilicification. Part 6. Poly-L-Histidine mediated synthesis of silica at neutral pH. *J. Inorg. Organomet. Polym. Mater.*, 13:49–53, 2003. doi: https://doi.org/10.1023/A:1022952931063.

[272] S. Stoll and A. Schweiger. EasySpin: a comprehensive software package for spectral simulation and analysis in EPR. *J. Magn. Reson.*, 178:42–55, 2006. doi: https://doi.org/10.1016/j.jmr.2005.08.013.

[273] H. Schwalbe, S. Grimshaw, A. Spencer, M. Buck, J. Boyd, C. Dob-

son, C. Redfield, and L. Smith. A refined solution structure of hen lysozyme determined using Residual Dipolar Coupling data. *Protein Sci.*, 10:677–688, 2001. doi: https://doi.org/10.1110/ps.43301.

[274] P. Bernado and M. Blackledge. Anisotropic small amplitude peptide plane dynamics in proteins from Residual Dipolar Couplings. *JACS*, 126(15):4907–4920, 2004. doi: https://doi.org/10.1021/ja036977w.

[275] M. Buck, J. Boyd, C. Redfield, D.A. MacKenzie, D.J. Jeenes, D.B. Archer, and C.M. Dobson. Structural determinants of protein dynamics: analysis of ^{15}N NMR relaxation measurements for main-chain and side-chain nuclei of hen egg white lysozyme. *Biochemistry*, 34 (12):4041–4055, 1995. doi: https://doi.org/10.1021/bi00012a023.

[276] Y. Kamatari, H. Yamada, K. Akasaka, J. Jones, C. Dobson, and L. Smith. Response of native and denatured hen lysozyme to high pressure studied by ^{15}N/^{1}H NMR spectroscopy. *Eur. J. Biochem.*, 268:1782–1793, 2003. doi: https://doi.org/10.1046/j.1432-1327.2001.02050.x.

[277] Y. Fang, N. Poulsen, M.B. Dickerson, Y. Cai, S.E. Jones, R.R. Naik, N. Kröger, and K.H. Sandhage. Identification of peptides capable of inducing the formation of titania but not silica via a subtractive bacteriophage display approach. *J. Mater. Chem.*, 18:3871–3875, 2008. doi: https://doi.org/10.1039/B806797J.

[278] C. Redfield and C.M. Dobson. Sequential proton NMR assignments and secondary structure of hen egg white lysozyme in solu-

tion. *Biochemistry*, 27(1):122–136, 1988. doi: https://doi.org/10.1021/bi00401a020.

[279] D. Marulanda, M.L. Tasayco, A. McDermott, M. Cataldi, V. Arriaran, and T. Polenova. Magic Angle Spinning solid-state NMR spectroscopy for structural studies of protein interfaces. Resonance assignments of differentially enriched Escherichia coli thioredoxin reassembled by fragment complementation. *JACS*, 126(50):16608–16620, 2004. doi: https://doi.org/10.1021/ja0464589.

[280] E. Ravera, S. Ciambellotti, L. Cerofolini, T. Martelli, T. Kozyreva, C. Bernacchioni, S. Giuntini, M. Fragai, P. Turano, and C. Luchinat. Solid-state NMR of PEGylated proteins. *Angew. Chem. Int. Ed.*, 55 (7):2446–2449, 2016. doi: https://doi.org/10.1002/anie.201510148.

[281] M. Knecht and T.R. Walsh. *Bio-inspired nanotechnology: From surface analysis to applications*. Springer, 2013. doi: https://doi.org/10.1007/978-1-4614-9446-1.

[282] D.B. van den Heuvel, T.M. Stawski, D.J. Tobler, R. Wirth, C.L. Peacock, and L.G. Benning. Formation of silica-lysozyme composites through co-precipitation and adsorption. *Front. Mater.*, 5:19, 2018. doi: https://doi.org/10.3389/fmats.2018.00019.

[283] S. Christiansen, N. Hedin, J. Epping, M. Janicke, Y. Del Amo, M. Demarest, M. Brzezinski, and B. Chmelka. Sensitivity considerations in polarization transfer and filtering using dipole-dipole couplings: Implications for biomineral systems. *Solid State Nucl. Magn. Reson.*, 29:170–82, 2006. doi: https://doi.org/10.1016/j.ssnmr.2005.10.010.

[284] P.K. Vinogradov, E. Madhu and S. Vega. Proton spectroscopy in solid state Nuclear Magnetic Resonance with windowed phase modulated Lee-Goldburg decoupling sequences. *Chem.Phys. Lett.*, 354(3):193–202, 2002. doi: https://doi.org/10.1016/S0009-2614(02)00060-X.

[285] J.P. Antoine, C. Chauvin, and A. Coron. Wavelets and related time-frequency techniques in Magnetic Resonance Spectroscopy. *NMR Biomed.*, 14(4):265–270, 2001. doi: https://doi.org/10.1002/nbm.699.

[286] G. Vivó-Truyols and P.J. Schoenmakers. Automatic selection of optimal Savitzky-Golay smoothing. *Anal. Chem.*, 78(13):4598–4608, 2006. doi: https://doi.org/10.1021/ac0600196.

[287] L. Chiron, M. Agthoven, B. Kieffer, C. Rolando, and M. Delsuc. Efficient denoising algorithms for large experimental datasets and their applications in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *PNAS*, 111:1385–1390, 2014. doi: https://doi.org/10.1073/pnas.1306700111.

[288] G. Laurent, W. Woelffel, V. Barret-Vivin, E. Gouillart, and C. Bonhomme. Denoising applied to spectroscopies - part I: concept and limits. *Appl. Spectrosc. Rev.*, 54(7):602–630, 2019. doi: https://doi.org/10.1080/05704928.2018.1523183.

[289] R. Ghanati, M.K. Hafizi, R. Mahmoudvand, and M. Fallahsafari. Filtering and parameter estimation of surface-NMR data using Singular Spectrum Analysis. *J. Appl. Geophys.*, 130:118 – 130, 2016. doi: https://doi.org/10.1016/j.jappgeo.2016.04.005.

[290] A. De Juan and R. Tauler. Multivariate Curve Resolution (MCR) from 2000: Progress in concepts and applications. *Crit. Rev. Anal. Chem.*, 36:163–176, 2006. doi: https://doi.org/10.1080/10408340600970005.

[291] J.A. Cadzow. Signal enhancement - A composite property mapping algorithm. *IEEE Trans. Signal Process.*, 36(1):49–62, 1988. doi: https://doi.org/10.1109/29.1488.

[292] R. Tauler, A. Smilde, and B. Kowalski. Selectivity, local rank, three-way data analysis and ambiguity in Multivariate Curve Resolution. *J. Chemometr.*, 9(1):31–58, 1995. doi: https://doi.org/10.1002/cem.1180090105.

[293] R. Tauler. Multivariate Curve Resolution applied to second order data. *Chemometr. Intell. Lab. Syst.*, 30(1):133–146, 1995. doi: https://doi.org/10.1016/0169-7439(95)00047-X.

[294] R. Tauler. Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by Multivariate Curve Resolution. *J. Chemometr.*, 15(8):627–646, 2001. doi: https://doi.org/10.1002/cem.654.

[295] H. William, Lawton, and E.A. Sylvestre. Self Modeling Curve Resolution. *Technometrics*, 13(3):617–633, 1971. doi: https://doi.org/10.1080/00401706.1971.10488823.

[296] L. Duponchel, W. Elmi-Rayaleh, C. Ruckebusch, and J.P. Huvenne. Multivariate Curve Resolution methods in Imaging Spectroscopy: influence of extraction methods and instrumental perturbations. *J.*

Chem. Inform. Comput. Sci., 43(6):2057–2067, 2003. doi: https://doi.org/10.1021/ci034097v.

[297] M. Garrido, F.X. Rius, and M.S. Larrechi. Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes. *Anal. Bioanal. Chem.*, 390(8):2059–2066, 2008. doi: https://doi.org/10.1007/s00216-008-1955-6.

[298] J. Jaumot, R. Tauler, and R. Gargallo. Exploratory data analysis of DNA microarrays by Multivariate Curve Resolution. *Anal. Biochem.*, 358(1):76 – 89, 2006. doi: https://doi.org/10.1016/j.ab.2006.07.028.

[299] R. Tauler, S. Lacorte, M. Guillamón, R. Cespedes, P. Viana, and D. Barceló. Chemometric modeling of main contamination sources in surface waters of Portugal. *Environ. Toxicol. Chem.*, 23(3):565–575, 2004. doi: https://doi.org/10.1897/03-176b.

[300] F. Puig-Castellví, I. Alfonso, and R. Tauler. Untargeted assignment and automatic integration of ^{1}H NMR metabolomic datasets using a Multivariate Curve Resolution approach. *Anal. Chim. Acta*, 964: 55–66, 2017. doi: https://doi.org/10.1016/j.aca.2017.02.010.

[301] J. Jaumot, V. Marchán, R. Gargallo, A. Grandas, and R. Tauler. Multivariate Curve Resolution applied to the analysis and resolution of two-dimensional [^{1}H, ^{15}N] NMR reaction spectra. *Anal. Chem.*, 76 (23):7094–7101, 2004. doi: https://doi.org/10.1021/ac049509t.

[302] W.L. Lawton and E.A. Sylvestre. Self Modeling Curve Resolution.

Technometrics, 13(3):617–633, 1971. doi: https://doi.org/10.2307/1267173.

[303] A. De Juan and R. Tauler. Multivariate Curve Resolution-Alternating Least Squares for spectroscopic data. In *Resolving Spectral Mixtures*, volume 30 of *Data Handling in Science and Technology*, pages 5–51. Elsevier, 2016. doi: https://doi.org/10.1016/B978-0-444-63638-6.00002-4.

[304] F. Bruno, R. Francischello, G. Bellomo, L. Gigli, A. Flori, L. Menichetti, L. Tenori, C. Luchinat, and E. Ravera. Multivariate Curve Resolution for 2D solid-state NMR spectra. *Anal. Chem.*, 92(6):4451–4458, 2020. doi: https://doi.org/10.1021/acs.analchem.9b05420.

[305] J.J. Helmus and C.P. Jaroniec. Nmrglue: an open source Python package for the analysis of multidimensional NMR data. *J. Biomol. NMR*, 55(4):355–367, 2013. doi: https://doi.org/10.1007/s10858-013-9718-x.

[306] W. Windig and J. Guilment. Interactive self-modeling mixture analysis. *Anal. Chem.*, 63(14):1425–1432, 1991. doi: https://doi.org/10.1021/ac00014a016.

[307] D. Marks and S. Vega. A theory for cross-polarization NMR of non-spinning and spinning samples. *J. Magn. Reson. A*, 118(2):157–172, 1996. doi: https://doi.org/10.1006/jmra.1996.0024.

[308] A.Z. Summers, C.R. Iacovella, O.M. Cane, P.T. Cummings, and C. McCabe. A transferable, multi-resolution coarse-grained model

for amorphous silica nanoparticles. *J. Chem. Theory Comput.*, 15(5): 3260–3271, 2019. doi: https://doi.org/10.1021/acs.jctc.8b01269.

[309] G. Pérez-Sánchez, S. Chien, J.R.B. Gomes, M.N.D.S. Cordeiro, S.M. Auerbach, P.A. Monson, and M. Jorge. Multiscale model for the templated synthesis of mesoporous silica: The essential role of silica oligomers. *Chem. Mater.*, 28(8):2715–2727, 2016. doi: https://doi.org/10.1021/acs.chemmater.6b00348.

[310] A. Centi and M. Jorge. Molecular simulation study of the early stages of formation of bioinspired mesoporous silica materials. *Langmuir*, 32(28):7228–7240, 2016. doi: https://doi.org/10.1021/acs.langmuir.6b01731.

[311] R.K. Hester, J.L. Ackerman, B.L. Neff, and J.S. Waugh. Separated Local Field Spectra in NMR: Determination of structure of solids. *Phys. Rev. Lett.*, 36:1081–1083, 1976. doi: https://doi.org/10.1103/PhysRevLett.36.1081.

[312] F.M. Marassi and S.J. Opella. Using Pisa pies to resolve ambiguities in angular constraints from PISEMA spectra of aligned proteins. *J. Biomol. NMR*, 23(3):239–242, 2002. doi: https://doi.org/10.1023/a:1019887612018.

[313] S. Jayanthi and K.V. Ramanathan. 2n-SEMA-a robust solid state Nuclear Magnetic Resonance experiment for measuring heteronuclear dipolar couplings in static oriented systems using effective transverse spin-lock. *J. Chem. Phys.*, 132(13):134501, 2010. doi: https://doi.org/10.1063/1.3336815.

[314] C.H. Wu and S.J. Opella. Proton-detected separated Local Field Spectroscopy. *J. Magn. Reson.*, 190(1):165 – 170, 2008. doi: https://doi.org/10.1016/j.jmr.2007.10.002.

[315] A. Ramamoorthy, Y. Wei, and D. Lee. Pisema solid-state nmr spectroscopy. In *Annual Reports on NMR Spectroscopy*, volume 52, pages 1–52. Academic Press, 2004. doi: https://doi.org/10.1016/S0066-4103(04)52001-X.

[316] P. Zhang, C. Dunlap, P. Florian, P.J. Grandinetti, I. Farnan, and J.F. Stebbins. Silicon site distributions in an alkali silicate glass derived by two-dimensional ^{29}Si Nuclear Magnetic Resonance. *J. Non-Cryst. Solids*, 204(3):294 – 300, 1996. doi: https://doi.org/10.1016/S0022-3093(96)00601-1.

[317] S. Hiller, I. Ibraghimov, G. Wagner, and V.Y. Orekhov. Coupled decomposition of four-dimensional NOESY spectra. *JACS*, 131(36): 12970–12978, 2009. doi: https://doi.org/10.1021/ja902012x.

[318] K. Kosiński, J. Stanek, M.J. Górka, S. Żerko, and W. Koźmiński. Reconstruction of non-uniformly sampled five-dimensional NMR spectra by signal separation algorithm. *J. Biomol. NMR*, 68:129–138, 2017. doi: https://doi.org/10.1007/s10858-017-0095-8.

[319] A. Shchukina, M. Urbańczyk, P. Kasprzak, and K. Kazimierczuk. Alternative data processing techniques for serial NMR experiments. *Concepts Magn. Reson. A*, 46A(2):e21429, 2017. doi: https://doi.org/10.1002/cmr.a.21429.

[320] J. Stanek, R. Augustyniak, and W. Koźmiński. Suppression of sampling artefacts in high-resolution four-dimensional NMR spectra using signal separation algorithm. *J. Magn. Reson.*, 214:91–102, 2012. doi: https://doi.org/10.1016/j.jmr.2011.10.009.

[321] Yoh Matsuki, Matthew T. Eddy, Robert G. Griffin, and Judith Herzfeld. Rapid three-dimensional MAS NMR spectroscopy at critical sensitivity. *Angew. Chem. Int. Ed.*, 49(48):9215–9218, 2010. doi: https://doi.org/10.1002/anie.201003329.

[322] N.M. Balsgart, M. Mulbjerg, Z. Guo, K. Bertelsen, and T. Vosegaard. High throughput identification and quantification of phospholipids in complex mixtures. *Anal. Chem.*, 88(4):2170–2176, 2016. doi: https://doi.org/10.1021/acs.analchem.5b03798.

[323] K. Chen and N. Tjandra. The use of Residual Dipolar Coupling in studying proteins by NMR. *Top. Curr. Chem.*, 326:47–67, 2011. doi: https://doi.org/10.1007/128_2011_215.

[324] G. Jeschke. DEER distance measurements on proteins. *Annu. Rev. Phys. Chem.*, 63:419–46, 2012. doi: https://doi.org/10.1146/annurev-physchem-032511-143716.

[325] M. Pannier, S. Veit, A. Godt, G. Jeschke, and H.W. Spiess. Dead-time free measurement of dipole-dipole interactions between electron spins. *J. Magn. Reson.*, 142:331–40, 2000. doi: https://doi.org/10.1006/jmre.1999.1944.

[326] G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J.E. Banham, C.R. Timmel, D. Hilger, and H. Jung. DeerAnalysis2006 -

a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.*, 30:473–498, 2006. doi: https://doi.org/10.1007/BF03166213.

[327] M. Zweckstetter and A. Bax. Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. *J. Biomol. NMR*, 20(4): 365–377, 2001. doi: https://doi.org/10.1023/A:1011263920003.

[328] G. Schröder, W. O'Dell, D. Myles, A. Kovalevsky, and F. Meilleur. IMAGINE: neutrons reveal enzyme chemistry. *Acta Crystallogr. D*, 74:778–786, 2018. doi: https://doi.org/10.1107/S2059798318001626.

[329] O. Kovalevskiy, R. Nicholls, F. Long, A. Carlon, and G. Murshudov. Overview of refinement procedures within REFMAC 5: Utilizing data from different sources. *Acta Crystallogr. D*, 74:215–227, 2018. doi: https://doi.org/10.1107/S2059798318000979.

[330] M. Zweckstetter and A. Bax. Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. *J. Biomol. NMR*, 23:127–37, 2002. doi: https://doi.org/10.1023/A:1016316415261.

List of Tables

1.1 *Details of the calculations of the magnetic susceptibility tensors [117].* 29

1.2 *Comparison of the experimental and QC-calculated (at the NEVPT2 level) magnetic susceptibility anisotropy tensors at 298 K for structures 1 and 4 [117].* 37

2.1 *MaxOR values for different combinations of conformations [186].* 68

4.1 *Data collection and refinement statistics parameters.* $^{\dagger}R_{merge} = \frac{\sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|}{\sum_{hkl}\sum_i I_i(hkl)}$, *where* $I_i(hkl)$ *is the mean intensity of the i-th observation of symmetry-related reflections hkl.* $^{\ddagger}R = \frac{\sum_{hkl} ||F_O| - |F_C||}{\sum_{hkl} |F_O|}$, *where* F_C *is the calculated protein structure factor from the atomic model (*R_{free} *was calculated with a randomly selected 5% of the reflections).* 91

4.2 *Processing parameters of the 2D spectra [304].* 110

4.3 *Position (PP), linewidth (LW), and intensity of the cross-peaks in the synthetic spectra [304].* 111

4.4 *S/N ratio in the different spectra across the different processing steps [304].* 113

6.1 *FC, PC, and diamagnetic contributions to the total QC-calculated shifts and experimental shifts for NiSAL-MeDPT.* 126

6.2 *FC, PC, and diamagnetic contributions to the total QC-calculated shifts and experimental shifts for NiSAL-HDPT.* 127

List of Figures

1.1 *Structures and ^1H NMR spectra acquired at 1200 MHz of NiSAL-MeDPT (panels a, b, c) and NiSAL-HDPT (panels d, e, f). The theory-based reassignment (black) is compared with the one proposed by Sacconi and LaMar [70] (grey) [65].* 21

1.2 *Structures of NiSAL-MeDPT (panel a) and optimized NiSAL-HDPT (panel b) [65].* 22

1.3 *Agreement between experimental and QC-calculated shifts for NiSAL-MeDPT and NiSAL-HDPT, with the assignment proposed by Sacconi and LaMar in 1967 (panels a and b) [70] and the theory-based reassignment (panels d and e). Comparison between initial (beige) and optimized (cyan) NiSAL-HDPT structures (panel c) and agreement between experimental and QC-calculated shifts for the NiSAL-HDPT optimized structure (panel f). The equation and the correlation coefficient are given in the figure, and the linear regression and the bisecting lines are shown in red and black, respectively. The experimental and QC-calculated shifts, together with their FC, PC, and diamagnetic contributions, of NiSAL-MeDPT and optimized NiSAL-HDPT are reported in Tables 6.1 and 6.2, respectively [65].* 23

1.4 *COSY spectrum of NiSAL-HDPT acquired at 80 MHz (panel a). 1D difference traces obtained by irradiating the signals at 229 ppm and -8.5 ppm (panel b, top and bottom, respectively). COSY spectrum of the aromatic region of NiSAL-HDPT acquired at 400 MHz (the areas shaded in grey are T1 noise and artifacts at the transmitter frequency arising from imperfections in 90° pulses) [65].* . 25

1.5 *X-ray CoMMP-12 structure with details of the active site (left). Reduced models of the active site (right): with the hydroxamate (model A) or the hydroxamic acid (model B, structure 1) moieties of the NNGH inhibitor (the different protonation state is marked with black circles), and the corresponding agreements between experimental and QC-calculated (at the NEVPT2 level) PCS [117].* 28

1.6 *Flowchart of the structural refinement procedure.* 33

1.7 *Agreement between raw experimental and best fit PCS for the X-ray structure [117].* 34
1.8 *Graphical representation of the degrees of freedom of the geometry optimization. The involved atoms are labeled and highlighted in green (DA stands for "Dummy Atom" and represents a fictitious atom centered in the intersection of the two dotted lines) (left). Only the moving molecules are colored (right) [117].* 35
1.9 *Comparison between structures 1 and 4 in terms of: geometry (left), agreement between experimental and QC-calculated (at the NEVPT2 level) PCS (top and bottom center), and isoPCS surfaces (calculated at ±1.5 ppm) for QC-calculated and experimental PCS (top and bottom right) [117].* 36
1.10 *Ab Initio Ligand Field Theory (AILFT) 3d orbitals with their energies (panel a) and splitting (panel b) based on SOC-CASSCF(7,5)/NEVPT2 calculations of 10 quartet and 40 doublet states (panel a). Grey and black bars correspond to structures 1 and 4, respectively (panel b).* 37
1.11 *Different orientations of structures 1 (magenta) and 4 (green) with the idealized trigonal bipyramidal geometry represented in yellow [117].* 38
1.12 *$2F_O$-F_C map of ZnMMP-12 with the original structure in magenta and the constrained one in green. The density is shown at 1.0σ [117].* 39

2.1 *Generic CaM conformation: the NTD is depicted as a solid ribbon, while the CTD is represented both explicitly as a transparent ribbon and implicitly as a triad of axes placed in its center of mass. The triad schematization, useful for simplifying the visualization of several conformations in the same picture, is oriented with respect to an arbitrary reference position, which corresponds to the axes (1,0,0; 0,1,0; 0,0,1) of an extended 1CLL conformation [163]. The Ca^{2+} ions are shown as light blue spheres in the native CaM used for SAXS experiments (panel a), the lanthanoid ion replacing one Ca^{2+} ion to obtain PCS and RDC is represented as an orange sphere (panel b), and the Gd^{3+} ion, coordinated to the DOTA-M tag for the DEER measurements, as a sea-green sphere (panel c) [164].* 51

2.2 *Comparison between the MaxOcc values obtained for 1000 conformations with NMR only (red, sorted in ascending order) and with different combinations of experimental datasets: SAXS (a), DEER (b), SAXS+DEER (c), NMR+SAXS (d), single DEER datasets (N53C-A103C (e), N53C-T110C (f), and T34C-T117C (g), respectively), and comparison between different combinations of DEER datasets (h) [164].* 56

2.3 *Comparison between the MaxOcc values obtained for 1000 conformations with different combinations of experimental datasets: NMR only (red), NMR+SAXS (green), NMR+SAXS+DEER (blue). The conformations are sorted by ascending NMR only order (panel a). Triad of axes representation of the 1000 conformations, color-coded according to the MaxOcc values (<5% blue, ≥34% red) obtained with NMR only (panel b) and NMR+SAXS+DEER (panel c) [164].* 57

2.4 *Explicit representation of the nine conformations with MaxOcc ≥30% obtained with the NMR+SAXS+DEER combination.* . . . 58

2.5 *Representation of the MaxOcc values for NMR only (panel a) and NMR+SAXS+DEER (panel b), color-coded according to the MaxOcc values (<5% blue, ≥34% red). Each conformation is located in the plot according to its distance in terms of angular (ρ) and cartesian (r) coordinates from the conformation which, in the NMR+SAXS+DEER analysis, has the highest MaxOcc value [164].* 59

2.6 *The capsid protein (top right) comprises NTD (green) and CTD (red) [183]. During capsid assembly, the NTD forms pentameric [184] (middle right) and hexameric [185] (bottom right) rings (with the NTDs shown in blue and green, respectively, and the CTDs shown in red in both oligomers). A model of the fully assembled capsid (left) comprises adjacent hexamers connected to each other via CTD dimers and exactly 12 pentamers are required to close the cone [182] (panel a). Distribution of positively (blue) and negatively (red) charged residues in the NMR structural ensemble [186] (panel b).* 62

2.7 *Agreement between experimental and back-calculated RDC for dimers of the NMR ensemble (PDB ID: 2KOD) in Pf1 (panel a) and in C12E5 (panel b). Distribution of the RDC q factors for the structures in the NMR family calculated using PALES with (red) and without (blue) tails. For comparison, the best fit value obtained from back-calculated RDC is shown in black (panel c). Agreement between experimental and PALES-predicted RDC for the capsid CTD dimer, without the C-terminal tail, for Pf1 (panel d) and C12E5 (panel e). Comparison between dimers obtained by HADDOCK and those belonging to the NMR family. The best agreement between experimental and predicted RDC is obtained for the dimers colored in red (Pf1) and in blue (C12E5) (panel f) [186].* . 64

2.8 *Agreement between predicted and experimental RDC for an ensemble of dimers possessing C-terminal tails generated by Xplor-NIH. The data for alignment in Pf1 and C12E5 are shown in panels a and b, respectively [186].* 66

2.9 *MaxOcc values, sorted by ascending value, for the highest 200 conformations out of the whole set of 9999 [186].* 66

2.10 *Two different views of the capsid CTD, containing 25 (panels a and b) and 100 (panels c and d) tails with the highest MaxOcc values. The conformations in group A or B (encircled) point in different directions with respect to the plane indicated by the black dotted line, positioned at the C_α of the last residue of the folded part (panel b). The tails in panels c and d have been calculated with two different RDC datasets: C12E5 only (green) and PF1+C12E5 (beige) [186].* 67

3.1 *Comparison of probe sensitivity for various sample (rotor) sizes. Shown are relative sensitivity per unit volume (dashed line) and relative sensitivity for a full rotor (solid line) (panel a) [219]. Trend of the price of commercial NMR spectrometers using superconducting magnet systems normalized by the field, at the time of market introduction of the 1.2 GHz instrument (source: Bruker Biospin) (panel b) [220].* 75

4.1 *Structures of TiBALDH, anatase titania [243] (panel a, top and bottom, respectively), H_4SiO_4 and silica [244] (panel b, top and bottom, respectively).* 79

4.2 *Structures of SOD1 (1UXM [249], panel a), cyt-c (1HRC [250], panel b), and lysozyme (1HEW [251], panel c).* 81

4.3 *^1H spectra of the titania-based composite: direct excitation spectrum (black), multiplied 30 times (red), and relaxation-filtered experiment (cyan) that shows the broader resonances associated with the protein (panel a). ^1H windowed-PMLG spectra of the titania-based composite acquired at 294 (red) and 269 (black) K (panel b).* 85

4.4 *^1H-^{15}N solution HSCQ spectrum of the free protein (red) superimposed to the ^1H-^{15}N HETCOR spectrum acquired in the solid state (grey).* 86

4.5 *Voltammetric scans obtained for Ti-BALDH/ cyt-c SPCE with PBS buffer, scan rate 10 mV/s (panel a). Comparison between Ti-BALDH/cyt-c SPCE and Ti-BALDH/lysozyme SPCE (blank) voltammetric scans (panel b).* 86

4.6 *Effect of the composite containing different cyt-c mutants and of a lysozyme+SOD1 solution (blank) on a pyrogallol solution upon sunlight irradiation. Each sample has been split in 4 wells to check whether the amount of material that is pipetted is reproducibly yielding the same amount of oxidation.* 87

4.7 *Scheme of the performed experiments on the silica-lysozyme composite.* 89

4.8 *^{1}H NMR spectra of lysozyme 10 mg/mL (red), lysozyme 10 mg/mL with silicic acid 100 mmol/dm^3 (blue), and their difference (green) (panel a). Enlargements of regions of the ^{1}H-^{15}N HSQC spectra of lysozyme 100 mg/mL (black), lysozyme 100 mg/mL with TiBALDH 100 mmol/dm^3 (red) [266]. The assignments are taken from references [275, 276].* 94

4.9 *HEWL crystals before and after the soaking with 100 mmol/dm^3 tetraoxosilicic(IV) acid (panel a), 100 mmol/dm^3 and pure TiBALDH (panels b and c, respectively).* 95

4.10 *Detail of the hydrogen bonds stabilizing the position of the titanium species through two water molecules represented as red spheres (distances are expressed in Å). The $2F_O$-F_C map is contoured at 1.0σ (blue) and 3.0σ (yellow) level [266].* 97

4.11 *Surface representation (calculated with Pymol APBS plugin) of the environment in which the titanium species (the red and white tetrahedron) is found, colored by electrostatic potential, showing chloride ions in green and sodium in purple [266].* 98

4.12 *SEM images of the silica-lysozyme composite with magnification factor 28.82 KX (left) and 37.06 KX (right), respectively.* 99

4.13 *^{1}NMR spectra of lysozyme solution 10 mg/mL in a mixture H_2O/D_2O 90:10 (green) and 98% D_2O (red). The assignments are taken from [278] (top panel). ssNMR spectrum of the silica-lysozyme composite (blue) and its 15 fold magnification (magenta) (bottom panel).* 100

4.14 *^{1}H-^{13}C CP ssNMR spectra of the composite (freeze dried and rehydrated in red and black, respectively) and of the pure lysozyme sediment (magenta).* 101

4.15 *^{29}Si direct excitation (DE) and cross-polarization (CP) ssNMR spectra of the composite.* 102

4.16 *^{1}H traces extracted from the 2D HETCOR ^{1}H-^{13}C (magenta) and ^{1}H-^{29}Si (black) spectra of the dry (panel a), rehydrated with H_2O (panel b), and with D_2O (panel c) sample.* 103

4.17 *EPR spectrum of the composite containing labeled lysozyme (Fig. 4.7, orange path).* 104

4.18 *Expected and observed behavior of the composite with respect to ionic strength and denaturation.* 105

4.19 *Synthetic spectrum without and with noise, result of the processing of the spectrum with MCR and with MCR+Cadzow denoising [304].* 112

4.20 *Synthetic spectra at different mixing times (0.5, 6, and 10 ms) with gaussian noise, coprocessed with MCR and with MCR+Cadzow [304].* 114

4.21 *Effect of denoising on a low S/N experimental spectrum: initial, reconstructed with MCR, and with MCR+Cadzow denoising [304].* 115

4.22 *Distribution of the intensities along the points of the row of the spectrum extracted at 50 ppm, which contains only noise, representing the noise in the initial spectrum (initial) and after application of combined denoising. Intensity in the "signal" spectrum* $\mathbb{C}\mathbb{S}^T$ *(left) and in the "noise" spectrum* $\mathbb{E}$ *(right). The red curves are the best fitting gaussian distributions that approximate the intensities [304].* 116

4.23 *Experimental spectra at different mixing times (0.5, 6, and 10 ms), processed with MCR and with MCR+Cadzow [304].* 117

4.24 *Effect of changing the order of the denoising steps on the same low S/N experimental spectrum [304].* 118

6.1 *ORCA output parameters for structure 1.* 128

6.2 *ORCA output parameters for structure 2.* 129

6.3 *ORCA output parameters for structure 3.* 130

6.4 *ORCA output parameters for structure 4.* 131

6.5 *Variations of the degrees of freedom between structures 1 and 4. The differences, sorted in descending order, are expressed relatively to the corresponding increment element* h_k*, with* k *= angle, dihedral, bond. The nomenclature and the graphical representation of the movement types are reported in Fig. 1.8 (left). Comparison between structures 1 and 4 with shown the coordination distances (right).* 132

Acknowledgements

I would like to thank Prof. Claudio Luchinat, Dr. Enrico Ravera and all my research group, as well as my colleagues and friends at CERM.

I would also like to thank all the collaborators from other Universities: Prof. Frank Neese and Dr. Lucas Lang from the Max-Planck-Institut für Kohlenforschung, Prof. Daniella Goldfarb from the Weizmann Institute of Science, Prof. Angela M. Gronenborn from the Pittsburgh School of Medicine, Prof. Vladimir K. Michaelis from the University of Alberta, and Prof. Gil Goobes from the Bar-Ilan University, Roberto Francischello from the University of Pisa, Dr. Alessandra Flori from Scuola Superiore Sant'Anna of Pisa, and Luca Menichetti from the National Research Council of Pisa. A special thanks, finally, goes to Dr. Elizaveta A. Suturina, my local supervisor at the University of Bath, where my research period as a visiting PhD student was prematurely interrupted due to the Covid-19 emergency.