University of Florence

Doctorate in Chemical Sciences

Cycle XXIII (2008-2010)

CHIM/03

# Structure and Dynamics of Complexes and Molecular Aggregates trough NMR: from Protein-Ligand Interactions to Nanobioreactors

*Ph.D. thesis of*

Daniela Lalli

| *Tutor* | *Coordinator* |
|---|---|
| Prof.ssa Paola Turano | Prof. Andrea Goti |

# Table of contents

**CHAPTER 4: Protein-metal ion interaction: Iron mineralization in ferritin**

**CHAPTER 5: A structure determination concept for large biological systems in ssNMR**

**CHAPTER 6: Conclusions**

# Introduction

## 1. 'Omics' sciences

Life as we know is characterized by a myriad organisms with which we share the planet. Every organism possesses a **genome** that contains the biological information needed to construct and maintain a living example of that organism. The initial product of genome expression is the **transcriptome**, a collection of RNA molecules derived from those protein-coding genes whose biological information is required by the cell at a particular time. The transcriptome is maintained by the process called transcription, in which individual genes are copied into RNA molecules. The second product of genome expression is the **proteome**, the cell's repertoire of proteins, which specifies the nature of the biochemical reactions that the cell is able to carry out (Fig. 1.1). The proteins that make up the proteome are synthesized by translation of the individual RNA molecules present in the transcriptome.
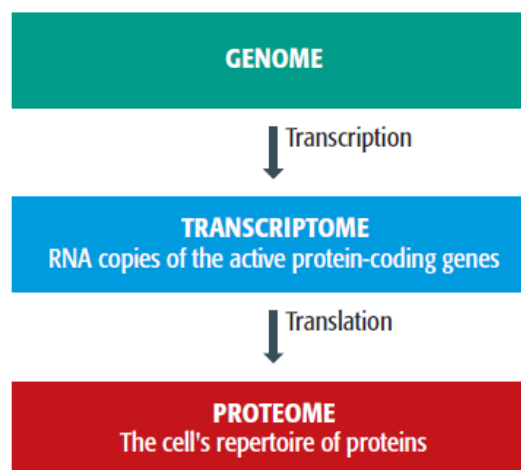


**Figure 1.1.** Correlation between 'omics' sciences.

All known biological processes are controlled by direct and specific molecular interactions involving DNA, RNA, proteins, metals and small molecules as well as biological membranes. The whole set of molecular interactions between biological entities in cells and organisms is called **interactome**, a word originally coined in 1999 by a group of French scientists.[1] Describing the contents, structure, function and

behavior of the interactome provides a link at the functional level between the genome, the proteome and the transcriptome worlds of different organisms. The structure and dynamic characterization of the interactome represents one of the major scientific challenges of the 21th century.

## 1.2. Historical basis

Genome sequencing has revolutionized all fields of life sciences. The development of genome sequencing technologies resulted in a wealth of sequence information, culminating, in 2001, in the completion of the first map of the human genome.[1,2] At present, genome sequences for thousands of organisms are available and the amount of data is increasing. From the genome sequences, it is possible to deduce the primary structure of essentially all of the proteins that a living organism can produce. The steps forward the genome sequencing are the analysis of the primary sequence data of essentially all of the proteins that a living organism can produce and afterwards the comprehension of the molecular bases of living processes trough the interactome comprehension. The most important interaction that the interactome takes into account comprise protein-DNA, protein-membrane, protein-small ligand, protein-metal and protein-protein interactions. The last three interactions are object of the present research work.

### 1.2.1. Biorecognition

Molecular recognition or biorecognition is at the heart of all biological processes. How it works is still a partially unknown and lots of questions are still not completely answered. One thing is clear: in all these cases a combination of noncovalent interactions are involved, namely ionic, hydrogen bonding and hydrophobic interactions. In addition, shape complementarity appears to play a pivotal role in the process of biorecognition.[3] Scientists are trying to understand how these forces are involved or affect a particular system, through which means to obtain a better understanding of the biorecognition process. For this purpose, many methods are available, including X-ray crystallography, Nuclear Magnetic Resonance spectroscopy (NMR), neutron diffraction, mass spectroscopy calorimetry and Cryo-electron

microscopy (cryo-EM).[4] Currently, the best approach to study and understand biorecognition is to determine the three-dimensional structure experimentally via X-ray crystallographic methods or NMR[5]. Alternative theoretical approaches, e.g., bioinformatics-based computational methods, still depend largely on previously determined structures[6].

### 1.2.2. Protein-Ligand Interactions

The biological function of a protein often depends on its interaction with ligand molecules. This is the case for enzyme-substrate, hormone receptor-hormone, cell surface antigen-receptor and for the ion channels-neurotransmitters interactions. It is clear that understanding biological function requires a precise knowledge of the protein-ligand recognition events at an atomic level. Therefore, many studies have been performed in order to accurately characterize such processes.

Of special interest are certainly protein-ligand interactions involved in the development of several diseases, such as HIV or cancer. Understanding the principles of such protein-ligand interactions allows the development of drug candidates able to inhibit the complex formation.

The identification of such drug candidates is time consuming and consists in several steps. At first, it is necessary to map the protein-ligand interaction of interest. Due to improvements in the techniques for structure determination, such as X-ray crystallography and NMR spectroscopy, and thanks to the large-scale structure-determination projects driven by genomics consortia, the number of 3D structures is increasing rapidly. When the 3D structure of the complex is not available, computational modeling approaches give a good response. Indeed in the case of protein-natural ligand complexes, it has been often found that the small molecules bind most readily to the deepest clefts on the protein surface.[7] Once the binding site of the target protein in complex with the substrate is known, the common strategy is strategy is to employ *in silico* approaches for the initial screening of a compound database ($>10^6$) to predict those that should bind the target protein in the region of interest.

In principle, there are several ways to construct novel ligands designed for a particular target that could subsequently become lead candidates. One approach is to build up a drug candidate, with typical molecular weight of 500 Daltons, from screening

a database of typically 1000-15000 compounds composed of smaller molecules (fragments) with molecular weight < 300 Daltons and good aqueous solubility.

Computational "hits" from this virtual screening will need to be validated experimentally trough biochemical assays or *in vitro* screenings. X-ray crystallography or NMR spectroscopy screenings it's possible to obtain information on the binding area of the ligands. A relatively small number (e.g. 1%) of top-ranking selective binders can be selected from this screening.

Often, a second screen is carried out to find a second fragment that will bind in close proximity to the first fragment. Use of a second screen has an advantage in that binding of the first fragment may select for a particular conformation of the target that can enable binding of a particular second fragment that would not have been found in the absence of the first. Individually, the fragments may bind to the target with $K_d$ values of only $10^{-4}$ to $10^{-3}$ M. However, by covalently linking the fragments, the additivity of the binding enthalpy and the favorable entropic contribution may pay off in a big way. In fact, sub-micromolar bi-dentate compounds can be found that are generally novel in structure.

Typically, the iterative loop of computational screening and structure determination (Fig. 1.2), is carried until the chemical leads affinity higher than the natural ligand's one. Once identified the lead compounds, it's necessary to test them *in vivo* by biochemical assays and cell based assays in order to prove the real drug's functionality.[7,8]
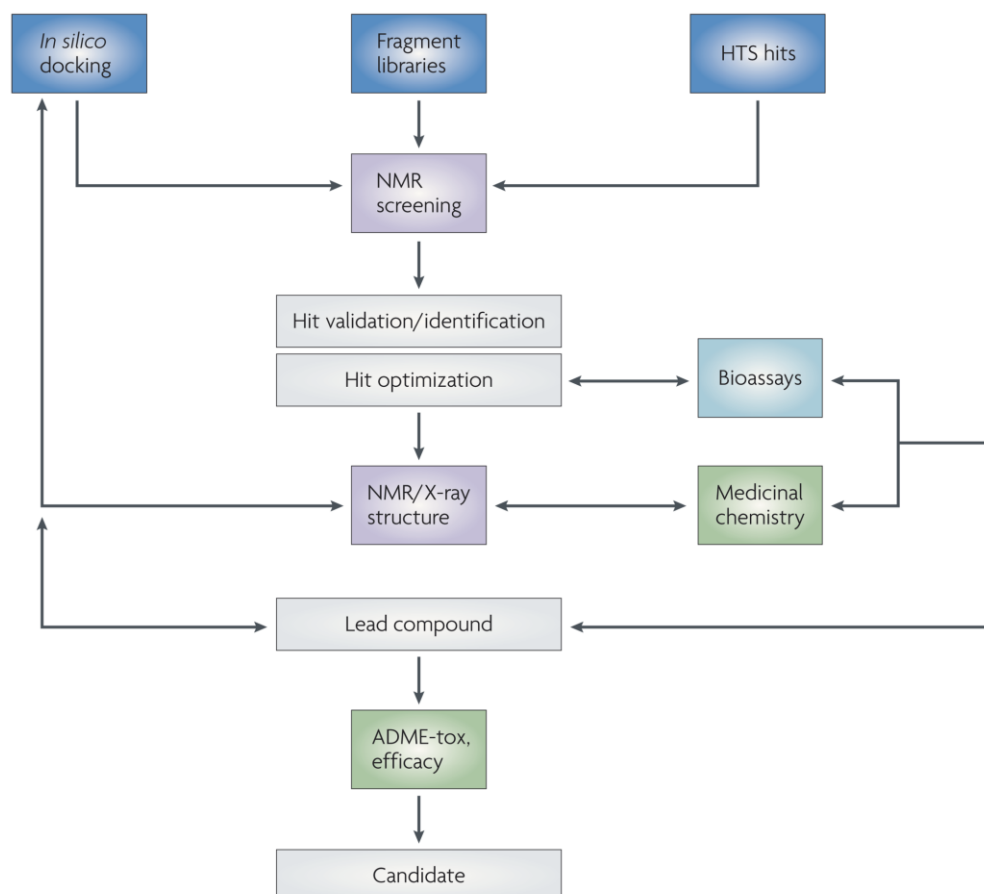
**Figure 1.2.** NMR spectroscopy can provide critical information at early stages of hit validation and identification. NMR measurements for binding studies can represent a key step to eliminate false positives from high-throughput (HTS) campaigns, to validate putative hits from *in silico* screens or to identify novel scaffolds in fragment-based programmes. NMR and X-ray crystallography can also provide unique information to subsequently guide hit-to-lead optimization. ADME-tox, absorption, distribution, metabolism, excretion and toxicity are preliminary tests for the candidate identification.

Beyond the protein-ligand interaction inhibition, it's necessary to mention the **protein-protein** interaction inhibition trough the development of candidate drugs able to bind the target protein in the same region of the partner. This topic is discussed in Chapter 2. The ability to identify the sites of a protein that can bind with high affinity to small, drug like compounds has been an important goal in drug design. The general lack of small molecule starting points for drug design, the typical flatness of the interface, the difficulty of distinguishing real from artefactual binding, and the size and character of typical small molecule libraries make the things difficult. Natural small molecules known to bind at protein-protein interfaces are rare, whereas drug friendly enzymes

often have small molecule substrates that can serve as templates for designing antagonists. One way to obtain a small molecule starting point for a protein-protein interface has been to map the epitope of one of the proteins onto a small peptide or peptidomimetic. The shape of a typical protein-protein interface adds to the difficulty of drug discovery.

Approximately 700-1500 $Å^2$ of surface area is buried on each side of the interface,[9] and the binding surfaces are relatively flat.[10,11] Therefore, the X-ray structures of protein-protein pairs do not usually show small, deep cavities that look like small molecule binding sites. However, it might not be necessary for a small molecule to cover the entire protein binding surface, because the subset of the interface that contributes to high affinity binding (the 'hot spot') is often much smaller.[12-14] In addition, for many protein-protein interactions, the apparent complementarity between the two surfaces involves a significant degree of protein flexibility and adaptivity.[15] Moreover, there might be binding site conformations, that are well-suited to small molecule binding, but not visible in a single crystal structure.[16-18]

### 1.2.3. Protein-Metal ion Interactions

Life has developed taking advantage of the minerals present in the earth. Therefore it is not surprising that living beings have developed the capability to use inorganic elements to perform specific physiological functions. More than 20 metals are essential for most species to carry out their function and they are involved in a lot of processes such as charge balance and electrolytic conductivity (Na, K), signaling (Ca), electron transfer (Fe, Cu), redox catalysis (Fe, Co, Ni, Cu), energy storage (Na, K, Fe) and biomineralization (Ca, Mg, Fe, Cu).

Some metals like Fe, Cu or Zn are strongly associated with proteins and form the so called metallo-proteins or metallo-enzymes, because, as enzymes usually do, they function as biological catalysts. Some others, like the non transition metal ions $Na^+$, $K^+$, $Mg^{2+}$ or $Ca^{2+}$ form complexes with proteins with a significantly smaller affinity constant. These latter systems are considered metal-protein complexes.

In multicellular organisms, sodium and calcium are found mostly outside the cellular compartment (extracellular), while potassium and magnesium are largely intracellular. Calcium and magnesium are often metal activators in proteins to which

they bind with relatively low affinity. Under appropriate circumstances, these metal ions induce conformational changes in the protein upon binding and in doing so they may transmit a signal e.g. the firing of neurons by rapid influx of sodium ions across a cell membrane or the regulation of intracellular functions by calcium binding proteins such as calmodulin.

Owing to the great advances in research we now know the structures of many components of the systems that biology has adapted through evolution to perform these essential functions. One of the most common system that nature adopts to associate a metal with a protein is the so called a metal cofactor. A third of all proteins contain metals. A cofactor is a chemical entity that is associated with a biological macromolecule and is necessary for the protein's biological function. For this reason the metal cofactor is usually called prosthetic group, indicating its ability to help the macromolecule to carry out its function. Typically metal cofactors catalyze reactions or help provide the required geometry of a protein.

Beside directly incorporating a variety of metal ions into proteins, the nature has also selected other inorganic ligands for the metal centre such as the iron-sulfur clusters and tetrapyrroles. These prosthetic groups are largely used in electron transfer proteins and redox enzymes as their redox potential can opportunely be tuned over large ranges.

Some metal ions are found deeply buried within proteins. Such metal ions are usually structural in function. Their interaction with the protein helps insure the optimal protein structure and contributes to the stability and appropriate behavior necessary to perform the physiological function. For example, the $Zn^{2+}$ ions in Zn fingers which are transcription factors necessary for the adoption of the proper shape of the protein, allowing its interaction with DNA. It is not currently known if the zinc ion plays more than a structural role in this proteins i.e. if the $Zn^{2+}$ concentrations are also used in some manner to regulate gene expression.

Beyond the biocoordination chemistry that generally focuses the attention on interactions between metals and ligands at the level of the coordination sphere, the biological process that produces minerals, called biomineralization, is an important subject (nowadays largely studied) that will be examine in Chapter 4. The research aims of biomineralization consist in the characterization of the structure and composition of the biomineral, in the comprehension of the functional properties of biominerals and elucidation of the process through which protein control the synthesis, construction and organization of inorganic mineral-based materials.

### *1.2.4. Protein-Protein Interactions*

Complete sequencing of genomes has indicated that a living organism can contain tens of thousands of protein-protein interactions (PPIs), which form a web of intricate communication networks.[19,20] PPIs are known to be the major components of a wide variety of cellular events such as the transport machinery across biological membranes, packaging of chromatin, muscle contraction, signal transduction and regulation of gene expression, just to name a few. Anomalous PPIs are implicated in a big number of diseases; monitoring them serves as tools to reveal the molecular basis of such diseases. The structural characterization of a PPI is a topic that will be discussed in Chapter 3.

It is now clear that many proteins interact with multiple partners, either simultaneously or separately, depending on availability and environment. The ability to bind multiple partners forms the basis of network complexity, which in turn is related to the extent of evolution of a particular organism.

The systematic identification of interactions for a given proteome was proposed as an informative functional strategy.[21,22] At the beginning, interactions involved in biological processes were characterized individually. This was a reductionist approach that suffered from a lack of information about time, space, and context in which the interactions occur in vivo. Further, a combined experimental and computational approach, was developed with the aim of building PPI maps.

The first global experimental analyses of PPIs was made on model organisms (yeast *Saccharomyces cerevisiae* and worm *Caenorhabditis elegans*)[23] using the yeast two-hybrid technique[24]. In parallel, a series of computational bioinformatic methods have been designed to predict PPIs. Yeast, worm and fly[25-28] PPIs were analyzed with this method. On the basis of experimental and computational analyses 28000 PPIs are estimated for yeast[19,29]. The big goal is to fully understand human molecular mechanisms of PPIs. Human proteome may have an order of magnitude more complexes than the yeast cell and the number of different complexes across all relevant genomes may be several times larger still.

An interdisciplinary approach that includes *in vivo* and *in vitro* techniques has been developed for this purpose.[30] Among the *in vivo* methods to identify PPIs network it is worth mentioning:

- the yeast two hybrid technique;

- the Protein-fragment Complementation Assay (PCA), also known as bimolecular fluorescence complementation assay;[31]
- fluorescence resonance energy transfer (FRET);

while between those *in vitro* the most important are:

- the **protein affinity chromatography**, where the target protein is coupled to a matrix under controlled conditions and used to select ligand proteins;

- the **immunoprecipitation**, where cell lysates are generated, antibody is added, the protein-antibody complex precipitates, the complex washed, and complexed proteins are eluted and analyzed;

- the **microarrays** where different proteins or specific DNA binding sequences immobilized on microchips are able to selectively bind protein partners or RNA fragments;

- the **library based methods** used to screen large libraries for genes or fragments of genes whose products may interact with a protein of interest;

- **computational prediction** of protein–protein interactions;

Concerning the biophysical techniques used to study PPIs the most important are light scattering, fluorescence spectroscopy, surface plasmon resonance and calorimetry, while the structure of macromolecular complexes can be solved with atomic resolution by X-ray crystallography and NMR spectroscopy depending on the strength of the complex, and with lower resolution but with less limitations due to the size of the complex by cryo-electron microscopy (cryo-EM).

The strength of a PPI can be described by the equilibrium dissociation constant ($K_d = [P] [L]/[PL]$ where $[P]$, $[L]$ and $[PL]$ are the equilibrium concentrations of protein, ligand and complexed state, respectively). The range of $K_d$ values observed in biologically relevant processes is extremely wide and can span over twelve orders of magnitude [16]. Despite this large difference in affinity, all protein–protein interactions maintain a high degree of **specificity** for their partners. The PPIs are often broadly classified as weak, or strong. Decades of extensive studies have provided significant understanding of a large fraction of PPIs in living cells; however, the vast majority of these belong to the strong category ($K_d < 10^{-6}$ M).[32] The weak PPIs, especially those with $K_d > 10^{-4}$ M, are poorly understood, despite the fact they might be crucial for mediating many important cellular events, such as transient and rapid assembly and/or reassembly of protein complexes for multi-component enzyme regulation, signal transduction turnover, etc..[32] One reason for the slow progress of studies on many weak

PPIs is the technical difficulty in characterizing them, especially those with $K_d > 10^{-4}$ M, because they usually do not result in a sufficient amount of complexes that can be directly detected *in vitro* or *in vivo* by conventional methods. NMR has been known for years as one of the most sensitive tools for examining weak but functionally important protein-target interactions at physiological conditions, which are normally undetectable by most biochemical methods.[33,34]

### 1.3. NMR as unique technique for weak interaction study

NMR is a powerful and reliable method for studying such interactions. In particular it's an ideal technique for studying weak and transient interactions which are ubiquitous in the crowded environment of a cell, but generally ignored by other approaches. The unique potential of NMR for studying transient protein complexes that interact weakly has been demonstrated in the past years. Unprecedented new insight was gained for various multi-domain proteins that play crucial roles in intra- and extra-cellular signaling,[35] the regulation of gene expression,[36] chaperone-protein complexes in kinase folding [Sreeramulu et al Schwalbe JBC 2008 kinase chaperone], and on metal mediated transient protein-protein interactions,[37] to name a few. Moreover, has been recently demonstrated the utility and the potential of monitoring the course of events in protein-protein interactions; for example the protein phosphorylation in living cells can be followed in real time.[38] In many cases, structural analysis by X-ray crystallography was not possible since crystallization failed or the crystallized protein constructs were strongly influenced by crystal packing. In others, although X-ray was successful in providing a structural view, NMR was necessary to relate it to the functional details of the post-translational regulatory mechanism.

The unique strengths of NMR for these studies (can be summarized as) that are not matched but are very complementary to other structural biology techniques are:

- mapping binding interfaces for protein-protein interactions and characterizing the competition and cooperativity of interactions;
- determining the structures and dynamics of (large) multicomponent protein complexes using NMR solution data, also by employing multidisciplinary hybrid approaches. NMR is unique in that it provides structural information for individual components (or domains) in a multi-

component system and in that it can follow the changes in dynamics upon complex formation;

- studying transient and weak interactions, which are critical for the regulation of cellular processes;

- applications of NMR to the study of protein-protein and protein-nucleic acid complexes of medical importance includes the analysis of misfolded, partially folded or intrinsically disordered proteins.

Complementary methods, like X-ray crystallography, SAXS, electron microscopy, cryo-EM and mass spectrometry are increasingly used together with NMR spectroscopy.

Reference List

1. Sanchez, C. *et al.* Grasping at molecular interactions and genetic networks in Drosophila melanogaster using FlyNets, an Internet database. *Nucleic Acids Res.* **27**, 89-94 (1999).

2. Macilwain, C. World leaders heap praise on human genome landmark. *Nature* **405**, 983-984 (2000).

3. Williams, D. H., Stephens, E., O'Brien, D. P. & Zhou, M. Understanding noncovalent interactions: ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes. *Angew. Chem. Int. Ed Engl.* **43**, 6596-6616 (2004).

4. Morikis, D. & Lambris, J. D. Physical methods for structure, dynamics and binding in immunological research. *Trends Immunol.* **25**, 700-707 (2004).

5. Lindorff-Larsen, K., Best, R. B., DePristo, M. A., Dobson, C. M. & Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **433**, 128-132 (2005).

6. Dwyer, M. A., Looger, L. L. & Hellinga, H. W. Computational design of a biologically active enzyme. *Science* **304**, 1967-1971 (2004).

7. Vajda, S. & Guarnieri, F. Characterization of protein-ligand interaction sites using experimental and computational methods. *Curr. Opin. Drug Discov. Devel.* **9**, 354-362 (2006).

8. Pellecchia, M. *et al.* Perspectives on NMR in drug discovery: a technique comes of age. *Nat. Rev. Drug Discov.* **7**, 738-745 (2008).

9. Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J Mol Biol* **285**, 2177-2198 (1999).

10. Bahadur, R. P., Chakrabarti, P., Rodier, F. & Janin, J. A dissection of specific and non-specific protein-protein interfaces. *J. Mol. Biol.* **336**, 943-955 (2004).

11. Nooren, I. M. & Thornton, J. M. Structural characterisation and functional significance of transient protein-protein interactions. *J. Mol. Biol.* **325**, 991-1018 (2003).

12. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1-9 (1998).

13. DeLano, W. L. Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.* **12**, 14-20 (2002).

14. Keskin, Z., Gursoy, A., Ma, B. & Nussinov, R. Principles of protein-protein interactions: What are the preferred ways for proteins to interact? *Chemical Reviews* **108**, 1225-1244 (2008).

15. Sundberg, E. J. & Mariuzza, R. A. Luxury accommodations: the expanding role of structural plasticity in protein-protein interactions. *Structure.* **8**, R137-R142 (2000).

16. Teague, S. J. Implications of protein flexibility for drug discovery. *Nature Rev. Drug Discov.* **2**, 527-541 (2003).

17. Luque, I. & Freire, E. Structural stability of binding sites: consequences for binding affinity and allosteric effects. *Proteins* **Suppl 4**, 63-71 (2000).

18. Ma, B., Shatsky, M., Wolfson, H. J. & Nussinov, R. Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein Sci.* **11**, 184-197 (2002).

19. Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147 (2002).

20. Ho, Y. *et al.* Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415**, 180-183 (2002).

21. Walhout, A. J. & Vidal, M. Protein interaction maps for model organisms. *Nat. Rev. Mol. Cell Biol.* **2**, 55-62 (2001).

22. Sharan, R., Ulitsky, I. & Shamir, R. Network-based prediction of protein function. *Mol. Syst. Biol.* **3**, 88 (2007).

23. Matthews, L. R. *et al.* Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". *Genome Res.* **11**, 2120-2126 (2001).

24. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246 (1989).

25. Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623-627 (2000).

26. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U. S. A* **98**, 4569-4574 (2001).

27. Reboul, J. *et al.* C. elegans ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat. Genet.* **34**, 35-41 (2003).

28. Giot, L. *et al.* A protein interaction map of Drosophila melanogaster. *Science* **302**, 1727-1736 (2003).

29. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U. S. A* **96**, 4285-4288 (1999).

30. Piehler, J. New methodologies for measuring protein interactions in vivo and in vitro. *Curr. Opin. Struct. Biol.* **15**, 4-14 (2005).

31. Michnick, S. W. Protein fragment complementation strategies for biochemical network mapping. *Curr. Opin. Biotechnol.* **14**, 610-617 (2003).

32. Nooren, I. M. A. & Thornton, J. M. Diversity of protein-protein interactions. *Embo Journal* **22**, 3486-3492 (2003).

33. Zuiderweg, E. R. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* **41**, 1-7 (2002).

34. Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H. & Schreiber, G. The molecular architecture of protein-protein binding sites. *Curr. Opin. Struct. Biol.* **17**, 67-76 (2007).

35. Wegener, K. L. *et al.* Structural basis of integrin activation by talin. *Cell* **128**, 171-182 (2007).

36. Qian, Y. Q. *et al.* NMR structure determination reveals that the homeodomain is connected through a flexible linker to the main body in the Drosophila Antennapedia protein. *Proc. Natl. Acad. Sci. U. S. A* **89**, 10738-10742 (1992).

37. Banci, L. *et al.* The Atx1-Ccc2 complex is a metal-mediated protein-protein interaction. *Nat. Chem. Biol.* **2**, 367-368 (2006).

38. Selenko, P. *et al.* In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. *Nat. Struct. Mol. Biol.* **15**, 321-329 (2008).

# Targeting S100B-p53 binding interface with small molecules

**Chapter**

**2**

## 2. Summary

In this chapter is reported the NMR screening of fragment libraries targeting the human S100B-p53 interaction site. This is a project in collaboration with Siena Biotech, a company that focuses its research in discovering new drugs for therapeutic intervention against neurodegenerative diseases and oncology. S100B represents an interesting pharmaceutical target because interacting with p53 it's able to inhibit its tumor suppressor activity. The identification of small molecule compounds able to bind S100B preventing S100B-p53 complex formation, restores p53 dependent tumor activity in cancer. Siena Biotech scientists performed the *in silico* screening that led to the identification of 280 novel prospective ligands for S100B. In CERM we carried out an *in vitro* screening by solution NMR that revealed specific binding at the p53 interaction site for a set of these compounds and confirmed their potential for further rational optimization. The X-ray structure determined for one of the binders revealed key intermolecular interactions paving the way for structure-based ligand optimization.

### 2.2. S100B

Calcium is a universal, intracellular second messenger that plays a regulatory role in processes or events such as the conduction and transmission of the nerve impulse, muscle contraction, cell motility, cell growth and differentiation, gene expression, cross-talk between different enzyme systems, apoptosis and necrosis.[1] Cells evolved classes of intracellular Ca(II)-binding proteins that act to regulate the levels of cytosolic Ca(II) and/or to transduce the intracellular Ca(II) signal.[2] The EF-hand protein family[3] is one of such classes of calcium-binding proteins displaying several modes of calcium signaling via interaction with target proteins.[4] The EF-hand calcium-binding motif is a helix-loop-helix motif that was discovered first in the crystal structure of parvalbumin and found again in calmodulin and troponin C.[5] One important subclass of EF-hand calcium-binding proteins is the S100 protein family whose name refers to their

solubility in a 100%-saturated solution of ammonium sulfate.[6] This family of low-molecular-weight (20-24 kDa) dimeric proteins comprises approximately 25 members which are expressed in a cell-specific manner and have been implicated roles in such as cellular growth and differentiation, cell-cycle regulation, and cytoskeletal function and structure.[7-9]

Of these, S100B has been one of the most intensely studied and well characterized. First isolated from brain in 1965 by Moore,[6] S100B has subsequently been found in a variety of other tissues and has been shown to be highly conserved through a variety of species.[10]

### 2.2.1. Structure

In 1996, the three-dimensional structures of the calcium-free (apo) forms of rat[11] and bovine S100B[12] were determined by solution NMR. Two years later, the first three-dimensional solution structure of human Ca(II)-S100B was also determined.[13] The comparison of this structure with those of rat and bovine proteins allows a detailed analysis of the calcium-induced conformational changes of S100B.

S100B is a symmetrical non-covalent homodimer in which the two monomers, of 91 amino acids each, are packed together in an antiparallel fashion. Each monomer in Ca-S100B is comprised of four α-helices (I, Leu3-Tyr17; II, Ser30-Glu39; III, Val52-Glu58; IV, Phe70-Phe88), three *loops* (I, Ser18-His25 (L1); II, Leu40-Glu49 (H); III, Asp61-Gly66 (L2)) and two short antiparallel β-strands (I, Lys26-Lys28; II, Glu67-Asp69) (Fig. 2.1).
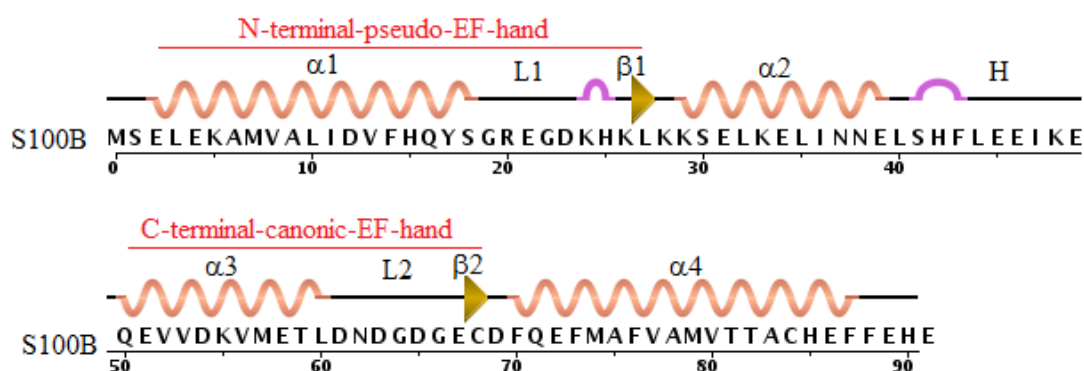


**Figure 2.1** Amino acidic sequence of the humanS100B.

Each S100B monomer is formed by two helix-loop-helix EF-hand calcium-binding motifs joined by a flexible linker region (Fig. 2.2). In Ca-S100B the **N-terminal-pseudo-EF-hand**, also called low affinity ($K_d$ = 200 mM) calcium-binding site, consists of the helix I, the calcium-binding loop L1 and β-strands I, and the helix II; while the **C-terminal-canonical-EF-hand**, known as the high-affinity ($K_d$ = 7–50 mM) calcium binding site comprises the helix III, the calcium-binding loop L2 and β-strands II, and the helix IV.[14-16] A linker region (hinge loop, H) connects helix II–III. Helix IV is followed by a C-terminal extension (Fig. 2.2). The two monomeric S100B subunits form a compact, globular, non-covalent homodimer in solution. The dimer interface primarily involves hydrophobic interactions between helices I of different monomers and helices IV of different monomers. Contacts between helices I of different monomers extend nearly the entire length of each helix (Glu2–Val13) (Fig. 2.2).
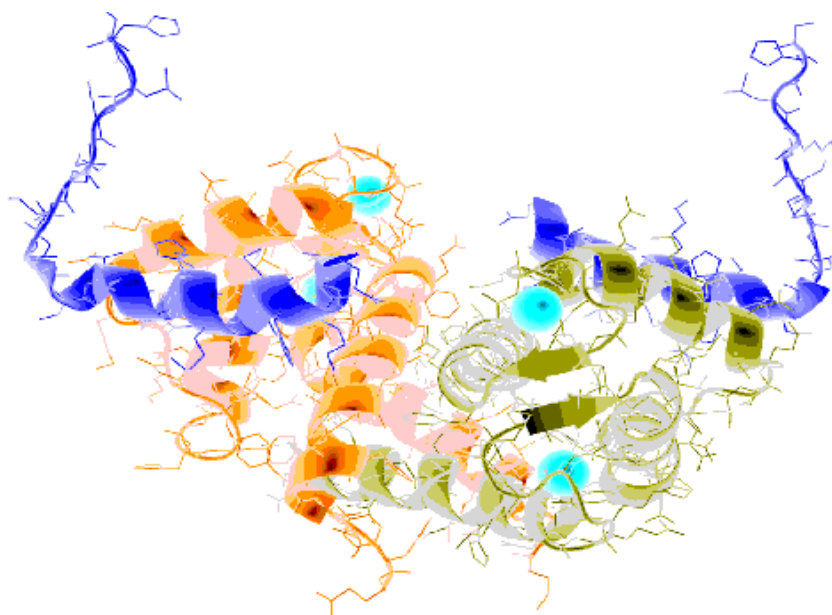


**Figure 2.2.** Schematic representation of the secondary structure of an S100 protein.

Comparing the NMR structures of human Ca(II)-S100B, rat and bovine apo-S100B, some structural differences are evident. The N-terminal helix-loop-helix region in Ca(II)–S100B (residues Val8–Glu39) is very similar to that of apo- rat and bovine S100B (Fig. 2.3a), while the C-terminal calcium-binding loop in S100B undergoes a dramatic change in conformation compared to that of rat apo- and bovine S100B (Fig. 2.3b). The superimposition of the entire calcium binding loop (Asp61–Glu72) and the same region of helix IV (Phe73–Phe76) for human Ca(II)–S100B, rat and bovine apo-

S100B reveals a reorientation of the β sheet and helix IV in Ca(II)-S100B compared to both rat apo-S100B (backbone pairwise rmsd ~4.06 Å) and bovine apo-S100B (backbone pairwise rmsd ~5.03 Å). The cause for this difference can be attributed to the first six residues of the calcium-binding loop. More specifically, rearrangement of the C-terminal loop in Ca(II)–S100B upon calcium binding brings the side chains of Asp61 and Asp63 into a more appropriate geometry to coordinate calcium. This conformational change is propagated through helix III, which becomes more perpendicular to helix IV as a consequence of a remarkable change of the inter helical angle between these two helices.[13,17-19] Owing to this change, the hinge region swings out. As a result of these structural changes, a cleft forms in each monomer, which is defined by residues in the hinge region, helices III and IV and the C-terminal extension, and is buried in apo-S100B monomer. Residues defining this cleft are believed to be important for the Ca(II)-dependent recognition of S100 target proteins.
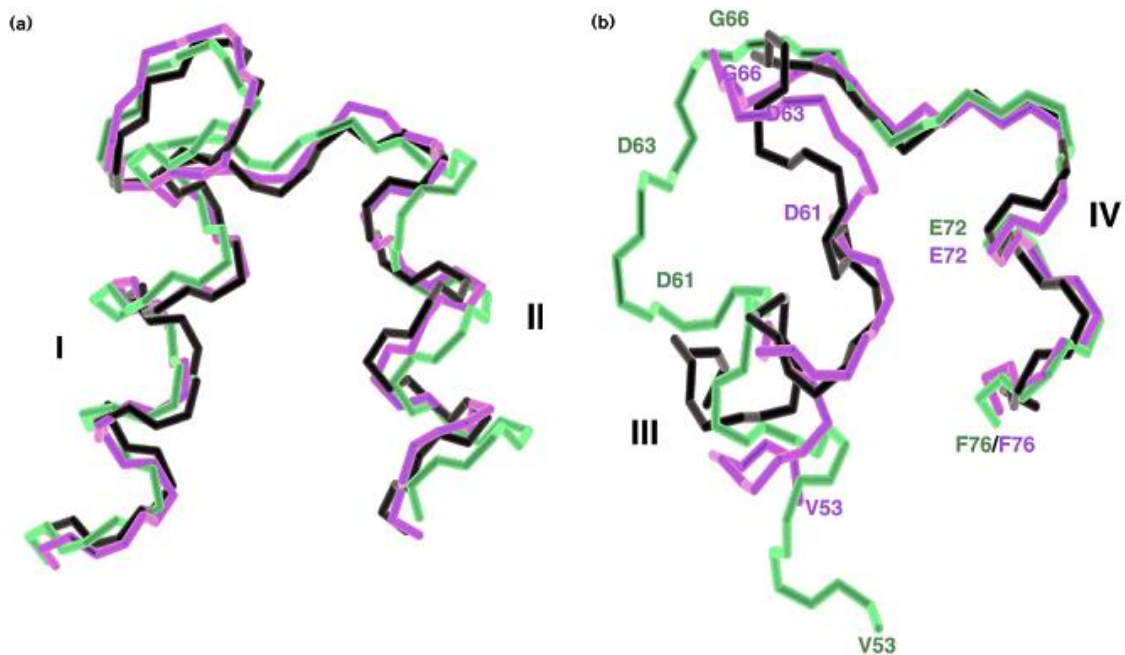


**Figure 2.3.** Comparison of the EF-hands of human Ca(II)–S100B, rat apo S100B and Ca(II)–calbindin.[20] (a) The N-terminal pseudo EF-hand in Ca(II)–S100B (green) and rat apo S100B (magenta), superimposed (N, Cα and CO atoms) with the analogous residues in the pseudo EF-hand of Ca(II)–calbindin (white). (b) The superposition (N, Cα and CO atoms) of the C-terminal EF-hands of Ca(II)–S100B (green) and rat apo S100B (magenta) with the analogous residues of Ca(II)–calbindin (white).

The hinge region and the C-terminal extension play a critical role in the interaction of S100B with several target proteins.[21-24] Thus, upon Ca(II) binding, each S100B monomer opens up to accommodate a target protein, and the dimer can bind target proteins on opposite sides. By this mechanism a S100B dimer functionally cross-links two homologous or heterologous target proteins (Fig. 2.4). Given the positions of the helices of one monomer relative to those of the other monomer (particularly helices I and IV and I' and IV') (Fig. 2.4), helix I' might participate in the formation of each of the two binding surfaces on a given S100 dimer.[25] This would explain why most S100 members form dimers; one S100 monomer is not enough for binding a target protein or, alternatively, target protein binding to S100 occurs with reduced strength.[10]
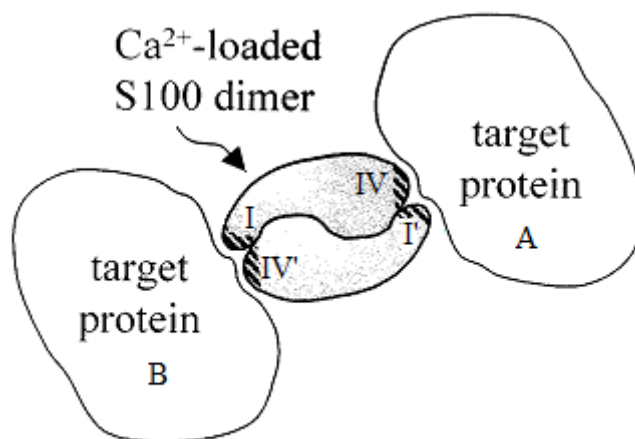


**Figure 2.4.** Schematic representation of the putative mode of interaction of an S100 dimer with target proteins. A can be equal or different from B.

The exposure to the solvent of hydrophobic regions in Ca(II)-S100B gives rise to the S100B association and precipitation at high ionic strength. Indeed, it was shown that calcium-binding to S100B has a strong salt dependence. In particular, S100B self-association occurs in the presence of calcium at higher ionic strengths, which gives rise to extensive broadening of NMR signals.[15,16,16,17] This observation is consistent with the increased exposure of hydrophobic residues in Ca(II)–S100B as suggested by fluorescence and absorption studies.[17]

*2.2.2. Function*

Most S100 proteins are generally expressed and distributed in a cell-definite fashion, indicating a conserved biological role. In this setting S100B is mainly concentrated in the nervous system and in adipose tissue.

S100B has been widely associated with the pathology of degenerative neurological disorders including Down syndrome, Alzheimer's disease, temporal lobe epilepsy, and brain damage after cardiac arrest.[26,27] Functionally, S100B regulates several cellular functions like protein phosphorylation, protein degradation, calcium homeostasis, cell locomotion, transcription factors, stimulation of enzymes activity, regulation of receptor function and cell proliferation and differentiation via interaction with transcription factors in the nucleus.[10,27]

In general, low levels of S100B have trophic effects, and higher levels are toxic, resulting in uncontrolled cell growth.[8,28,29] Increased levels of S100B are found in astrocytic tumors,[30] T cell leukemia,[31] renal cell tumors,[32] malignant mature T cells,[33] and rat renal cell tumor lines.[32] In addition, elevated S100B levels have prognostic value for assessing the progression of malignant melanoma including recurrence of disease, disease progression to more advanced stages, and metastatic potential of the melanoma in the later stages leading to the suggestion that S100B may have a role in this neuropathological disease.[34-36]

Since S100B undergoes a calcium-sensitive conformational change, it's able to interact with a variety of other target proteins; up to now more than 20 known partners. For example, calcium-binding to S100B regulates interactions with tubulin[37] and glial fibrillary acidic protein (GFAP),[38] which are important for cellular architecture, inhibits the phosphorylation of neuromodulin,[39] myristoylated alanine-rich C kinase substrate (MARCKS)[40] and tau proteins[41] through interaction with the substrate, and binds the C-terminus of p53 inhibiting its tumor suppressor function.[23,42] In this study I will focus the attention on the S100B-p53 interaction.

## 2.2.3. Interaction: S100B-p53

The p53 protein plays an important role in preventing cancer development. This protein has been called the "guardian of the genome" [43] as it prevents inappropriate cell proliferation and maintains genome integrity following genotoxic stress.[44,45] Presently, p53 is known to play a key role in practically all types of human cancers, and the mutation or loss of the p53 gene can be identified in more than 50% of all human cancer cases worldwide.

Wild-type p53 protein contains 393 amino acids and is composed of several structural and functional domains (Fig. 2.4):

i) the **N-terminus** containing an amino-terminal domain (residues 1-42) and a proline-rich region (residues 61-94), required for trans-activation activity,

ii) the **central core** domain (residues 102-292), able to bind DNA,

ii) the **C-terminus** containing a tetramerization domain (residues 324-355), and a regulatory domain (residues 363-393), required for trans-activation activity.[44,46,47]



**Figure 2.5.** Schematic representation of the p53 structure.

p53 is inactive in normal, unstressed cells, but becomes active under various intracellular and extracellular stimuli, such as DNA damage, heat shock and oncogene overexpression. The p53 activation is regulated by post-translational mechanisms such as phosphorylation, acetylation, and prolyl isomeration,[48,49] or by PPI.[50] The latter mechanism promotes p53 structural changes and modulates its DNA binding domain affinity to DNA. In addition, tetramerization was found to be relevant in p53 activation,[51] p53 binds to its DNA response elements most efficiently as a tetramer, then

a monomer. The p53 active form bound to DNA induces the expression of downstream genes and thereby arrests cell cycle, repairs damaged DNA, or induces apoptosis.[52]

The majority of cancer-associated p53 mutations cluster within the DNA binding domain and disrupt the p53-DNA binding activity.[43] However, recent studies have identified mutations also within the tetramerization domain that are associated with cancer.[53]

S100B is able to bind the C-terminus of p53 and inhibit its tumor suppressor function. Evidence for S100B-p53 interaction in vivo comes from co-immunoprecipitation experiments with S100B performed in human primary malignant melanoma cancer cells.[42] Upon binding p53, S100B inhibits phosphorylation of the tumor suppressor at the C-terminus and disrupts the p53 tetramers,[21,54] two steps required for p53 transcriptional activity.[55,56] Thus, high levels of S100B in cancer cells such as malignant melanoma likely contribute to uncontrolled cell growth by down-regulating p53.[57]

The interaction between S100B and p53 is a calcium dependent process.[21,58]. A comparison of the 3D structures of calcium-free S100B (apo-S100B), calcium-bound S100B (holo-S100B), and holo-S100B bound to a peptide derived from the C-terminal negative regulatory domain illustrates the details of the calcium-dependent S100B-p53 interaction.[23] Specifically, calcium binding to S100B causes a conformational rearrangement, upon which several residues form a hydrophobic patch and become solvent-exposed and available for interaction with the target polypeptide (Fig. 2.6).[59]

In particular, several hydrophobic residues on helix III, helix IV, and loop II are exposed due to a large conformational change in the second EF-hand domain of S100B (Fig. 2.6). Are these newly exposed residues that form a mini-hydrophobic patch on holo-S100B and participate in the S100B-p53 binding interface (Fig. 2.6). In the absence of calcium, these same hydrophobic residues are buried in the core of the S100B subunit, and p53 cannot bind S100B, even at mM concentrations[21] (Fig. 2.6).
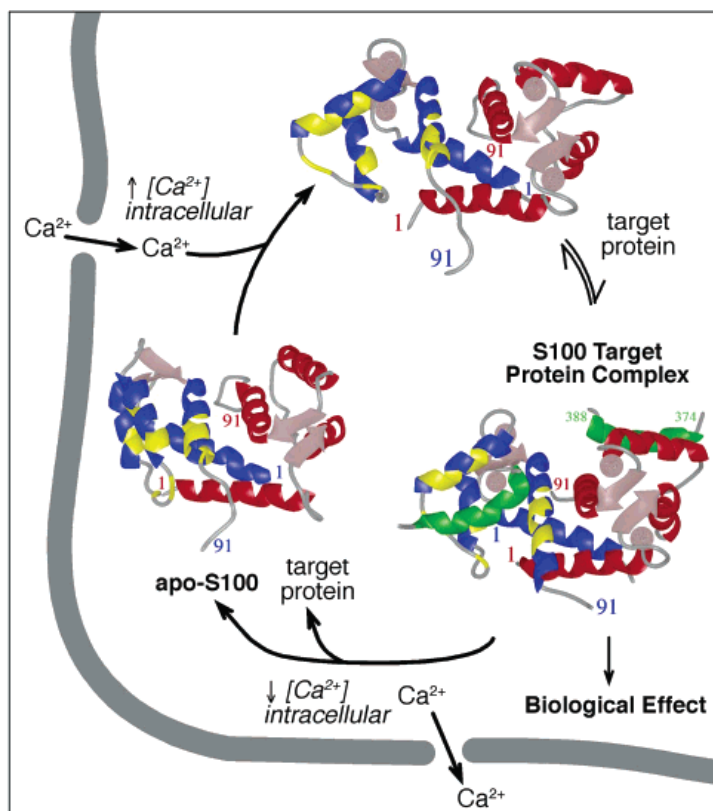
**Figure 2.6.** Diagram illustrating the Ca(II)-dependent conformational change in S100B necessary to bind target proteins. In the ribbon diagrams of S100B, subunits are shown in red and blue, and the target protein is shown in green. Residues that have intermolecular correlations between S100B and the target protein are colored yellow on one subunit of S100B in each ribbon diagram to illustrate the Ca(II)-dependent conformational change necessary to expose the target protein binding site on S100B.

The key interactions between the two partner proteins have been identified by NMR via the solution structure of rat S100B in complex with the C-terminal negative regulatory domain of p53 (residues 367-388) as involving 21 residues from in helices III and IV and the hinge loop of S100B (Fig. 2.7).[23] Among these residues, the most important in terms of interaction are Leu44, Glu45, Val52, Val56, Met79, Val80, Glu86 and Phe87 whose protons are shown in blue in Figure 2.7.
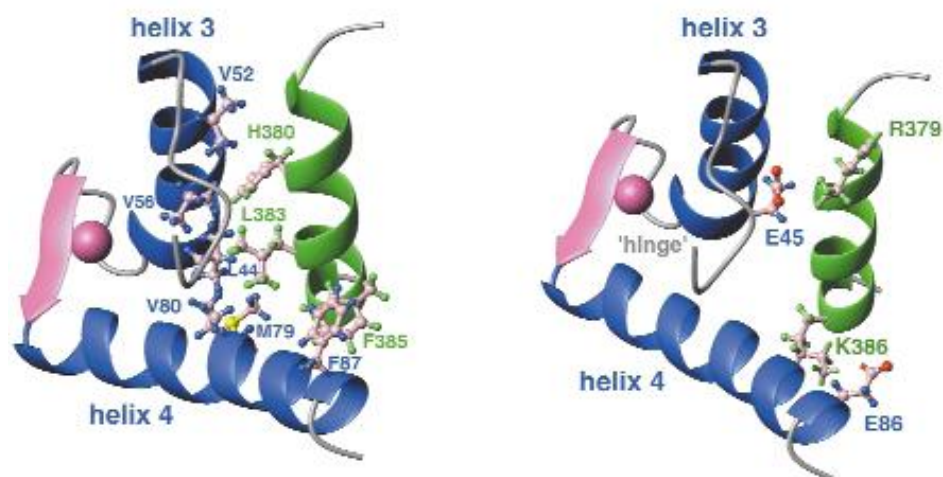
**Figure 2.7.** Positioning of side chain residues in the p53 peptide-S100B complex. Ribbon diagram displaying the side chains involved in the hydrophobic binding pocket.

The same region was identified to bind TRTK-12,[60] a twelve-residue peptide derived from the actin capping protein CapZ with sub-micromolar affinity for S100B ($K_d$ = 200-300 nM).[61] The residue involved in the interaction are Leu40, Leu44, Glu45 and Ile47, from the hinge loop, Val52, Lys55, Val56 and T59, from the helix $\alpha$3 and Met79, Val80 and Ala83 from the helix $\alpha$4 (Fig. 2.8).
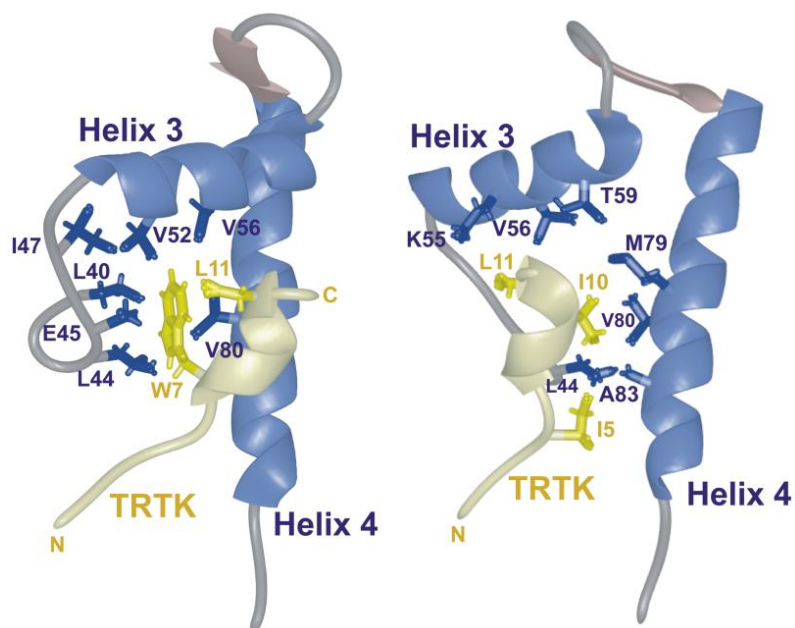


**Figure 2.8.** Side-chains of S100B and TRTK-12 that are involved in the peptide–protein interface.

There is no sequence similarity between p53 and TRTK-12 peptides, which in fact occupies the hydrophobic pocket of S100B a in a different orientation with respect to TRTK-12 (Fig. 2.9).
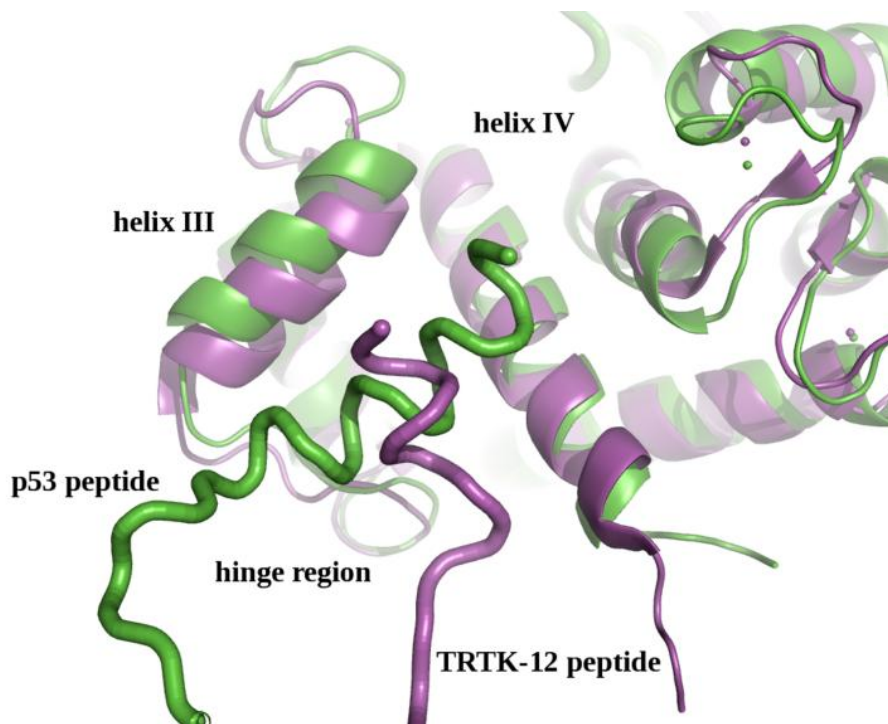


**Figure 2.9.** Side-chains of S100B and TRTK-12 that are involved in the peptide–protein interface.

The availability of high-resolution structural information of bond form of S100B provides the possibility of applying structure-based approaches to identify small molecule compounds that bind to this hydrophobic patch on holo-S100B and inhibit the calcium-dependent S100B-p53 interaction. There are a many studies concerning this topic: Weber and coworkers[62] identified few small molecules binding human S100B using a NMR-based screening, while Arendt et al. identified inhibitors of the highly homologous bovine protein with a similar methodology.[63]

### 2.3. Aim of the project

In the present study we have performed an *in vitro* screening of virtual hits, selected by an *in silico* screening performed by Siena Biotech, with the aim to identify suitable scaffolds for the development of inhibitors of PPIs of human S100B. The challenge of targeting PPIs resides on the large and flat contact surface area for the complex between the two partner proteins (~1500-3000 Å$^2$).[64] The lack of well-defined binding sites and "druggable" cavities represents a major issue for planning *in silico* virtual screening experiments.[65] Identification of 'hot spots' is a non trivial task,[66] moreover the small molecule may bind trapping protein conformations not represented in the "frozen" X-ray structure.

Siena Biotech used the frames present in the available NMR structures for *in silico* screening, in order to consider the various conformations of both target and/or ligand as a model for system flexibility and compound selection.

A set of selected virtual hits were then screened in CERM by NMR studies and led to the identification of a number of weak ($K_d$ in the range of 0.1-1.4 mM) S100B ligands, characterized by novel chemo types. The X-ray structure of the complex between S100B and one of the identified fragments was determined as starting point for structure-based drug design.

### 2.4. Methods

NMR is a powerful tool for assessing interactions between protein and ligands. Over recent years NMR has gained importance as a versatile tool for industrial drug research. Its greatest potential in this field lies in the information that it can reveal about intermolecular interactions at various levels.[67] NMR screening exploits the fact that many NMR observables change even upon temporary binding because of global and local effects. One example of application is the qualitative screening of pharmaceutical libraries to assess binding. NMR experiments, such as the $^1$H NMR relaxation measurements, are sensitive to the overall molecular motion of test compounds, which is very different for free versus bound ligands. Thus, these simple approaches can be used to validate ligand binding and/or to identify potential ligands in mixtures of test compounds. Another approach is the chemical shift mapping, where interacting residues

can be identified and mapped on the surface of a known protein structure. Indeed, a simple parameter, the NMR chemical shift, is highly sensitive to the exact environment of the nucleus and therefore yields information about whether a small molecule binds to a target protein, what parts of the small molecule are interacting and to which part of the macromolecular target the small molecule is bound. A further application could be the modeling and sometimes the determination of the three-dimensional structure of a protein–ligand complex. The variety of readily measurable NMR parameters allows several applications of NMR in drug discovery, including assessing target druggability, pharmacophore identification, hit validation, hit optimization and potentially structure-based drug design.

The advantages offered by NMR for drug research can be summarized as follow:

1. NMR can detect the weakest ligand-target interactions of any method so far and thus can discover initial hits with even mM dissociation constants.

2. NMR is a universal screening technique that, unlike bioassays, requires no knowledge of a protein's function and therefore no target-specific setup.

3. NMR concomitantly enables a determination of binding constants.

4. NMR spectrally separates individual components, allowing the direct screening of mixtures from natural sources or combinatorial chemistry.

5. NMR gives structural information for both target and ligand with atomic resolution for the subsequent optimization of weak initial hits into strongly binding drug candidates.

### 2.4.1. NMR ligand-based screening

**Does a compound bind to the target protein?** The NMR ligand-based screening experiments answer this question. The exquisite sensitivity of ligand NMR parameters in the bound and free states constitutes the basis for all ligand-based techniques.

The typical implementation of ligand-based methods compares the NMR parameters of a mixture of compounds in the presence and absence of the target molecule. This approach renders the molecular weight of the receptor molecule not subject to size limitations. In fact, the most powerful ligand-based approaches become more sensitive when dealing with larger receptors. Additionally, ligand observation bypasses the need to produce milligram quantities of isotope labeled protein. Depending

on the approach, less than 1 mg of unlabeled protein is required for these experiments. An obvious disadvantage of ligand-based approaches is the inability to localize the binding site of the small molecule hits on the receptor.

One of the most utilized techniques is the WaterLOGSY (Water-Ligand Observed by Gradient SpectroscopY) experiment.[68]

### 2.4.1.1. WaterLOGSY

The WaterLOGSY[69] technique represents a powerful method for primary screening in the identification of compounds interacting with macromolecules. In the WaterLOGSY experiment the bulk water magnetization selectively perturbed is partially transferred to the free ligand via the ligand–protein complex. The occurring transfer of magnetization is therefore water → receptor → ligand (Fig. 2.10).
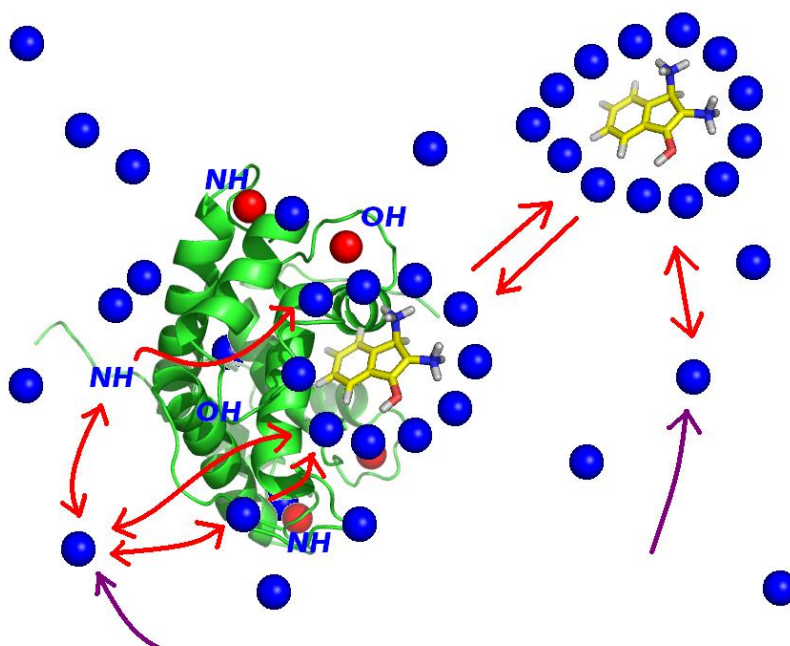


**Figure 2.10:** Magnetization transfer mechanisms underlying WaterLOGSY. Magnetization transfer from bulk water to ligand occurs via labile receptor protons within and remote from the ligand-binding site as well as from long-lived water molecules within the binding pocket.

Inverted water magnetization can be transferred to the bound ligands via three simultaneous pathways.

- One pathway involves direct $^{1}$H-$^{1}$H cross-relaxation between the bound ligand and "bound" water molecules at the binding site. Such water molecules are bound in the sense that their residence times on the surface of the receptor exceeds the receptor rotational correlation time. As a result, the rate constants governing dipole-dipole intermolecular cross-relaxation between the bound ligand and water molecules are negative and tend to bring the ligand magnetization to the same inverted state as that of the water.

- A second pathway is direct cross-relaxation with exchangeable receptor NH and OH protons within the binding site. Chemical exchange of these protons with those of bulk water inverts their magnetization. These NH/OH then propagate inversion to the bound ligand protons via intermolecular dipole-dipole cross-relaxation.

- The third pathway involves indirect cross-relaxation with remote exchangeable NH/OH protons via spin diffusion. The inverted magnetization is then relayed to other non labile spins via spin diffusion.

The bound ligand interacts directly or indirectly with inverted water spins via dipole-dipole interactions with sufficiently long rotational correlation times $\tau_c$ to yield negative cross-relaxation rates. By contrast, nonbinders' dipolar-dipolar interactions with water have much shorter $\tau_c$, leading to positive cross-relaxation rates. As a consequence, binders and nonbinders display WaterLOGSY peak intensities of opposite sign, thus providing an easy means to discriminate between them (Fig. 2.11). Therefore, the WaterLOGSY technique allows binding compounds to pick up the bulk water inversion while residing in the receptor binding site.

Another mechanism exists for the transfer of magnetization to free ligands, which is via chemical exchange of bulk water with exchangeable ligand protons, and consequent intermolecular ligand dipole-dipole cross relaxation. This effect has the potential to complicate interpretation of data but can be easily correct by collecting control WaterLOGSY spectra for free ligands.
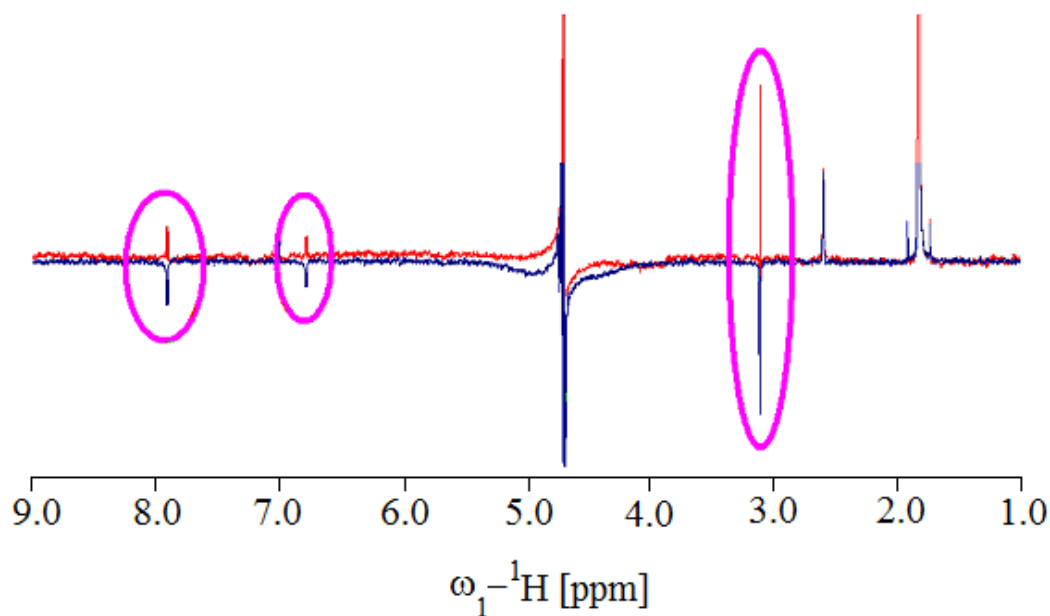
**Figure 2.11.** Water-LOGSY spectrum of the free ligand (red) and of the ligand in the presence of a protein (blue). The signals belonging to the binding ligand are highlighted in magenta.

The major aspects of the Water-LOGSY technique are summarized in Table 2.1.

|  | Water-LOGSY |
| --- | --- |
| *Large protein (>30 kDa)* | Yes |
| *Small protein (<10 kDa)* | No |
| *Amount of protein (nM) at 500 MHz* | $\approx 25$ |
| $K_d$ *tight binding* | 100 pM |
| $K_d$ *weak binding* | $\approx 10$ mM |

**Table 2.1.** Water-LOGSY major characteristics.

*2.4.2. NMR protein-based screening*

**Where is the ligand localized on the protein surface?** The NMR target-observed screening answers this question. Observation on the target is neither restricted by the ligands' size nor by an upper affinity limit. It also immediately reveals different binding sites, enables direct distinction between specific and unspecific binding and provides a wealth of structural information for ligand optimization. This information, however, can only be extracted if the spectral assignments for the target are known. Inherent limitations often preclude observation on the large target molecule, although relaxation-optimized techniques combined with appropriate extent of sample deuteration may be of benefit. The most used protein based technique is based on the chemical shift perturbation.

### 2.4.2.1. Chemical shift perturbations

One of the advantages of NMR spectroscopy is its ability to efficiently map the interaction surface of a protein with ligands.[70] This is based on the fact that the chemical shift of a particular nucleus is a very sensitive function of changes in its chemical environment.

The most used NMR mapping technique is based on Chemical Shift Perturbations (CSPs). This procedure exploits differences in chemical shift between free and bound protein in $^{15}$N-$^1$H and two-dimensional correlation spectra of the labeled target upon titration of a unlabeled ligand. The interaction causes environmental changes on the protein interfaces and, hence, affects the chemical shifts of the nuclei in this area (Fig. 2.12A). The quantification of the chemical shift variation for each protein residue upon binding of the ligand requires the calculation of an average weighted $^1$H-$^{15}$N chemical shift difference following the equation:[71]

$$\Delta\delta \ (NH) = \sqrt{\frac{\left[\Delta\delta\left(^1H\right)\right]^2 + \left[\Delta\delta\left(^{15}N\right)/5\right]^2}{2}} \tag{2.1}$$

The average weighted CSP values for each backbone amide are usually plotted as a function of the residue number, thus providing the so called Garrett plot, as exemplified in Figure 2.12B. The analysis of the plot provides a per-residue characterization of the protein-ligand interface (Fig. 2.12C), thus allowing the study of local effects that are associated with ligand binding and the identification of the binding area on the protein. This procedure is of course possible only if the assignment of backbone amide resonances is available. One drawback of CSPs is the possible reflection of indirect effects linked to structural changes induced by ligand binding.
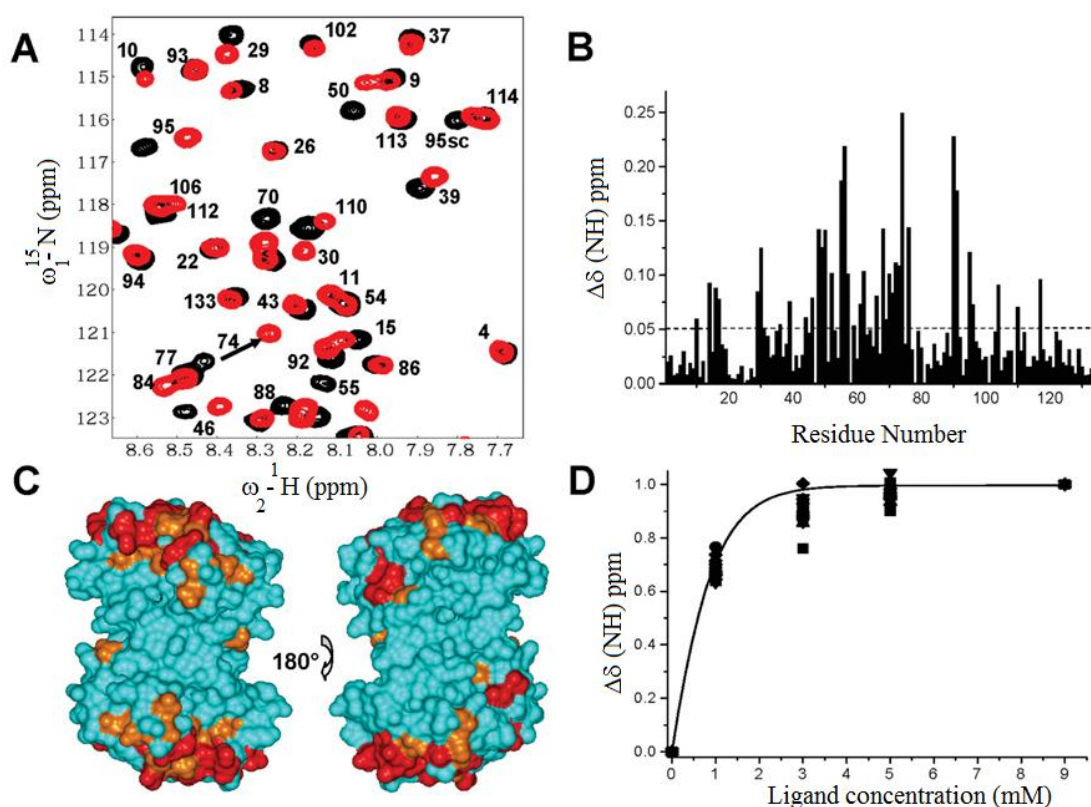


**Figure 2.12.** Theoretical example of chemical shift mapping. (A) Superimposition of $^{15}N$–$^{1}H$ HSQC (Heteronuclear Single-Quantum Coherence)[72] spectra of protein without ligand (black cross-peaks) and protein with ligand (red cross-peaks). (B) $^{15}N$ and $^{1}H$ weighted chemical shift changes between protein without ligand and with ligand are plotted against the amino acid sequence of protein. (C) Residues of the protein that have been most shifted by binding to lactose are highlighted in red and orange. (D) The fractional change in protein HSQC weighted chemical shifts is plotted against the concentration of the ligand.

CSPs also allows quantification of the kinetics, specificity, stoichiometry, and binding affinities. Plotting the $\Delta\delta$ (NH) of each residue as function of the ligand concentration (mM) and fitting the curve with the function 2.2, it is possible to calculate the dissociation constant ($K_d$) (Fig. 2.12D).

$$y = \frac{Bx}{K_d + x} \qquad\qquad (2.2)$$

where B is the limit value $\Delta\delta$ (NH) corresponding to the maximum concentration of the protein in complex with the ligand, $y = \Delta\delta$ (NH) and x represents the ligand concentration (mM).

The chemical shift approach is arguably one of the most robust, reliable and reproducible ligand binding assays currently available. However, although shift perturbation measurements reveal which residues are located at an intermolecular interface, they do not in themselves provide information about the relative orientation of the two partners. X-ray crystallography or NMR spectroscopy provide such information. The following data represents an example of a combined NMR-X-ray approach where solution NMR was used to screen potential ligands and to identify their binding areas on the protein surface, while crystallography provided structural information on the complex.
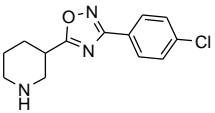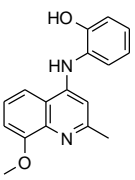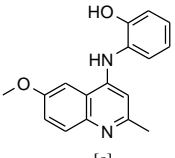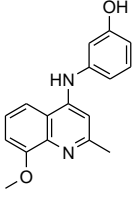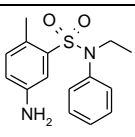
**2.5. Results**

*2.5.1. Virtual screening*

The fragment library that Siena Biotech scientists used for virtual screening purposes was generated starting from a data (6 M compounds) of commercially available compounds. The collection was filtered by the rule of three (molecular weight <300, Hydrogen Bond Acceptor (HBA) and Hydrogen Bond Donor (HBD) < 3).[73] Undesired and reactive groups, e.g., acyl chlorides, azides, aldehydes, hydrazones, etc. were also removed from the collection. Furthermore, an insolubility model[74] was applied in order to filter out potentially insoluble compounds. The resulting set of 123,000 fragments was submitted to virtual screening.

Three complementary *in silico* approaches 280 compounds were selected.

*2.5.2. NMR screening*

NMR screening of the 280 compounds selected by in silico approaches was performed using ligand-based WaterLOGSY experiments.[68,75] 123 molecules showed peak intensity variation > 20% in the presence of protein and were classified in different classes of ligand strength on the basis of the observed signal intensity changes. Peak intensity variations <20% were considered not meaningful and typical for non interacting molecules; reduction of the original signal intensity in the 20-60% range was taken as an indication of a weak binder; strong reduction or absence (i.e. variations in intensity in the 60-100% range) of the signals was taken as an indication that the molecule is a medium affinity ligand; peak intensities of opposite sign with respect to the original spectrum were considered typical for strong binders. Based on the extent of the observed effect, therefore, 11 molecules could be classified as strong binders, 15 as medium-strength binders, and 97 as weak binders. The strong and medium binders were used to titrate $^{15}$N-labelled protein samples in $^{15}$N-HSQC spectra. In all cases, fast exchange between free and bound forms was observed on the NMR chemical shift time scale. Following the chemical shift perturbations (equation (2.1)) of the backbone amide signals during the titration with the organic compounds, it was possible to identify the binding area. Based on this type of analysis, 5 selective binders were identified, with

average chemical shift variations $\geq 0.03$ ppm. The five molecules correspond to three different scaffolds, as summarized in the first column of Table 2.2.

| Molecular FormulaLibrary code | MW | Estimated $K_d$ (mM) | Ligand Efficiency (kcal mol-1) Min-Max |
|---|---|---|---|
| 205A[a] | 263.7 | 0.5-1.0 | 0.25-0.23 |
| 684A[b] | 280.3 | 0.6-1.2 | 0.21-0.19 |
| 683A[c] | 280.3 | 0.1-0.5 | 0.26-0.21 |
| 728A[b] | 280.3 | 0.1-1.0 | 0.26-0.19 |
| 829[d] | 290.4 | 0.6-1.4 | 0.22-0.19 |

[a] Compound selected via docking (S100B-TRTK12)

[b] Compound selected via structure-based pharmacophore approach (S100B-TRTK12)

[c] Compound selected via ligand-based pharmacophore

[d] Compound selected via structure-based pharmacophore (S100B-p53).

**Table 2.2.** NMR screening results.

For these ligands estimates of $K_d$ were obtained plotting the weighted average chemical shifts variations of perturbed residues as a function of the concentration of the added ligand and found to be in the 0.1-1.4 mM range (Table 2.2). Calculated ligand efficiencies for these compounds are somewhat smaller than the minimal 0.3 kcal mol[-1]

threshold value[76] and the number of non-hydrogen atoms is relatively high (18-21 non-hydrogen atoms). Nevertheless they still fall in a class suitable for optimization. For example, a remarkable improvement in ligand efficiency from 0.26 to 0.37 kcal mol$^{-1}$ has been reported upon addition of two non-hydrogen atoms to a compound of 31 non-hydrogen atoms.[77] Structural information on the binding mode is expected to help for an efficient progression of their optimization.[78] This goal was here pursued by a combined NMR/X-ray approach.

A detailed analysis of the chemical shift perturbations in $^{15}$N-HSQC spectra reveals that there is common binding region for all 5 ligands, defined by the hinge loop, the C-terminal part of the helix α4 and the helix α3 (Fig. 2.13), that corresponds to the broad interaction area with p53.
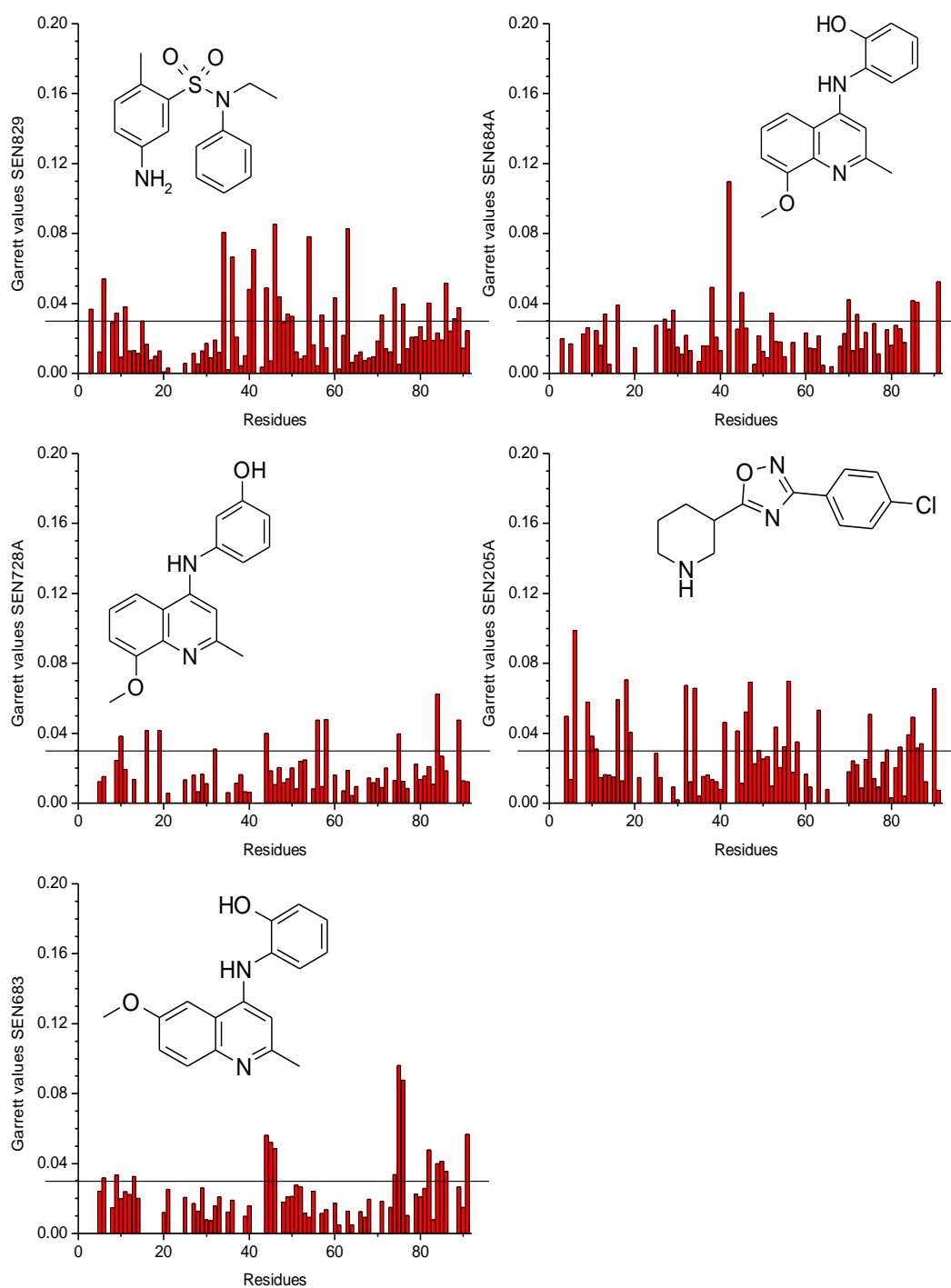
**Figure 2.13.** Plots of the weighted average ${}^{1}$H and ${}^{15}$N chemical shift variation (Garrett values) for the most relevant compounds.

In order to better prove that binding of our ligands really involves spots at the target protein-protein binding area, the high affinity TRTK-12 peptide was used as test compound, to verify that its binding is able to displace the lower affinity small molecules. Binding of TRTK-12 to S100B results in the disappearance of a number of resonances in the $^1$H-$^{15}$N HSQC spectrum of the isolated protein, and growth of new signals (Fig. 2.14A).



**Figure 2.14.** (A) Superimposition of the $^{15}$N-HSQC spectra of S100B (black trace) and S100B-TRTK-12 complex in the 1:1 ratio (green trace). (B) Superimposition of the $^{15}$N-HSQC spectra of S100B (black trace) and S100B-SEN205A complex 1:5 ratio (red trace). (C) Superimposition of the $^{15}$N-HSQC spectra of the S100B-TRTK-12 complex, 1:1 ratio (green trace) and the S100B-SEN205A complex at a 1:5 ratio (red trace). (D) Superimposition of the $^{15}$N-HSQC spectra of the S100B-TRTK-12 complex at a 1:1 ratio (green trace) and S100B-SEN205A complex titrated with TRTK12 at a

1:5:1 ratio (magenta). Very low intensity peaks are below the displayed threshold in some of the spectra.


This behavior is typical for a slow exchange regime, on the NMR chemical shift time scale, between free and bound protein forms. Slow exchange on the chemical shift NMR time scale is consistent with the much larger affinity of TRTK-12 for S100B with respect to our ligands. Consistently, maximum chemical shift differences between resonances (calculated according to equation (2.1)) of the same residue in the free and TRTK-12-bound state are larger (0.3 ppm) than the chemical shift variations measured in the presence of small ligands (0.18 ppm).

Residues changing their position in the NMR spectrum of S100B bound to TRTK-12 are many and clustered in the p53 interaction area, constituted by the hinge loop and part of the helices III, IV and I (Fig. 2.9). The large number of involved residues indicates a wide contact area between the peptide and the protein. The ensemble of residues affected by TRTK-12 binding invariably includes those experiencing chemical perturbations in the presence of our ligands. This fact already indicates that binders resulting from our screening are indeed targeting spots at the S100B/TRTK-12 interaction surface. Competitive displacement of our ligands upon addition of TRTK-12 was demonstrated for three molecules representative of the three identified scaffolds i.e., SEN205A, SEN683A, and SEN829. The complex between SEN205A and S100B was obtained using a 5-fold excess of the small molecule with respect to the concentration of the monomeric S100B (Fig. 2.14B). Addition of TRTK-12 to this solution induces formation of a new species, in slow-exchange with the pre-existing one. Addition of 0.5 equivalents of TRTK-12 is sufficient to fully displace the small molecule. Upon addition of 1 equivalent of TRTK-12 a spectral fingerprint identical to that of the 1:1 S100B/TRTK-12 adduct (Fig. 2.14D) is obtained. Therefore, TRTK-12 displaces SEN205A from its binding site and substitutes it. The same behavior was obtained for SEN683A, and SEN829 and is reported in Figure 2.15 and 2.16.
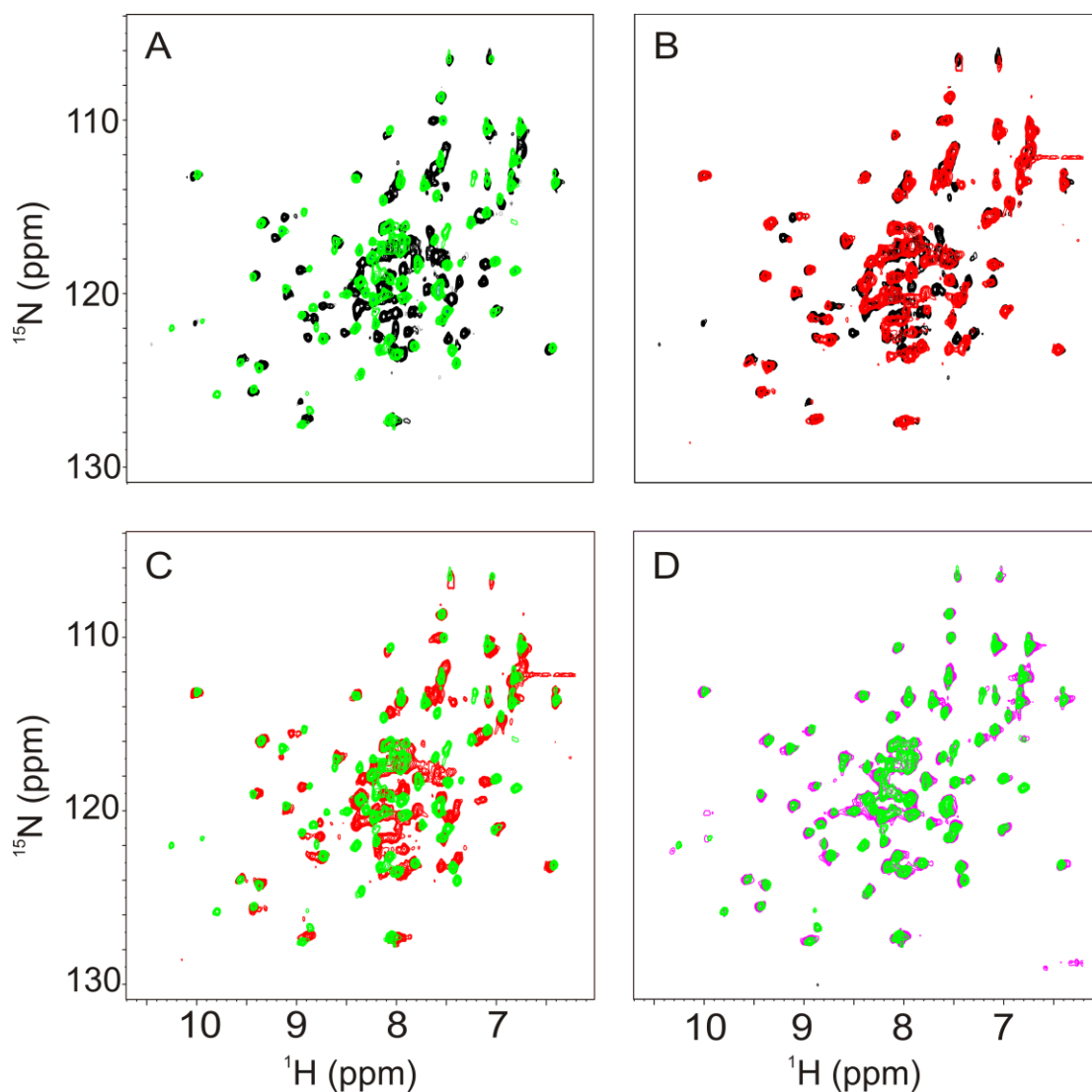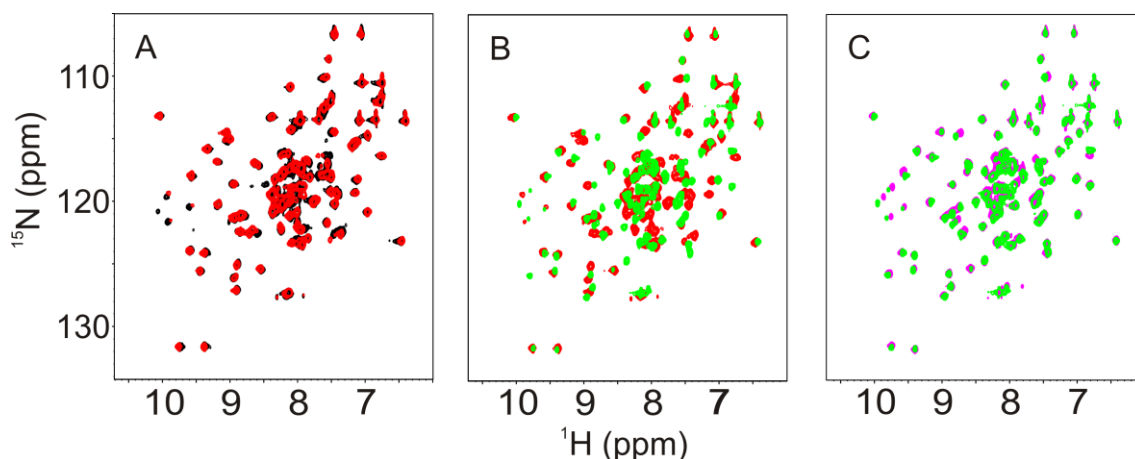
**Figure 2.15.** (A) Superimposition of the $^{15}$N-HSQC spectra of S100B (black trace) and S100B-SEN683A complex, 1:4 ratio (red trace). (B) Superimposition of the $^{15}$N-HSQC spectra of the S100B-TRTK-12 complex, 1:1 ratio (green trace) and the S100B-SEN683A complex,1:4 ratio (red trace). (C) Superimposition of the $^{15}$N-HSQC spectra of the S100B-TRTK-12 complex, 1:1 ratio (green trace) and S100B-SEN683A complex titrated with TRTK12 in the 1:4:1 ratio (magenta). Very low intensity peaks are below the displayed threshold in some of the spectra.
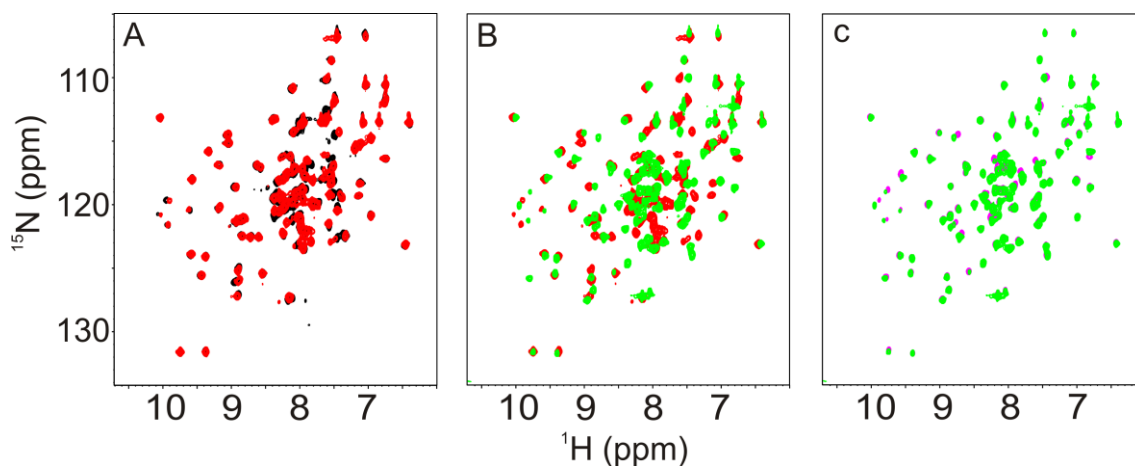


**Figure 2.16.** (A) Superimposition of the $^{15}$N-HSQC spectra of S100B (black trace) and S100B-SEN829 complex, 1:2 ratio (red trace). (B) Superimposition of the $^{15}$N-HSQC spectra of the S100B-TRTK-12 complex, 1:1 ratio (green trace) and the S100B-SEN829 complex, 1:2 ratio (red trace). (C) Superimposition of the $^{15}$N-HSQC spectra of the S100B-TRTK-12 complex, 1:1 ratio (green trace) and S100B-SEN829 complex titrated with TRTK12 in the 1:2:1 ratio (magenta). Very low intensity peaks are below the displayed threshold in some of the spectra.

*2.5.3. Protein crystallography*

The structure of S100B-SEN205A complex has been determined by X-ray crystallography by the research group of Mangani from Siena, using a reflection data set collected at the EMBL X11 beamline c/o DESY, Hamburg. The structure has been easily solved by molecular replacement and clearly shows the presence of a S100B homodimer in the crystal asymmetric unit. The refined model consists of two S100B independent subunits where only the last 2 residues at the C-terminus are missing in chain A and the last residue is missing in chain B. The final model consists of 179 residues (Met0 to Glu89 (His90 in chain B)), 4 calcium ions and 96 water molecules. The Ramachandran plot shows that 91.7% residues are in preferred regions, 8.3% in allowed regions and none in the disallowed regions.

The electron density in the cavity delimited by helices III and IV (Fig. 2.17) in both subunits is consistent with the dimensions of the SEN205A fragment. The region appearing as the SEN205A binding site is not involved in any crystal contacts so that the molecule binding is not influenced by lattice interactions.
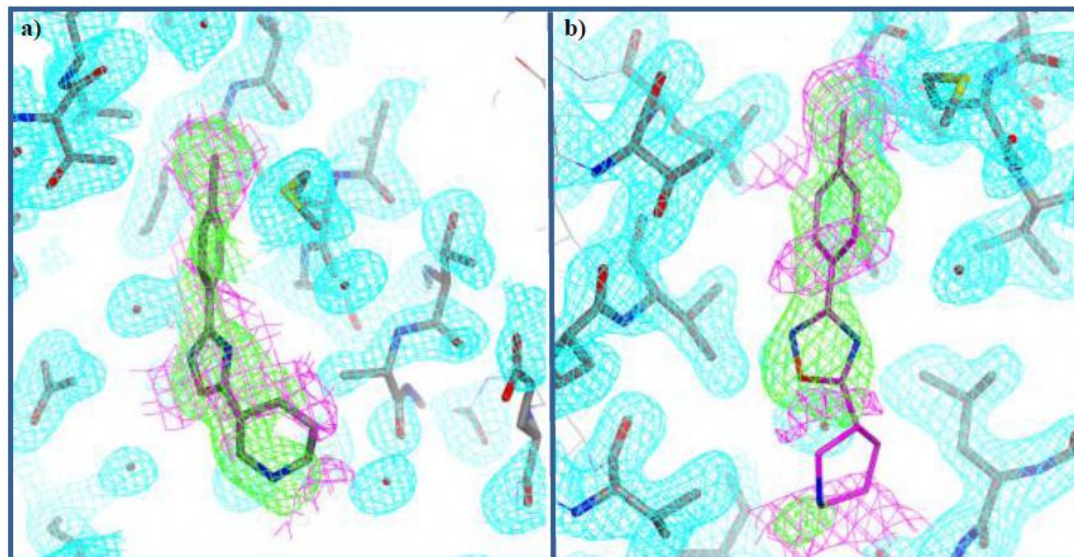


**Figure 2.17.** Electron density maps of the fragment in the two protein subunits. Protein maps are in cyan, ligand maps are in green. Positive map peaks of the protein refined without the ligand are represented in magenta.

Ligand fitting to the density, followed by refinement of the complex, showed a well defined orientation of the fragment within the binding site.

The orientations of SEN205A appear to be different between the two S100B subunits (Fig. 2.18).
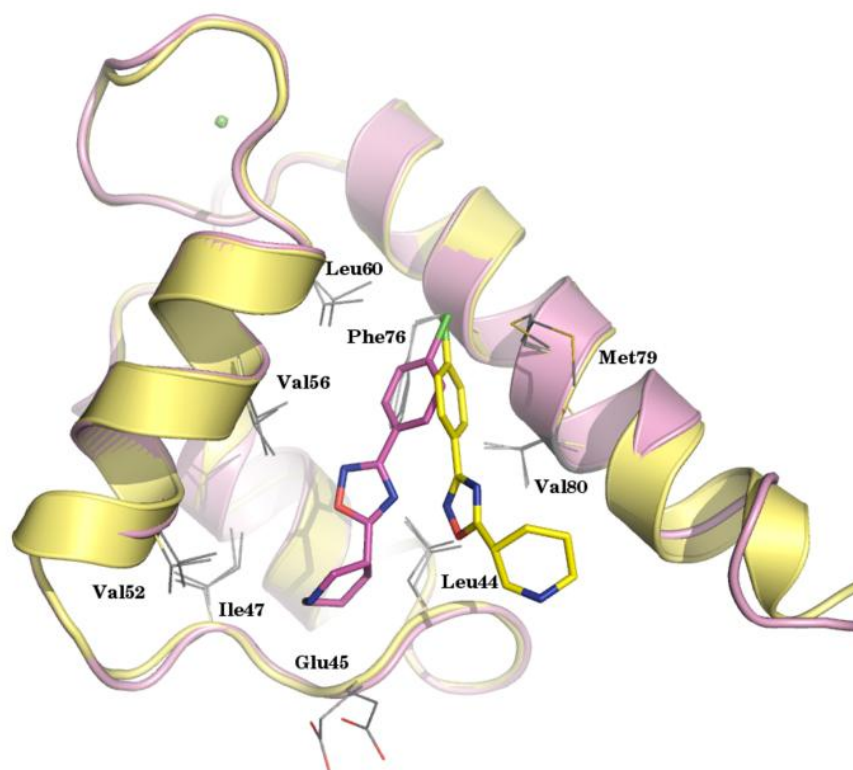


**Figure 2.18.** Different fragment orientation present in the crystal structure. The two S100B monomers with bound fragments (chain A, in pale yellow, chain B in light pink and bound fragments in yellow and magenta, respectively) are overlaid on backbone atoms, showing two different orientation of compound SEN205A within the same binding site of the two subunits.

A cross-check with NMR experimental data highlighted that the location of residues showing NH-chemical shift variations is consistent with the binding mode determined for chain B interaction site. The hydrophobic moiety of SEN205A fits with a good shape complementarity into the hydrophobic pocket determined by residues Leu32, Ile36, Leu44, Ile47, Val52, Leu60 and Phe76. In particular, the rotation of Met79 side chain opens a small cavity where the chlorine substituent of the phenyl is placed. The piperidine ring protrudes toward the protein residues of the flexible hinge loop, which is mainly composed of polar residues.

## 2.6. Discussion

Analysis of S100B complexes with binding peptides highlighted that the calcium-dependent S100B-protein interaction is characterized by a wide contact surface with a hydrophobic central region and polar surrounding residues (Fig. 2.10). This kind of PPI between an amphypatic α-helix on a protein and a hydrophobic cleft on its partner is similar to other therapeutically relevant interactions such as Bcl-$x_L$-Bak (See section 3.2.4).

Small molecule modulation of PPIs for therapeutic intervention represents a challenging task. Fesik et al. have recently reported[79] the discovery of a potent inhibitor of the anti-apoptotic protein Bcl-$x_L$ via a fragment-based SAR by NMR approach. Identification of fragment targeting hot spot has thus proved to be a successful strategy for developing PPI inhibitors. Moreover molecular fragments would adapt more easily to the intrinsic flexibility of protein-protein interface.

The virtual screening method allowed the selection of a relatively small number of virtual hits to be submitted to biophysical screening, by rationally sampling a set of 280 fragments from a large library of 123,000 compounds. The 5 most active compounds originated from our *in vitro* approaches represents the 2% of the screened library, thus confirming the validity/effectiveness of NMR-based screening approach.

It can be observed that compounds SEN683A, SEN684A and SEN728A are close analogues (Table 2.2).

It is worth noting that 4 out of the 5 compounds identified originated from models based on the S100B structure bound to the most potent ligand TRTK-12. Analysis of S100B complexes showed that, although the C-terminal part of p53 and TRTK-12 peptides interacts with S100B broadly in the same binding site, an induced fit effect is evident and S100B conformation varies following specific peptide binding (Fig. 2.9). TRTK-12 shows a higher shape complementarity to S100B surface. NMR-estimated $K_d$ and ligand efficiency values for the identified molecules are low. However, the five binders identified by NMR really target the TRTK-12 binding site and show potentialities for future optimization.

**2.7. Experimental section**

*2.7.1. Protein samples and organic molecules*

All materials were obtained from commercial suppliers and used without further purification. Unlabelled and ^{15}N-enriched human S100B for NMR and X-ray studies was purchase from ProtEra srl. Potential ligands selected by virtual screening for further biophysical characterization were purchased from Asinex, Chembridge, ChemDiv, Enamine and Specs.; the TRTK-12 peptide was purchased from Mimotopes Pty Ltd.

*2.7.2. NMR experiments*

Unlabelled and ^{15}N-enriched S100B samples were prepared in acetate buffer 20 mM pH 6.0, and 20 mM $CaCl_2$. Samples for WaterLOGSY (Water Ligand Observed via Gradient SpectroscopY)[68] experiments were 20 μM in protein monomer concentration and 800 μM in candidate ligand concentration. Samples for ^{15}N-HSQC[72] and ^{15}N NOESY-HSQC[80] spectra were 300 μM in monomeric protein concentration. NMR spectra were recorded at 298 K. WaterLOGSY experiments were run on a Bruker AVANCE 700 MHz spectrometer equipped with a 5 mm triple-resonance TXI, Z-gradient probe, and an automatic sample changer. For each compound, a reference 1D WaterLOGSY spectrum of the compound alone and a 1D WaterLOGSY spectrum in the presence of the protein were recorded. The initial set up of the WaterLOGSY experiment was performed on pentamidine, a molecule known to bind S100B at the p53 site with high affinity[81]; this provided us an internal control of the reliability of our methodological approach. WaterLOGSY NMR experiments employed a 2 ms selective rectangular 180° pulse at the water signal frequency and a NOE mixing time of 2 s. ^{15}N HSQC and 3D ^{15}N NOESY-HSQC (mixing time of 110 ms) experiments were acquired on a Bruker AVANCE 800 MHz spectrometer, equipped with a cryoprobe. The acquisition parameters for these NMR spectra are provided as Supporting Information (Table 2.3).

| Experiments | Dimension of acquired data (nucleus) | | | Spectral width (ppm) | | |
|---|---|---|---|---|---|---|
| | $t_1$ | $t_2$ | $t_3$ | $F_1$ | $F_2$ | $F_3$ |
| WaterLOGSY[a] | 16384($^1$H) | | | 12.0 | | |
| $^1$H-$^{15}$N-HSQC[b] | 100($^{15}$N) | 1024($^1$H) | | 14.3 | 40.0 | |
| $^{15}$N-edited [$^1$H-$^1$H]-NOESY[b] | 256($^1$H) | 40($^{15}$N) | 1024($^1$H) | 16.0 | 40.0 | 16.0 |

[a] Data acquired on a 700 MHz spectrometer.
[b] Data acquired on a 800 MHz spectrometer equipped with a cryo-probe (TXI, 5-mm)
All the spectra were processed using the standard Bruker software (XWINNMR) and analyzed by the CARA program.

**Table 2.3.** Acquisition parameters for the NMR experiments, acquired at 298 K.

The chemical shift change (in ppm) of the individual amide pairs was defined as the weighted average $^1$H and $^{15}$N chemical shift variations, according to the equation (2.1).

$K_d$ values in the case of fast exchange between free and bound forms were calculated by plotting the weighted average $^1$H and $^{15}$N chemical shifts of affected residues as a function of fragment concentration during the titration and fitted considering the one site binding mode. $K_d$ values in the case of slow exchange between free and bound forms were calculated by plotting the decrease in intensity of the resonances of the free form as a function of peptide concentration and fitted considering the one site binding mode.[82] The ligand efficiency (LE) was also calculated on the basis of the identified $K_d$ values through the following equation (2.3):[76,83]

$$LE = \frac{\Delta G}{N_{non-hydrogen-atoms}} \approx -\frac{RT \ln K_d}{N_{non-hydrogen-atoms}} \qquad (2.3)$$

**2.8. Conclusion**

Starting from an *in silico* screening, preformed by Siena Biotech, that lead to the identification of a library of 280 fragments that targets the p53-S100B binding interface, we performed an *in vitro* NMR screening that confirmed 5 selective binders. The estimated $K_d$ is in the range of 0.1-1.4 mM and the ligand efficiency in the range of 0.19-0.26 kcal mol$^{-1}$. Although weak binders, these molecules target the binding site of TRTK-12 peptide, thus validating the reliability of the approach.

These results are currently being utilized to develop the fragments into potent lead compounds via structure-based optimization.

Reference List

1. Berridge, M. J. The versatility and complexity of calcium signalling. *Novartis Found Symp* 52-64 (2001).

2. Kretsinger, R. H. Structure and evolution of calcium-modulated proteins. *CRC Crit. Rev. Biochem.* **8**, 119-174 (1980).

3. Kretsinger, R. H. & Nockolds, C. E. Carp Muscle Calcium-binding Protein. II. Structure Determination and General Description. *J. Biol. Chem.* **248**, 3313-3326 (1973).

4. Porumb, T., Crivici, A., Blackshear, P. J. & Ikura, M. Calcium binding and conformational properties of calmodulin complexed with peptides derived from myristoylated alanine-rich C kinase substrate (MARCKS) and MARCKS-related protein (MRP). *Eur. Biophys. J.* **25**, 239-247 (1997).

5. Kawasaki, K. & Kretsinger, R. H. Calcium-binding proteins 1: EF-hands. *Protein* **1**, 343-517 (1994).

6. Moore, B. W. A soluble protein characteristic of the nervous system. *Biochem. Biophys. Res. Commun.* **19**, 739-744 (1965).

7. Zimmer, D. B., Cornwall, E. H., Landar, A. & Song, W. The S100 protein family: history, function, and expression. *Brain Res. Bull.* **37**, 417-429 (1995).

8. Fano, G. *et al.* The S-100: a protein family in search of a function. *Prog. Neurobiol.* **46**, 71-82 (1995).

9. Schafer, B. W. & Heizmann, C. W. The S100 family of EF-hand calcium-binding proteins: functions and pathology. *Trends Biochem. Sci.* **21**, 134-140 (1996).

10. Donato, R. S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles. *Int. J. Biochem. Cell Biol.* **33**, 637-668 (2001).

11. Drohat, A. C. *et al.* Solution structure of rat apo-S100B(beta beta) as determined by NMR spectroscopy. *Biochemistry* **35**, 11577-11588 (1996).

12. Kilby, P. M., Van Eldik, L. J. & Roberts, G. C. The solution structure of the bovine S100B protein dimer in the calcium-free state. *Structure* **4**, 1041-1052 (1996).

13. Smith, S. P. & Shaw, G. S. A novel calcium-sensitive switch revealed by the structure of human S100B in the calcium-bound form. *Structure* **6**, 211-222 (1998).

14. Isobe, T. & Okuyama, T. The amino-acid sequence of S-100 protein (PAP I-b protein) and its relation to the calcium-binding proteins. *Eur. J. Biochem.* **89**, 379-388 (1978).

15. Mani, R. S., Shelling, J. G., Sykes, B. D. & Kay, C. M. Spectral studies on the calcium binding properties of bovine brain S-100b protein. *Biochemistry* **22**, 1734-1740 (1983).

16. Baudier, J., Glasser, N. & Gerard, D. *J. Biol. Chem.* **261**, 8192-8203 (1986).

17. Baudier, J. & Gerard, D. Ions binding to S100 proteins: structural changes induced by calcium and zinc on S100a and S100b proteins. *Biochemistry* **22**, 3360-3369 (1983).

18. Drohat, A. C., Baldisseri, D. M., Rustandi, R. R. & Weber, D. J. Solution structure of Calcium-bound rat S100B(bb) as determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **37**, 2729-2740 (1998).

19. Rety, S. *et al.* The crystal structure of a complex of p11 with the annexin II N-terminal peptide. *Nat. Struct. Biol.* **6**, 89-95 (1999).

20. Kordel, J., Skelton, N. J. & Chazin, W. J. High-resolution structure of calcium-loaded calbindin $D_{9k}$. *J. Mol. Biol.* **231**, 711-734 (1993).

21. Rustandi, R. R., Drohat, A. C., Baldisseri, D. M., Wilder, P. T. & Weber, D. J. The Ca(2+)-dependent interaction of S100B(beta beta) with a peptide derived from p53. *Biochemistry* **37**, 1951-1960 (1998).

22. Rustandi, R. R., Baldisseri, D. M., Drohat, A. C. & Weber, D. J. Structural changes in the C-terminus of Ca2+-bound rat S100B (beta beta) upon binding to a peptide derived from the C-terminal regulatory domain of p53. *Protein Sci.* **8**, 1743-1751 (1999).

23. Rustandi, R. R., Baldisseri, D. M. & Weber, D. J. Structure of the negative regulatory domain of p53 bound to S100B(betabeta). *Nat. Struct. Biol.* **7**, 570-574 (2000).

24. Garbuglia, M. *et al.* Role of the C-terminal extension in the interaction of S100A1 with GFAP, tubulin, the S1. *Biochem. Biophys. Res. Commun.* **254**, 36-41 (1999).

25. Kilby, P. M., Van Eldik, L. J. & Roberts, G. C. Identification of the binding site on S100B protein for the actin capping protein CapZ. *Protein Sci.* **6**, 2494-2503 (1997).

26. Heizmann, C. W., Fritz, G. & Schafer, B. W. S100 proteins: structure, functions and pathology. *Front Biosci.* **7**, d1356-d1368 (2002).

27. Mrak, R. E., Sheng, J. G. & Griffin, W. S. Glial cytokines in Alzheimer's disease: review and pathogenic implications. *Hum. Pathol.* **26**, 816-823 (1995).

28. Van Eldik, L. J. & Griffin, W. S. S100 beta expression in Alzheimer's disease: relation to neuropathology in brain regions. *Biochim. Biophys. Acta* **1223**, 398-403 (1994).

29. Mariggio, M. A., Fulle, S., Calissano, P., Nicoletti, I. & Fano, G. The brain protein S-100ab induces apoptosis in PC12 cells. *Neuroscience* **60**, 29-35 (1994).

30. Camby, I. *et al.* Supratentorial pilocytic astrocytomas, astrocytomas, anaplastic astrocytomas and glioblastomas are characterized by a differential expression of S100 proteins. *Brain Pathol.* **9**, 1-19 (1999).

31. Takahashi, K. *et al.* Natural killer (NK) activity of cultured S100 beta-positive T-leukemia cells. *Virchows Arch. B Cell Pathol. Incl. Mol. Pathol.* **59**, 159-164 (1990).

32. Takashi, M. *et al.* Elevated concentrations of the beta-subunit of S100 protein in renal cell tumors in rats. *Urol. Res.* **22**, 251-255 (1994).

33. Suzushima, H., Asou, N., Hattori, T. & Takatsuki, K. Adult T-cell leukemia derived from S100 beta positive double-negative (CD4- CD8-) T cells. *Leuk. Lymphoma* **13**, 257-262 (1994).

34. Garbe, C. *et al.* Diagnostic value and prognostic significance of protein S-100beta, melanoma-inhibitory activity, and tyrosinase/MART-1 reverse transcription-polymerase chain reaction in the follow-up of high-risk melanoma patients. *Cancer* **97**, 1737-1745 (2003).

35. Jury, C. S., McAllister, E. J. & MacKie, R. M. Rising levels of serum S100 protein precede other evidence of disease progression in patients with malignant melanoma. *Br. J. Dermatol.* **143**, 269-274 (2000).

36. Banfalvi, T. *et al.* Use of serum 5-S-CD and S-100B protein levels to monitor the clinical course of malignant melanoma. *Eur. J. Cancer* **39**, 164-169 (2003).

37. Donato, R., Giambanco, I. & Aisa, M. C. Molecular interaction of S-100 proteins with microtubule proteins in vitro. *J. Neurochem.* **53**, 566-571 (1989).

38. Bianchi, R., Giambanco, I. & Donato, R. S-100 protein, but not calmodulin, binds to the glial fibrillary acidic protein and inhibits its polymerization in a Ca(2+)-dependent manner. *J. Biol. Chem.* **268**, 12669-12674 (1993).

39. Sheu, F. S., Huang, F. L. & Huang, K. P. Differential responses of protein kinase C substrates (MARCKS, neuromodulin, and neurogranin) phosphorylation to calmodulin and S100. *Arch. Biochem. Biophys.* **316**, 335-342 (1995).

40. Lin, L. H., Van Eldik, L. J., Osheroff, N. & Norden, J. J. Inhibition of protein kinase C- and casein kinase II-mediated phosphorylation of GAP-43 by S100 beta. *Brain Res. Mol. Brain Res.* **25**, 297-304 (1994).

41. Baudier, J. *et al.* Comparison of S100b protein with calmodulin: interactions with melittin and microtubule-associated tau proteins and inhibition of phosphorylation of tau proteins by protein kinase C. *Biochemistry* **26**, 2886-2893 (1987).

42. Lin, J. *et al.* Inhibition of p53 transcriptional activity by the S100B calcium-binding protein. *J. Biol. Chem.* **276**, 35037-35041 (2001).

43. Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. P53 Mutations in Human Cancers. *Science* **253**, 49-53 (1991).

44. Vousden, K. H. & Lu, X. Live or let die: the cell's response to p53. *Nat. Rev. Cancer* **2**, 594-604 (2002).

45. Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307-310 (2000).

46. Slee, E. A., O'Connor, D. J. & Lu, X. To die or not to die: how does p53 decide? *Oncogene* **23**, 2809-2818 (2004).

47. Bode, A. M. & Dong, Z. G. Post-translational modification of p53 in tumorigenesis. *Nature Reviews Cancer* **4**, 793-805 (2004).

48. Zheng, H. *et al.* The prolyl isomerase Pin1 is a regulator of p53 in genotoxic response. *Nature* **419**, 849-853 (2002).

49. Zacchi, P. *et al.* The prolyl isomerase Pin1 reveals a mechanism to control p53 functions after genotoxic insults. *Nature* **419**, 853-857 (2002).

50. Samuels-Lev, Y. *et al.* ASPP proteins specifically stimulate the apoptotic function of p53. *Mol. Cell* **8**, 781-794 (2001).

51. Chene, P. The role of tetramerization in p53 function. *Oncogene* **20**, 2611-2617 (2001).

52. el-Deiry, W. S. Regulation of p53 downstream genes. *Semin. Cancer Biol.* **8**, 345-357 (1998).

53. Ribeiro, R. C. *et al.* An inherited p53 mutation that contributes in a tissue-specific manner to pediatric adrenal cortical carcinoma. *Proc. Natl. Acad. Sci. U. S. A* **98**, 9330-9335 (2001).

54. Wilder, P. T., Rustandi, R. R., Drohat, A. C. & Weber, D. J. S100B(betabeta) inhibits the protein kinase C-dependent phosphorylation of a peptide derived from p53 in a Ca2+-dependent manner. *Protein Sci.* **7**, 794-798 (1998).

55. Mayr, G. A. *et al.* Serine phosphorylation in the NH2 terminus of p53 facilitates transactivation. *Cancer Res.* **55**, 2410-2417 (1995).

56. Liang, S. H. & Clarke, M. F. Regulation of p53 localization. *Eur. J. Biochem.* **268**, 2779-2783 (2001).

57. Mocellin, S., Zavagno, G. & Nitti, D. The prognostic value of serum S100B in patients with cutaneous melanoma: a meta-analysis. *Int. J. Cancer* **123**, 2370-2376 (2008).

58. Delphin, C. *et al.* Calcium-dependent interaction of S100B with the C-terminal domain of the tumor suppressor p53. *J. Biol. Chem.* **274**, 10539-10544 (1999).

59. Buer, J. *et al.* Elevated serum levels of S100 and survival in metastatic malignant melanoma. *Br. J. Cancer* **75**, 1373-1376 (1997).

60. Inman, K. G., Baldisseri, D. M., Miller, K. E. & Weber, D. J. Backbone Dynamics of the Calcium-Signaling Protein apo-S100B as Determined by 15N NMR Relaxation. *Biochemistry* **40**, 3439-3448 (2001).

61. Ivanenkov, V. V., Jamieson, G. A., Jr., Gruenstein, E. & Dimlich, R. V. Characterization of S-100b binding epitopes. Identification of a novel target, the actin capping protein, CapZ. *J. Biol. Chem.* **270**, 14651-14658 (1995).

62. Markowitz, J. *et al.* Identification and characterization of small molecule inhibitors of the calcium-dependent S100B-p53 tumor suppressor interaction. *J. Med. Chem.* **47**, 5085-5093 (2004).

63. Arendt, Y. *et al.* Fragment docking to S100 proteins reveals a wide diversity of weak interaction sites. *ChemMedChem* **2**, 1648-1654 (2007).

64. Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001-1009 (2007).

65. Zhong, S., Macias, A. T. & MacKerell, A. D., Jr. Computational identification of inhibitors of protein-protein interactions. *Curr. Top. Med. Chem.* **7**, 63-82 (2007).

66. DeLano, W. L. Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.* **12**, 14-20 (2002).

67. Pellecchia, M., Sem, D. S. & Wuthrich, K. NMR in drug discovery. *Nat. Rev. Drug Discov.* **1**, 211-219 (2002).

68. Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M. & Stockman, B. J. WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *J. Biomol. NMR* **21**, 349-359 (2001).

69. Dalvit, C. *et al.* Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J. Biomol. NMR* **18**, 65-68 (2000).

70. Zuiderweg, E. R. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* **41**, 1-7 (2002).

71. Grzesiek, S. *et al.* The solution structure of HIV-1 Nef reveals an unexpected fold and permits delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase. *Nat. Struct. Biol.* **3**, 340-345 (1996).

72. Sklenar, V., Piotto, M., Leppik, R. & Saudek, V. Gradient-tailored water suppression for $^1$H-$^{15}$N HSQC experiments optimized to retain full sensitivity. *J. Magn. Reson. Ser. A* **102**, 241-245 (1993).

73. Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'rule of three' for fragment-based lead discovery? *Drug Discov. Today* **8**, 876-877 (2003).

74. Lamanna, C., Bellini, M., Padova, A., Westerberg, G. & Maccari, L. Straightforward recursive partitioning model for discarding insoluble compounds in the drug discovery process. *J. Med. Chem.* **51**, 2891-2897 (2008).

75. Lepre, C. A., Moore, J. M. & Peng, J. W. Theory and Applications of NMR-Based Screening in Pharmaceutical Research. *Chem. Rev.* **104**, 3641-3675 (2004).

76. Kuntz, I. D., Chen, K., Sharp, K. A. & Kollman, P. A. The maximal affinity of ligands. *Proc. Natl. Acad. Sci. USA* **96**, 9997-10002 (1999).

77. Lafleur, K., Huang, D., Zhou, T., Caflisch, A. & Nevado, C. Structure-based optimization of potent and selective inhibitors of the tyrosine kinase erythropoietin producing human hepatocellular carcinoma receptor B4 (EphB4). *J. Med. Chem.* **52**, 6433-6446 (2009).

78. Carr, R. A., Congreve, M., Murray, C. W. & Rees, D. C. Fragment-based lead discovery: leads by design. *Drug Discov. Today* **10**, 987-992 (2005).

79. Petros, A. M. *et al.* Discovery of a potent inhibitor of the antiapoptotic protein Bcl-xL from NMR and parallel synthesis. *J. Med. Chem.* **49**, 656-663 (2006).

80. Wider, G., Neri, D., Otting, G. & Wüthrich, K. A Heteronuclear Three-Dimensional NMR Experiment for Measurements of Small Heteronuclear Coupling Constants in Biological Macromolecules. *J. Magn. Reson.* **85**, 426-431 (1989).

81. Charpentier, T. H. *et al.* Divalent metal ion complexes of S100B in the absence and presence of pentamidine. *J. Mol. Biol.* **382**, 56-73 (2008).

82. Meyer, B. & Peters, T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew. Chem. Int. Ed.* **42**, 864-890 (2003).

83. Hopkins, A. L., Groom, C. R. & Alex, A. Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* **9**, 430-431 (2004).

# Structural model for
# Bcl-x$_L$-cytochrome c complex

## 3. Summary

In this chapter will be presented a NMR-derived model structure of the adduct between the human anti-apoptotic protein Bcl-x$_L$ and the human cytochrome c. Both proteins are involved in the programmed cell death called apoptosis. During the apoptotis mediated by mitochondria, cytochrome c is released into the cytosol where it initiates reactions that lead to the cell death.[1] *In vitro* studies shown that cytochrome c interacts specifically with Bcl-x$_L$, thus making the protein unable to perform this death reaction cascade and blocking apotosis.[2] The model structure of the adduct, reported in the present study, provides insights at the atomic level on the mechanism by which cytochrome c translocated to cytosol can be intercepted, such that the apoptosome is not assembled. Moreover, this model allows the definition of the protein-protein interaction surface and reveals key intermolecular contacts that identify new potentially druggable spots on cytochrome c and Bcl-x$_L$.

### 3.1. Apoptosis

The life and death of cells must be balanced if tissue homeostasis has to be maintained. Cell death is responsible of the maintenance of the cell populations in tissues of multicellular organisms upon physiological and pathological conditions. Cells have an intrinsic mechanism of self destruction called *programmed cell death* or *apoptosis*. The term *programmed cell death* was introduced to propose that cell death is not accidental but follows a sequence of controlled steps leading to self-destruction. *Apoptosis* (a greek term having the meaning "falling off or dropping off" in analogy to leaves falling off trees or petals dropping off flowers) had been coined in 1972 in order to describe the morphological processes leading to controlled cellular self-destruction.[3] Apoptosis is also called "clean death". In fact, during this process the cell is fragmented into compact membrane-enclosed structures, called *apoptotic bodies* that are engulfed

by macrophages and thus are removed from the tissue without damaging the surrounding cells or causing an inflammatory response, as occurs in necrosis (Fig. 3.1).[4]

Apoptosis has a central role in maintaining tissue homeostasis, in the nervous system development, in the elimination of the old cells and in the immune system. Defect in apoptosis can result in cancer, autoimmune diseases and spreading of viral infections, while neurodegenerative diseases are caused or enhanced by excessive apoptosis.[5]
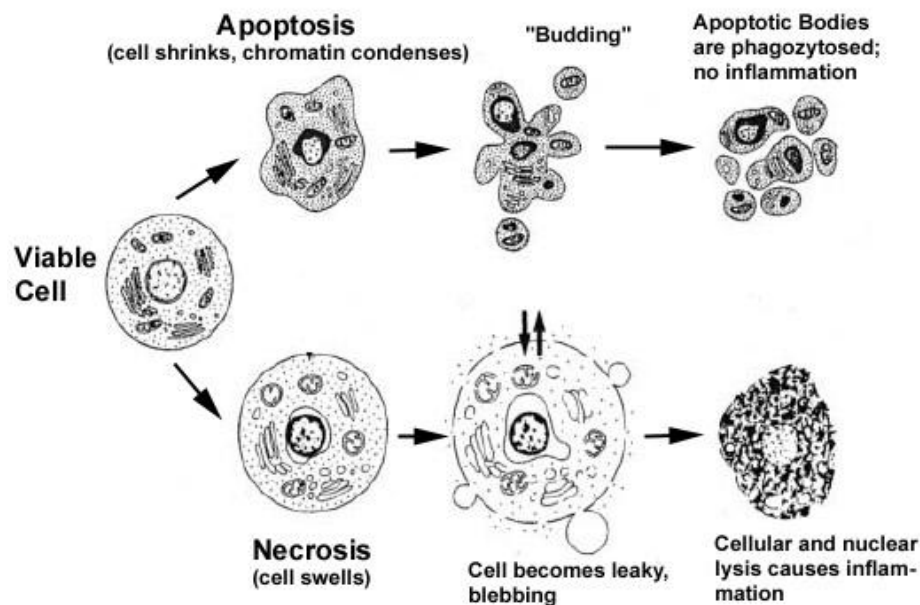


**Figure 3.1.** Hallmarks of the apoptotic and necrotic cell death process.[6]

### 3.1.1. Molecular mechanisms of apoptosis signaling pathways

In principle, two alternative pathways that initiate apoptosis exist: one, called *extrinsic pathway*, is mediated by death receptors on the cell surface, the other is mediated by mitochondria and is known as the *intrinsic pathway*. In both pathways, specific proteases, called *caspases*, are activated leading to the biochemical and morphological changes that are characteristic of apoptosis (Fig. 3.2).

In the *intrinsic pathway* the death receptors, activated by their natural death ligands, form a death-inducing signaling complex that activates the *initiator* procaspases-8 by autocatalytic cleavage (Fig. 3.2). In some type of cells, the signal coming from the activated death receptor does not generate a caspase signaling cascade

strong enough for execution of cell death. In this case, the signal needs to be amplified via mitochondria-dependent apoptotic pathways. The link between the caspase signaling cascade and the mitochondria is provided by BID, a Bcl-2 family member. BID is cleaved by caspase-8 and in its truncated form (tBID) translocates to the mitochondria where it induces the release of cytochrome c and other mitochondrial pro-apoptotic factors into the cytosol (Fig. 3.2).[7]
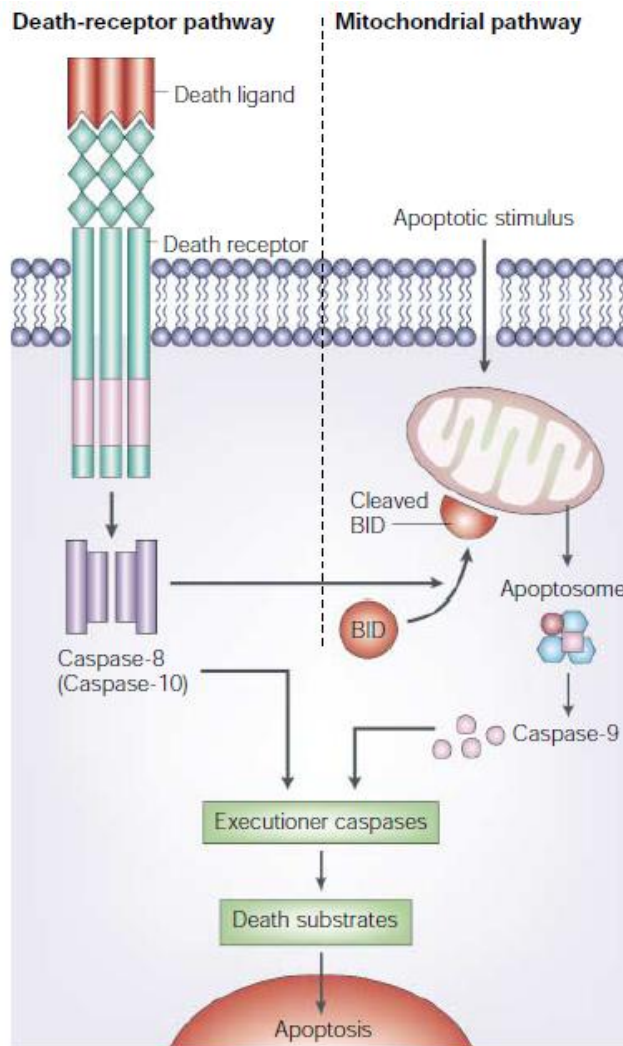


**Figure 3.2.** The two main apoptotic signaling pathways.

In the *extrinsic pathway*, death at the mitochondrial level is initiated by stimulation of the mitochondrial membrane, and proceeds via release of cytochrome c and other apoptogenic factors from the intermembranous space of mitochondria (Fig. 3.2). In the cytosol, cytochrome c forms a complex with Apoptotic Protease Activating Factor-1 (Apaf-1), ATP and the inactive initiator caspase procaspase-9. Within this complex, known as the apoptosome, caspase-9 is activated (Fig. 3.3).[8] Activated caspase-9 subsequently initiates a caspase cascade involving downstream executioner caspases ultimately resulting in cell death (Fig. 3.2).[9]
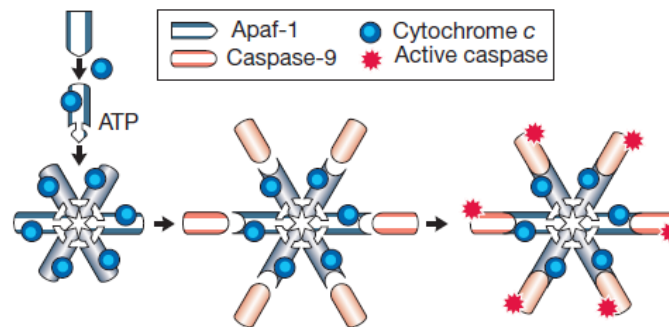


**Figure 3.3.** Mitochondrial pathway of caspase activation and apoptosome formation.

### 3.1.2. Extrinsic pathway: regulatory mechanisms in apoptosis

How cytochrome c manages to cross the mitochondrial outer membrane is not yet exactly known, but is known that the Bcl-2 family is involved in the regulation of this process. For example, addition of pro-apoptotic Bcl-2 family members to isolated mitochondria is sufficient to induce cytochrome c release.[10,11] Several hypotheses have been advanced but none of them has been proven definitively.[10,12] There are three basic models that describes this regulation process (Fig. 3.4).

1) Pro-apoptotic Bcl-2 members form channels that facilitate protein transport. Based on the structural similarity of Bcl-$x_L$ to the pore-forming subunit of bacterial toxin, it has been suggested that Bcl-2 proteins might form channels into the outer mitochondrial membrane by inserting and following conformational changes[13] Bcl-2 family members indeed can insert into synthetic lipid bilayers, oligomerize, and form channels with discrete conductance.[14]

2) Pro-apoptotic Bcl-2 family members interact with other mitochondrion proteins (as the Voltage Dependent Anion Channel (VDAC)) either to generate a pore or to modulate mitochondrial homeostasis.[15]

3) Pro- and anti-apoptitic Bcl-2 family members heterodimerize to form channels.
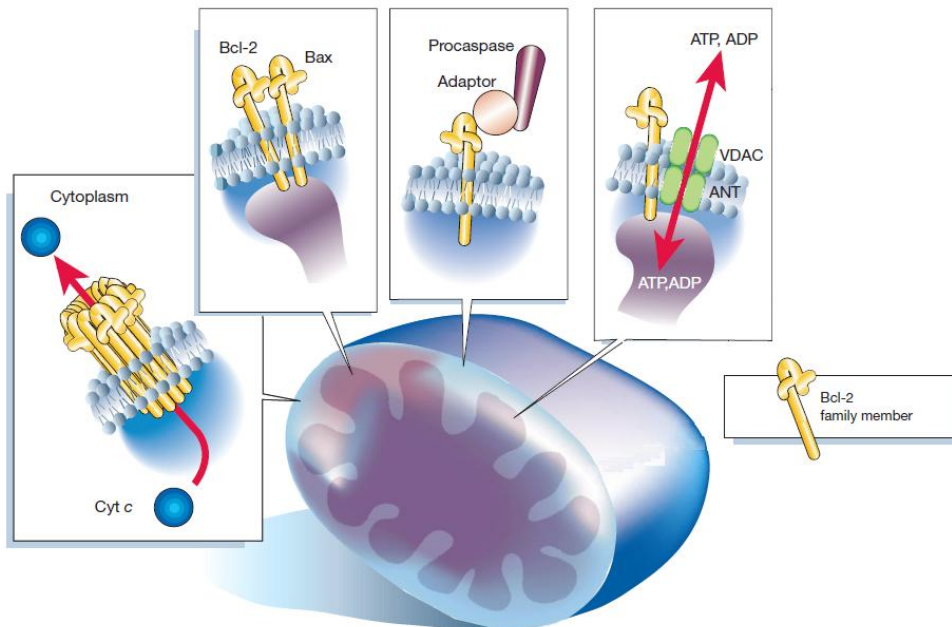


**Figure 3.4.** Possible mechanisms of action of Bcl-2 family members.

How can this processes occur? Central to this question are Bax and Bak. Bax is a cytosolic monomer that during apoptosis changes its conformation, integrates into the outer mitochondrial membrane and oligomerizes, forming channels on the outer mitochondrial membrane.[16]

In contrast, anti-apoptotic Bcl-2 members sequester pro-apoptotic Bcl-2 proteins by binding to their BH3 domains and thereby ultimately prevent Bax or Bak activation/ oligomerization and consequently inhibit mitochondrial pro-apoptotic events. The balance between the expression of anti- and pro-apoptotic Bcl-2 family members determines the cell death or survive.

### 3.2. Bcl-x$_L$

Human Bcl-x$_L$ is a member of the Bcl-2 family proteins. The founder of this family is the *BCL-2* proto-oncogene found in the human cells affected from B type Leukemia called B Cell Leukemia-2. From these cells was isolated the first gene codifying for an apoptotic protein.[17] The Bcl-2 family plays a central role in the regulation of apoptotic cell death. At least 15 Bcl-2 family members have been identified in mammalian cells and several others in viruses.[18] They can be divided in pro-apoptotic and anti-apoptotic proteins according to their function. All members show α-helical secondary structure and present high structural homology; they possess up to four conserved motifs BH1, BH2, BH3 and BH4 known as Bcl-2 homology (BH) domains. Between the anti-apoptotic members two main groups can be distinguished: one contains at least BH1 and BH2 domains (A1, Mcl-1), the other one has all four BH domains (Bcl-2, Bcl-x$_L$, Bcl-w). Similarly, the pro-apoptotic proteins can be divided in two subfamilies: Bax, Bak and Bok, which contain BH1, BH2 and BH3 resembling Bcl-2 family closely, while the seven other known mammalian killers (Bik, Blk, Hrk, BNIP3, Bim$_L$, Bad, Bid) possess only the central short BH3 domain (Fig. 3.5) and they are unrelated to any other known proteins.
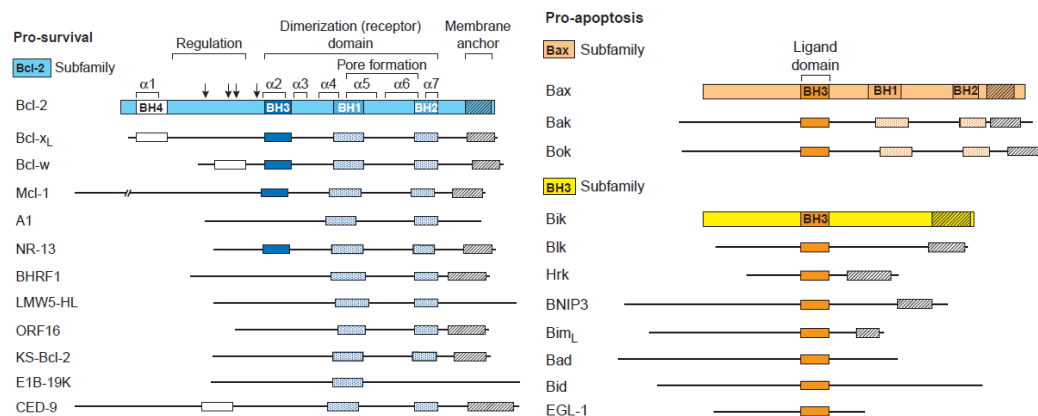


**Figure 3.5.** The Bcl-2 family: structural homology schematization.

Deletion and mutagenesis studies argue that the amphipathic α-helical BH3 domain serves as critical death domain. These BH3 domain proteins may well represent the physiological antagonist of the anti-apoptotic proteins. These proteins are able to interact forming homo- and hetero-dimeric complexes. The ratio of pro- and anti-apoptotic proteins modulate the sensitivity of a cell to apoptosis.

During the last few years the interest with respect to Bcl-x$_L$ has incredibly grown because it plays a central role in the studies focused on the apoptosis mechanism comprehension.

### 3.2.1. Structure

The 3D structure of human Bcl-x$_L$ was determined by both X-ray crystallography and NMR in 1996.[13] The structure consists in two central hydrophobic α-helices (α5 e α6) which are about 30 Å in length, flanked on one side by α3 and α4, and on the other by α1, α2 and α7 (Fig. 3.6).
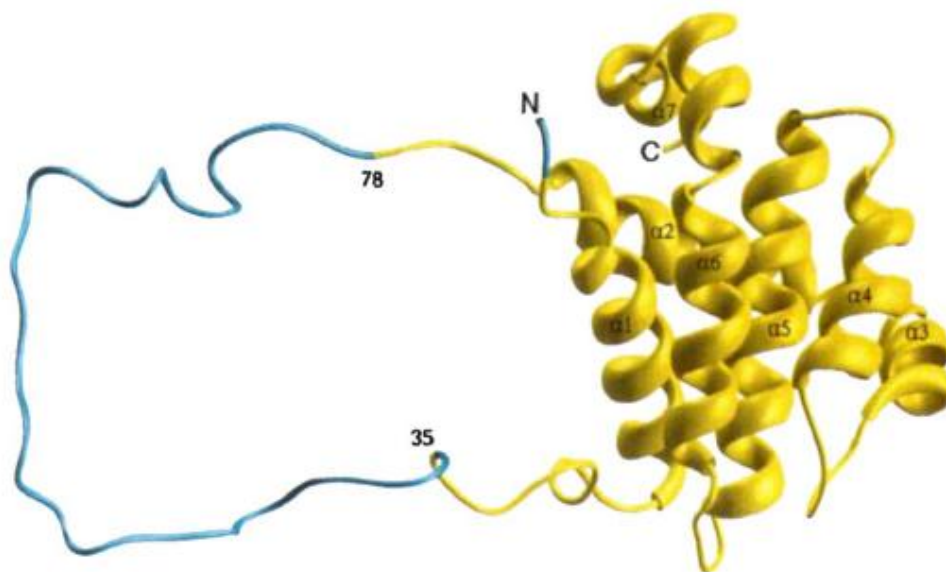


**Figure 3.6.** Three dimensional structure of human Bcl-x$_L$ (PDB code 1LXL).

The two central helices containing predominantly hydrophobic residues, are arranged in an antiparallel fashion, with a crossing angle 23°. Helix α6 contains Pro180 responsible of the helix direction. The C-terminal end contains hydrophobic residues that makes the protein able to anchor to the mitochondrial membrane. The helix α7 is connected to α6 by an irregular *turn* composed of two glycines (Gly186 e Gly187), which are largely conserves in the Bcl-2 family. The N-terminal helix forms hydrophobic interactions with α2, α5 and α6 and seems to be important for stabilizing the protein structure. A flexible *loop* of 52 residues (Ser28-Val80), connects helices α1 and α2. It's a low homology structure region in the Bcl-2 family and was found to be not essential for the anti-apoptotic activity.[19] The three functionally important Bcl-2 homology regions (BH1, BH2 e BH3) are in close spatial proximity and form an elongated hydrophobic cleft that may represent the binding site for other Bcl-2 family members. It has been demonstrated that the first two domains are essential for the anti-apoptotic activity of Bcl-$x_L$; mutations within these regions abrogate the pro-survival activity blocking the heterodimerization with other members of the Bcl-2 family (for example Bad e Bak) that promote the programmed death.[20] While BH3 has been found to be essential for the killer activity of the death-promoting proteins.[21] Here reported the aminoacidic sequence of human Bcl-$x_L$ where the BH1, BH2 and BH3 domains are colored respectively in yellow, red and green (Fig. 3.7).
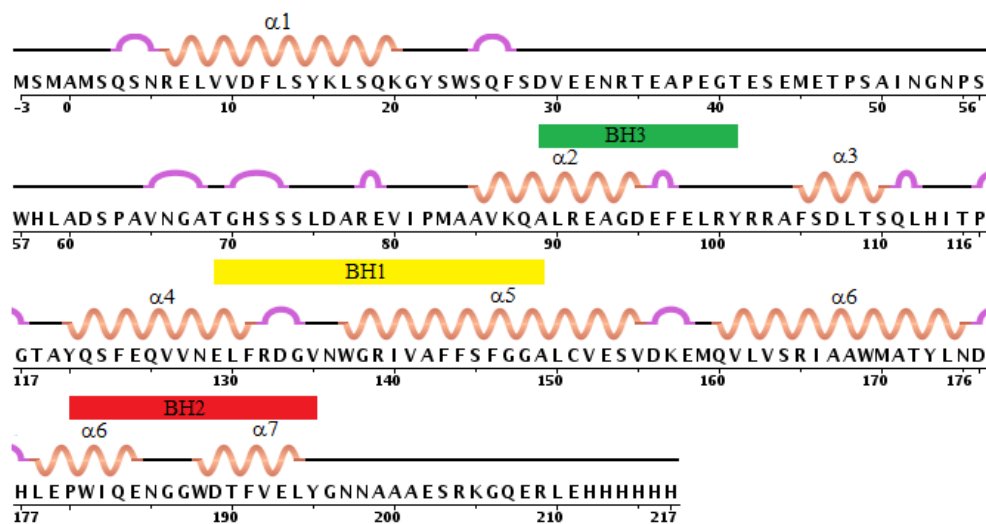


**Figure 3.7.** Aminoacidic sequence of the human Bcl-$x_L$.

A three dimensional structure is here reported to show that these domains are in close spatial proximity (Fig. 3.8) and form an elongated hydrophobic cleft in the protein able to interact with pro-apoptotic proteins as Bak as reported in the following paragraph.
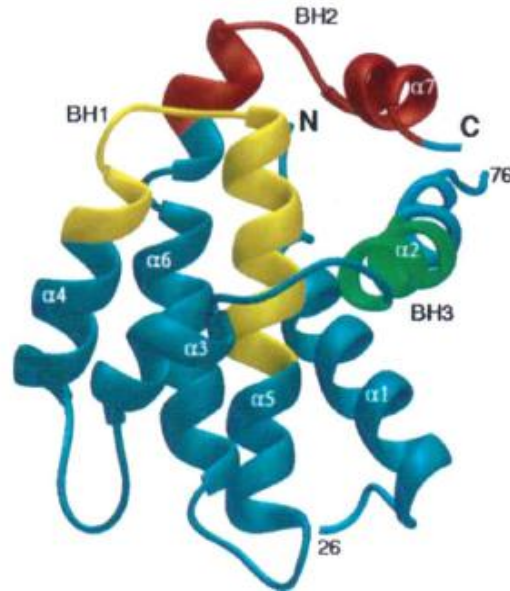


**Figure 3.8.** BH1, BH2 and BH3 domains are displaced forming an hydrophobic pocket.

Indeed, mutations of a highly conserved glycine (Gly138), part of this hydrophobic region, inhibit the death-repressor activity of Bcl-$x_L$ and block binding to Bax.[22] This could be due to amino acids with bulkier side chain that can potentially block the access to this pocket.

### 3.2.3. Function

Bcl-$x_L$ occupies a central role in the cellular programmed death regulation with an anti-apoptotic function, it is over expressed in tumors, it contributes to the cancerogenic cells growth and it is responsible of the cancer resistance to the therapy.[23]

Moreover, Bcl-$x_L$ resembles the membrane insertion domain found in bacterial toxin[24] where these membrane insertions domains contain two central helices consisting of polar residues, surrounded by five amphypathic helices able to form pores in the

membranes where they localize.[25] By analogy Bcl-$x_L$, as all the Bcl-2 family members, should be able to form pores in the membranes where it localize.

It has been shown that Bcl-$x_L$ is able to form functional ion channels when added to synthetic membranes *in vitro*.[26] However, the properties of these channels vary considerably among the family members. Bcl-$x_L$ may form small conductance channels in mitochondrial membranes that counterbalance the effect of Bax by maintaining electrical homeostasis and membrane integrity or, alternatively, their presence may prevent channel formation by Bax.[27]

### 3.2.3. Interaction: Bcl-$x_L$-Bak

As discussed above Bcl-$x_L$ inhibits programmed cell death, while Bak promotes apoptosis. Interactions between these two proteins modulate the sensitivity of a cell to apoptosis (3, 4). Bcl-$x_L$ and Bak are anchored to the mitochondrial membrane as shown in Figure 3.9.
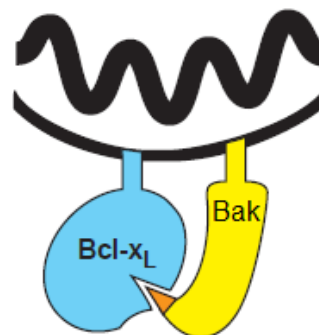


**Figure 3.9.** Model for Bcl-$x_L$-Bak interaction.

To understand how Bcl-x$_L$ interacts with Bak inhibiting its ability to promote cell death, the solution NMR structure of Bcl-x$_L$ complexed with a 16-residue peptide derived from the BH3 region of Bak is reported (residues 72-87).[28] The Bak peptide binds in a hydrophobic cleft formed by the BH1, BH2, and BH3 regions of Bcl-x$_L$ (Fig. 3.10).
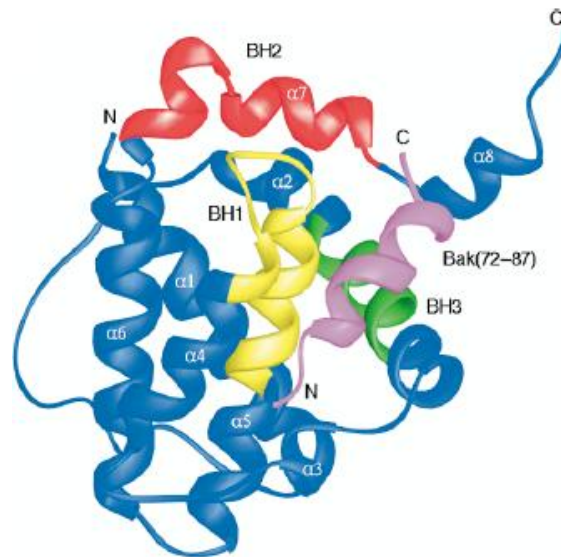


**Figure 3.10.** Ribbon representation of the 3D structure for the complex. The BH1, BH2, and BH3 regions of Bcl-x$_L$ are shown in yellow, red, and green, respectively. The Bak peptide is shown in magenta.

The NH$_2$-terminal residues of the peptide interacts with the BH1 region of Bcl-x$_L$ (Val126, Glu129, Leu130, and Phe146), whereas the C-terminal portion of the Bak peptide interacts predominantly with residues in the BH2 and BH3 regions (Phe97, Arg100, Tyr101, and Phe105). The hydrophobic side chains of the peptide (Val74, Leu78, Ile81, and Ile85) point into a hydrophobic cleft of Bcl-x$_L$ (Fig. 3.11) and stabilize complex formation. In addition to these hydrophobic interactions, the charged side chains of the Bak peptide (Arg76, Asp83, and Asp84) are close to oppositely charged residues of Bcl-x$_L$ (Glu129, Arg139, and Arg100, respectively) (Fig. 3.11).
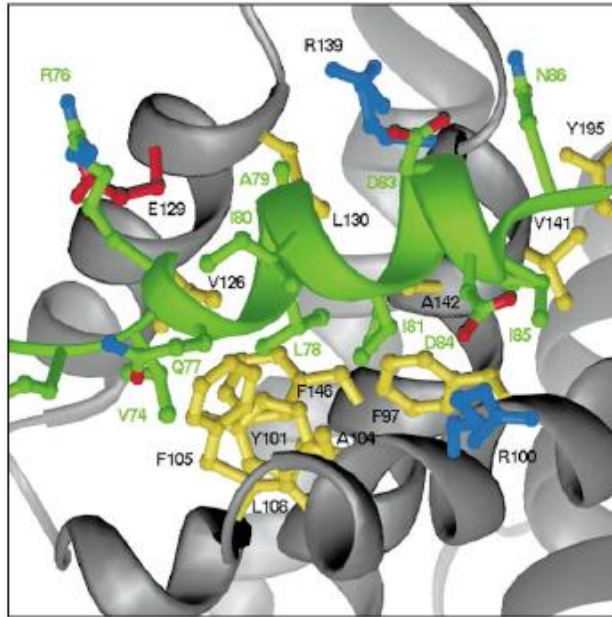
**Figure 3.11.** Binding pocket representation of Bcl-x$_L$ bound to the Bak peptide. Protein and peptide side chains are in yellow and green respectively.

In summary the molecular interactions that stabilize the Bcl-x$_L$–Bak peptide complex likely reflect the important interactions that occur between the full-length proteins. Only the BH3 region of Bcl-x$_L$ is critical for the ability to inhibit pro-apoptotic generation,[21,29] while several regions of the death-inhibiting proteins participate in their antiapoptotic activity and heterodimerization.

### 3.3. Cytochrome c

Cytochromes are among the most studied and popular protein among biochemists and biophysicists.[30] This is due to their red color, which makes protein purification easier, their small size and their high solubility. Moreover, their relatively high thermodynamic stability together with high helical content and the presence of the heme cofactor have allowed cytochromes to be studied through a variety of spectroscopic techniques.

In the early 1920s, David Keilin discovered cytochromes becoming intrigues by the changing distribution of color in tissue of the larval form of a fly that lives in the stomach of the horse. The larva carries a considerable amount of a hemoglobin like pigments that confer it a red color. During metamorphosis the red color disappears from

the young fly reappearing in the flight muscles of the adult. The absorption spectrum of the flight muscles gives four bands: three bands at 604, 566 and 550 nm already known from the hemocromes spectra, and the fourth at 520 nm. The proteins responsible of this absorption spectra were termed *cytochrome* for cellular pigment and were shown to undergo oxidation and reduction followed by a change in the protein color. On the basis of the different cytochromes absorption spectra three type of cytochromes were identified as a, b and c type. I will take into account the cytochrome c type. Among those, four types of cytochromes c were classified by sequence, three dimensional structures, number and type of hemes, position of the axial iron ligands, and redox potential.[31] In the present work I'll focus the attention on the mithocondrial cytochrome c that belongs to the type 1.

Type 1 cytochrome c are usually small proteins (8-20 kDa), generally water-soluble with a single heme, although multi-heme proteins are also included in this family. In the type 1 cytochrome c the prosthetic group is constituted by a type c heme defined by the presence of a single Cys-Xaa-Xaa-Cys-His (CXXCH) signature for heme attachment.

Heme is a rigid almost planar molecule comprising four pyrrole groups linked by methene bridges to form a tetrapyrrole ring (Fig. 3.12).
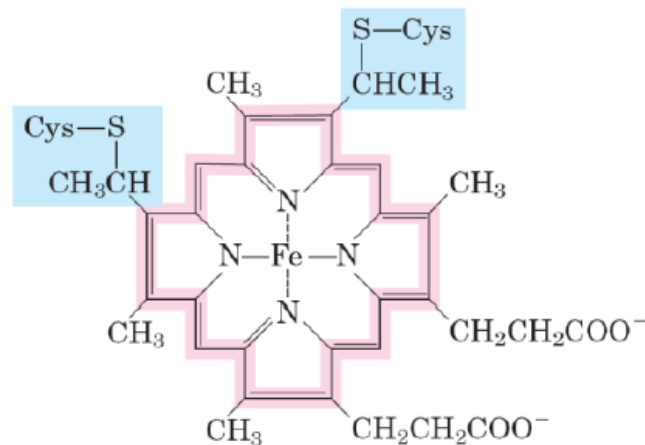


**Figure 3.12.** Structure of heme c of c-type cytochromes.

The cytochrome c is covalently bound to the polypeptide backbone of the molecule trough thioether linkages at the vinyl side chains to cysteins 14 and 17 in the polypeptide chain. The prosthetic group contains one iron ion esa-coordinated by four nitrogen of the heme tetrapyrrole ring and with two axial protein residues. For mitochondrial cytochrome c this ligands are His18 and Met80 (Fig. 3.13). The ion iron is able to accept and release electrons cycling between the oxidized and reduced state, thus conferring to the protein an electron transport activity. A series of observations highlighted chemical and physiochemical differences between the oxidized and reduced states of cytochrome c suggesting that the structure of the ferrocytochrome c was stronger than that of ferricytochrome c. This was attributed to an oxidation-state dependent dynamics change that precisely regulates the thermodynamic and the kinetic of the redox reaction.
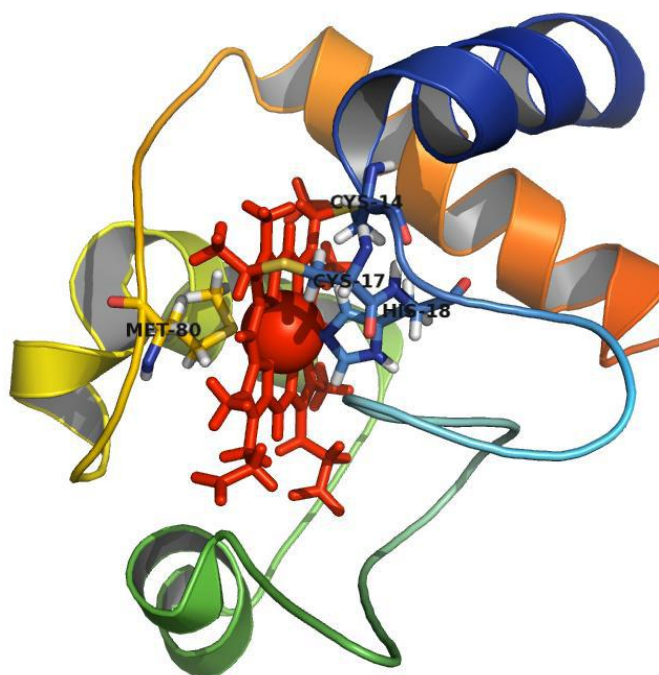


**Figure 3.13.** Cytochrome c covalently bound to the heme prosthetic group, shown as red sticks, through cysteins 14 and 17. The iron atom, shown as a red ball, is axially coordinated by Met80 and His18.

*3.3.1. Structure of mithocondrial cytochrome c*

The first three-dimensional structure of cytochrome c has been solved in the 1962. Several additional structures appeared during the later years as the one of mitochondrial cytochrome c in the 1970s.[32] Three structural elements are found in all cytochrome c experimental structures: the N- and C-terminal α-helices (respectively, helix α1 and α5 in mitochondrial cytochromes), as well as the long helix (helix α3, also called the 60's helix in mitochondrial cytochromes) preceding the short helix and the loop containing the second axial ligand to the heme iron, which is always a methionine[33-35]. The structure is stabilized by hydrogen bonds formed between heme-propionate and Arg38, Tyr48 and Trp59 and by other propionate and Ser49 Asp50 and Thr78. A number of conserved basic residues have been shown to be important to function, namely Lys8, 13, 25, 27, 72, 73, 79, 86 and 87. The nine basic residues surround the heme edge where electron transfer usually takes place.

*3.3.2. Function of mithocondrial cytochrome c*

Cytochrome c is loosely associated with the inner membrane of the mithocondrion and is an essential component of the electron transport chain. Cytochrome c has several functions: it is essentially an electron transfer protein, mainly involved in aerobic as well as anaerobic respiration, in mammalian cells, it is involved in apoptosis[36] (see paragraph 3.1 for details) and it can catalyze several reactions as hydroxilation, aromatic oxidation and shows peroxidase activity, in particular, it is not uncommon that cytochrome c is fused to redox enzymes and constitutes an entry/exit point for electrons in the catalytic cycle of the enzyme.[37]

3.3.2.1. Cytochrome c: an electron transport protein

The electron transport chain is the most important process to generate energy in mitochondria and many prokaryotes. Energy is derived from the transfer of electrons through a series of membrane-bound enzyme complexes and conserved as a transmembrane proton gradient, which is eventually used for the ATP synthesis. In

particular electrons move from an electron donor (NADH or QH$_2$) to a terminal electron acceptor (O$_2$) via a series of redox reactions. These reactions are coupled to the creation of a proton gradient across the mitochondrial inner membrane mediated by proton pumps. The resulting transmembrane proton gradient is used to synthesize ATP trough the ATP synthase.

The electron transport chain is constituted by four complexes embedded in the inner mitochondrial membrane (Fig.3.14).
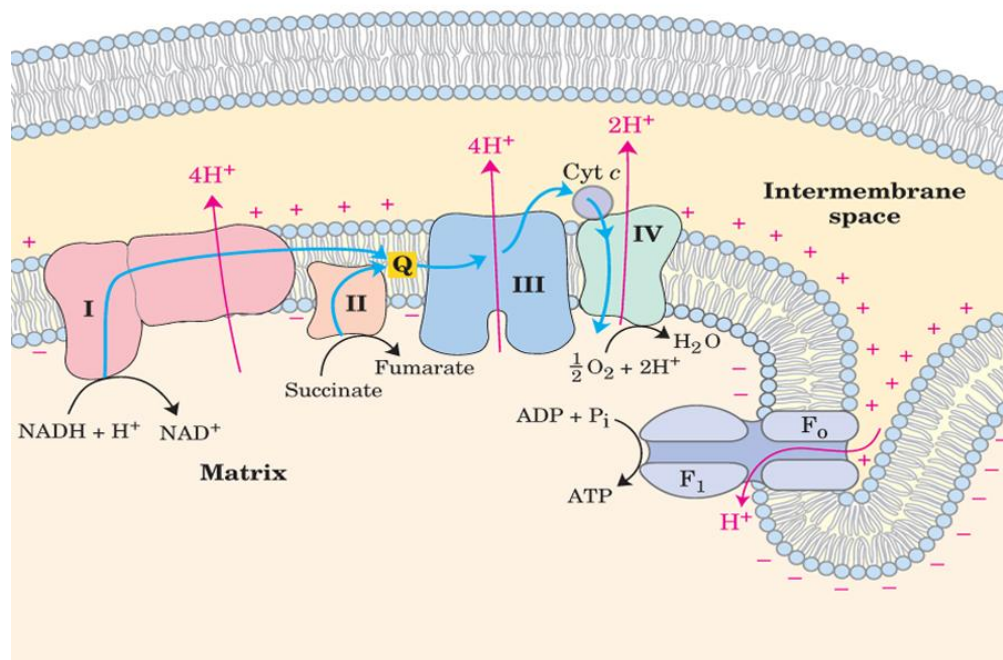


**Figure 3.14.** Schematic picture of the enzymes of the mitochondrial inner membrane involved in the terminal steps of aerobic respiration.

In complex I, called NADH Ubiquinone Oxidoreductase, two electrons are removed from NADH and transferred to a lipid-soluble carrier, Ubiquinone (Q) and four protons (H$^+$) are pumped across the membrane.

In Complex II, called Succinate Dehydrogenase, additional electrons originating from succinate are delivered into the quinone pool.

In Complex III, called cytochrome bc1 complex, two electrons are removed from ubiquinol (QH$_2$) and transferred to two molecules of cytochrome c located within the intermembrane space. At the same time, four protons are translocated across the membrane, contributing to the proton gradient.

$$QH_2 + 2\text{cytochrome c } (Fe^{3+}) + H^+_{IN} \rightarrow Q + 2\text{cytochrome c } (Fe^{2+}) + 4H^+_{OUT}$$
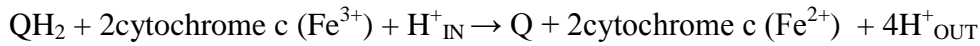
In Complex IV, called cytochrome c oxidase (CCO) four electrons are removed from four molecules of cytochrome c and transferred to molecular oxygen $O_2$, producing two molecules of water. At the same time, four protons are translocated across the membrane, contributing to the proton gradient.

$$4\text{cytochrome c } (Fe^{2+}) + O_2 + 8H^+_{IN} \rightarrow 4\text{cytochrome c } (Fe^{3+}) + 2H_2O + 4H^+_{OUT}$$
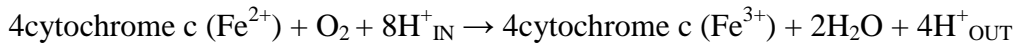
The electron transfer coupled to the generation of a electrochemical gradient across the membrane drives ATP synthesis by the ATP synthase complex.[38]

### 3.3.3. Interactions

Efficient electron transfer between cytochrome c and its counterparts in the respiratory chain requires rapid adduct formation and rapid product dissociation as well as the achievement of proper orientation of the partner proteins in the transient adduct to optimize the electron transfer rate. Such requirements are reflected in the nature of key interactions involving surface residues surrounding the heme crevice of cytochrome c. The optimal interfacial arrangement is tuned by hydrophobic interactions among short range contacts. The transient nature of the complexs is assured by the possibility to switch on and off the potential electrostatic interactions among residues of different sign surrounding the contact central region on the two proteins. Long-range recognition of the partners is driven by non-specific electrostatic interactions that rely on the presence of large patches of different sign on the two protein surfaces.

- **cytochrome c and cytochrome bc1**

    In the crystal structure of the complex between cytchrome c and cytochrome bc1 from Saccharomyces cerevisiae (Fig. 3.15A) the interaction with the soluble domain cytochrome c1 of the enzyme is mainly mediated by non polar contacts involving residues Thr12, Arg13, Val28 and Ala81 on cytochrome c from Saccharomyces cerevisiae. Weak

polar interaction involving Lys79 and Lys86 are present, while additional electrostatic interactions (involving for example cytochrome c Lys87) have been proposed to modulate intermediate states and the unbinding step (pdb code 1KYO).[39] The soluble domain of cytochrome c1 is negatively charged at neutral pH, thus complementary to the positively charged soluble cytochrome c. However, the structure of the complex suggests that despite the high electrostatic charge present on both proteins and the good complementary character in this respect between the two partners, the most stable configuration of the adduct is predominantly determined by non polar contacts.

- **cytochrome c and CCO**

Cytochrome c interacts with subunit II of CCO, which is composed of two transmembrane helices and a β-barrel globular domain facing the intermembrane or periplasmic space, and contains the catalytic group composed by two electronically coupled copper ions so called $Cu_A$ site. A structural model of the adduct between cytochrome c-552 and the $Cu_A$ subunit of CCO from Paracoccus denitrificans has been proposed (Fig. 3.15B) (pdb code 1ZYY).[40]  In the various conformers of the complex common hydrpophobic patches are found, while different networks of electrostatic intermolecular interactions are established within negatively charged residues of CCO and the positively charged of cytochrome c-522. In particular, the hydrophobic interactions involve residues Ala16, Val26, Ala79, Phe80, and Ala81 on  the cytochrome c-522, while the electrostatic interactions involve Asp and Glu contiguous to the CCO's central hydrophobic surface area and Lys residues (namely Lys13, Lys15, Lys19, Lys70, Lys74 and Lys77) surrounding the heme crevice on the cytochrome c-522.[41]

- **cytochrome c and cytochrome c peroxidase (CCP)**

A three-dimensional structure of the adduct between yeast-iso-1-cytochrome c and CCP is available (pdb code 2pcb).[42] The region of cytochrome c in direct contact with CCP is relatively similar to that involved in the interaction with cytochrome c1 subunit of the cytochrome bc1, described above. Also in the CCP-cytochrome c complex, hydrophobic interactions are the predominant forces holding the complex together (Fig. 3.15C). On the side of cytochrome c they involve residues Leu9, Arg13, Gln16, Cys17, Ala81, Phe82, Gly83 and Lys86. The side chains of Asn70, Lys73 and Lys87 are potentially involved in hydrogen bonds and/or salt bridges.
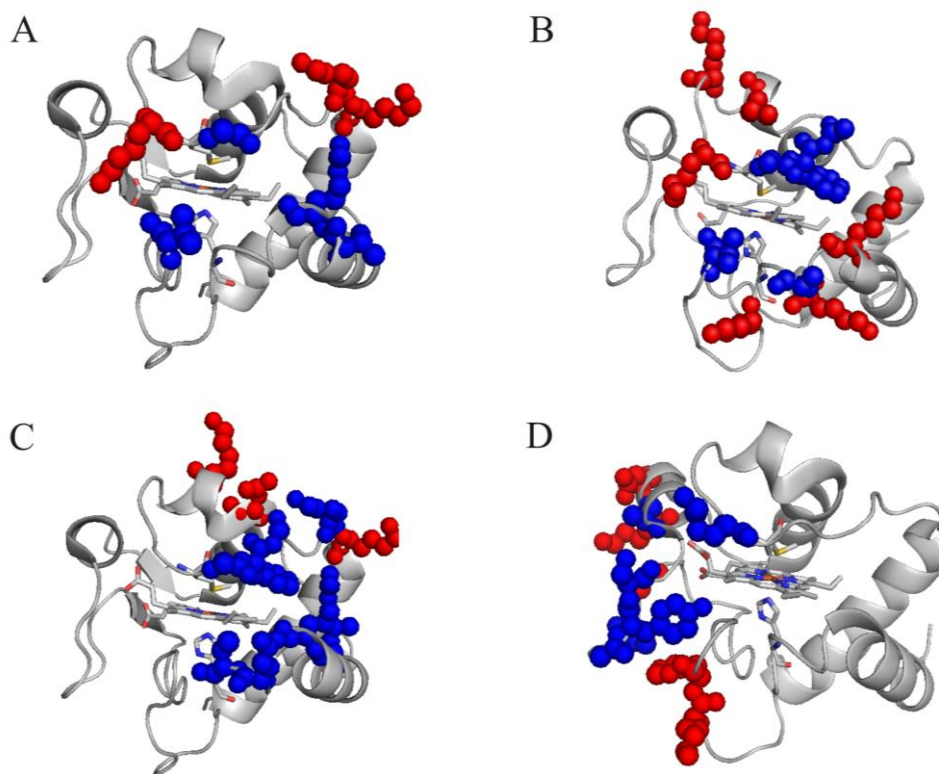


**Figure 3.15.** (A) Residues in S. cerevisiae cytochrome c involved in the interaction with cytochrome bc1 involved in hydrophobic contacts are represented as blue spheres, those giving rise to electrostatic interactions are red spheres, whereas Lys87, proposed as potentially involved in salt bridges of transient states, is shown in cyan; (B) Residues in cytochrome c522 involved in hydrophobic contacts with CCO are shown as blue sphere, while the surrounding Lys potentially involved in alternating

electrostatic interactions are red spheres; (C) Residues of interest for the interaction between yeast iso-1-cytochrome c and cytochrome c peroxidase; (D) Residues in human cytochrome c involved in hydrophobic contacts with Bcl-x$_L$ are shown as blue spheres, while the amino acids involved in electrostatic/H-bond interactions are represented as red spheres; the colour coding is the same as for panel (A). In the four panels, cytochrome c is represented with an orientation where the "loop face" points towards the observer.

In all the adducts presented above both electrostatic and hydrophobic interactions play an important role in the complex formation. The initial recognition it's supposed to be driven by the shape and the potential of the negative cluster, and would work long-range, whereas the hydrophobic interaction acting short-range would fine-adjust the redox partners for optimal electron transfer.

Moreover, cytochrome c exposes to all its partners a common area that coincides with the solved exposed porphyrin ring. This region is the most suitable contact surface where an electron can be transferred to an acceptor.

No structural information exist on the interactions between cytocrome c and proteins involved in the apoptosis process.

### 3.4. Aim of the project

Apoptosis normally eliminates cells with damaged DNA or aberrant cell cycle. Pro-survival proteins are therefore potentially oncogenic. Clarifying how the Bcl-2 family governs apoptosis might provide the ability to control the apoptotic threshold.

Conventional cytotoxic therapy indirectly induces apoptosis, but more effective outcomes should be achieved by direct activation of the apoptotic machinery. Promising approaches include impairing expression of pro-survival proteins or identifying drugs that inhibits their action. The identification of interfaces between partner molecules provides targets for pharmacological intervention: the protein-protein interaction surface between Bcl-x$_L$ and cytochrome c here identified may offer one of these targets.

## 3.5. Methods

Nuclear magnetic resonance (NMR) spectroscopy is unique among the methods available for three-dimensional structure determination of proteins at atomic resolution. Since proteins perform their physiological functions biological fluids, knowledge of the molecular structures in solution is highly relevant. In the NMR experiments, solution conditions such as the temperature, pH and salt concentration can be adjusted close to physiological conditions.

In addition to protein structure determination, NMR applications include investigations of dynamic features, as well as structural studies of interactions between proteins supported also by molecular docking programs, such as HADDOCK (High Ambiguity Driven protein-protein DOCKing).[52]

### 3.5.1. Resonance assignment: triple resonance expermients

In the sequence-specific resonance assignment the use of multidimensional triple resonance NMR experiments for the protein resonance assignment has become standard.[43,44] For proteins with less than 200 amino acids, triple resonance NMR data analysis is, in many cases, a routine task. The resonance assignment can be divided into two steps, i.e., backbone sequential assignment and side chain assignment. I will focus the attention on the former.

Based on the primary sequence, the sequential assignment is usually achieved by comparing multidimensional NMR spectra providing the necessary correlations to identify spin systems within each amino acid (intra-residue) and then to link them in a sequence-specific manner (inter-residue).

The information about intra- and inter-residual correlation is obtained using the triple resonance experiments uniformly $^{13}$C and $^{15}$N labelled proteins, based on well-defined heteronuclear scalar couplings along the polypeptide chain (Fig.3.16).[45-48]
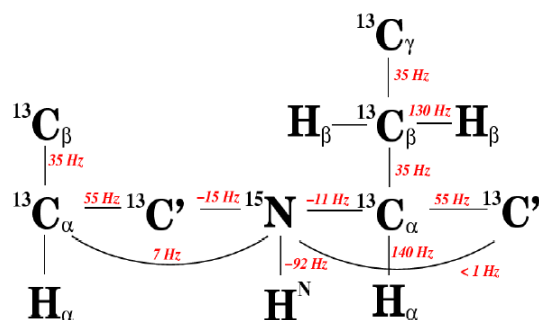
**Figure 3.16.** Spin system of the peptide backbone and the size of the $^1$J and $^2$J coupling constants that are used for magnetization transfer in $^{13}$C-, $^{15}$N-labelled proteins.

The most popular sequential assignment strategy relies on a set of experiments HNCA and HN(CO)CA, (Fig. 3.17). The names of the experiments indicates all nuclear spins that are involved in the coherence transfer pathway exploiting scalar couplings and frequency labelling in the various dimensions of the experiments.
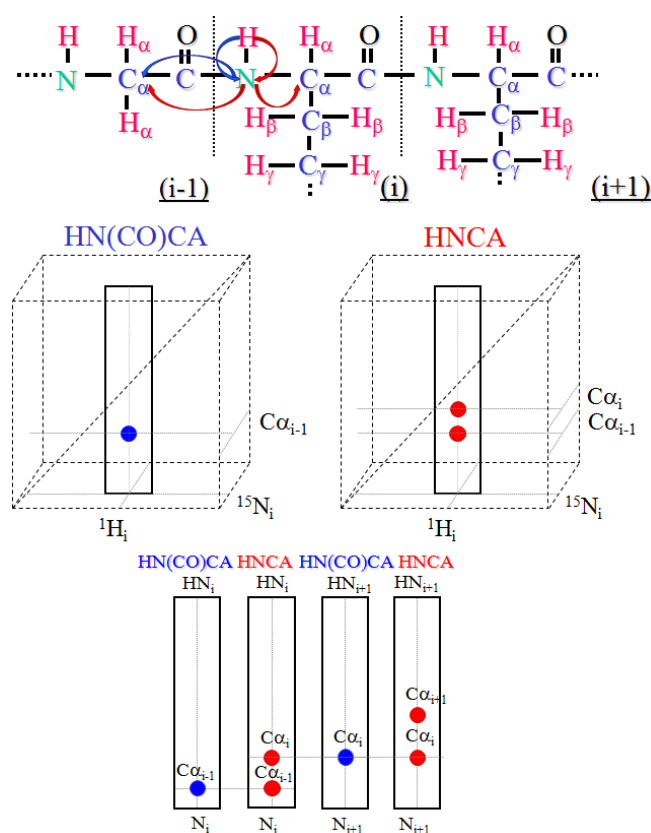


**Figure 3.17.** Schematic representation of the 3D heteronuclear NMR spectra HNCA and HN(CO)CA.

The information about amino acid type is obtained from the combination of the Cα chemical shifts, which provide a  first indicator of the amino acid type and facilitates the sequential assignment.[49] The HNCO 3D experiment can then be used to complete backbone assignment by the identification of the C' chemical shift such as HNCO and HN(CO)CA.

The most important double- and triple-resonance experiments used for sequential and side-chain resonance assignment and the key correlations that can be detected in each experiment are summarized in Table 3.1.

| Experiment | Correlations |
|---|---|
| CBCANH | $C_i^\beta / C_i^\alpha - N_i - HN_i$ |
| | $C_{i-1}^\beta / C_{i-1}^\alpha - N_i - HN_i$ |
| CBCA(CO)NH | $C_{i-1}^\beta / C_{i-1}^\alpha - N_i - HN_i$ |
| HNCA | $C_i^\alpha - N_i - HN_i$ |
| | $C_{i-1}^\alpha - N_i - HN_i$ |
| HNCO | $C_{i-1}' - N_i - HN_i$ |
| HN(CO)CA | $C_{i-1}^\alpha - N_i - HN_i$ |
| HN(CA)CO | $C_i' - N_i - HN_i$ |
| | $C_{i-1}' - N_i - HN_i$ |

**Table 3.1:** Correlations observed in the double and triple resonance experiments used for sequential and side chain assignment ($^1$H approach).

### 3.5.2. Molecular Docking for the study of protein-protein interaction

The traditional approach to solve macromolecular complexes by NMR spectroscopy requires the collection of proton-proton nOes (nuclear Overhauser effects), which can be translated into distances. When the structures of the separate biomolecules are already known, a structure of the complex can be build up by selectively analyzing the intermolecular NOEs. In addition to these intermolecular NOEs, also orientational restraints like residual dipolar couplings (RDCs), heteronuclear relaxation rates, or

paramagnetic pseudo-contact shifts can be used, that precisely define the orientation of the two macromolecules in the complex. These methods require, however, extensive series of experiments and detailed analysis.

Another way of studying complex formation by NMR is chemical shift perturbation mapping, where the chemical shifts changes of given atoms are monitored in the course of the formation of the complex (see Section 2.4.3.2). The information obtained by CSP analysis, however, is only qualitative; it does not allow to unambiguously identify pairs of atoms that are in contact in the complex and it does not provide precise distance or orientation information, and has therefore not been commonly used in structure calculations.

For this reason, chemical shift perturbation are not directly used for structure calculations, but are combined with computational methods that predicts at a structural level the assembly of two separate biomolecules into their biologically relevant complex. HADDOCK[52] is one example of computational methods that make use of biochemical and/or biophysical interaction data resulting from NMR titration experiments or mutagenesis data, introduced as Ambiguous Interaction Restraints (AIRs), to drive the docking process. After calculation, the structures are ranked according to their intermolecular energy, that is, sum of electrostatic, van der Waals, and AIR energy terms.

HADDOCK protocol defines active and passive residues. The active residues are all residues showing a significant chemical shift perturbation after the formation of the complex, with a solvent accessibility higher than 50% in the free-form protein; the passive residues are those neighboring the active ones and have a solvent accessibility usually higher than 50%.

The docking protocol consists of three consecutive steps:

- rigid body minimization driven by interaction restraints, in which 1000 structures of the complex are generated;
- the best 200 structures in terms of total intermolecular energy are further submitted to the semi-flexible simulated annealing in which side-chains and backbone atoms of the interface residues are allowed to move;
- final refinement in Cartesian space in water.

Finally the 200 structures obtained are clustered using a threshold of 7.5 Å of RMSD among the structure of the cluster.

Clusters containing at least 4 structures were considered and analyzed on the basis of HADDOCK score value (HADDOCK score: 1.0 $E_{vdw}$+0.2 $E_{elc}$+0.1 $E_{dist}$+1.0 $E_{solv}$, where the four terms are respectively: van der Walls Energy, Electrostatic Energy, Distance Restrains Energy and Desolvation Energy), RMSD and Buried Surface Area.

### 3.6. Results & Discussion

Chemical shift changes provide a highly sensitive tool for identifying the residues that play a role in PPIs as already discussed for S100B. NMR chemical shift perturbations of backbone amides in Bcl-$x_L$ and cytochrome c reveal that the two proteins form detectable amounts of an adduct. The observed chemical shift variations are small (Fig. 3.18 and Fig. 3.19) but increase in a saturable manner upon titration (Fig. 3.20).



**Figure 3.18.** Plot of the chemical shift variation (Garrett values) of the backbone cytochrome c amide signals in the cytochrome c – Bcl-$x_L$ adduct.

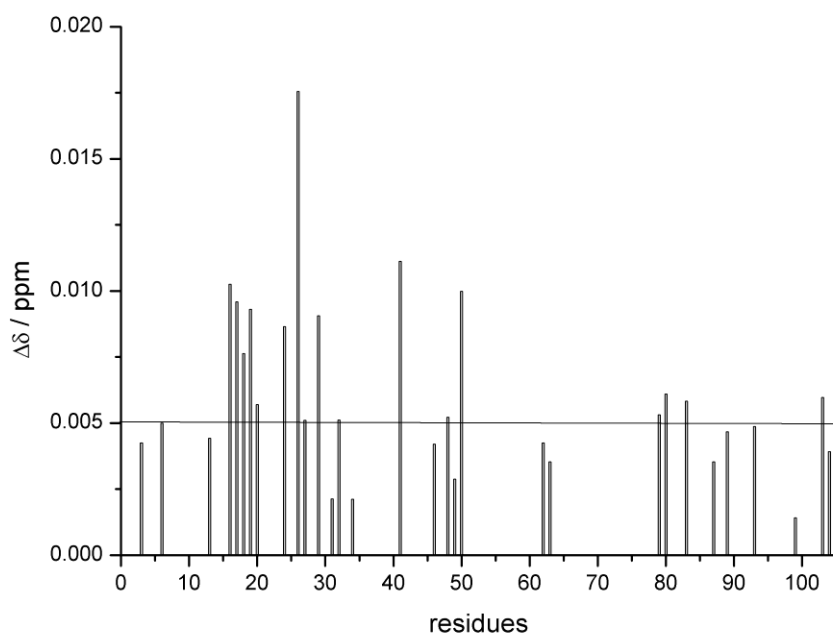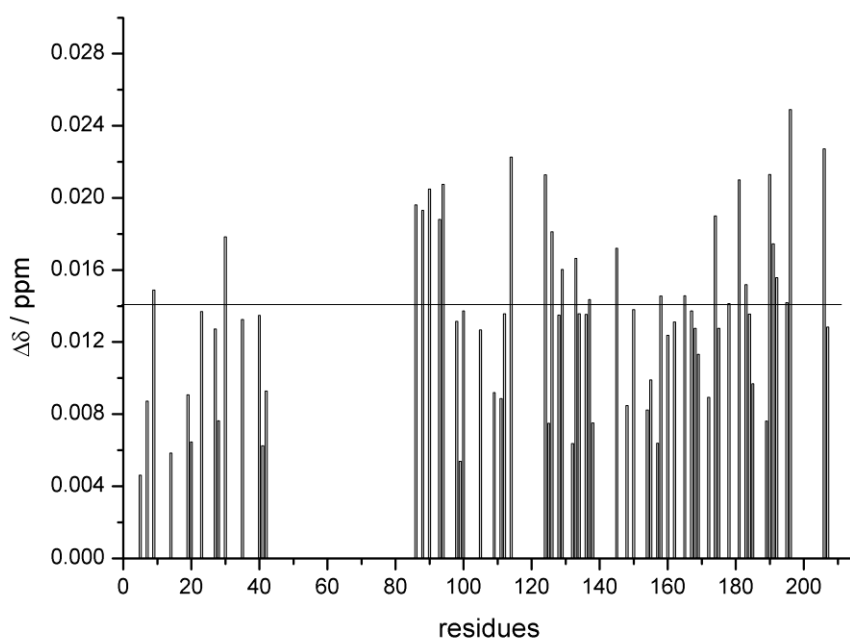**Figure 3.19.** Plot of the chemical shift variation (Garrett values) of the backbone Bcl-x$_L$ amide signals in the cytochrome c – Bcl-x$_L$ adduct.
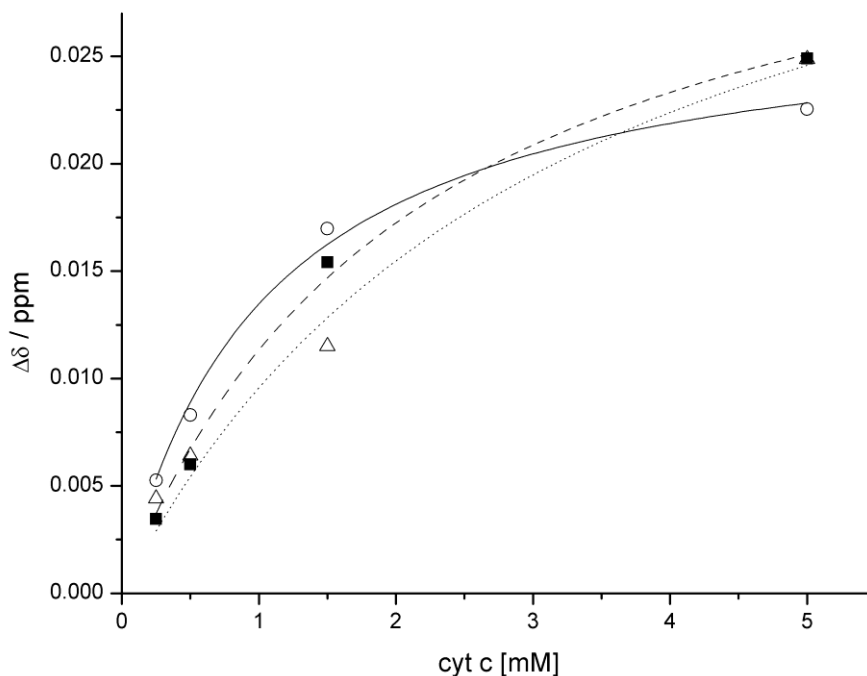


**Figure 3.20.** Fitting of the weighted average chemical shift variations of three perturbed residues (L90, G94, G200) of the $^{15}$N-enriched Bcl-x$_L$ as a function of the concentration of the unlabeled cytochrome c.

77

The interaction between cytochrome c and Bcl-x$_L$ has been reported to be strongly dependent on ionic strength:[2] in 50 mM phosphate buffer, the K$_d$ of 1.2 10$^{-7}$ M at 80 mM NaCl increases by nearly 12-fold in the presence of 600 mM NaCl. The relatively high concentrations required for the NMR experiments of these two heavily charged proteins (total charge: -14 for Bcl-xL and +9 for cytochrome c) contribute to the increase of the overall ionic strength of the solution, setting us farther from the optimal conditions for the complex formation. Consistently, the K$_d$ values estimated from our chemical shift data (Fig. 3.20), although measured at 50 mM phosphate buffer and 150 mM NaCl, are of the order of 1*10$^{-3}$ M. The low affinity of the complex combined with the intrinsic low solubility of Bcl-x$_L$ prevented us from achieving protein concentrations in solution higher than 500 μM for the anti-apoptotic protein, that would have provided larger amount of the of the bound state and therefore larger chemical shift perturbations.

An overall increase in $^{15}$N transverse relaxation rate, R$_2$, values is observed upon titration of Bcl-x$_L$ with cytochrome c, which is consistent with an increase in the overall tumbling correlation time upon complex formation.[50] An accurate measure of $^{15}$N R$_2$ in the complex, however, is hampered by the low stability of Bcl-x$_L$ caused by local sample heating associated to this type of measurements.

Residues whose chemical shift values are affected by the presence of the partner molecule, when mapped on the proteins' surface, result confined to well defined areas, suggesting the formation of a specific, albeit transient, complex. The restraints derived from the NMR experiments were used as input data for docking calculations on the human cytcochrome c - Bcl-x$_L$ system with the program HADDOCK and unequivocally define the docking face for both proteins.

The obtained ensemble of structural models is constituted by a well defined cluster (Table 3.2) of 128 conformers with BB RMSD of 0.8±0.5 Å from the overall lowest energy structure. The dominant contribution to the total interaction energy comes from the electrostatic term. This is consistent with the experimental finding that the interaction affinity is reduced by an increase in ionic strength.[2]

| N ° of structures within the cluster | van der Waals energy (kcal/mol) | Electrostatic energy (kcal/mol) | Desolvatation energy (kcal/mol) | Restrains violations | Overall RMSD (Å) | Buried surface area (Å$^2$) |
|---|---|---|---|---|---|---|
| 128 | -59.6±8.4 | -535.9±27.6 | 39.2±3.3 | 194.7±58.84 | 0.8±0.5 | 2040.1±81.5 |
| 21 | -39.5±5.7 | -371.2±43.7 | 39.2±6.8 | 141.5±48.88 | 11.5±0.1 | 1599.6±121.6 |
| 6 | -40.2±5.3 | -371.7±31.8 | 45.0±10.4 | 116.7±44.73 | 11.8±0.1 | 1458.1±98.2 |
| 11 | -43.7±6.0 | -291.4±34.4 | 40.7±10.0 | 114.2±52.48 | 8.9±0.2 | 1285.5±40.7 |

**Table 3.2.** Structural statistics calculated over all structures for the 4 clusters obtained by HADDOCK.

A buried surface area of the order of about 2,000 Å$^2$ is identified, which contrasts with the short -lived nature of the complex, for which values < 1,200 Å$^2$ would be expected. A similar situation has been already reported for the cytochrome c − Cu$_A$ adduct and explained in terms of a biased picture resulting from the docking procedure;[40] dense networks of intermolecular contacts are provided in the same structural model as if they could be contemporarily present whereas reasonably only a fraction of them is actually formed on average. Consistently with this view, the relatively large restraints violation energy hints that none of the calculated structures satisfies the experimental constraints altogether. Observed chemical shift perturbations in solution reflect the average effect of various interconverting adducts with slightly different binding contacts, as summarized in Table 3.3.

| Bcl-x$_L$ | Cyt | Contacts | Frequency |
|---|---|---|---|
| Glu96 (Oε1, Oε2) | Lys53 (Hζ1,Hζ2,Hζ3) | Hydrogen Bonds | 141 |
| Arg100 (HH12, HH21, H22) | Gly41 (CO) | Hydrogen Bonds | 64 |
| Tyr101 (HH) | Ala43 (CO) | Hydrogen Bonds | 31 |
| Tyr101 (Cε1,Cγ,Cζ) | Ala43 (Cα,Cβ) | Non Bond | 195 |
| Asp133  (Oδ2) | Lys25 (Hζ1,Hζ2,Hζ3) | Hydrogen Bonds | 87 |
| Asp133 (Oδ1, Oδ2) | His26 (HN) | Hydrogen Bonds | 36 |
| Asn136 (Hδ21, Hδ22) | His26 (CO) | Hydrogen Bonds | 44 |
| Asn136 (Cβ,Cγ) | Tyr46 (Cδ1) | Non Bond | 82 |
| Trp137 (Cβ) | Ser47 (Cβ) | Non Bond | 43 |
| Gly138 (CO) | Gly45 (CO) | Non Bond | 54 |
| Gly138 (Cα) | Tyr46 (CO,Cα) | Non Bond | 128 |
| Thr190 (Cγ2) | Lys79 (Cε) | Non Bond | 42 |
| Phe191 (Cε1) | Ser47 (Cβ) | Non Bond | 32 |
| Leu194 (CO) | Ala50 (NH) | Hydrogen bond | 47 |
| Leu194 Backbone (CO) | Ala50 (Cβ) | Non Bond | 32 |
| Tyr195 (OH) | Lys53 (Hζ1,Hζ2,Hζ3) | Hydrogen Bonds | 48 |
| Tyr 195 (Cα,Cδ1) | Ala50 (Cβ) | Non Bond | 72 |
| Ser203 (Hγ) | Asn54 (Oδ1) | Hydrogen Bonds | 45 |

**Table 3.3** Intermolecular contacts Non Bond and Hydrogen Bond calculated all the model structures of solutions obtained by HADDOCK and repetition frequency >30.

Considering all the identified contacts (Fig. 3.21), they define a large and flat contact area between the two partner proteins.
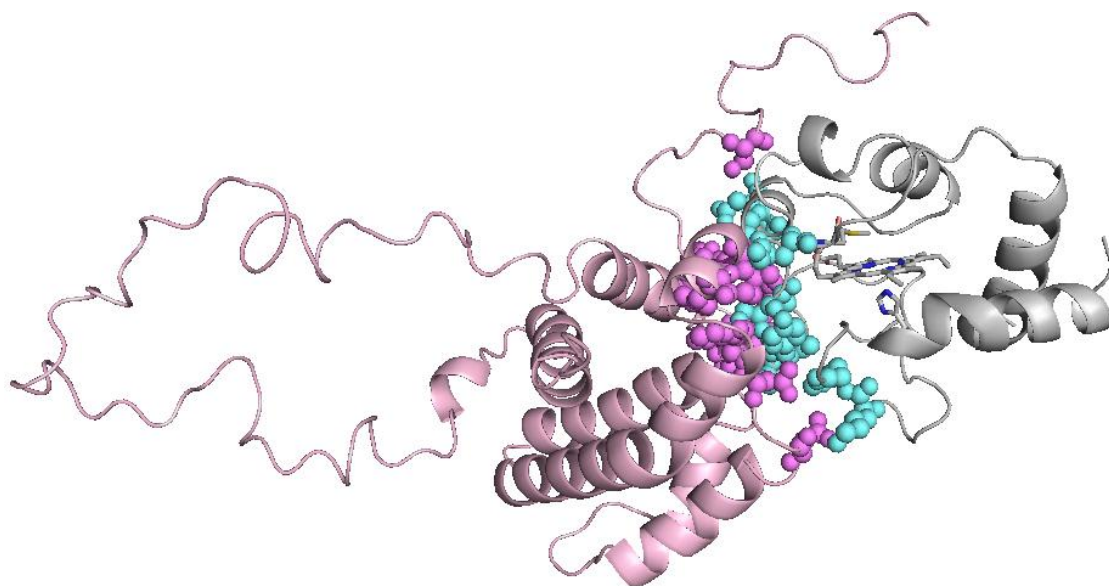


**Figure 3.21.** Residues on cytochrome c (cyan) and Bcl-x$_L$ (pink) that are involved in intermolecular contacts in our structural model. They have been mapped on one of the lowest energy structure of our cluster of 128 conformers.

Residues on human cytochrome c involved in contacts with Bcl-x$_L$ are located on the loop1, α2–helix and loop 2 (Fig. 3.15A). Interestingly, the only known pro-apoptotic mutant of cytochrome is G41S,[51] i.e. a variant bearing a mutation on a residue of the β-strand found to form an H-bond with Arg100 of Bcl-x$_L$ in 64 out of 128 structures of our ensemble. In contrast, those on yeast cytochrome c or on cytochrome c522 in complex with the electron transfer partners are centered on the heme crevice defined by the two loops, where atoms of the porphyrin ring become partially exposed (Fig. 3.15B-D). Such differences might account for the different primary and secondary structures of the examined cytochromes as well as for the different cytochrome c functions. In fact, when cytochrome c acts as electron transfer protein needs to have the porphyrin ring as close as possible to the redox partner. This is possible only from the porphyrin ring solvent exposed area. While the interaction of cytochrome c with Bcl-x$_L$ does not require any electron transfer, trailing off the need of the solvent exposed heme edge. For this reason cytochrome c in the anti-apoptotic complex shares only a few contact residues with the electron transport adducts. Indeed the core interactions are here centered on the left side (according to the view of Fig. 3.15A) of the heme crevice.

Involvement of the loops probably still reflects the need of conformational adaptability to facilitate an induced fit.

The structure of Bcl-x$_L$ consists of seven helices (Fig. 3.6) and a long flexible loop, spanning residues 45 to 84.[13,28] The C-terminal part contains a hydrophobic tail proposed to constituted the anchoring point in the membrane bound form. At the basis of this short tail the protein fold forms a large and flat surface (Fig. 3.22), that in the membrane-bound form should be oriented towards the mitochondrion. Residues in contact with cytochrome c are all located in this area. In particular: Glu96 and Tyr101 are on helix-3; Glu129 and Arg139 are the second-last and the first residue, respectively, of helix-4 and helix-5, which are antiparallel to each other and perpendicular to helix-3; residues 133-138 are on the loop connecting helix-4 to helix-5; Trp181, Glu184 and Asn185 are on helix-7 and Thr190, Glu193 and Leu194 on helix-8, two short helices roughly parallel to helix-3; finally, the last two residues forming contacts are Tyr195, immediately after helix-7 and Ser203 at the basis of the C-term tail.
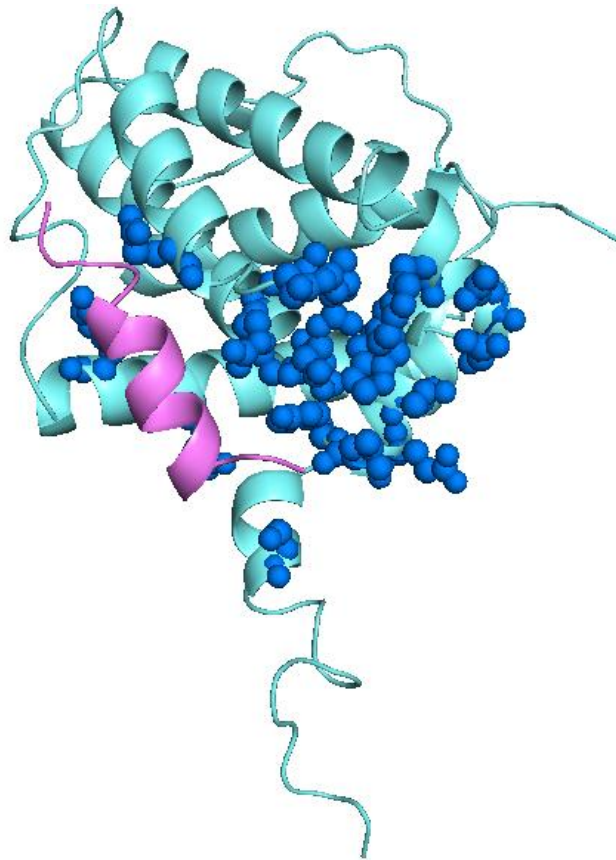


**Figure 3.22.** Ribbon representation of the structure of Bcl-x$_L$: the putative transmembrane hydrophobic tail point towards the bottom part of the picture. Residues involved in contacts with cytochrome c are displayed using a spacefill representation.

The Bak peptide is shown in magenta and its interaction area has only a few contact points with that defined for cytochrome c.

Their spatial location with respect to the anchoring tail suggests that cytochrome c is captured by the protein just at its entrance into the cytosolic space.

Arg139, whose mutation into Glu has been reported to inhibit the antiapoptotic activity of Bcl-$x_L$,[28] is involved in the interaction with cytochrome c and also with the Bak peptide; otherwise the contact surfaces residues of Bcl-$x_L$ with the two counterparts do not coincide. Complexation of Bcl-$x_L$ with the pro-apoptotic Bak peptide(s) has been reported to occur through an extended interaction with the hydrophobic cleft of Bcl-$x_L$ defined by helices 3 and 4; in addition a few charged side chains of opposite signs on the two partners are facing each other.[28]

The non-coincidence of the contact surface areas in the two adducts may provide hints for differently targeting the pro-apoptotic and the anti-apoptotic protein-protein interactions.

While our structural model seems reliable, one could question about the relevance of such a weak complex for blocking the apoptosome formation. In considering this aspect one should take into account the fact that in our NMR experiments both cytochrome c and Bcl-$x_L$ can free diffuse in three dimensions.

The anti-apoptotic process of sequestration of cytochrome c by Bcl-$x_L$ can be seen as a baid and prey process, where Bcl-$x_L$ acts as baid when anchored to the external mitochondrial membrane and therefore has restrict motions. The prey cytochrome c is "fished" once released in the cytosol, where in principle it may still be a three dimensional diffusant, but the proximity of the mitochondrial membrane may still influence its diffusion modes. Reducing the dimensionality of the recognition process between the two protein may lead to a sensible increase in binding efficiency.

### 3.7. Experimental part

#### 3.7.1. Protein samples

Full length human cytochrome c was expressed and purified as reported in the literature[52] in the unlabelled and $^{15}$N-labeled form.

The Bcl-$x_L$ construct used in our experiments contains residues 1-209 and lacks the C-terminal hydrophobic tail. The construct also has four additional N-term residues (numbers -3 to 0).

Unlabelled, $^{15}$N-labelled and $^{13}$C,$^{15}$N labeled forms of the protein were used for different NMR experiments.

Typical protein concentrations for NMR experiments were in the 50 μM to 5 mM range, in 50 mM sodium phosphate buffer at pH 7.3, 150 mM NaCl, 1mM DTT and with 10% $D_2O$ for lock.

#### 3.7.2. NMR experiments

All NMR spectra were acquired at 300 K using Bruker Advance spectrometers operating at proton frequencies of 500, 700, 800 and 900 MHz, all equipped with cryoprobes. Table 3.4 summarizes the NMR experiments performed.

| Experiments | Dimension of acquired data (nucleus) | | | Spectral width (ppm) | | |
|---|---|---|---|---|---|---|
| | $t_1$ | $t_2$ | $t_3$ | $F_1$ | $F_2$ | $F_3$ |
| $^1$H-$^{15}$N-HSQC[a] | 200 or 256 ($^{15}$N) | 1024($^1$H) | - | 13.0 | 40.0 | - |
| HNCA[b] | 200($^{13}$C) | 40($^{15}$N) | 1024($^1$H) | 32.0 | 36.0 | 14.0 |
| HN(CO)CA[b] | 200($^{13}$C) | 40($^{15}$N) | 1024($^1$H) | 32.0 | 36.0 | 14.0 |
| $^{15}$N $R_2$[c] | 200($^{15}$N) | 2048($^1$H) | | 32 | 14 | |

[a] Spectra acquired at 500, 800 or 900 MHz. All the spectrometers are equipped with cryo-probe (TXI 5-mm).
[b] Data acquired using a 700 MHz spectrometer equipped with a cryo-probe (TXI 5-mm)
[c] Data acquired using a 500 MHz spectrometer equipped with a cryo-probe (TXI 5-mm)
All the spectra were processed using the standard Bruker software (Topspin version 2.0) and analyzed by means of the CARA program.[53]

**Table 3.4.** Acquisition parameters for the NMR experiments; all spectra were recorded at 300 K.

### 3.7.3. Interaction studies

Titrations of ^{15}N human cytochrome c with unlabelled Bcl-x$_L$ and titrations of ^{15}N-Bcl-x$_L$ with unlabelled cytochrome c were followed through ^{1}H-^{15}N HSQC. Looking at the ^{15}N-enriched Bcl-x$_L$ the system was studied until a ratio of Bcl-x$_L$:cytochrome c 1:10. Looking at the ^{15}N-enriched cytochrome c we could reach a cytochrome:Bcl-x$_L$ ratio of 1:20.

Assignment of Bcl-x$_L$: Backbone resonance assignments of Bcl-x$_L$ were performed through conventional multidimensional NMR techniques based on triple resonance experiments, as summarized in Table 3.4. The assignment was carried out starting from the reported assignment on BMRB entry 6578,[54] that refers to a dimeric form of the protein lacking the 45-84 flexible loop. We have accomplished 84 and 80% assignment of the C$\alpha$ and HN backbone resonances, respectively.

### 3.7.4. R$_2$ measurements

The generalized increase in [15]N R$_2$ relaxation rates of Bcl-x$_L$ was used to monitor the increase in average molecular size in the presence of  2-fold and 2-fold excess of cytochrome c. The experimental details are provided in Table 3.3. The local overheating typical of R$_2$ measurements affects the stability of Bcl-x$_L$, as revealed by ^{1}H-^{15}N HSQC experiments recorded in an interleaved manner during R$_2$ experiments. The effect becomes more important in the presence of cytochrome c and is proportional to its concentration. Nevertheless an overall increase in R$_2$, consistent with an increase in the correlation time for tumbling was observed.

### 3.7.5. Chemical shift mapping

The interaction between cytochrome c and Bcl-x$_L$ was monitored through chemical shift changes of the signals from the backbone amide moieties, whose magnitude increased upon increasing concentration of the titrant (Fig. 3.18 and Fig. 3.19). The extent of the changes was quantified through the equation (2.1).

$K_d$ were obtained by plotting the weighted average chemical shift variations of perturbed residues on the $^{15}N$-enriched Bcl-$x_L$ as a function of the concentration of the unlabeled partner cytochrome c (Fig. 3.20) and were found to be in the 1-3 mM range.

### 3.7.6. Model structure calculations

A structural model of the cytochrome c − Bcl-$x_L$ adduct was obtained using HADDOCK program.[55] HADDOCK calculations were started with the coordinates of human cytochrome c (PDB id: 1J3S) and human Bcl-$x_L$ (PDB id: 1LXL). The starting structure for cytochrome c is actually the first member of an ensemble of 20 NMR conformers.

The active residues defined for the present calculation are listed in Table 3.5.

| Bcl-$x_L$ | cyt c |
|---|---|
| Q88, E129, D133, T190, F191[a], G196 | Q16, V20, G24, H26, G29, G41, A50, K79, M80, V83 |

a The residue was considered active although it does not result solvent accessible from NACCESS calculations based on the solution structure 1LXL.

**Table 3.5.** Haddock active residue of cytochrome c and Bcl-$x_L$.

The solvent accessibility has been calculated with the program NACCESS.

Four clusters containing 128, 21, 6, and 11 structures, were obtained through the HADDOCK docking procedure. The structural statistics calculated over all structures of each cluster are shown in Table 3.2. Cluster 1 is by far the best in terms of RMSD and energy values. It unequivocally defines the docking face for both proteins and their relative orientations.

### 3.8. Conclusions

Apoptosis normally eliminates cells with damaged DNA or aberrant cell cycle. Pro-survival proteins are therefore potentially oncogenic. Clarifying how the Bcl-2 family governs apoptosis might provide the ability to control the apoptotic threshold.

Conventional cytotoxic therapy indirectly induces apoptosis, but more effective outcomes should be achieved by direct activation of the apoptotic machinery. Promising approaches include impairing expression of pro-survival proteins or identifying drugs that inhibits their action. The identification of interfaces between partner molecules provides targets for pharmacological intervention: the protein-protein interaction surface between Bcl-$x_L$ and cytochrome c here identified may offer one of these targets.

## Reference List

1. Li, P. *et al.* Cytochrome c and dATP-dependent formation of Apaf-1/caspase-9 complex initiates an apoptotic protease cascade. *Cell* **91**, 479-489 (1997).

2. Yadaiah, M., Rao, P. N., Harish, P. & Bhuyan, A. K. High affinity binding of Bcl-xL to cytochrome c: possible relevance for interception of translocated cytochrome c in apoptosis. *Biochim. Biophys. Acta* **1774**, 1370-1379 (2007).

3. Kerr, J. F., Wyllie, A. H. & Currie, A. R. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br. J. Cancer* **26**, 239-257 (1972).

4. Leist, M. & Jaattela, M. Four deaths and a funeral: from caspases to alternative mechanisms. *Nat. Rev. Mol. Cell Biol.* **2**, 589-598 (2001).

5. Fadeel, B., Orrenius, S. & Zhivotovsky, B. Apoptosis in human disease: a new skin for the old ceremony? *Biochem. Biophys. Res. Commun.* **266**, 699-717 (1999).

6. Van, C. S. & Van Den, B. W. Morphological and biochemical aspects of apoptosis, oncosis and necrosis. *Anat. Histol. Embryol.* **31**, 214-223 (2002).

7. Luo, X., Budihardjo, I., Zou, H., Slaughter, C. & Wang, X. Bid, a Bcl2 interacting protein, mediates cytochrome c release from mitochondria in response to activation of cell surface death receptors. *Cell* **94**, 481-490 (1998).

8. Acehan, D. *et al.* Three-dimensional structure of the apoptosome: implications for assembly, procaspase-9 binding, and activation. *Mol. Cell* **9**, 423-432 (2002).

9.  Slee, E. A. *et al.* Ordering the cytochrome c-initiated caspase cascade: hierarchical activation of caspases-2, -3, -6, -7, -8, and -10 in a caspase-9-dependent manner. *J. Cell Biol.* **144**, 281-292 (1999).

10. Adams, J. M. & Cory, S. The Bcl-2 protein family: arbiters of cell survival. *Science* **281**, 1322-1326 (1998).

11. Antonsson, B. & Martinou, J. C. The Bcl-2 protein family. *Exp. Cell Res.* **256**, 50-57 (2000).

12. Gross, A., McDonnell, J. M. & Korsmeyer, S. J. BCL-2 family members and the mitochondria in apoptosis. *Genes Dev.* **13**, 1899-1911 (1999).

13. Muchmore, S. W. *et al.* X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. *Nature* **381**, 335-341 (1996).

14. Alirol, E. & Martinou, J. C. Mitochondria and cancer: is there a morphological connection? *Oncogene* **25**, 4706-4716 (2006).

15. Shimizu, S., Narita, M. & Tsujimoto, Y. Bcl-2 family proteins regulate the release of apoptogenic cytochrome c by the mitochondrial channel VDAC. *Nature* **399**, 483-487 (1999).

16. Antonsson, B., Montessuit, S., Lauper, S., Eskes, R. & Martinou, J. C. Bax oligomerization is required for channel-forming activity in liposomes and to trigger cytochrome c release from mitochondria. *Biochem. J.* **345 Pt 2**, 271-278 (2000).

17. Boise, L. H. *et al.* bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. *Cell* **74**, 597-608 (1993).

18. Chao, D. T. & Korsmeyer, S. J. BCL-2 family: regulators of cell death. *Annu. Rev. Immunol.* **16**, 395-419 (1998).

19. Sedlak, T. W. *et al.* Multiple Bcl-2 family members demonstrate selective dimerizations with Bax. *Proc. Natl. Acad. Sci. U. S. A* **92**, 7834-7838 (1995).

20. Yang, E. *et al.* Bad, a heterodimeric partner for Bcl-XL and Bcl-2, displaces Bax and promotes cell death. *Cell* **80**, 285-291 (1995).

21. Boyd, J. M. *et al.* Bik, a novel death-inducing protein shares a distinct sequence motif with Bcl-2 family proteins and interacts with viral and cellular survival-promoting proteins. *Oncogene* **11**, 1921-1928 (1995).

22. Yin, X. M., Oltvai, Z. N. & Korsmeyer, S. J. BH1 and BH2 domains of Bcl-2 are required for inhibition of apoptosis and heterodimerization with Bax. *Nature* **369**, 321-323 (1994).

23. Reed, J. C. Bcl-2 and the regulation of programmed cell death. *J. Cell Biol.* **124**, 1-6 (1994).

24. Parker, M. W. & Pattus, F. Rendering a membrane protein soluble in water: a common packing motif in bacterial protein toxins. *Trends Biochem. Sci.* **18**, 391-395 (1993).

25. London, E. Diphtheria toxin: membrane interaction and membrane translocation. *Biochim. Biophys. Acta* **1113**, 25-51 (1992).

26. Minn, A. J. *et al.* Bcl-x(L) forms an ion channel in synthetic lipid membranes. *Nature* **385**, 353-357 (1997).

27. Reed, J. C., Jurgensmeier, J. M. & Matsuyama, S. Bcl-2 family proteins and mitochondria. *Biochim. Biophys. Acta* **1366**, 127-137 (1998).

28. Sattler, M. *et al.* Structure of Bcl-xL-Bak peptide complex: recognition between regulators of apoptosis. *Science* **275**, 983-986 (1997).

29. Zha, H., ime-Sempe, C., Sato, T. & Reed, J. C. Proapoptotic protein Bax heterodimerizes with Bcl-2 and homodimerizes with Bax via a novel domain (BH3) distinct from BH1 and BH2. *J. Biol. Chem.* **271**, 7440-7444 (1996).

30. Butt, W. D. & Keilin, D. Absorption spectra and some other properties of cytochrome c and of its compounds with ligands. *Proc. Roy. Soc. London* **B512**, 429-458 (1962).

31. Ambler, R. P. Sequence variability in bacterial cytochromes. *Biochim. Biophys. Acta* **1058**, 42-47 (1991).

32. Swanson, R. *et al.* Tuna cytochrome c at 2.0 A resolution. I.Ferricytochrome structure analysis. *J. Biol. Chem.* **252**, 759-775 (1977).

33. Chothia, C. & Lesk, A. M. Helix movements and the reconstruction of the haem pocket during the evolution of the cytochrome *c* family. *J. Mol. Biol.* **182**, 151-158 (1985).

34. Benini, S., Rypniewski, W., Wilson, K. S., Van Beeumen, J. & Ciurli, S. Crystal structure of oxidized *Bacillus pasteurii* cytochrome $c_{553}$ at 0.97 Å resolution. *Biochemistry* **39**, 13115-13126 (2000).

35. Bartalesi, I., Rosato, A. & Zhang, W. Hydrogen exchange in a bacterial cytochrome *c*: a fingerprint of the cytochrome *c* fold. *Biochemistry* **42**, 10923-10930 (2003).

36. Kluck, R. M., Bossy-Wetzel, E., Green, D. R. & Newmeyer, D. D. The release of cytochrome *c* from mitochondria: a primary site for Bcl-2 regulation of apoptosis. *Science* **275**, 1132-1136 (1997).

37. Pettigrew, G. W. & Moore, G. R. *Cytochromes c; Biological Aspects*. Springer-Verlag, Berlin (1987).

38. Crofts, A. R. The cytochrome bc1 complex: function in the context of structure. *Annu. Rev. Physiol.* **66**, 689-733 (2004).

39. Lange, C. & Hunte, C. From the Cover: Crystal structure of the yeast cytochrome bc1 complex with its bound substrate cytochrome c. *Proc. Natl. Acad. Sci. USA* **99**, 2800-2805 (2002).

40. Bertini, I., Cavallaro, G. & Rosato, A. A structural model for the adduct between cytochrome c and cytochrome c oxidase. *J. Biol. Inorg. Chem.* **10**, 613-624 (2005).

41. Salje, J., Ludwig, B. & Richter, O. M. Is a third proton-conducting pathway operative in bacterial cytochrome c oxidase? *Biochem. Soc. Trans.* **33**, 829-831 (2005).

42. Pelletier, H. & Kraut, J. Crystal structure of a complex between electron transfer partners, cytochrome c peroxidase and cytochrome c. *Science* **258**, 1748-1755 (1992).

43. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. Magn. Reson.* **89**, 496-514 (1990).

44. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progr. NMR Spectrosc.* **34**, 93-158 (1999).

45. Bax, A. & Ikura, M. An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue. *J. Biomol. NMR* **1**, 99-104 (1991).

46. Ikura, M., Kay, L. E. & Bax, A. A novel approach for sequential assignment of $^1$H, $^{13}$C and $^{15}$N spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **29**, 4659-4667 (1990).

47. Grzesiek, S. & Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. Magn. Reson.* **99**, 201-207 (1992).

48. Grzesiek, S. & Bax, A. Improved 3D Triple-Resonance NMR Techniques Applied to a 31 KDa Protein. *J. Magn. Reson.* **96**, 432-440 (1992).

49. Grzesiek, S. & Bax, A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291-6293 (1992).

50. Cantor, R. C. & Schimmel, P. R. *Biophysical chemistry*. Freeman,W.H., San Francisco (1980).

51. Morison, I. M. *et al.* A mutation of human cytochrome c enhances the intrinsic apoptotic pathway but causes only thrombocytopenia. *Nat. Genet.* **40**, 387-389 (2008).

52. Wegerich, F., Turano, P., Allegrozzi, M., Moehwald, H. & Lisdat, F. Cytochrome c mutants for superoxide biosensors. *Anal. Chem.* **81**, 2976-2984 (2009).

53. Keller, R. & Wüthrich, K. A New Software for the Analysis of Protein NMR Spectra. 2002. Ref Type: Personal Communication

54. Petros, A. M., Fesik, S. W. & Olejniczak, E. T. 1H, 13C and 15N resonance assignments of a Bcl-xL/Bad peptide complex. *J. Biomol. NMR* **32**, 260 (2005).

55. de Vries, S. J., van, D. M. & Bonvin, A. M. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5**, 883-897 (2010).

# Iron mineralization in ferritin
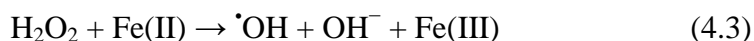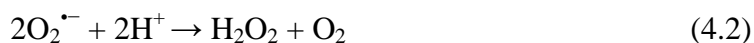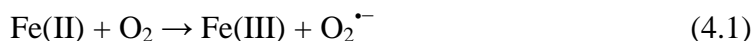
**Chapter**

# 4

## 4. Summary

In this chapter the ferritin-iron interaction study as an example of the protein-metal interaction will be presented. Ferritins are multimeric nanocage proteins present in most living organisms throughout evolution. Ferritins are iron storage and detoxification proteins, since their role is to store a non-toxic, water-soluble, yet bioavailable iron core and to release it whenever the cell needs. The biomineralization process that occurs inside the protein is a complicated multistep process long studied. The first step of the iron storage occurs at the ferritin catalytic ferroxidase site where two iron(II) ions react with dioxygen to form diferric species. These iron(III) complexes, once formed, migrate into the protein cavity where the iron nuclei of crystalline mineral form and aggregate to give the biomineral core. The structure of the iron(III) clusters as well as their pathway from the ferroxidase site to the central cavity, was still unknown and is subject of NMR studies at CERM. In addition, the crystal structure of ferritin with Fe(III) present in the active site, difficult to obtain because of the transient nature of substrate/enzyme interactions, was solved during this research study thanks to the collaboration between our research group and Mangani's group from the University of Siena.

The large size of ferritin hampers the standard solution NMR experiments for sequence specific assignment described in Section 3.4. Immobilization of the protein in the solid crystals overcame the problems arising from the rapid transverse relaxation and Magic Angle Spinning (MAS) solid state NMR (ssNMR)[1] provided spectra that allowed ferritin sequence specific assignment.

## 4.1. The importance of iron for biological systems

Iron approximately constitutes one third of our planet's mass and it is the fourth most abundant element in Earth's crust. Iron is an essential element for most organisms since serves as cofactor in several enzymes and is an important component of many cellular processes including respiration, electron transport processes, energy

metabolism, DNA synthesis and gene regulation. Despite its abundance, in physiological conditions it is poorly available. This is due to the fact that the stable oxidation state of the metal is Fe(III), that at neutral pH values forms insoluble hydroxy-aquo-complexes that are in equilibrium with a free Fe(III) concentration around $10^{-18}$ M. Some organisms need higher iron concentration, bacteria, for example, need iron concentrations of $10^{-7}$ M for growth. Moreover, free iron in the presence of oxygen can cause the one electron reduction of the lattice resulting in the formation of superoxide radical (equation 4.1) which can accept another electron and two protons to produce hydrogen peroxide (equation 4.2). On the other hand, Fe(II) can react with hydrogen peroxide to generate the hydroxyl radical that's a very reactive and damaging species.

$$Fe(II) + O_2 \rightarrow Fe(III) + O_2^{\bullet -} \qquad (4.1)$$

$$2O_2^{\bullet -} + 2H^+ \rightarrow H_2O_2 + O_2 \qquad (4.2)$$

$$H_2O_2 + Fe(II) \rightarrow {}^{\bullet}OH + OH^- + Fe(III) \qquad (4.3)$$

The sum of the last two equations produces the hydroxyl anion, the hydroxyl radical and $O_2$ (equation 4.4). This reaction occurs in the presence of a catalytic amount of iron.[2]

$$H_2O_2 + O_2^{\bullet -} \rightarrow OH^- + O_2 \qquad (4.4)$$

This hydroxyl radicals can damage a number of cellular components, like DNA, proteins and membrane lipids. Aerobic organisms, therefore, have developed strategies for iron uptake delivering and storage that protect cells from the potentially toxic effect of free iron and radical chemistry. There are three general mechanisms adopted in nature to acquire, to solubilize and store iron in a non toxic form: chelation, reduction and acidification. The most common chelating agent are called **siderophores**. They have high affinity for iron due to their multidentate ligand character. In addition they have high affinity for Fe(III) over other metals such that siderophores are very efficient in specifically mobilize iron even in the presence of other metal ions. The second strategy for solubilizing iron is its reduction to Fe(II) generically mediated by a class of enzymes called **ferrireductase**. This represents the mechanism that mammalian cells use for iron uptake. Once reduced, iron is transported into the cell by $Fe^{2+}$ specific transporters. The third strategy is acidification of the extracellular environment that

cause an increase of iron concentration. This strategy is commonly used by plants in case of iron deficiency.

### 4.1.1. Cellular iron uptake in mammalian

Mammalian cells can uptake iron as heme-iron or nonheme-iron. Uptake of nonheme-iron from the lumen if the mammalian intestine is mediated by a Fe(II) uptake system. Iron is first reduced by cell surface ferrireductases or by natural reductants and then taken into the intestinal epithelial cells. Once in the epithelial cells, the iron is exported across their membrane into the blood by **ferroportin**. Before its release into the serum, iron is oxidized and then bound by **transferrin** ($K_d = 10^{-22}$), that delivers Fe(III) to the cells of most tissue in the body. Iron uptake from transferrin into other cells types combine the three strategies discussed before as summarized in Figure 4.1. Once inside the cell Fe(III) is reduced by a ferrireductase and transported into the cytoplasm where it reaches its cellular targets (Fig. 4.1). Intracellular iron pathways are not fully understood.
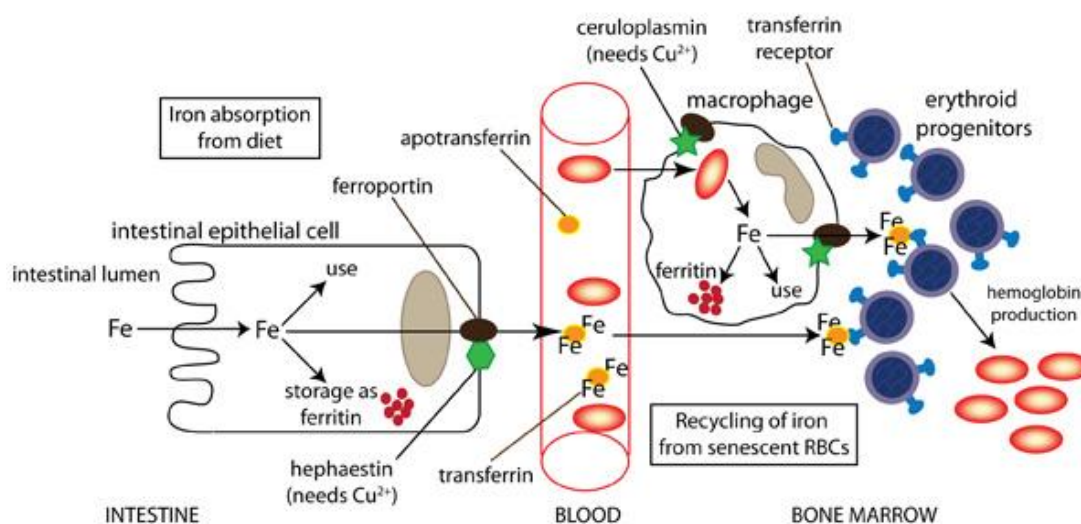


**Figure 4.1.** Mammalian iron uptake and storage.

*4.1.2. Intracellular iron storage and biomineralization in mammalian*

Organisms posses an efficient storage mechanism for iron accumulation. More in general, storing a metal ion inside a protein has the benefit to concentrate it without the toxic consequences that the accumulation of minerals might otherwise entail. Indeed, many of the storing mechanisms of metal ions also play a role in detoxifying toxic metals. The primary site of iron storage found in vertebrates, plants, some fungi and bacteria is ferritin.

## 4.2. Ferritins

Ferritins are members of the broad superfamily of iron storage and detoxification proteins found ubiquitariously in nature.[3] They have highly conserved supramolecular nanostructures designed to sequester thousands of iron atoms in a mineralized and biologically available form and to release them in a controlled manner. In the process of uptaking and storing iron, ferritins catalyze the contemporary oxidation of two Fe(II) trough a single dioxygen/hydrogen peroxide molecule by a mechanism that avoids the odd electron oxidation of iron(II) and the associated production of superoxide and hydroxyl radicals harmful to the cell.[2,4] The dioxygen/hydrogen peroxide molecules migrate into the cavity where biomineralization reaction occurs (equation 4.5):

$$2Fe(II) + O_2 + (H_2O)_{x+3} \rightarrow Fe_2O_3(H_2O)_x + 4H^+ + H_2O_2 \qquad (4.5)$$
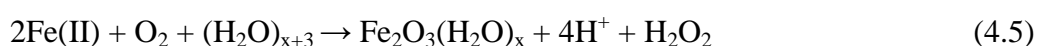
Thus, the ferritins are not only able to efficiently harvest and accumulate iron, but they also help to protect the cell against oxidative stress minimizing "free iron" which contributes to the generation of reactive oxygen species.[5] Another important property of ferritin is the protection of DNA from oxidative damage.[6]

In this respect, the importance of ferritin is illustrated by embryonic lethality of gene deletion in mammals[7] and resistance to oxidants in animals, plants, bacteria and archea.[8]

### 4.2.1. Mini and maxi ferritins

Ferritins are formed by the self-assembling of multiple α-helix bundle subunits oriented almost parallel to the protein surface and forming a ring around a cage that occupies about 30% of the total protein volume.[9] It has been shown that they use oxygen to make the iron concentrates.[10]

Ferritins are generally classified in two sub-families called **maxi**- and **mini**-ferritins (Fig. 4.2).
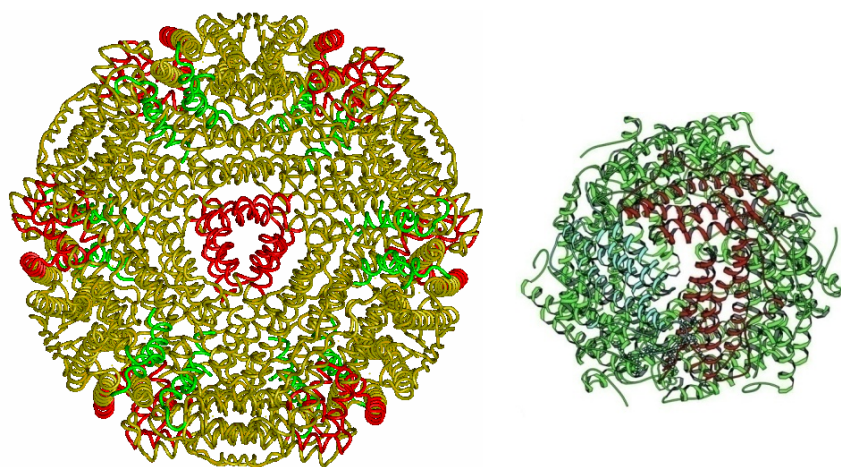


**Figure 4.2.** Maxiferritin (right) and mini-ferritin (left) ribbon representation.

Maxi-ferritins of animals, plants, and bacteria are constituted by 24 subunits of 20 kDa each assembled in a supramolecular complex of 480 kDa able to host 4500 iron atoms. The cage has an external diameter of 12 nm and a diameter cavity of 8 nm. Mini-ferritins of bacteria and archaea are formed by 12 subunits that form a protein cage of 240 kDa able to host up to 500 iron atoms. The nanocage has an external diameter of 9 nm and an internal one of 5 nm. They were first discovered during starvation and named Dps (DNA protection during starvation) proteins. In contrast to maxiferritins where dioxygen is used to concentrate iron, the miniferritins use iron to detoxify dioxygen or peroxide and protect DNA from damage.[11,12] I will focus my attention on the maxi-ferritins.

*4.2.2. Ferritin: type of chains*

The single ferritin subunits are four-helix bundles which typically consist of four α-helices packed in a coiled-coil arrangement. The four α -helices that run parallel and antiparallel one to another, are designated as A, B, C, and D. A long loop connects the B to the C α -helix, while two short loops connect the others.[13] The long loop and the helices A and C are solvent exposed, while B and D helices flank the inner cavity. In maxi-ferritins, a short fifth helix forms a turn with an axes of about 60° with respect to the four-helix bundle axis; it is the most phylogenetically variable part of the sequence (Fig. 4.3).
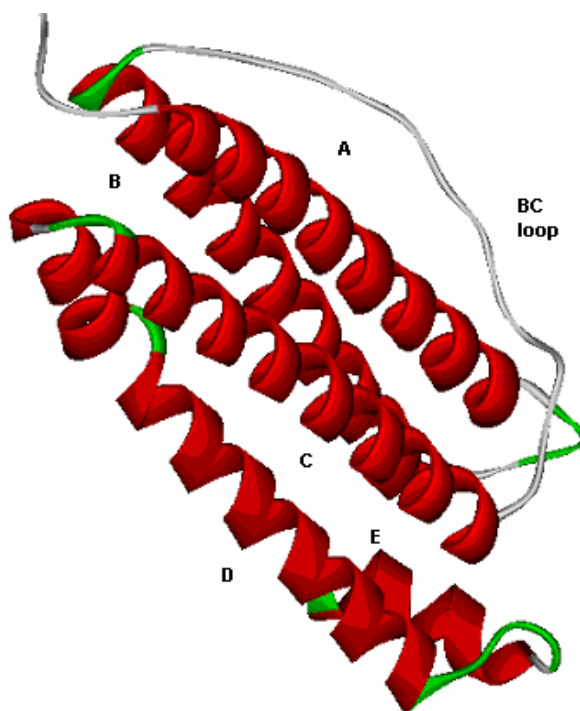


**Figure 4.3.** Ribbon diagram of the bull frog M ferritin (pdb code 1MFR) showing the relationship between the helices A, B, C, D and E, inclined at ∼60° to the four-helix bundle axes and the BC loop.

The single subunit can be constituted by a number of amino acid in the 154-185 range. Plant and bacterial ferritins are homopolymers, whereas mammalian ferritins are composed of two subunit types, the H (heavy ≈ 21 kDa)  and L (light ≈ 19 kDa) chains where the first is a few residues longer than the second one. The H chain contains the forroxidase site responsible for the iron oxidation, while the L chain does not show

97

ferroxidase activity but is contributes to the iron incorporation inside the iron core exposing carboxyl groups into the cavity.[14] In mammals both subunits co-assemble in different proportions in a tissue-specific manner.[15,16] Human H- and L-chains are of similar size (182 and 174 residues respectively) and share 55% amino acid sequence identity. Beyond the H and L chains, a so-called M (middle) subunit type exists. Its molecular weight is just in the middle between the H and the L subunit and it presents ferroxidase activity.

H and M subunits contain a set of residues that is highly conserved in the majority of ferritins. These residues constitute the catalytic oxidoreductase (ferroxidase) sites that is situated in the center of each of the four-helix bundles (Fig. 4.15A). At the catalytic site two Fe(II) once bound, are oxidized by dioxygen to form diferric oxo complexs precursors of the hydrated ferric oxide mineral.

### 4.2.3. Ferritin shell: axes and channels

Within the single subunits, a big network of interactions involving both hydrophobic and hydrophilic interactions are established. Moreover, ferritin subunits self-assemble to form a symmetrical and extremely stable ferritin spherical shell. The outcome of the assembly of 24 subunits, is a shell that has the approximate geometry of a rhombic  dodecahedron, the faces of which consist of two subunits related by a 2-fold symmetry axis at its centre (e.g., I and II or III and VI Fig. 4.4).
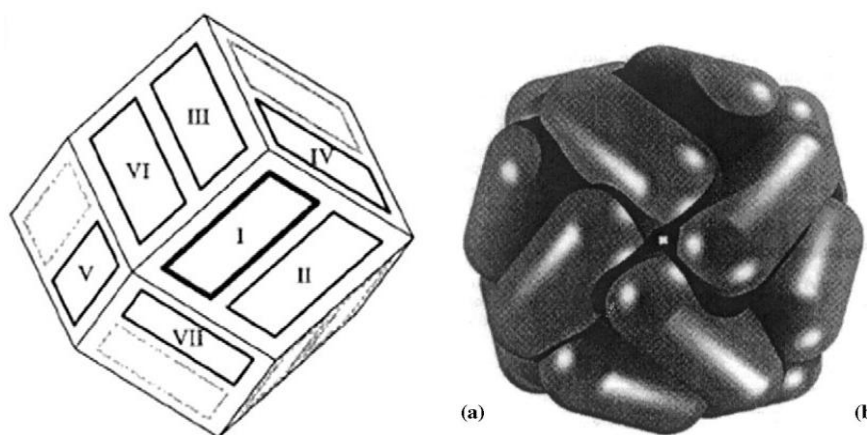


**Figure 4.4.** (a) Schematic diagram illustrating the relation between symmetry related subunits. (b) Assembled ferritin viewed from the fourfold axis.

Beyond the six 2-fold axes, eight 3-fold axes and six 4-fold axes are present (Fig. 4.5). The long side of subunit I interacts with the ends of subunits III and VI, while the two ends of subunit I also interact with subunits IV and VII. Most of the intersubunit interactions, involving both hydrophobic and hydrophilic interactions, lie within these interfaces, between the 3-fold and 4-fold axes.

This wide network of inter- and intra- subunit interactions explains the extremely high ferritin stability. Indeed ferritin can be safely heated up to 80 °C,[17] stands 6 M guanidine at neutral pH[18] and allows pH variation from 3 to 11 without unfolding. Moreover, intersubunit and intrasubunit interactions are so similar[19] that subunit dissociation and subunit unfolding usually coincide.

There are two clearly discernable channels leading from the outside to the inside of the protein shell along the 4-fold and 3-fold axes of symmetry. The six channels along the 4-fold axes (Fig. 4.5B) are constituted by four helices of four neighboring subunits. This gives a long apolar channel, 12.5 Å in length, and relatively narrow  it is accessible to a 2 Å diameter probe.[20] The eight 3-fold axes (Fig. 4.5C) are funnel-shaped with a wider entrance at the outside of the molecule and a narrow passage (3.4 Å wide and about 6 Å depth). The amino acid residues which neighbor and line these channels include residues from the C-terminus of helix C and those of the N-terminus of helix D. These hydrophilic residues (117–137) are highly conserved in all ferritins. Most data indicate that iron enters and exits the cavity of animal ferritins via these 3-fold channels[4] while the hydrophobic channels are thought to be involved in diffusion of oxygen and hydrogen peroxide.
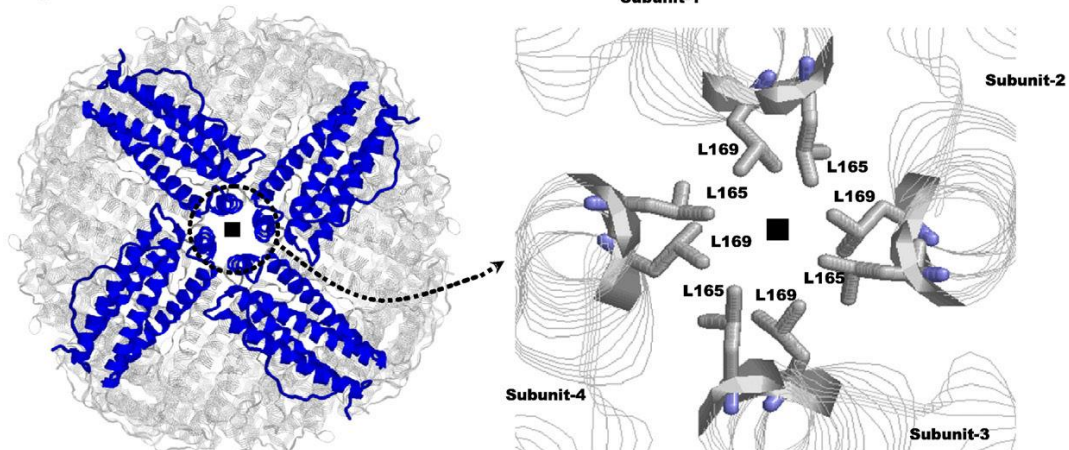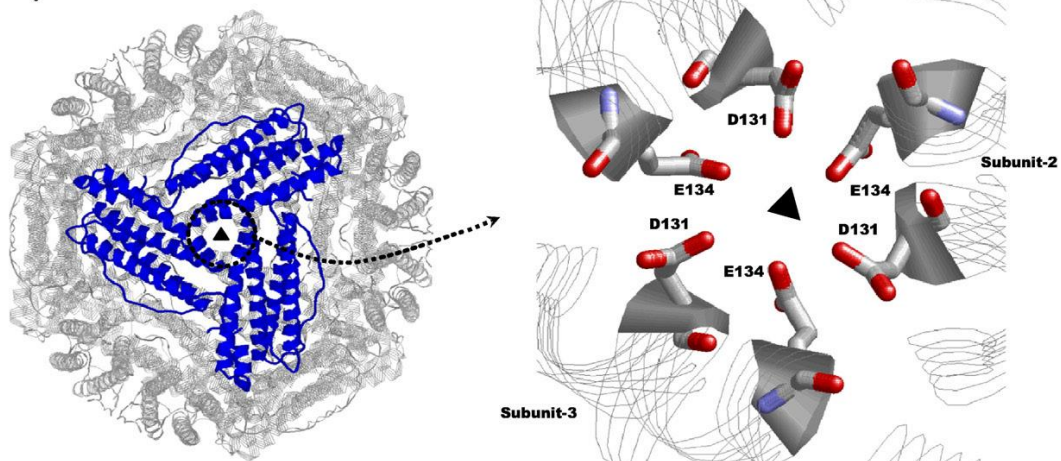
**Figure 4.5.** Full ferritin shell (left panel) and expanded views (right panel) of three channel types: (A) 4-fold, (B) 3-fold, and (C) 2-fold channels with key residues indicated. The rotation symmetry related subunits is highlighted in blue in the left panels.
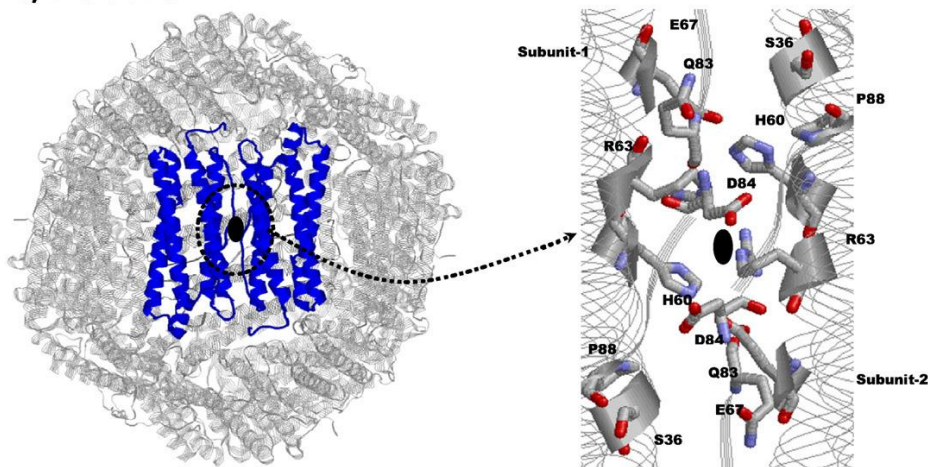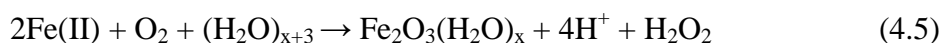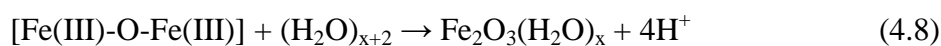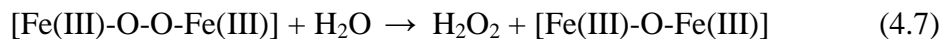
### 4.3. Iron storage in maxi-ferritins

Thousands (up to 4500) of iron atoms can be accommodated in ferritin minerals inside the protein cage. A large number of studies were carried out to understand the mechanism of iron uptake and deposition in ferritin. It has been proposed a multistep mechanism:[21]

1) two iron(II) ions bind at the protein oxidoreductase (ferroxidase) sites followed by a rapid oxidation trough dioxygen to form diferric peroxo products (equation 4.6) that immediately evolve in diferric oxo complexes (equation 4.7);

2) the resulting diferric oxo complexes migrates into the protein cavity following a pathway that since now was unknown;

3) in the protein cavity, a stable nucleus of crystalline mineral (ferrihydrite $Fe_2O_3(H_2O)_x$) starts to form (equation 4.8); the growing mineral core provides additional nucleation sites onto which incoming Fe(II) can be further deposited and oxidized.

$$2Fe^{2+} + O_2 \rightarrow [Fe(III)\text{-}O\text{-}O\text{-}Fe(III)] \qquad (4.6)$$

$$[Fe(III)\text{-}O\text{-}O\text{-}Fe(III)] + H_2O \rightarrow H_2O_2 + [Fe(III)\text{-}O\text{-}Fe(III)] \qquad (4.7)$$

$$[Fe(III)\text{-}O\text{-}Fe(III)] + (H_2O)_{x+2} \rightarrow Fe_2O_3(H_2O)_x + 4H^+ \qquad (4.8)$$

—————————————————————————————

$$2Fe(II) + O_2 + (H_2O)_{x+3} \rightarrow Fe_2O_3(H_2O)_x + 4H^+ + H_2O_2 \qquad (4.5)$$
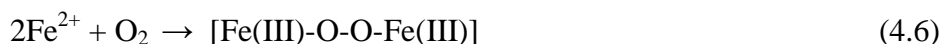
In accordance with equation 4.6, stopped-flow kinetics, resonance Raman spectroscopy, Mössbauer spectroscopy, and EXAFS have established that diferric peroxo product are formed during the first 50 ms after Fe(II) oxidation by $O_2$ at the ferroxidase site[22] and are 1,2-$\mu$ peroxo di Fe(III) intermediate having an unusually short Fe-Fe distance of 2.54 Å. This first intermediate subsequently decays within 150 ms to one or more diferric oxo intermediate(s) and small clusters whereupon $H_2O_2$ is produced in a stoichiometric amount and released into solution. Finally, the diferric oxo dimers and clusters migrate into the protein cavity where they lead to the formation of large polynuclear aggregates and the mineral core itself (equation 4.8). This reaction *in vitro* requires from days to months. This can be due to the slow kinetic of the mineral aggregation reaction that involves formation and breaking of bonds between iron products and protein residues.

Once a central metal core is formed into the protein cavity (up to 200 iron atoms), the iron oxidation occurs directly on the mineral surface as stopped-flow absorption spectrometry, UV spectrometry, and electrode oximetry revealed.[21]

What about the two sub-products $H^+$ and $H_2O_2$? From the overall reaction shown in equation 4.5 is clear that for each Fe(II) $2H^+$ are produced. Protons have to diffuse out of the protein in order to avoid fast pH decreasing and hydrolysis of peptide bonds. Concerning $H_2O_2$ it has been shown that it leaves only partly the ferritin shell, while the other part migrates inside the cavity where it reacts with Fe(II) helping the mineral formation (equation 4.9).[23]

$$2Fe^{2+} + H_2O_2 + 2H_2O \rightarrow 2FeOOH_{(core)} + 4H^+ \qquad (4.9)$$

### 4.3.1. Active site

As the first X-ray structures of ferritins began to appear, it became clear that ferritins were members of the larger family of binuclear non-heme iron enzymes. Indeed, the ligand set of iron in ferritins, as well as the mechanism of iron oxidation, oxygen activation and the diferric peroxo intermediate formation,[24-26] are similar to those found in other binuclear non-heme iron enzymes.[27]

Binuclear non-heme iron enzymes catalyze a variety of reactions using a largely conserved ligand set at the active site. Among these, the most studied are ribonucleotide reductase (R2), which converts ribonucleotide into deoxyribonucleotides, $\Delta^9$-desaturase ($\Delta^9$d), which inserts a double bond into a fatty acid chain, and methane monooxygenase (MMO), which catalyzes the oxidation of methane to methanol.

A crystal structure of ferritin with Fe(II) present in the active site is difficult to obtain because of the transient nature of substrate/enzyme interactions. However, a considerable number of ferritins X-ray structures co-crystallized with metal ion homologues for Fe(II) (e.g. with Mg(II),[28] Ca(II)[20] or Zn(II)[29]) are available, permitting comparisons with the diiron cofactor proteins.

In frog M ferritin, Mg(II) (Fig. 4.6A) has a the same ligand motif to that of the corresponding Fe(II) in $\Delta^9$d or R2 (Fig. 4.6C), consisting of two carboxylate and one histidine ligand-EEH and a quite similar on with respect to the MMO (Fig. 4.6B) that presents a two carboxylate and one histidine ligand-DEH motif. For most diiron

proteins this motif is generally repeated at the second iron site, while for ferritin such ligands are substantially different and consist of two glutamates (one is the Glu58 bridge), a glutamine, and an aspartate residue–EQD. Here, the histidine is replaced with a weaker aspartate ligand, and the unusual presence of the glutamine ligand prevents the formation of a second carboxylate bridge between the metals, a structural feature thus far present in all other nonheme diiron sites.
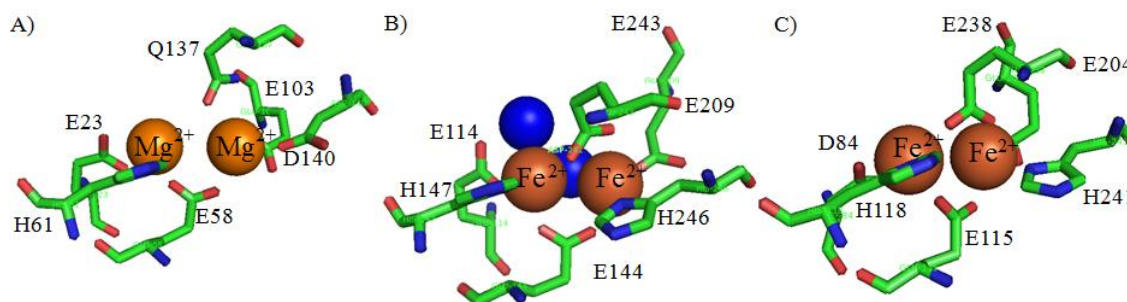


**Figure 4.6.** Crystal structures of the active sites (A) M ferritin crystallized with Mg(II) (pdb code 1MFR); (B) methane monooxygenase (pdb code 1FYZ) crystallized with Fe(II); (C) Ribonucleotide Reductase (pdb code 1PFR) crystallized with Fe(II).

This structural catalytic site differences reflect the diverse fate of the diferric species in ferritin versus non-heme iron enzymes. As reported in the Figure 4.7 in ferritin (left), the peroxodiferric complex decays to diferric oxo or hydroxo precursors that are translocated from the catalytic to the biomineralization sites and hydrogen peroxide is released. In diiron proteins the resting state has to be restored without any uptake of new iron. Rapid freeze-quench (RFQ) EPR, and Magnetic Circular Dichroism (MCD) studies shown that the peroxodiferric complex evolves into a tyrosyl radical that it's not released from the catalytic site.[30]
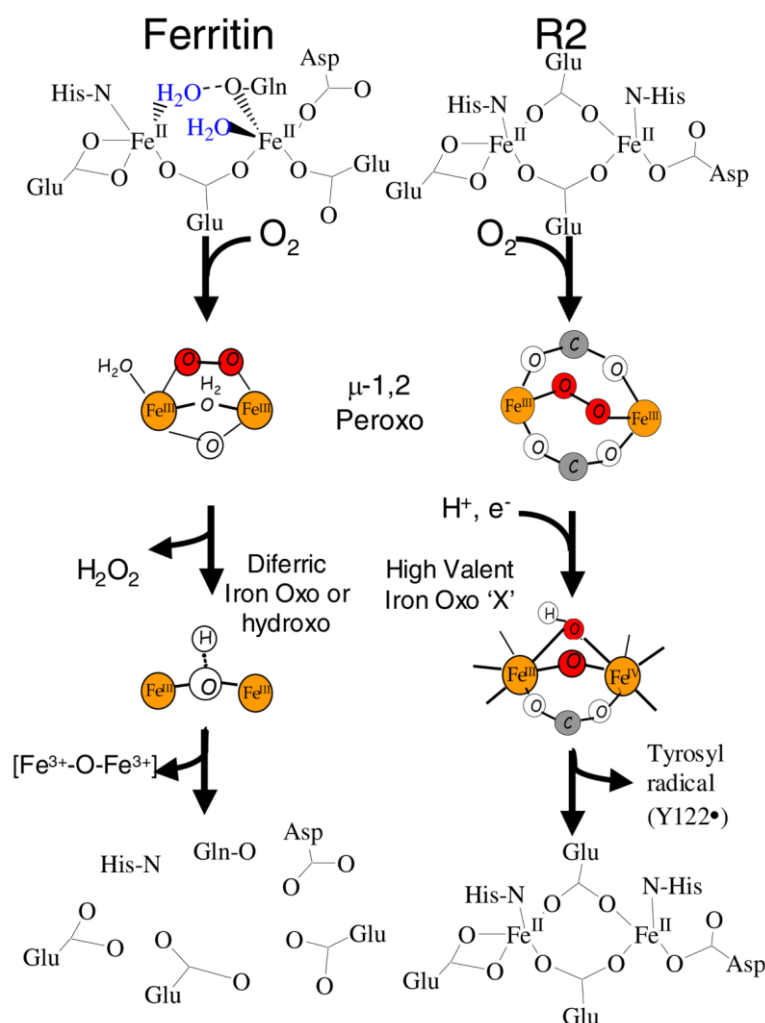
**Figure 4.7.** Left: proposed reaction pathway for M ferritin resulting in release of ferric-oxo mineral precursor from the active site, assuming equal rates of dioxygen binding and ET for all the types of diiron sites. The two bound waters may provide the protons necessary for formation of hydrogen peroxide. The water derived bridge present in the peroxo species may exist as a hydroxo or aquo bridge.[25,31] Right: reaction pathway for R2 ribonucleotide reductase leading to a high valent intermediate and generation of a stable tyrosyl radical (Y122•).[30]

In conclusion, the key differences between the diferrous substrates in ferritin and the diferrous cofactor sites in MMO, R2 and $\Delta^9$d are: 1) the single glutamate bridge between the bound Fe (II) atoms, 2) two water ligands, one coordinated to each Fe(II), 3) the weaker ligand field, and 4) the concerted Fe(II) binding mechanism and cooperativity.

### 4.3.2. Nucleation site

Once the diferric oxo complexes are formed, they migrate into the protein cavity where the iron (III) biomineral (ferrihydrite) starts to form (equation 4.8). It has been shown that the mineralization step occurs in the inner surface of the protein shell where the nucleation sites are localized. These mineralization centers are made up of the highly conserved diamond cluster of carboxylate residues clustered at the subunit dimer interfaces.[32,33]

The ferritin core is a hydrated ferric oxide mineral phase (ferrihydrite) of variable crystallinity and composition. The structural and magnetic properties of the iron-core have been investigated in detail by high-resolution transmission electron microscopy, Extended X-ray absorption fine structure (EXAFS), electron diffraction and Mossbauer spectroscopy.[34]

The iron core appears as a nanoparticle of 8 nm of diameter encapsulated in the protein shell. Each iron atom is surrounded by 6 oxygen atoms at an average distance 2 Å which are likely in a distorted octahedral (Fig. 4.8a) or hexagonal (Fig. 4.8b) arrangement and also it has 7 iron neighbors at an average distance of 3 Å. Depending from the composition and the rate of growing, the structure for the iron core is more or less ordered.[35]



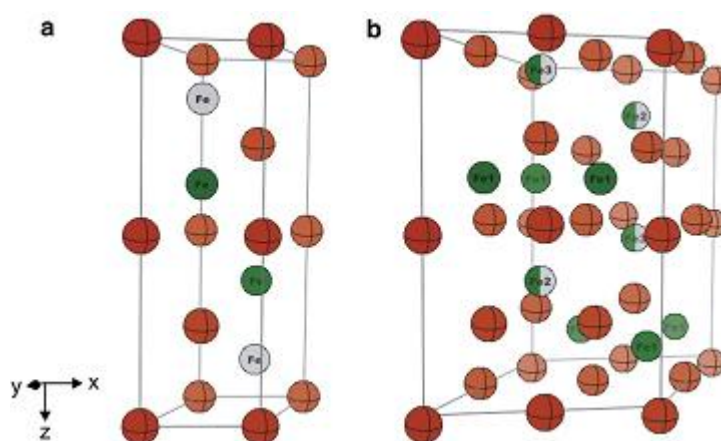**Figure 4.8.** Two proposed structural models for the arrangement of iron (grey/green) and oxygen (brown) in ferrihydrite. (a) Double-hexagonal component:[36] Fe atoms randomly occupy 50% of the octahedral sites with the constraint that neighboring Fe sites are not occupied simultaneously. (b) Hexagonal unit cell:[37] Partially filled circles represent 40–45% vacancies in the Fe2 (octahedral) and Fe3 (tetrahedral) sites.

### *4.3.3. Iron uptake*

When iron concentration in cells is high, ferritin acts also as an iron scavenger by sequestering excess cellular iron. The mechanism by which cytosolic iron is delivered to ferritin is still unknown. However, two years ago it has been found that human poly (rC)-binding protein (PCBP1) increased the amount of iron loaded into ferritin when expressed in yeast. Thus, PCBP1 may function as a cytosolic iron chaperone in the delivery of iron to ferritin.[38]

Moreover, recent electrostatic potential energy calculations, molecular diffusion, and thermodynamic and kinetics measurements of animal ferritin support the proposal that Fe(II) uses the threefold channels for entry.[39,40,40] The negative outer entrance  (Fig. 4.9) is surrounded by patches of positive potential, and this arrangement leads to electrostatic fields directing cations towards the channel entrance, the 3-fold route. The region of negative potential extends through the 3-fold channel to the interior of the molecule, such that $Fe^{2+}$ translocation through the channel would be driven by an electrostatic gradient. Alterations of the conserved amino acid residues that line these channels (Fig. 4.5B) decreased the capacity of H-chain ferritins to bind and oxidize Fe(II), documenting the role of the 3-fold channel in the entry of divalent cations into the interior of the protein. Inhibition kinetics experiments employing Zn(II) and Tb(III) as probes of Fe(II) oxidation indicated that the mechanism of inhibition occurs through blocking the passage of Fe(II) to the ferroxidase center, suggesting that the 3-fold channels are the main pathway of iron(II) entry in ferritin.[41] Figure 4.9 illustrates the putative pathway into the protein via the 3-fold channel where residues Thr135, His136, and Tyr137 are proposed to be involved in guiding iron from the inner opening of the channel to the ferroxidase centers of the protein.
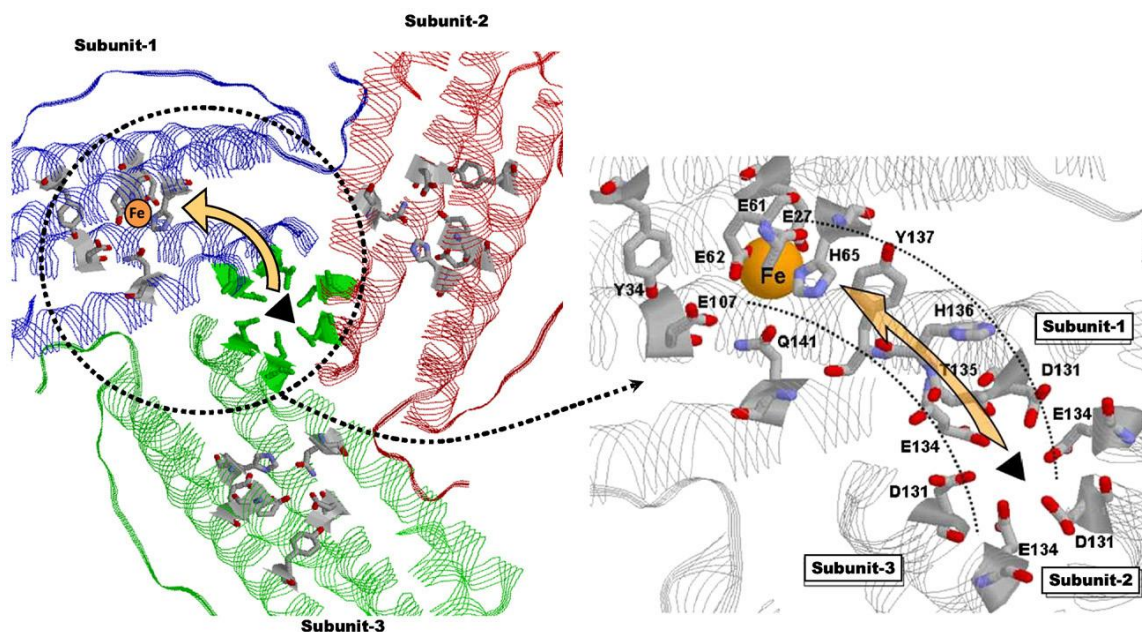
**Figure 4.9.** Internal view of the three-fold channel of ferritin (left panel). The right panel shows a putative Fe(II) pathway from the three-fold channel to the ferroxidase center of the protein.

### 4.3.4. Iron release

Ferritin releases iron in a controlled fashion. In response to decreasing cytosolic iron concentration, ferritin receives external signals that promote the immediate dissolution of the iron biomineral.

The mechanism of iron release from ferritin is still unknown. Two models are proposed for recovering ferritin iron *in vivo*. In the first model, cytoplasmic ferritin is incorporated into the lysosome, followed by protein degradation, iron mineral dissolution, and iron export to the cytoplasm.[42] It's still not clear how ferritin enters lysosomes and how iron crosses the lysosome membrane. The second model is based on the reversible unfolding/folding of ferritin pores that regulates the access of reductant and chelating agents allowing Fe(II) release from ferritin.[43,44] Ferritin pores are formed from six helices and present four highly conserved amino acids (Figure 4.10). that can be seen as analogous to the gated channels formed by membranes proteins involved in the control of ion flow.[42,45] Selective destabilization of the ferritin pore structure, independently of global ferritin structure, was observed by crystallography, site-direct mutagenesis and Circular Dichroism (CD) spectroscopy. Ferritin pores unfolding has

been observed at 56 °C, at millimolar concentration of urea or guanidine and after mutation of some residues in proximity to the pore. In particular, it has been shown that the substitution of the set of conserved residues increases mineral chelation rates. These residues are an interhelix (C-D) leucine-leucine pair, an interloop (C/D-B/C) aspartate-asparagine ion pair, and a short C-D loop.

It is possible that biological regulators recognize the pore gates and/or the folded/unfolded states to protect the ferric iron mineral from cellular reductants until the iron is needed for synthesis of iron cofactors.
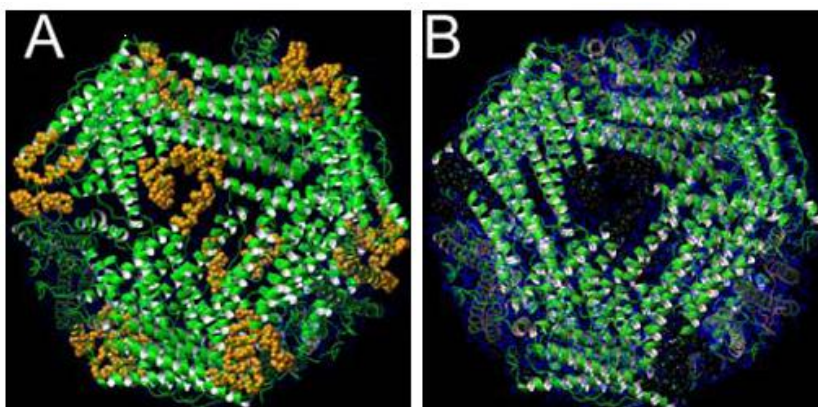


**Figure 4.10.** A) closed pores, B) open pores gates and unfolded pores.

## 4.4. Aim of the project

Here is reported the study of the iron(III) pathway from the ferroxidase site to the central cavity of the protein. To manage the investigation of such a big system a new NMR strategy was developed. The strategy is based on $^{13}C$-$^{13}C$ solution NOESY experiments combined with solid state proton driven $^{13}C$-$^{13}C$ spin diffusion and 3D coherence transfer experiments. In this way 75% of amino acids were recognized and 35% sequence specific assigned. Paramagnetic broadening, induced by iron(III) species in solution $^{13}C$-$^{13}C$ NOESY spectra, localized the iron within each subunit and traced the progression to the central cavity. Moreover, the crystal structure of iron-loaded ferritin has been solved.

## 4.5. Methods

^{1}H direct-detection NMR experiments are at the basis of the solution structure determination for many proteins and biological complexes. Complete structure determinations by solution NMR have generally been restricted to proteins of relatively low molecular weight ($< 30$ kDa). The approach fails for large protein (MW $> 30$ kDa) or paramagnetic proteins, proteins regions affected by exchange processes or unfolded proteins. While in the last case, a "solution structure" does not exist, in the other situations the method fails because of the short transverse relaxation times ($T_2$) that characterize these systems. Indeed, relaxation rates ($T_2^{-1}$ or $R_2$) and thus signal linewidths ($\Delta v$), increase with the size of the macromolecule: $\Delta v_{1/2} = 1/(\pi T_2)$, broadening signals beyond detection.

For proteins of $< 100$ kDa,[46] novel techniques using cross-correlated relaxation between dipole-dipole interactions and chemical shift anisotropy (CSA) have overcome the limitations imposed by transverse relaxation, allowing the detection of some sets of signals.[47] Nevertheless, for systems larger than 100 kDa, the use of [1]H–[15]N triple resonance experiments fails, because the coherence transfer relying on scalar couplings suffers for the reduced transverse relaxation times. Examples are transverse relaxation-optimised spectroscopy (TROSY),[48] cross-correlated relaxation-enhanced polarization transfer (CRINEPT),[49] that in combination with extensive protein deuteration, have successfully been used for the detection of amide resonances in systems up to 900 kDa,Fiaux, 2002 12702 /id} but do not allow assigning proteins.

### 4.5.1. Carbon detection experiments in solution NMR

Beside ^{1}H-detected NMR, ^{13}C-direct detection experiments, have been developed with significant results.[50-52] These experiments take advantage of the lower gyromagnetic ration of ^{13}C with respect to ^{1}H, which reflects in less efficient relaxation mechanisms and therefore longer $T_2$ values.[53]

On the other hand, [13]C-^{13}C NOESY-type experiments are not influenced by the negative effect of the molecular size. Indeed, the magnetization transfer phenomena at the basis of NOESY experiments occur when the magnetization is along the z-axis, and are therefore not affected by transverse relaxation.[54] During the NOESY mixing time

two processes are operative: the cross-relaxation, that is responsible of the magnetization transfer through dipolar coupling, and the longitudinal relaxation which restores the magnetization to the equilibrium values. The longitudinal relaxation times are substantially longer than transverse relaxation times in large proteins. The cross relaxation increases with molecular weight, being directly proportional to the rotational correlation time of the molecule.[55] Therefore the NOESY intensities in high molecular weight systems gain from both processes.

Nevertheless, $^{13}C$-$^{13}C$ NOESY-type experiments do not allow the sequence-specific assignment because sequential Cα-Cα connectivities are usually too long to be observed while those between carbonyls and carbonyls and Cα of sequential amino acids, which are as short as 2.4-2.9 Å, at the current stage of spectral resolution could not be determined with certainty given the large linewidth of carbonyl resonances. For this reason it has been necessary to develop new techniques, such as solid state NMR, that permit the assign of such a big system.

### 4.5.2. Magic Angle Spinning Solid State NMR

Over the last decades, solid-state NMR spectroscopy (ssNMR) has emerged as a key technique for chemical analysis, structure determination and dynamic investigation of organic, inorganic and biological systems.[56] This rapid development is mainly due to advances in instrument hardware design, stable magnets, sample isotope labeling and novel experimental methodologies. In the past, ssNMR revealed structural details of materials such as polymers, cement, petroleum, coal, glasses, organometallic compounds, catalysts, and many more,[57] not accessible by other spectroscopic methods. Today, a major challenge is to characterize at an atomic resolution the three dimensional structure and dynamics of biomolecules such as insoluble proteins, immobilized biological assemblies, membrane-embedded proteins,[58] fibrillar aggregates[59] or intact viruses.[60]

In solution, molecules tumble quickly enough to average out the orientationally dependent anisotropic interactions (such as Chemical Shift Anisotropy (CSA), heteronuclear and homonuclear dipolar couplings and, for spin with S > 1/2, quadrupolar couplings) giving spectra with relatively sharp transitions. By contrast, solids do not tumble enough and anisotropic or orientation-dependent interactions are

observed in the spectrum. Thus anisotropic general approaches for obtaining high resolution ssNMR spectra have been developed: the use of oriented samples and the MAS technique. In the latter, the sample is spun around an axis tilted at an angle of 54.73° (the magic angle) with respect to the static magnetic field[61]. In this work, we have focused on MAS methods for structural studies. MAS gives sufficiently resolved spectra and allows the detection of the isotropic chemical shifts making the implementation of the well-established solution NMR strategies for resonance assignment possible. Moreover, the ssNMR assignment can be exploited to obtain information about secondary structure motifs, as common in solution NMR.

### 4.5.2.1. Basic tools in solid-state MAS NMR

The anisotropic nuclear interactions that are of interest in solid-state NMR spectroscopy can be described by second-rank tensors which couple the nuclear magnetic moment to the pertinent field and together form the nuclear spin Hamiltonian H:

$$H = H_Z + H_{RF} + H_{CS} + H_D + H_J + H_Q \qquad (4.10)$$

$H_Z$ is the Zeeman term, which describes the interaction between a nuclear spin I and the external field $B_0$:

$$H_Z = \mathbf{I} \, Z \, \mathbf{B}_0 \qquad (4.11)$$

where I and $B_0$ are represented by vectors and Z is a second-ranked tensor.

Transitions between the Zeeman states are induced by an oscillating field $B_1$ and are described by the $H_{RF}$ term:

$$H_{RF} = \mathbf{I} \, Z \, \mathbf{B}_1 \qquad (4.12)$$

where the $B_1$ field is generally thousand times smaller than the $B_0$ field.

The remaining terms of the Hamiltonian give rise to the characteristic features of an NMR spectrum. These are the chemical shift Hamiltonian $H_{CS}$, the dipolar

Hamiltonian $H_D$, the J-coupling Hamiltonian $H_J$ and the quadrupolar coupling Hamiltonian $H_Q$. The chemical shift Hamiltonian $H_{CS}$ can be written as:

$$H_{CS} = \gamma \, \mathbf{I} \, \sigma \, \mathbf{B}_0 \qquad\qquad (4.13)$$

where $\sigma$ is the shielding tensor and describes the effect of the electron distribution around the nuclear spin. Because the electron distribution is not uniform, the chemical shift interaction depends on the orientation of the nucleus with respect to $B_0$. In a first-order perturbation theory, the truncated chemical shift Hamiltonian $H_{CS}$ for a single spin becomes:

$$H_{CS} = \gamma \, \mathbf{I}_z \, \sigma_{zz}^{LF} \, \mathbf{B}_0 \qquad\qquad (4.14)$$

The index LF indicates that $\sigma_{zz}^{LF}$ is the zz element of the $\sigma$ matrix in its laboratory-frame (LF) representation.

The dipolar Hamiltonian $H_D$ describes the through-space coupling between two nuclear spins $\mathbf{I}^I$ and $\mathbf{I}^J$:

$$H_D = \mathbf{I}^I D \, \mathbf{I}^J \qquad\qquad (4.15)$$

where $D$ is the dipolar coupling tensor, which describes how the field due to the spin I varies with the orientation of the I-J internuclear vector in the applied field ($\theta_{IJ}$). The dipolar coupling has also an $r^{-3}$ distance dependence, which is source of distance restraints in NMR. The truncated dipolar Hamiltonian is:

$$H_D = - \left(\mu_0/4\pi\right) \, \hbar \, \Sigma_I \, \Sigma_J \, (\gamma^I \gamma^J / r_{IJ}^3) * (1/2) * (3\cos^2(\theta_{IJ}) - 1) \, 2 * I_z^I I_z^J \qquad\qquad (4.16)$$

J-couplings are dipolar couplings that are mediated through bonds by the electrons.[62] In the solid-state J-couplings are much smaller than the dipolar interactions, nevertheless solid state NMR experiments have recently been presented, which utilize J-couplings to establish homonuclear[63,64] and heteronuclear[65] correlations.

The quadrupolar coupling Hamiltonian $H_Q$ is non-zero only for nuclei with spins greater than 1/2 and is proportional to the electric-field gradient tensor $V$. The chemical shift tensor $\sigma$, the dipolar interaction tensor D and the electric-field gradient tensor $V$ contain all a $(3\cos^2\theta - 1)$ orientation dependence factor. Consequently, in a MAS NMR

experiment, chemical shift anisotropy, dipolar and quadrupolar interactions can be averaged by mechanically spinning the sample about an axis aligned at 54.7° with respect to the external magnetic field, as depicted in Figure 4.11. When the sample is at the magic-angle, the anisotropic factor vanishes and under fast rotation the anisotropic broadening is removed. The isotropic shift interactions and the isotropic J couplings are left and high-resolution NMR spectra can be recorded, as in solution NMR.
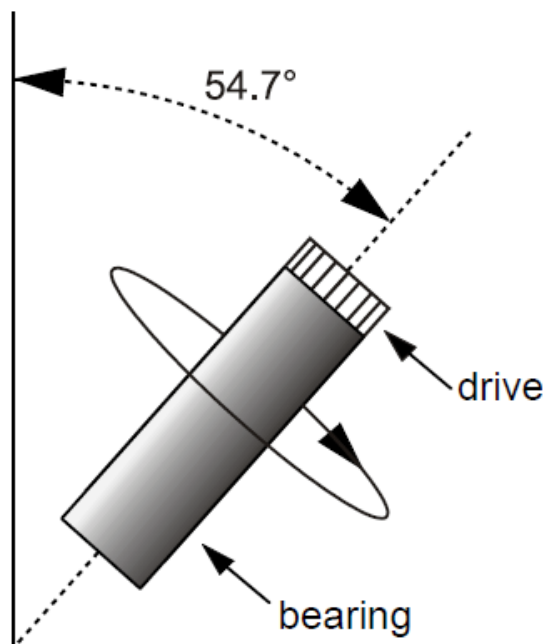


**Figure 4.11**. Schematic representation of the magic-angle spinning (MAS) technique. The rotor, that contains the sample, is tilted at 54.7° respect to the external magnetic field. By the combined application of bearing and driving pressures, the sample rotates around itself, at the desired frequency.

4.5.2.2. Cross Polarization transfer (CP)

Proton would be in principle the best nucleus to be studied in ssNMR due to its large gyromagnetic ratio ($\gamma_H$) and its high natural abundance. In an NMR experiment, the sensitivity, i.e. the signal-to-noise ratio (S/N), depends on both these parameters and, of all naturally occurring nuclei, the proton has the best sensitivity. In solution NMR, indeed, proton detected experiments are used to define protein structure, as discussed in the previous chapter. However, a large $\gamma_H$ implicates large homonuclear

dipolar $^1$H-$^1$H couplings and gives a homogenously broadened spectrum unsuitable also for line narrowing techniques such as MAS and homonuclear dipolar decoupling. Under the best homonuclear decoupling sequence, the proton resonances in solid are at least an order of magnitude broader than in solution. Different techniques have been developed to overcome this problem.

Up to now the most adopted method consists in heteronuclear detection. So far, all standard ssNMR experiments have relied on $^{13}$C and $^{15}$N detection. Low-$\gamma$ nuclei detection in the presence of MAS and high power proton decoupling gives high resolution spectra. The sensitivity can be improved chemically using isotope enrichment and with spectroscopic tools using Cross-Polarization (CP). CP transfers magnetization from proton to the low-$\gamma$ nuclei ($^{13}$C or $^{15}$N) as shown in Figure 4.12.
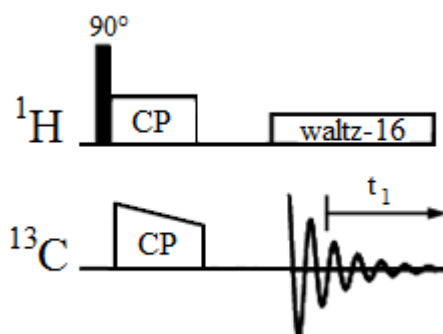


**Figure 4.12.** Pulse program for a $^1$H-$^{13}$C CP with dipolar coupling during the acquisition time.

The signal enhancement resulting from CP is due to two factors. First, the larger gyromagnetic ratio of protons (e.g. $\gamma_H$ is 4 times bigger than $\gamma_C$) creates a larger $^1$H polarization which is transferred to the low-$\gamma$ nucleus. Second, the repetition time of the experiment is determined by the shorter $^1$H relaxation time relative to low-$\gamma$ spins and hence the experiment repetition rate can be increased. For typical $^1$H-$^{13}$C CP experiment, these two factors can easily result in a 10-fold increase in sensitivity. In a CP experiment, the field strengths in the rotating frame are set to the Hartmann-Hahn condition, as depicted in Figure 4.13.
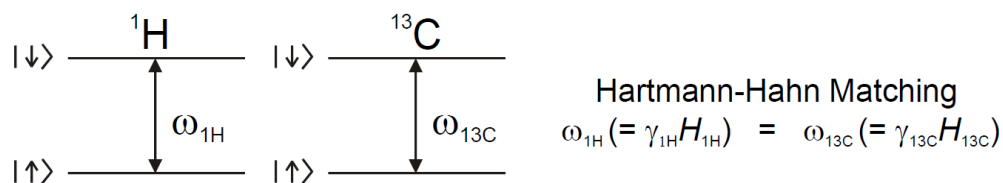
**Figure 4.13.** The classic Hartmann-Hahn matching condition for a non-spinning sample, that allows transfer of polarization from $^1$H to low-$\gamma$ nuclei, such as $^{13}$C. For a spinning sample, the MAS frequency has to be taken into account and the matching condition becomes: $\omega^1H - \omega^{13}C = \pm n\,\omega_r$.

The standard pulse program for ssNMR protein resonance assignment starts with a $^1$H excitation 90°-pulse and then magnetization is transferred from protons to the low frequency channel (e.g. $^{13}$C) by the matched pair of CP pulses.

### 4.5.2.3. Assignment strategy

In ssNMR the common protein resonance assignment procedure established in the last few years consists in the identification of the amino acid spin systems and of the intraresidue backbone amide nitrogen, followed by the linking of the different spin systems for the sequential assignments. The three-dimensional chemical shift correlation experiments used for this purpose are the dipolar assisted rotational resonance (DARR),[64] the NCACX, the NCOCX and the CANCO.[65,66] The DARR experiment is based on the spin diffusion effects along amino acids side chains making the identification of the side chains of each amino acid possible. In the NCACX experiment it is possible to link each spin systems to its nitrogen, while in the NCOCX experiment each side chain is connected with the nitrogen of the following residue. Finally, the CANCO experiment links the carbonyl of each residue to the C$\alpha$ and nitrogen of the following residue. In particular, a peak in the NCACX resonates at the frequencies $N_i$-C$\alpha_i$-CX$_i$, where CX$_i$ can be one of the $^{13}$C nuclear spins of residue *i* (C$'_i$, C$\alpha_i$, C$\beta_i$ etc.) as shown in Figure 4.14. A peak in the NCOCX resonates at the frequencies of $N_{i+1}$-C$'_i$-CX$_i$. A peak in the CANCO resonates at the frequencies of C$\alpha_{i+1}$-N$_{i+1}$-C$'_i$. Therefore, a strip at a given $^{15}$N frequency in the NCACX permits the identification of intraresidue (residue *i*) C$\alpha$, C$\beta$ and C$'$. Comparison with the NCOCX

strip allows the identification of the backbone $^{15}$N frequency for the following amino acid (*i+1*). With this $^{15}$N frequency, in the CANCO we can identify the Cα of the (*i+1*) residue (Fig. 4.14).
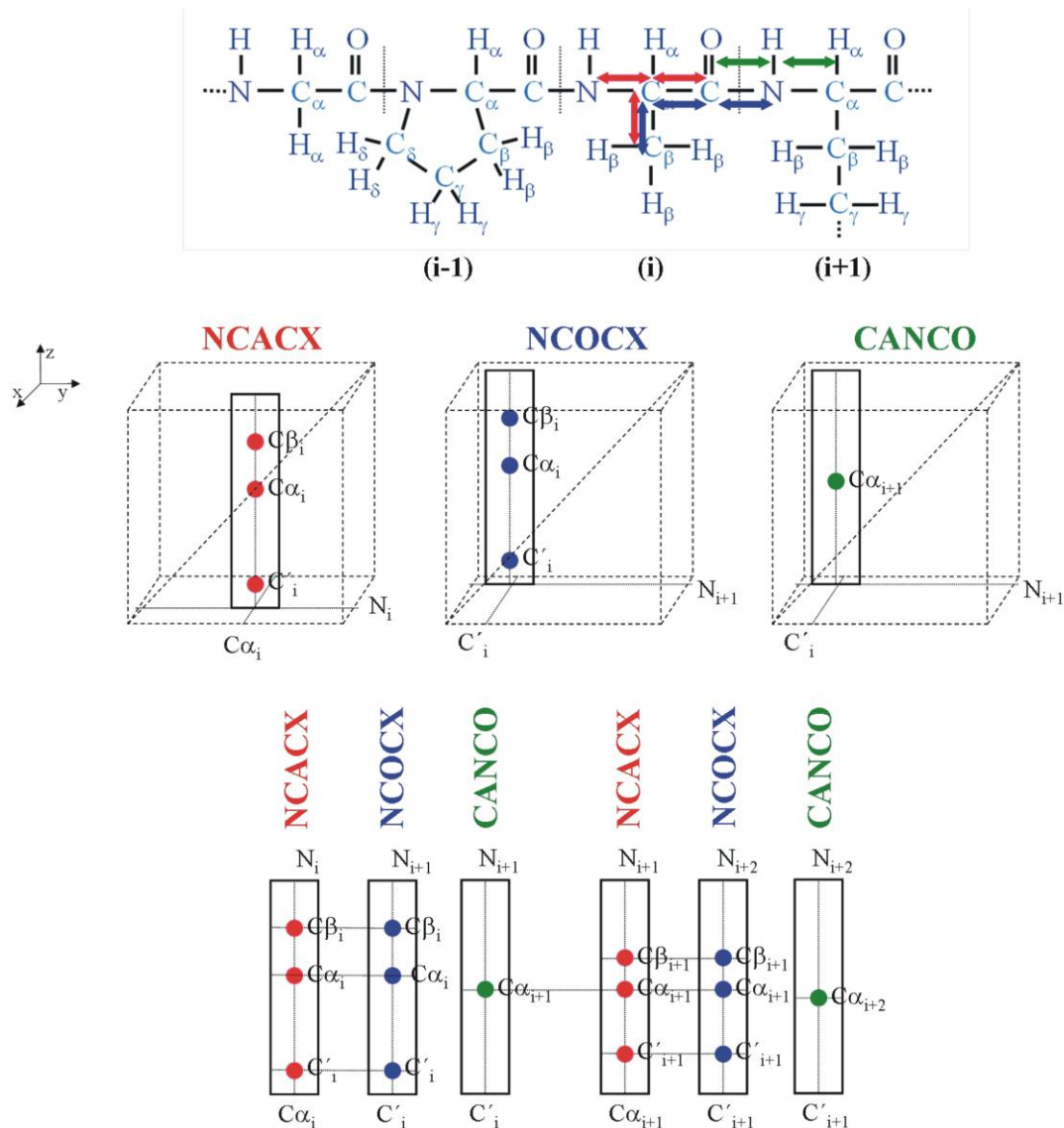


**Figure 4.14.** Schematic representation of the possible connectivities detected in the 3D experiments NCACX (red), NCOCX (blue) and CANCO (green) used for the sequential assignment at the solid state. Independently of the acquisition dimension, the 3D spectra are oriented with the $^{15}$N dimension along the x axis. Each strip is a portion of a $^{13}$C-$^{13}$C plane (y, z), centered at the $^{13}$C frequency reported in the bottom of each spectrum, extracted from a plane at a $^{15}$N frequency. The vertical strips are then visualized next to one another and allow to build-up sequential patterns.

### 4.6. Results

The pathway of the ferroxidase products to the mineralization site into the protein cavity was so far unknown. Simple inspection of the protein cage does not show an obvious path from the active site to the mineralization cavity. Therefore, we sought the path by solution NMR spectroscopy by taking advantage of the paramagnetic effects induced by the iron(III) species released at the catalytic site, on the resonances of nearby residues. The unequivocal identification of the signals of amino acids to be used as probes for the paramagnetic effects required the development of a new NMR approach because of the high molecular weight of the assembled protein (480 kDa). Indeed, the large size of ferritin causes slow tumbling the molecule in solution and prevents $^1$H NMR signal detection. We applied $^{13}$C-$^{13}$C NOESY in solution for the identification of about 75% of amino acid spin systems.[67] On the other hand slow molecular tumbling induces rapid transverse relaxation and hampers solution NMR experiments for sequence specific assignment based on coherence transfer. Immobilization of the protein in the solid state overcomes the problems arising from the rapid transverse relaxation, and magic angle spinning (MAS) solid state NMR[1,68] provided relatively well resolved spectra (Fig. 4.15A).
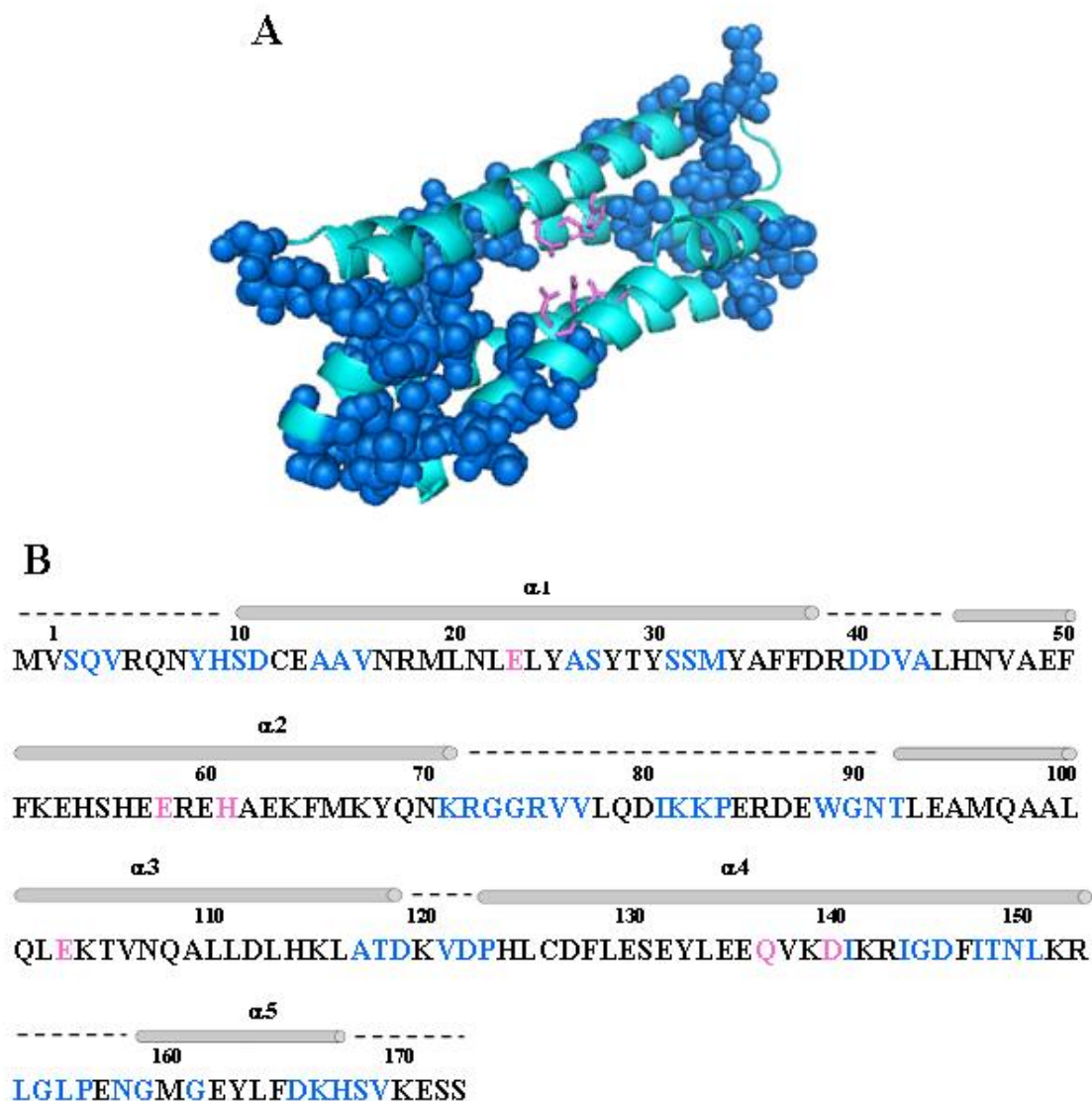
**Figure 4.15.** Structure of the 24-mer bullfrog ferritin (A). The primary and secondary structures (taken from PDB 1MFR) of bullfrog M ferritin subunits are shown in panel (B). Residues belonging to the ferroxidase site are indicated by magenta characters. Residues whose resonances are sequence specifically assigned are indicated in blue.

We used the protein resonance assignment procedure for ssNMR described in Section 4.1.4, that provides direct connectivities between these backbone atoms on adjacent residues (Figs. 4.16-4.17).
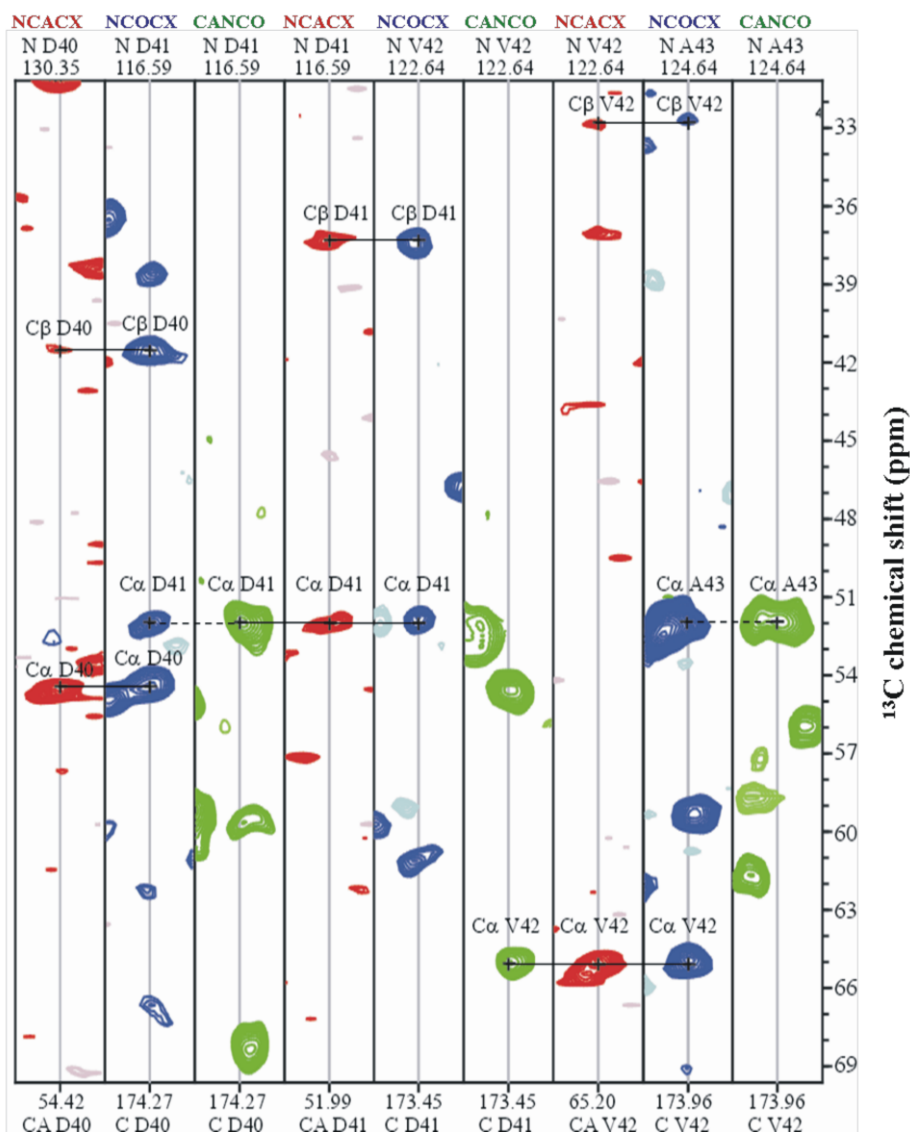
**Figure 4.16.** An example of a sequential walk for residues 40-43 through strips constructed from NCACX (red traces), NCOCX (blue traces), and CANCO (green traces) experiments. The type of spectrum from which strips are extracted is reported on the top of each strip, together with the $^{15}$N frequency of the plane from which they are extracted. Each strip is centered at the $^{13}$C frequency reported on the bottom. Strips in the CANCO and NCACX with the same $^{15}$N chemical shift and containing the same C$\alpha$ allow the sequential assignment through the C'. Peaks in the NCOCX connected to those of CANCO through dashed lines are sometimes observed, as reported by Oschkinat and coworkers.[69]
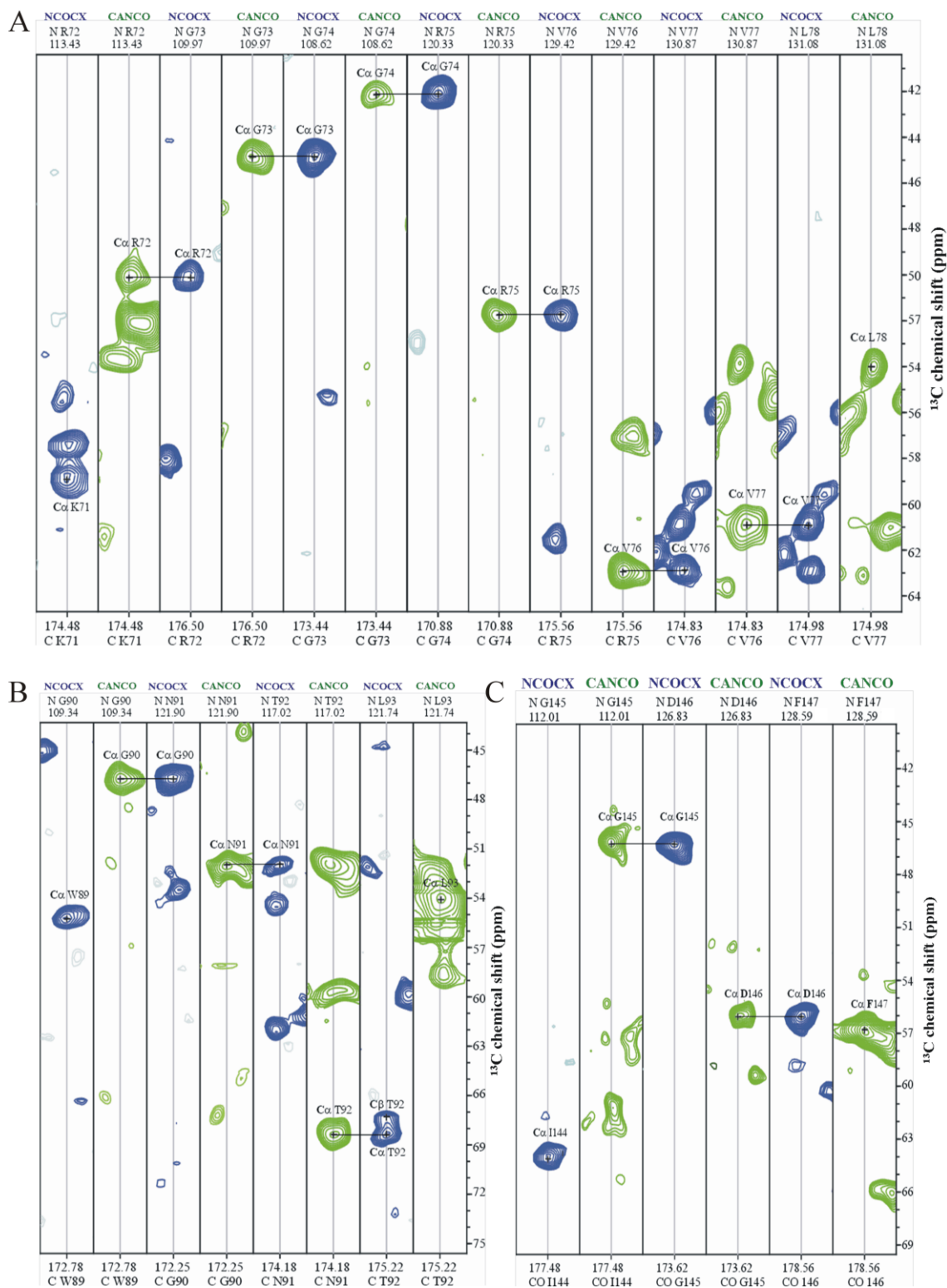
**Figure 4.17.** Examples of the sequential walk connecting residues 71-78 (A), 89-93 (B), and 144-147 (C). The representation is the same as in Figure 4.16, but for sake of clarity, only the NCOCX and CANCO strips are shown.

As a result, 59 residues, distributed throughout each subunit, were sequence-specific assigned (Fig. 4.15A/B). The chemical shift index is consistent with their location on the expected secondary structure elements, while the use of ferritin expressed in [2-$^{13}$C]-labeled glycerol confirmed the chemical nature of the assigned residues.[69] The transfer of sequence-specific assignments derived from solid state data to the solution $^{13}$C-$^{13}$C NOESY is relatively straightforward if one takes into account the deuterium isotope shift[70-72] and overlays the spectra searching for the best superimposition of the aliphatic region (Fig. 4.18), which is characterized by a good resolution both in solution and in the solid state.



**Figure 4.18.** Superimposition of the aliphatic region of the ferritin $^{13}$C-$^{13}$C NOESY spectrum acquired in solution (blue trace) and of the solid state $^{13}$C-$^{13}$C DARR spectrum (red trace) permits the transfer of the solid-state assignments to solution data and vice versa. The two 2D maps were overlaid by looking for the best superimposition of the aliphatic part.

Our strategy is based on the recognition of side chain spin patterns. The several factors that affect the completeness of the assignment are *i*) the intrinsically low chemical shift dispersion related to the four-helix bundle structure of the monomers,[73] *ii*) the primary sequence of ferritin (Fig. 4.15A) that has large numbers of residues of the same type (the most abundant amino acids are Leu19, Glu18, Lys13, and Asp13, and account for 36% of the total amino acids); *iii*) the extremely narrow distribution in

[13]C chemical shifts within each of the most abundant amino acids, and within the Lys/Leu and Glu/Asp groups; *iv*) the difficulty in discriminating between Glu and Gln and between Asp and Asn. More extensive assignments will require repetition of the experiments using samples prepared with selective 'unlabelling' of different amino acids,[74] and is beyond the scope of this report. Nevertheless, the present results represent a large step forward in the assignment of big molecular complexes.

The location of the assigned residues on different structural elements (Fig. 4.15A/B) makes the group of 59 identified residues structurally representative. Many of the assigned residues (62%) are on protein loops, because of the intrinsically lower chemical shift dispersion characteristic of amino acids located on the α-helices.[73] However residues assigned on α-helices are distributed throughout the subunit: helix 1 - residues 10, 11, 14-16, 26, 27, 31, 32, helix 3 - 117, 118, 119; helix 4- 141, 144-146, 148-151, and helix 5 -, 161, 163, 167-169. The set of signals for the sequence assigned amino acids was used to monitor the fate of the iron(III) products in solution after iron(II) is added and reacts with dioxygen at the active sites.

When iron(II) and oxygen substrates are converted to ferric species in ferritin, selective changes are observed of the [13]C-^{13}C solution NOESY (Fig. 4.19).
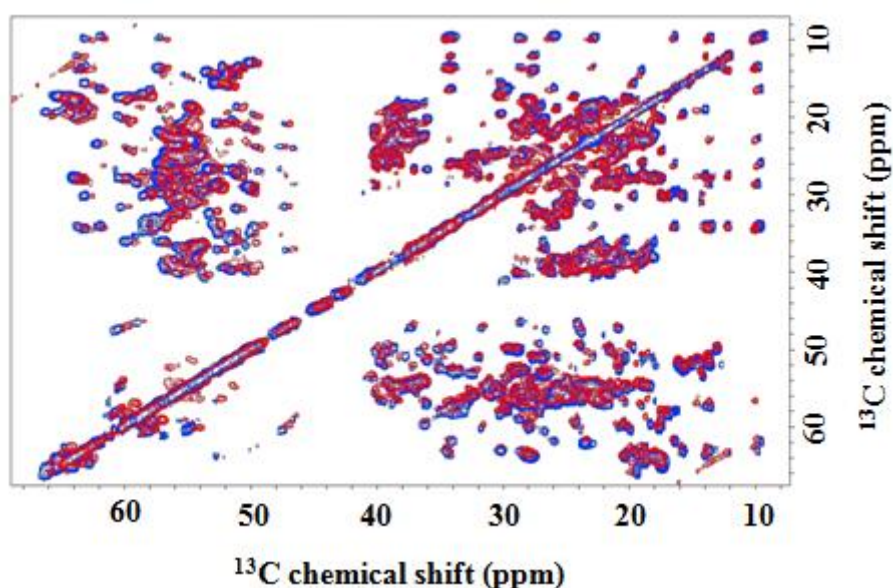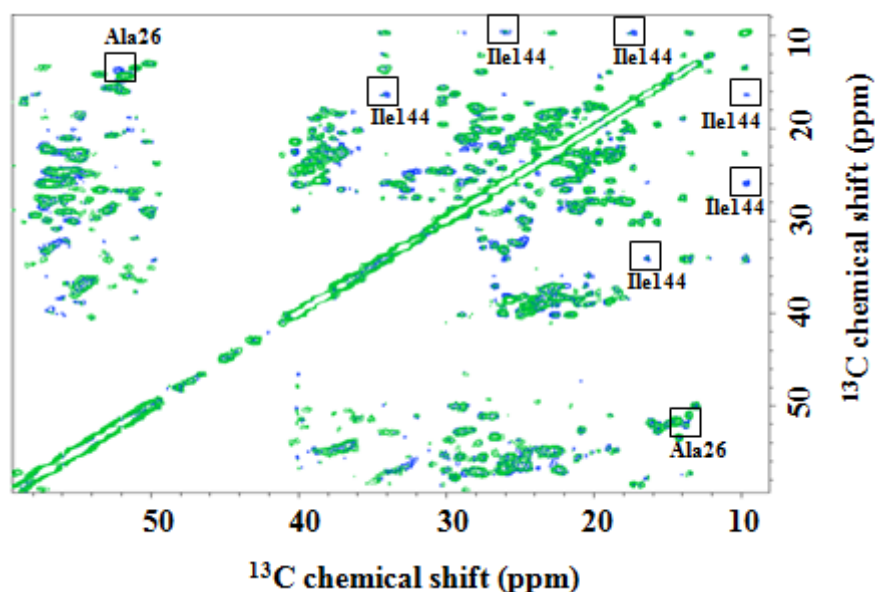


**Figure 4.19.** Superimposition of the aliphatic region of the ferritin [13]C-^{13}C NOESY spectra acquired in solution before (blue) and after (green) the addition of 1 equivalent of iron (II) (2 iron(II)/active site; 48 iron(II)/nanocage); resonances disappearing upon formation of iron (III) products are labeled.

The titration used saturating amounts of ferrous substrate for each catalytic cycle, beginning with the addition of 48 atoms of iron(II) per ferritin nanocage, i.e., 2 iron(II) ions per ferroxidase site. The iron(II) ions and oxygen substrates are immediately converted into iron(III) species, which is accompanied by the disappearance of some resonances in well resolved spectral regions, well outside the noise level in the spectrum of the apoprotein. This behavior is consistent with the presence of paramagnetic species that cause broadening beyond detection of resonances of nearby residues.[75] Except for the disappearance of a few cross peaks, the $^{13}$C-$^{13}$C NOESY spectra are identical with and without iron(III), indicating localized paramagnetic effects. Good indicators of the location of the product are provided by the well resolved side chain peaks of isoleucine residues; all four of the isoleucines present in ferritin are assigned. Among the four isoleucine residues, only the resonances of the side chain of Ile144 disappeared at this stage of the titration, indicating that the ferric products leave the catalytic site and migrate towards a new site close to Ile144, which has a side chain protruding towards the interior of the helix bundle. Ile141, by contrast, adjacent to the ferroxidase site and unaffected at this stage of the titration, has a side chain directed towards the exterior of the bundle, towards the internal cavity of the cage. Paramagnetic broadening is also induced in Ala26, another residue near the active site with a side chain pointing toward the interior of the bundle. These effects suggest that the ferric products leave the catalytic site and move inside the 4-helix bundle, from the ferroxidase site towards the short fifth helix at the C-terminus of each subunit (Fig. 4.20A). Addition of another two iron ions per active site for a total of 96 per nanocage, affects a larger number of residues. Besides Ile144 and Ala26, Trp89, Thr92, Ile148, Leu156 are also broadened beyond detection (Fig. 4.20A).
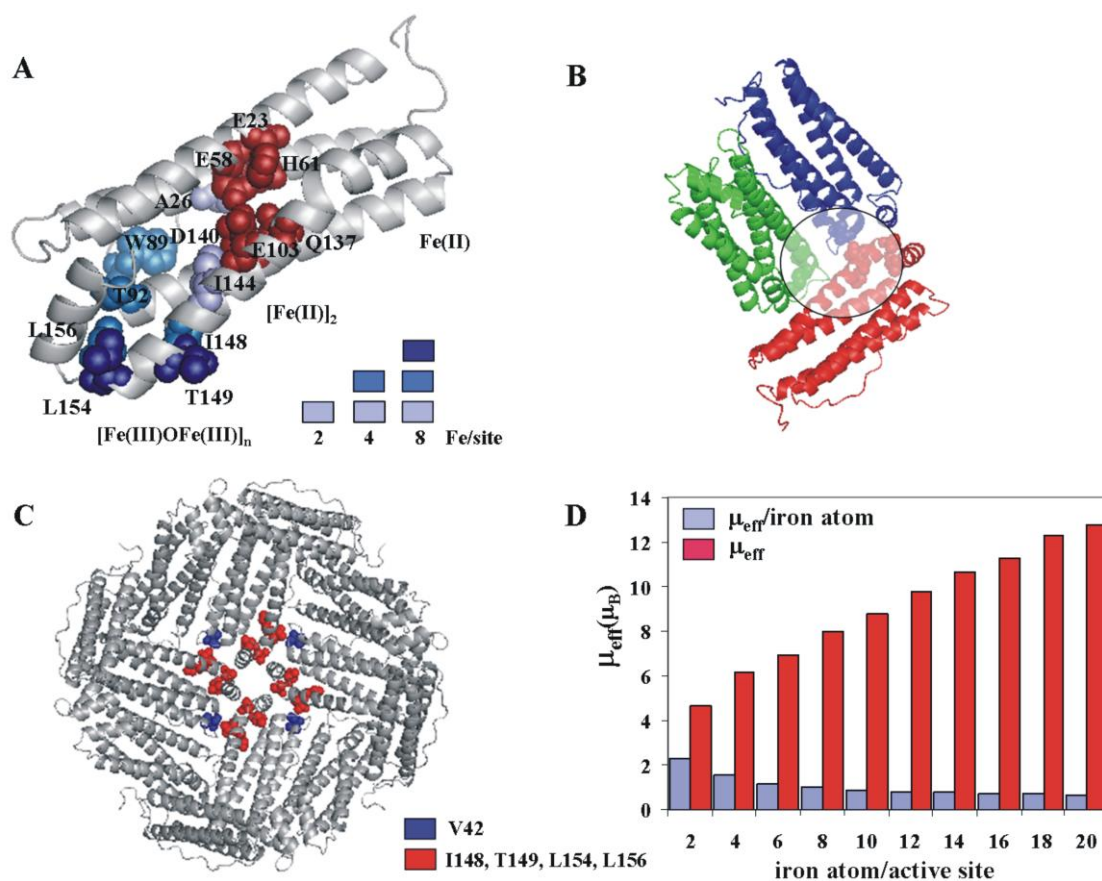
**Figure 4.20.** Tracing the iron channel in ferritin by paramagnetic effects. (A) NMR resonances that disappear in $^{13}C$-$^{13}C$ NOESY solution spectra, as the iron:protein ratio increases, are mapped onto the ribbon structure of a ferritin subunit as colored spheres (1 equivalent of iron = 2 iron(II)/active site = 48 iron(II)/nanocage). 1 equivalent: light blue, 2 equivalents: blue, 4 equivalents: dark blue. (B) View of the internal surface of the ferritin protein cage showing the relative spatial relationship near channel exits into the cavity; iron(III) products emerging from the channel of the red subunit have paramagnetic effects on residue Val42 in the blue subunit and on Arg72 and Gly74 in the green subunit, after 4 equivalents of iron are added. Paramagnetically broadened residues are shown as spheres. (C) Ferritin channel exits in four adjacent subunits surround the 4-fold axes of the ferritin protein cage: Val42 (blue spheres); Ile148, Thr149, Leu154, and Leu156 (red spheres). (D) Effective magnetic moments per subunit ($\mu_{eff}$, red bars) and average magnetic moment per iron atom ($\mu_{eff}$/iron atom, blue bars) obtained by Evans measurements at increasing concentrations of iron.

The side chains of these residues, as for Ala26 and Ile144, are located inside the helix-bundle, but farther from the ferroxidase site. Adding a total of four more iron ions per active site, in two additions, for a total of 192 iron atoms per nanocage, results in the disappearance of additional resonances, e.g. Thr149 and Leu154 (Fig. 4.20A). These residues, however, are located outside the helix bundle at the junction of the channel with the inner surface of the mineralization cavity and probably represent the gate for iron exit from the channel into the internal protein cavity. Thr149 and Leu154 are also at interfaces with other subunits, and, at this level of the titration, other residues in close contact to iron(III) complexes are in different subunits. For example the resonances of Val42 in one subunit, which is almost in contact with Thr149 on a perpendicular subunit (Fig. 4.20B), disappear. The relative position of the subunits in the protein architecture is such that multiple exits are close enough to nucleate mineralization inside the central compartment through fusion of the emerging products (Fig. 4.20C).

The first addition of iron (48/nanocage; 2/active site) causes no hyperfine shift in any of the resolved resonances, consistent with the presence of iron(III) species characterized by low anisotropy, which are incapable of influencing nuclei at distances larger than those broadened beyond detection. With 96-192 iron atoms a few resonances experience small chemical shifts ($\leq 0.2$ ppm). These signals belong to amino acids that have some signals broadened by paramagnetic effects or that are close in space to amino acids with signals broadened beyond detection.

To explore the aggregation state of the ferric products in ferritin after oxidation of iron(II), bulk magnetic susceptibility in solution, was measured at different steps in the iron titration (Fig. 4.20D, Table 4.1), by the Evans method. [76,77]

| N° of iron atoms per ferritin subunit | $\chi_M$ (m$^3$ mol$^{-1}$) | $\mu_{eff}$ ($\mu_B$) | $\mu_{eff}$ per iron atom ($\mu_B$) |
|---|---|---|---|
| 2 | $1.12 \pm 0.11\ 10^{-7}$ | $4.49 \pm 0.22$ | $2.24 \pm 0.11$ |
| 4 | $1.75 \pm 0.18\ 10^{-7}$ | $5.61 \pm 0.28$ | $1.40 \pm 0.07$ |
| 6 | $2.61 \pm 0.26\ 10^{-7}$ | $6.85 \pm 0.34$ | $1.14 \pm 0.06$ |
| 8 | $3.43 \pm 0.35\ 10^{-7}$ | $7.85 \pm 0.39$ | $0.98 \pm 0.05$ |
| 10 | $4.38 \pm 0.44\ 10^{-7}$ | $8.87 \pm 0.45$ | $0.89 \pm 0.04$ |
| 12 | $5.32 \pm 0.54\ 10^{-7}$ | $9.78 \pm 0.49$ | $0.81 \pm 0.04$ |
| 14 | $6.40 \pm 0.65\ 10^{-7}$ | $10.70 \pm 0.54$ | $0.76 \pm 0.04$ |
| 16 | $7.37 \pm 0.74\ 10^{-7}$ | $11.50 \pm 0.58$ | $0.72 \pm 0.03$ |

**Table 4.1.** Magnetic properties (per subunit) of iron-loaded ferritin at different degrees of metallation, as deduced from Evans measurements at 298 K.

The species formed upon addition of the first two irons per active site has an effective magnetic moment $\mu_{eff}$ of 4.49 ± 0.22 $\mu_B$ i.e., significantly lower than expected for two non-coupled high-spin iron(III) centers (4.5 $\mu_B$ *vs.* 8.3 $\mu_B$) and indicative of the presence of an antiferromagnetically coupled dimer,[78,79] as also observed by Mössbauer spectroscopy.[31] The small magnetic susceptibility is consistent with a μ-oxo complex.[78] This species accounts for the observed paramagnetic line broadening on nearby residues (estimated to be within 5-6 Å). Addition of the second two irons per active site is associated with a $\mu_{eff}$ of 5.61 ± 0.28 $\mu_B$ i.e., an average $\mu_{eff}$ of 1.4 ± 0.07 $\mu_B$ per iron. The decrease in the average $\mu_{eff}$ per iron atom from 2.24 ± 0.11 $\mu_B$ to 1.40 ± 0.07 $\mu_B$ suggests the presence of a four-iron(III) cluster with antiferromagnetically coupled metal ions. The presence of four iron ions induces paramagnetic line broadening of a greater number of amino acids, as observed in Figure 4.20A. Further additions of pairs of iron atoms per active site are accompanied by a continual increase in total $\mu_{eff}$ and decrease in the average $\mu_{eff}$ per iron atom, as expected for antiferromagnetically coupled clusters of increasing nuclearity. The $\mu_{eff}$ of 7.85 ± 0.39 $\mu_B$ obtained with 8 iron/subunit is comparable to the value reported for the $Fe_8(\mu_4\text{-}O)_4$ cluster.[80]

Insights on the kinetics of products formation could be obtained with Evans measurements as a function of time. The magnetic moments in Figure 4.20D are equilibrium values (Table 4.2).

| N° of iron atoms per ferritin subunit | $t_{stab}$ (hours) | $\mu_{eff}$ per iron atom ($\mu_B$) t=0 | $\mu_{eff}$ per iron atom ($\mu_B$) t=$t_{stab}$ |
|---|---|---|---|
| 2 | 1.5 | 2.36 ± 0.12 | 2.24 ± 0.11 |
| 4 | 24 | 1.56 ±0.08 | 1.40 ± 0.07 |
| 6 | 24 | 1.24 ± 0.06 | 1.14 ± 0.06 |
| 8 | 24 | 1.05 ± 0.05 | 0.98 ± 0.05 |
| 10 | 24 | 0.93 ± 0.05 | 0.89 ± 0.04 |
| 12 | 24 | 0.85 ± 0.04 | 0.81 ± 0.04 |
| 14 | 24 | 0.78 ±0.04 | 0.76 ± 0.04 |
| 16 | 24 | 0.74 ± 0.04 | 0.72 ± 0.03 |

**Table 4.2.** Time-dependence of the $\mu_{eff}$ values per iron atom of iron-loaded ferritin at different degrees of metallation, observed by Evans measurements at 298 K.

Equilibration of the dimeric ferric products produced by the first equivalent of iron is fast (about 1.5 hours). As more iron(II) is added, there is an initial increase in magnetic susceptibility attributable to the formation of a new dimer, followed by a slow

decrease until a stable value is reached in about 24 hours. This behavior suggests that the first product at the catalytic site is always a dimer, which then interacts with other products already present within the 4-helix bundle channel.

### 4.7. Discussion

The 4-helix bundle structure of the ferritin subunits provides the scaffold that hosts the ferroxidase sites and also guides the catalytic products towards the mineralization site. Based on the observed paramagnetic effects on residues lining the internal face of the 4-helix bundle, the functional channel corresponds to the interior of the 4-helix bundle. Iron catalytic products proceed along the 20 Å –long path inside the 4-helix bundle of each subunit, roughly parallel to the protein surface, until they reach the C-terminal edge, where the short fifth helix, angled about 60° from the axes of the bundle, creates a kink that may guide the iron(III) clusters into the mineralization cavity. The iron channel we observed is much longer than current models which predict a short hop across the protein cage, between the active site and the cavity.

Formation and migration of ferric products at the active sites of the 24 subunits appear to be synchronized in all the subunits at all titration steps. After the first catalytic turnover (2Fe/catalytic center), the iron(III) complexes have only moved a few angstroms into the channel from the active centers. Two Ile residues are close to the active site, Ile141 and Ile144, with Ile141 closer to the metal ions at the catalytic center. (Distances between the metal ion coordinated to Asp140/Gln137/Glu103 and Ile141 nuclei are C$\alpha$ 7.7 Å, C$\beta$ 8.0 Å, C$\delta$1 7.6 Å, C$\gamma$1 7.1 Å, whereas distances between the same metal ion and I144 are C$\alpha$ 10.2 Å, C$\beta$ 9.2 Å, C$\delta$1 7.1 Å, C$\gamma$1 8.2 Å; PDB 1MFR). If the iron(III) ions produced at the catalytic site maintained the same coordination of the ferrous substrate, Ile141 should be more affected by the paramagnetic effects. However, resonances of Ile141 are unchanged while those of Ile144, further from the active site, are broadened beyond detection. A possible driving force for the change in the coordination environment of the iron upon oxidation may be interaction with Tyr30 and His54; distances between the metal ion coordinated to Asp140/Gln137/Glu103 in the active site and potential donor atoms of these residues are: Tyr30 O$\eta$ 5.9 Å; His54 N$\delta$1 4.8 Å and N$\delta$2 6.1Å (PDB 1MFR). Migration of the iron(III) products from the active site, which as we now know is into the long channel of each subunit, is required

to complete turnover at the active site. In the channel, a coordination site with different affinities for substrate ferrous (lower than the active site) and ferric product (higher than the active site) is needed for the migration of the diferric species. This mechanism is consistent with the reported tenfold reduction in iron oxidation rate upon substitution of Tyr with Phe at position 30.[81] When a second equivalent of iron(II) is added (2 irons/active site), NMR showed paramagnetic species penetrating further into the subunit channels. In addition, the resonances of Ala26 and Ile144 remained broadened beyond detection suggesting that the second cycle of diferric products are at the same site of the diferric products from the first catalytic cycle. The ferritin iron channels contain the products of up to 4 catalytic turnovers, before iron reaches the central cavity. Equilibrium among different coordination states, with comparable and possibly increasingly lower affinity, is needed for the migration of the iron(III) species, when new diferric products are formed and released at the active site.

The magnetic susceptibility data suggest interactions between diferric reaction products in the protein channels, and indicate iron aggregation in the protein cage. Newly formed iron(III) dimers from the active site, upon addition of more iron(II) equivalents, react with earlier iron(III) products, on a time scale of the order of several hours, possibly by hydrolysis of water coordinated to ferric ions. The time required may reflect the slow kinetics of ligand exchange typical of iron(III), and may be ascribed to the establishment of interactions, although weak, between the ferric products  and the several potential ligands present in the channel. In addition to the already discussed Tyr30 and His54, other potential ligands are Met33, Thr92, Met96, Glu163, and Ser170.

The exits of the ferritin channels, where iron(III) catalytic products migrate from active sites to the cavity are close enough, one to the other, to promote fusion/hydrolysis of the emerging multimeric iron(III) complexes into larger biomineral. The multisubunit structure of ferritin and the symmetry properties of the protein nanocage, thus, are essential elements for the concentration of iron products in the nanocage cavity.

By flash-freezing ferritin crystals after 30' of soaking in concentrated solutions of iron(II) under aerobic conditions, we have been able to detect iron(III) product-dimers inside the 4-helix bundle channel. Crystallographic data collected at the Zn-edge resulted in a 2.7 Å resolution structure. Anomalous difference electron density maps acquired on such crystals invariably and clearly present two iron(III) ions in a coordination environment similar, but not identical, to that for the magnesium-bound form (Fig. 4.21).
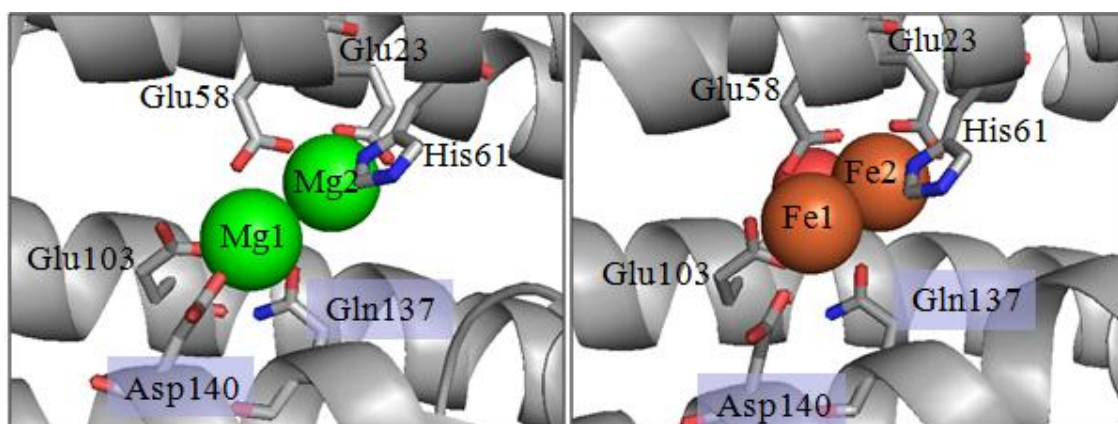
**Figure 4.21.** Ferroxidase sites of BF M-ferritin with Mg (left) and Fe (right).

As summarized in Table 4.3, each magnesium ion in the Mg2-ferritin is bound to four donor atoms from protein residues. For the so-called Mg(1) they are Oε1 and Oε2 of Glu23 (which acts a bidentate), the Nδ1 of His61, and the Oε1 of Glu58. The latter residue bridges the two magnesium ions, coordinating Mg(2) through its Oε2 atom. Mg(2) also binds to Asp140 Oδ2, Glu103 Oε1, and Gln137 Oε1. No solvent molecules are reported inside the 4-helix bundle channel, although most probably water molecules should complete the coordination sphere of the two magnesium ions.

In the 30'-ferric-ferritin, Fe(1) binds to Oε1 of Glu23 (which acts a monodentate ligand), to Nδ1 of His61, and to Oε1 of Glu58, which is again acting as a bridging ligand between the two metal ions; Fe(2) binds to Oε2 of Glu58, and to Glu103, which now becomes an asymmetric bidentate ligand. On the other hand distances with Gln137 Oε1 and Asp140 Oδ2 and Oδ1 are all too long for coordination bonds. In the first coordination sphere of the iron ions a bridging oxygen is present in 23 out of 24 subunits, and apical oxygen atoms on both Fe(1) and Fe(2) are found in 15 subunits and in 24 subunits, respectively. The metric Gln137 Oε1 is at a distance which is too long to give rise to a coordination bond (3.4 Å) and too short to imply a H-bonded water molecule. Analogous considerations applies to Asp140, whose Oδ2 is within H-bonding distance from the apical water bound to Fe(2)( 2.5 ± 0.9 Å).

No further ligands are detected in the present structure, although expected to achieve the hexacoordination typical for iron(III). In addition, spectroscopic data indicate that an oxo/hydroxo bridge should be present in the ferric dimers products. These consideration can be reconciled with the present structural data according to the

two following hypothesys: *(i)* the bridging oxygen belongs to an hydroxide and Gln137 acts as a loosely bound bridging ligand in trans position with respect to the bridging Glu58 or *(ii)* a μ-oxo-bridging ligand is present but not detectable with the actual resolution. The observed iron-iron distance of 3.2 Å is typical for (μ-oxo)bis(μ-carboxylato)diiron(III) compounds.

The six coordination sphere assumed by the two iron ions described above would account for the detachment from the Glu23 Oε2 (for Fe(1)) and Asp140 Oδ2 and possibly Gln137 Oε1 (for Fe(2)), which are the protein ligands of magnesium.

In any case, the Fe(2) ion is systematically and significantly displaced with respect to Mg(2) by more than 1.2 Å. These changes in coordination bonds translate into a movement of the dimetal center away from helix α4 and towards helices α1 and α2.

These structural data are fully consistent with the published solution NMR results. Ile141, adjacent to the Mg(2) ligands Gln137 and Asp140, is not affected by any paramagnetic broadening in the NMR spectra of the diiron species. On the other hand, the resonances of the side chain of Ile144, which is farther from the magnesium ions, disappear at this stage of the titration. This observation lead us to propose that the iron(III) ions produced at the catalytic site should have a different coordination with respect to the magnesium structure. If iron(III) would simply replace magnesium(II), indeed, Ile141 should be more affected by the paramagnetic effects. On the contrary, resonances of Ile141 are unchanged while those of Ile144, farther away from the magnesium binding site, are broadened beyond detection.

We propose here that this situation represents the ferroxidase site in the product conformation, where the ferric dimers are found in a coordination geometry that is better suited for their higher oxidation state.

The main differences between the reagent and the product conformation are mainly due to a translation of Fe(2) that, in its movement, gets closer to Tyr30 (Fe(2)-Tyr30 Oη about 5 Å vs. Mg(2)-Tyr30 Oη about 6 Å), a residue proposed to play a key role for the kinetics of the iron oxidation. This movement is accompanied by some reorientation of the side chain of Asp140, which also shows different conformation is the different subunits (see in Table 4.3 the larger errors with respect to the other amino acids for the distances measured from the two iron ions). On the other hand the protein backbone structure is essentially identical in the Mg2 and Fe2 structures (RMSD of 0.38

Å superimposing the 24-mer and 0.26 ± 0.01 Å superimposing the single subunits). This movement may provide room for the incoming new iron(II) substrates.

| | Mg(1) | Mg(2) | Fe(1) | Fe(2) |
|---|---|---|---|---|
| Glu23 Oε1 | 2.4 | | 2.3±0.2 | |
| Glu23 Oε2 | 2.5 | | 3.2±0.2 | |
| Glu58 Oε1 | 2.0 | 3.7 | 2.1±0.1 | 3.3±0.1 |
| Glu58 Oε2 | 3.7 | 2.5 | 3.4±0.2 | 2.0±0.1 |
| His61 Nδ1 | 2.7 | 4.7 | 2.8±0.1 | 4.5±0.1 |
| Glu103 Oε1 | | 2.9 | | 2.5±0.2 |
| Glu103 Oε2 | | 3.2 | | 2.4±0.2 |
| Gln137 Oε1 | 3.3 | 2.7 | 3.4±0.1 | 3.4±0.1 |
| Asp140 Oδ2 | | 2.5 | | 3.7±0.4 |
| | | | | |
| Fe-oxygen bridging OH/$H_2O$ | | | 2.8±0.2 | 2.7±0.2 |
| Fe-oxygen $H_2O(1)$ | | | 2.9±0.5 | |
| Fe-oxygen $H_2O(2)$ | | | | 2.9±0.4 |
| **Mg(1)-Mg(2)** | 4.0 | | | |
| **Fe(1)-Fe(2)** | | | 3.2±0.1 | |

**Table 4.3.** Summary of the main structural features of the Mg2 and Fe2 bound forms of M ferritin: distances (Å) from potential donor atoms and inter-metal distances (Å) are reported for the magnesium and iron ions in the two crystal structures. The reported errors provide a measure of the differences within the 24 subunits of the Fe2 structure. The Mg2 structure consists of 24 identical subunits.

Crystal structures of the diferric products of crystal obtained with different soaking times (30', 60', 90', 120' before flash freezing) were solved and refined, although at lower resolution (3.7-4.1 Å). However, anomalous difference maps clearly

show the ferroxidase site always occupied by two iron ions. Although they reasonably represents different stages of catalytic turnover process leading to the production of increasing amounts of oxidation products, the features of the coordination environment of this pair of iron(III) ions are very similar and invariably indicate the translation of the diferric center towards the α1 and α2 helices.

### 4.8. Experimental section

#### 4.8.1. Protein expression and purification

Frog (*R. catesbieana*) M ferritin protein was expressed from a pET3a plasmid.[81] Uniformly $^{13}$C, $^{15}$N-labelled protein for solid state, and $^{2}$H, $^{13}$C, $^{15}$N-labelled protein (deuteration > 90%) for solution studies were expressed in minimal medium with label algal hydrolysate as described in.[67] Unlabelled protein was used for Evans magnetic susceptibility measurements. A ([2-$^{13}$C] glycerol, $^{15}$N) ferritin sample was prepared according to the published procedure.[64] The crystals were grown by the sitting drop method: 2 μl of a 14 mg/ml protein solution was mixed with an equal volume of 3 M Na formate. The drop was equilibrated with 3 M Na formate at 18 °C. Trigonal crystals formed within a week.

#### 4.8.2. Protein crystallization for solid state NMR

A reaction tube was filled with 1 mL of a solution of PEG 3350 at 25% w/v in Na$_2$CO$_3$ 100 mM pH 6.5. A 0.2 mL drop of 40 mg/mL ferritin was deposited on the upper surface of this solution. The solution was kept at room temperature for 24 hours before the precipitate was separated by centrifugation. In this form, about 50 mg protein was transferred into a 4-mm MAS rotor.

### 4.8.3. Protein crystallization for X-ray crystallography

Anomalous diffraction data have been collected at ESRF beamline BM14-1. Before data collection Bull frog (BF)-ferritin crystals have been soaked on a solution at pH 6.5 containing 10 mM $FeSO_4$, 3 M Na formate under aerobic conditions and flash frozen in liquid nitrogen after 30-60-90-120 minutes soaking time. Anomalous diffraction data on each crystal have been collected both at the Fe K-edge (white line at 1.737 Angstrom) and below the Fe K-edge (at 1.759 Å) in order to exclude any other anomalous scatterer besides Fe. Each crystal provided a complete dataset to 3.7 Å resolution.

Datasets at 2.7 Å resolution have been collected on ESRF beamline ID14-4 at the Zn K-edge (1.292 Å). All structures have been solved by molecular replacement in the trigonal space group P31 2 1 (space group n. 152, cell parameters: a=210.23, b=210.23, c=320.76), manually rebuilt and refined with REFMAC5 (CCP4 suite). The crystal asymmetric unit contains a whole BF-ferritin molecule (24 subunits), thus allowing us the analysis of 24 independent subunits for each crystal.

Anomalous difference maps computed with the refined phases, unambigously show the presence of the Fe ions bound to BF-ferritin at different sites.

Iron(II) at the catalytic site of ferritin is immediately oxidized to iron(III), producing a colored protein form with a typical electronic absorption spectrum. The same color was observed in our crystals. However, photoreduction of ferric ions due to synchrotron radiation is always possible. The effect of the synchrotron radiation at the Zn K-edge on ferric ions bound at the catalytic site of the protein was checked acquiring EXAFS data at ESRF beamline ID14-4 at the Zn K-edge.

### 4.8.4. Solid state MAS NMR spectra

All the NMR experiments were performed at a field of 16.4 T (700 MHz [1]H Larmor frequency, 176.0 MHz [13]C Larmor frequency) on a Bruker Avance 700 wide-bore spectrometer equipped of a 4 mm CP-MAS probe. All the experiments were acquired at MAS frequency of 9 kHz. Accurate determination of sample temperatures in solid state NMR with MAS is known to be problematic, mainly due to frictional and RF heating that make the internal sample temperature significantly different from the

measured temperature outside the MAS rotor.[82] In order to define the best temperature for our measurements we acquired preliminary data over the 273-298 K range of measured temperature values outside the MAS rotor. The best resolution, signal to noise, and agreement with solution chemical shift values occurred for data acquired at a measured temperature of 298 K. Standard sequences were used. In all experiments a ramped cross polarization (CP) from protons was used.[65] The relaxation delays were of 3-4 s. Acquisition times in the direct dimension were in the range of 15-35 ms. For 2D $^{13}$C-$^{13}$C correlation experiments, PDSD and DARR were employed for $^{13}$C-$^{13}$C mixing.[83,84] For the $^{15}$N-$^{13}$C correlation experiments (NCO providing $N_iC'_{i-1}$ correlations and NCA providing $N_iC\alpha_i$ correlations) a double CP [85] sequence was utilized, in which the $^{15}$N-$^{13}$C transfer was achieved using the TAM pulse on $^{13}$C channel. The above mentioned building blocks (NCO or NCA and DARR) were combined to obtain the 3D $N_i$-$C\alpha_i$-$CX_i$ (NCACX) and 3D $N_i$-$C'_{i-1}$-$CX_{i-1}$ (NCOCX) experiments[66] Finally a 3D $C\alpha_{i+1}$- $N_{i+1}$-$C'_i$ (CANCO) experiment[66] was acquired. For 2D DARR experiments, mixing times of 20 and 40 ms were used. The $^1$H radiofrequency field strength during mixing was matched to the MAS speed to satisfy the n=1 condition. Each 2D $^{13}$C-$^{13}$C spectrum was acquired with 32 scans, 3744 points in the direct dimension, and 1024 experiments in the indirect dimension and spectral widths of $305 \times 305$ ppm. The 2D NCO and NCA experiments were acquired with 256 scans, 512 points in the direct dimension, and 160-256 experiments in the indirect dimension and spectral widths of $149 \times 141$ ppm respectively. The 3D NCACX and 3D NCOCX were recorded with 32 scans, 2710 points in the direct dimension, 80 experiments in the $^{15}$N indirect dimension, and 64 in the $^{13}$C indirect dimension (spectral widths of $354 \times 45 \times 20$ ppm). Both were acquired with a 22 ms DARR mixing time. The 3D CANCO was recorded with 128 scans for each experiment, 2710 points in the direct dimension, 32 experiments in the $^{15}$N dimension and 54 in the $^{13}$C indirect (spectral widths of $354 \times 45 \times 30$ ppm). All pulse sequences were implemented with 100 kHz spinal64 $^1$H decoupling [84] during indirect evolution and acquisition periods.

Spectra were processed with the program TopSpin (Bruker). The 2D and 3D maps were analyzed with the CARA program.[86]

*4.8.5. Solution NMR spectra*

**$^{13}$C-$^{13}$C NOESY.** Ferritin solutions, 3 mM in subunits (125 µM in nanocages) in 100 mM MOPS, pH 7.0, were used to acquire $^{13}$C-$^{13}$C NOESY maps[54,67] at increasing iron concentrations. Freshly prepared solutions of ferrous sulfate in 1 mM HCl were added to the protein solution at a ratio of 2 iron atoms/active site (48 iron atoms/protein nanocage). The conditions were comparable to those used previously for solution Mössbauer and EXAFS spectroscopies[24,25] which showed complete and fast oxidation of iron. In order to facilitate homogenous diffusion of iron and oxygen, protein samples were titrated according to the following procedure: *i*) taking out the sample from the NMR tube; *ii*) adding the iron(II) solution and gently pipetting up and down the entire solution; *iii*) spinning the solution. EXAFS data (to be published) confirmed that the iron products obtained with this procedure were all in the iron(III) oxidation state. Our samples were incubated at 298 K for 2 days before collecting the $^{13}$C-$^{13}$C NOESY spectra. After measurements were made on the sample with 48 iron atoms/nanocage, a second aliquot of 2 iron atoms/active site was added and the process repeated until 192 iron atoms/nanocage were analyzed through the $^{13}$C-$^{13}$C NOESY. The $^{13}$C-$^{13}$C NOESY maps at each step of the titration were acquired on a 16.4 T Bruker AVANCE 700 spectrometer equipped with a triple-resonance TXO cryoprobe optimized for $^{13}$C direct-detection, at 298 K, using two different mixing times: 500 ms on the full spectral width (200 ppm), for 18 h, and 150 ms on the aliphatic region (90 ppm) in order to increase resolution, for 16 h to achieve good signal-to-noise ratios.

*4.8.6. Magnetic susceptibility measurements*

The magnetic susceptibility of ferritin with increasing amounts of iron was measured by the modified Evans method[76,77] as previously reported.[87] Coaxial NMR tubes were used with *tert*-butyl alcohol and 1,4-dioxane as internal references. The paramagnetic and diamagnetic protein samples were prepared from the same stock solution i.e., 320 µM in subunit monomer (13.3 µM in protein nanocages) in 100 mM MOPS, pH 7 with 5 mM ter-butyl alcohol and 5 mM 1,4-dioxane. The protein solution was split into two aliquots. The inner capillary contained the diamagnetic apoferritin solution. Freshly prepared solutions of ferrous sulfate in 1 mM HCl were added to the

second aliquot of ferritin in steps of 2 iron/active site (48 iron atoms /protein nanocage) in the outer tube. Addition of iron to the protein solution followed the procedure described for the $^{13}$C-$^{13}$C NOESY. Ten additions of two iron(II)/subunit (up to a total of 20 iron atoms/subunit) were analyzed. The shifts of the proton signals of the two reference molecules were measured at on the 16.4 T Bruker AVANCE 700 spectrometer (the same used for $^{13}$C-$^{13}$C NOESY) at 298 K. The set of Evans measurements was repeated twice. An additional set was measured on the 18.7 Bruker AVANCE 800 spectrometer. The inner-outer tube peak separation ($\Delta\delta$ expressed in ppm) for each standard was measured and assigned to the bulk susceptibility shift. The paramagnetic contribution to the molar susceptibility of the solute ($\chi_M^{PARA}$) was related to the bulk susceptibility shift $\Delta\delta$ as indicated in the following equation:

$$\Delta\delta = C_M \, \chi_M^{PARA} / 3 \tag{4.17}$$

in which $C_M$ is the mM concentration of the protein and $\chi_M^{PARA}$ is given in m$^3$ mol$^{-1}$. The magnetic moment in solution ($\mu_{eff}$) was then calculated according to equation (4.18):

$$\mu^2_{eff} = \chi_M^{PARA} \, 3 \, k \, T / N_A \, \mu_0 \tag{4.18}$$

where k is the Boltzman constant, T is absolute temperature, $N_A$ is Avogadro's number, and $\mu_0$ is the vacuum permeability.

### 4.9. Conclusion

NMR revealed functional iron channels in ferritin that facilitate hydrolysis and guide the multinuclear ferric mineral precursors to the biomineralization cavity.

The NMR approach of combining solid state experiments for sequence specific assignment of key residues and solution $^{13}$C-$^{13}$C NOESY is a strategy that can provide chemical information at atomic resolution on the mechanism of action of other large molecular assemblies, in addition to the ferritin nanocage.

The X-ray structure of ferritin iron loaded showed the ferroxidase site in the product conformation.

Reference List

1.  Pintacuda, G. *et al.* Solid-state NMR of a paramagnetic protein: assignment and study of human dimeric oxidized Cu(II), Zn(II) superoxide dismutase. *Angew. Chem. Int. Ed.* **46**, 1079-1082 (2007).

2.  Zhao, G., Arosio, P. & Chasteen, N. D. Iron(II) and hydrogen peroxide detoxification by human H-chain ferritin. An EPR spin-trapping study. *Biochemistry* **45**, 3429-3436 (2006).

3.  Harrison, P. M. & Arosio, P. The ferritins: molecular properties, iron storage function and cellular regulation. *Biochim. Biophys. Acta* **1275**, 161-203 (1996).

4.  Bou-Abdallah, F. *et al.* Facilitated Diffusion of Iron(II) and Dioxygen Substrates into Human H-Chain Ferritin. A Fluorescence and Absorbance Study Employing the Ferroxidase Center Substitution Y34W. *J. Am. Chem. Soc.* **130**, 17801-17811 (2008).

5.  Galaris, D. & Pantopoulos, K. Oxidative stress and iron homeostasis: mechanistic and health aspects. *Crit Rev. Clin. Lab Sci.* **45**, 1-23 (2008).

6.  Chiancone, E., Ceci, P., Ilari, A., Ribacchi, F. & Stefanini, S. Iron and proteins for iron storage and detoxification. *Biometals* **17**, 197-202 (2004).

7.  Ferreira, C. *et al.* Early embryonic lethality of H ferritin gene deletion in mice. *J. Biol. Chem.* **275**, 3021-3024 (2000).

8.  Hintze, K. J. & Theil, E. C. DNA and mRNA elements with complementary responses to hemin, antioxidant inducers, and iron control ferritin-L expression. *Proc. Natl. Acad. Sci. U. S. A* **102**, 15048-15052 (2005).

9.  Pereira, A. S. *et al.* Direct spectroscopic and kinetic evidence for the involvement of a peroxodiferric intermediate during the ferroxidase reaction in fast ferritin mineralization. *Biochemistry* **37**, 9871-9876 (1998).

10. Waldo, G. S. & Theil, E. C. Formation of iron(III)-tyrosinate is the fastest reaction observed in ferritin. *Biochemistry* **32**, 13262-13269 (1993).

11. Zhao, G. *et al.* Iron and hydrogen peroxide detoxification properties of DNA-binding protein from starved cells. A ferritin-like DNA-binding protein of Escherichia coli. *J. Biol. Chem.* **277**, 27689-27696 (2002).

12. Kauko, A., Haataja, S., Pulliainen, A. T., Finne, J. & Papageorgiou, A. C. Crystal structure of Streptococcus suis Dps-like peroxide resistance protein Dpr: implications for iron incorporation. *J. Mol. Biol.* **338**, 547-558 (2004).

13. Crichton, R. R. & Declercq, J. P. X-ray structures of ferritins and related proteins. *Biochim. Biophys. Acta* **1800**, 706-718 (2010).

14. Santambrogio, P., Levi, S., Cozzi, A., Corsi, B. & Arosio, P. Evidence that the specificity of iron incorporation into homopolymers of human ferritin L- and H-chains is conferred by the nucleation and ferroxidase centres. *Biochem. J.* **314 ( Pt 1)**, 139-144 (1996).

15. Theil, E. C. Ferritin: structure, gene regulation, and cellular function in animals, plants, and microorganisms. *Annu. Rev. Biochem.* **56**, 289-315 (1987).

16. Andrews, S. C. *et al.* Structure, function, and evolution of ferritins. *J Inorg Biochem* **47**, 161-174 (1992).

17. Liu, X., Jin, W. & Theil, E. C. Opening protein pores with chaotropes enhances Fe reduction and chelation of Fe from the ferritin biomineral. *Proc. Natl. Acad. Sci. USA* **100**, 3653-3658 (2003).

18. Listowsky, I., Blauer, G., Enlard, S. & Betheil, J. J. Denaturation of horse spleen ferritin in aqueous guanidinium chloride solutions. *Biochemistry* **11**, 2176-2182 (1972).

19. Trikha, J., Theil, E. C. & Allewell, N. M. High resolution crystal structures of amphibian red-cell L ferritin: potential roles for structural plasticity and solvation in function. *J. Mol. Biol.* **248**, 949-967 (1995).

20. Hempstead, P. D. *et al.* Comparison of the three-dimensional structures of recombinant human H and horse L ferritins at high resolution. *J. Mol. Biol.* **268**, 424-448 (1997).

21. Bou-Abdallah, F. *et al.* mu-1,2-Peroxobridged di-iron(III) dimer formation in human H-chain ferritin. *Biochem. J.* **364**, 57-63 (2002).

22. Bou-Abdallah, F., Zhao, G., Mayne, H. R., Arosio, P. & Chasteen, N. D. Origin of the unusual kinetics of iron deposition in human H-chain ferritin. *J. Am. Chem Soc.* **127**, 3885-3893 (2005).

23. Zhao, G. *et al.* Multiple pathways for mineral core formation in mammalian apoferritin. The role of hydrogen peroxide. *Biochemistry* **42**, 3142-3150 (2003).

24. Pereira, A. S. *et al.* Direct spectroscopic and kinetic evidence for the involvement of a peroxodiferric intermediate during the ferroxidase reaction in fast ferritin mineralization. *Biochemistry* **37**, 9871-9876 (1998).

25. Hwang, J. *et al.* A short Fe-Fe distance in peroxodiferric ferritin: control of Fe substrate versus cofactor decay? *Science* **287**, 122-125 (2000).

26. Moenne-Loccoz, P. *et al.* The ferroxidase reaction of ferritin reveals a diferric mu-1,2 bridging peroxide intermediate in common with other O2-activating non-heme diiron proteins. *Biochemistry* **38**, 5290-5295 (1999).

27. Liu, X. & Theil, E. C. Ferritin reactions: direct identification of the site for the diferric peroxide reaction intermediate. *Proc. Natl. Acad. Sci. USA* **101**, 8557-8562 (2004).

28. Ha, Y., Shi, D., Small, G. W., Theil, E. C. & Allewell, N. M. Crystal structure of bullfrog M ferritin at 2.8 A resolution: analysis of subunit interactions and the binuclear metal center. *J. Biol. Inorg. Chem.* **4**, 243-256 (1999).

29. Toussaint, L., Bertrand, L., Hue, L., Crichton, R. R. & Declercq, J. P. High-resolution X-ray structures of human apoferritin H-chain mutants correlated with their activity and metal-binding sites. *J. Mol. Biol.* **365**, 440-452 (2007).

30. Mitic, N., Clay, M. D., Saleh, L., Bollinger, J. M., Jr. & Solomon, E. I. Spectroscopic and electronic structure studies of intermediate X in ribonucleotide reductase R2 and two variants: a description of the FeIV-oxo bond in the FeIII-O-FeIV dimer. *J. Am. Chem Soc.* **129**, 9049-9065 (2007).

31. Jameson, G. N. *et al.* Stoichiometric production of hydrogen peroxide and parallel formation of ferric multimers through decay of the diferric-peroxo complex, the first detectable intermediate in ferritin mineralization. *Biochemistry* **41**, 13435-13443 (2002).

32. Wade, V. J. *et al.* Influence of site-directed modifications on the formation of iron cores in ferritin. *J. Mol. Biol.* **221**, 1443-1452 (1991).

33. Granier, T. *et al.* Evidence of new cadmium binding sites in recombinant horse L-chain ferritin by anomalous Fourier difference map calculation. *Proteins* **31**, 477-485 (1998).

34. Mann, S., Williams, J. M., Treffry, A. & Harrison, P. M. Reconstituted and native iron-cores of bacterioferritin and ferritin. *J. Mol. Biol.* **198**, 405-416 (1987).

35. Michel, F. M. *et al.* Reactivity of ferritin and the structure of ferritin-derived ferrihydrite. *Biochim. Biophys. Acta* **1800**, 871-885 (2010).

36. Drits, V. A., Sakharov, B. A., Salyn, A. L. & Manceau, A. Structural model for ferrihydrite. *Clay Minerals* **28**, 185-207 (1993).

37. Michel, F. M. *et al.* The structure of ferrihydrite, a nanocrystalline material. *Science* **316**, 1726-1729 (2007).

38. Shi, H., Bencze, K. Z., Stemmler, T. L. & Philpott, C. C. A cytosolic iron chaperone that delivers iron to ferritin. *Science* **320**, 1207-1210 (2008).

39. Wang, Z. *et al.* Structure of human ferritin L chain. *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 800-806 (2006).

40. Yang, X., Arosio, P. & Chasteen, N. D. Molecular diffusion into ferritin: pathways, temperature dependence, incubation time, and concentration effects. *Biophys. J.* **78**, 2049-2059 (2000).

41. Bou-Abdallah, F., Arosio, P., Levi, S., Janus-Chandler, C. & Chasteen, N. D. Defining metal ion inhibitor interactions with recombinant human H- and L-chain ferritins and site-directed variants: an isothermal titration calorimetry study. *J. Biol. Inorg. Chem* **8**, 489-497 (2003).

42. Radisky, D. C. & Kaplan, J. Iron in cytosolic ferritin can be recycled through lysosomal degradation in human fibroblasts. *Biochem. J.* **336 ( Pt 1)**, 201-205 (1998).

43. Takagi, H., Shi, D., Ha, Y., Allewell, N. M. & Theil, E. C. Localized unfolding at the junction of three ferritin subunits. A mechanism for iron release? *J. Biol. Chem.* **273**, 18685-18688 (1998).

44. Jin, W., Takagi, H., Pancorbo, B. & Theil, E. C. "Opening" the ferritin pore for iron release by mutation of conserved amino acids at interhelix and loop sites. *Biochemistry* **40**, 7525-7532 (2001).

45. Gibbons, D. L. *et al.* Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus. *Nature* **427**, 320-325 (2004).

46. Frueh, D. P. *et al.* Non-uniformly sampled double-TROSY hNcaNH experiments for NMR sequential assignments of large proteins. *J. Am. Chem. Soc.* **128**, 5757-5763 (2006).

47. Wider, G. NMR techniques used with very large biological macromolecules in solution. *Methods Enzymol.* **394**, 382-398 (2005).

48. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated $T_2$ relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci. USA* **94**, 12366-12371 (1997).

49. Riek, R., Wider, G., Pervushin, K. & Wüthrich, K. Polarization Transfer by Cross-Correlated Relaxation in Solution NMR with Very Large Molecules. *Proc. Natl. Acad. Sci. USA* **96**, 4918-4923 (1999).

50. Bermel, W. *et al.* Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. *Angew. Chem. Int. Ed.* **44**, 3089-3092 (2005).

51. Bermel, W., Bertini, I., Felli, I. C., Kümmerle, R. & Pierattelli, R. [13]C direct detection experiments on the paramagnetic oxidized monomeric copper, zinc superoxide dismutase. *J. Am. Chem. Soc.* **125**, 16423-16429 (2003).

52. Bermel, W. *et al.* Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *J. Am. Chem. Soc.* **128**, 3918-3919 (2006).

53. Fischer, M. W. F., Zeng, L. & Zuiderweg, E. R. P. Use of [13]C-^{13}C NOE for the assignmetn of NMR lines of larger labeled proteins at larger magnetic fields. *J. Am. Chem. Soc.* **118**, 12457-12458 (1996).

54. Bertini, I., Felli, I. C., Kümmerle, R., Moskau, D. & Pierattelli, R. $^{13}$C-$^{13}$C NOESY: an attractive alternative for studying large macromolecules. *J. Am. Chem. Soc.* **126**, 464-465 (2004).

55. Neuhaus, D. & Williamson, M. *The nuclear Overhauser effect in structural and conformational analysis*. VCH, New York (1989).

56. *Encyclopedia of Magnetic Resonance*. Grant, D. M. & Harris, R. K. (eds.) (Encycl. of Nuclear Magn. Reson.,2002).

57. *Encyclopedia of Magnetic Resonance*. Grant, D. M. & Harris, R. K. (eds.) (Encycl. of Nuclear Magn. Reson.,1996).

58. Schneider, R. *et al.* Solid-state NMR spectroscopy applied to a chimeric potassium channel in lipid bilayers. *J. Am. Chem. Soc.* **130**, 7427-7435 (2008).

59. Wasmer, C. *et al.* Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science* **319**, 1523-1526 (2008).

60. Lorieau, J. L., Day, L. A. & McDermott, A. E. Conformational dynamics of an intact virus: Order Parameters for the coat protein of Pf1 bacteriophage. *Proc. Natl. Acad. Sci. USA* **105**, 10366-10371 (2008).

61. Andrew, E. R., Bradbury, A. & Eades, R. G. Removal of Dipolar Broadening of Nuclear Magnetic Resonance Spectra of Solids by Specimen Rotation. *Nature* **182**, 1659 (1958).

62. *Solid-state NMR spectroscopy: principles and applications*. (2002).

63. Ernst, R. R., Bodenhausen, G. & Wokaun, A. *Principles of Nuclear Magnetic Resonance in one and two dimensions*. Oxford University Press, London (1987).

64. Takegoshi, K., Nakamura, S. & Terao, T. C-13-H-1 dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem. Phys. Lett.* **344**, 631-637 (2001).

65. Pauli, J., Baldus, M., van Rossum, B., de Groot, H. & Oschkinat, H. Backbone and side-chain 13C and 15N signal assignments of the alpha-spectrin SH3 domain by magic angle spinning solid-state NMR at 17.6 Tesla. *Chembiochem.* **2**, 272-281 (2001).

66. Li, Y., Berthold, D. A., Frericks, H. L., Gennis, R. B. & Rienstra, C. M. Partial (13)C and (15)N chemical-shift assignments of the disulfide-bond-forming enzyme DsbB by 3D magic-angle spinning NMR spectroscopy. *ChemBioChem* **8**, 434-442 (2007).

67. Matzapetakis, M., Turano, P., Theil, E. C. & Bertini, I. $^{13}$C-$^{13}$C NOESY spectra of a 480 kDa protein: solution NMR of ferritin. *J. Biomol. NMR* **38**, 237-242 (2007).

68. Etzkorn, M. *et al.* Secondary structure, dynamics, and topology of a seven-helix receptor in native membranes, studied by solid-state NMR spectroscopy. *Angew. Chem.* **46**, 459-462 (2007).

69. Castellani, F. *et al.* Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420**, 98-102 (2002).

70. Venters, R. A., Farmer II, B. T., Fierke, C. A. & Spicer, L. D. Characterizing the use of perdeuteration in NMR studies of large proteins: 13C, 15N and 1H assignments of human carbonic anhydrase II. *J. Mol. Biol.* **264**, 1101-1116 (1996).

71. Garrett, D. S. *et al.* Solution structure of the 30 kDa N-terminal domain of enzyme I of the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR. *Biochemistry* **36**, 2517-2530 (1997).

72. Gardner, K. H., Rosen, M. K. & Kay, L. E. Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry* **36**, 1389-1401 (1997).

73. Zhang, H., Neal, S. & Wishart, D. S. RefDB: a database of uniformly referenced protein chemical shifts. *J. Biomol. NMR* **25**, 173-195 (2003).

74. Atreya, H. S. & Chary, K. V. Selective 'unlabeling' of amino acids in fractionally 13C labeled proteins: an approach for stereospecific NMR assignments of CH3 groups in Val and Leu residues. *J. Biomol. NMR* **19**, 267-272 (2001).

75. Bertini, I., Luchinat, C. & Parigi, G. *Solution NMR of paramagnetic molecules.* Elsevier, Amsterdam (2001).

76. Evans, D. F. The determination of the paramagnetic susceptibility of substances in solution by nuclear magnetic resonance. *J. Chem. Soc.* 2003-2005 (1959).

77. Phillips, W. D. & Poe, M. Contact Shifts and Magnetic Susceptibilities in Iron-Sulfur Proteins as determined from NMR spectra. *Methods Enzymol.* **24**, 304-317 (1972).

78. Mukherjee, R. N., Stack, T. D. P. & Holm, R. H. Angle dependence of the properties of the [Fe2X]4+ bridge unit (X = O, S): structures, antiferromagnetic coupling, and properties in solution. *J. Am. Chem. Soc.* **110**, 1850-1861 (1988).

79. Dawson, J. W. *et al.* Magnetic susceptibility study of hemerythrin using an ultrasensitive magnetometer. *Biochemistry* **11**, 461-465 (1972).

80. Baran, P. *et al.* Synthesis, Characterization, and Study of Octanuclear Iron-Oxo Clusters Containing a Redox-Active $Fe_{4O4-cubane\ core}$. *Inorg. Chem.* **47**, 645-655 (2008).

81. Fetter, J., Cohen, J., Danger, D., Sanders-Loehr, J. & Theil, E. C. The influence of conserved tyrosine 30 and tissue-dependent differences in sequence on ferritin function: use of blue and purple Fe(III) species as reporters of ferroxidation. *JBIC* **2**, 652-661 (1997).

82. Thurber, K. R. & Tycko, R. Measurement of sample temperatures under magic-angle spinning from the chemical shift and spin-lattice relaxation rate of 79Br in KBr powder. *J. Magn. Reson.* **196**, 84-87 (2009).

83. Szeverenyi, N. M., Sullivan, M. J. & Maciel, G. E. Observation of spin exchange by 2D FT ^{13}C CP/MAS. *J. Magn. Reson.* **47**, 462-475 (1982).

84. Fung, B. M., Khitrin, A. K. & Ermolaev, K. An improved broadband decoupling sequence for liquid crystals and solids. *J. Magn. Reson.* **142**, 97-101 (2000).

85. Schaefer, J., McKay, R. A. & Stejskal, E. O. Double-cross-polarization NMR of solids. *J. Magn. Reson.* **34**, 443-447 (1979).

86. Keller, R. & Wüthrich, K. A New Software for the Analysis of Protein NMR Spectra. 2002.
Ref Type: Personal Communication

87. Bertini, I., Luchinat, C., Turano, P., Battaini, G. & Casella, L. The magnetic properties of myoglobin as studied by NMR spectroscopy. *Chemistry - A European Journal* **9**, 2316-2322 (2003).

# A structure determination concept for large biological systems in ssNMR

**Chapter**

**5**

## 5. Summary

In this chapter will be discussed a novel high- resolution magic-angle spinning (MAS) solid-state-NMR (ssNMMR) method to improve protein assignment.

Up to now $^{13}$C-detection experiments represent the common strategy for ssNMR protein assignment resulting in complete structure elucidation of several peptides and small proteins.[1-3] However, this technique is still limited by problems of broad resonances, due to the $^{13}$C-$^{13}$C scalar coupling, resolution due to possible signal overlay in proteins beyond 8 kDa and to the limited chemical shift range typical of $^{13}$C resonances, and low sensitivity, that intrinsically characterize heteronuclear (the low-$\gamma$ nuclei) detection spectroscopy.

In the past few years, two methodologies have been proposed to overcome these problems. One is based on **proton detection**, the other one on **deuterium spectroscopy** in perdeuterated proteins. In both techniques resolution and sensitivity are improved with respect to the carbon detection experiments.

We now propose two new 3D ssNMR experiments, namely DCC and DCH, as a new strategy for protein assignment, counting the advantages of proton detection methods.

### 5.1. Introduction

Until now, all standard ssNMR experiments have relied on heteronuclear detection ($^{13}$C), as discussed previously (Section 4.5.3.4.). Proton, which is widely used in solution NMR because of its high natural abundance and high gyromagnetic ratio ($\gamma_H$), does not play a big role in solid-state NMR. In conventional solid-state NMR, indeed, $^{1}$H detection is usually less favorable, as large dipolar couplings lead to substantial line broadening. So far, the most used technique for protein assignment was based on correlation between heteronuclei like $^{13}$C or $^{15}$N, employing NCACX,

NCOCX and CANCO type experiments (Section 4.5.3.4.). The assignment of proteins beyond 8 kDa fails with this procedure mainly due to the broad resonances caused by $^{13}C$-$^{13}C$ scalar coupling.

Recently, two main approaches have been developed to overcome these problems. The first one is a $^{13}C$ spin dilution technique based on carbon detection experiments that solves the resolution and overlap problems. It consists in a **reduced labeling** scheme achieved by using [1,3-$^{13}C$]glycerol or [2-$^{13}C$]glycerol as carbon source for protein expression.[1] In this way, the bonded carbons are not simultaneously labeled and a lower number of connectivities is observed. Therefore, the carbon spectra are less crowded and the assignment is simplified. Moreover the carbon lines are significantly narrower with respect to spectra recorded on uniformly $^{13}C$ enriched samples because the carbon dilution suppresses or completely removes line-broadening due to $^{13}C$-$^{13}C$ scalar-couplings. Another advantage consists in the detection of weak couplings that contain structural information, thanks to the reduction of the number of strong dipolar couplings between connected nuclei.

The second approach consists in the development of **proton detected experiments** with $^{1}H$ narrow line. Currently, there are three general ways to destroy homogeneous interactions in MAS solid-state NMR. The first approach is based on quenching the proton dipolar coupling by spinning the sample at very high rotation frequencies around the magic angle. Improvements in MAS technology have made high frequency ($\approx$ 70 kHz) sample spinning possible.[4,5] A second spectroscopic approach involves the design of homonuclear decoupling sequences which allows the achievement of proton line widths in the order of 170–200 Hz.[6-8] The third approach involves dilution of the strongly coupled proton spin systems by deuteration. I will focus the attention on this last method.

### 5.1.1. Proton detected experiments in perdeuterated proteins

Suppression of strong $^{1}H$-$^{1}H$ dipolar couplings can be achieved chemically by **perdeuteration**. With this approach, all non-exchangeable proton sites are occupied by deuterium atoms while exchangeable sites are back substituted with protons. Deuterium has a gyromagnetic ratio that is a factor of 6.5 smaller compared to the proton one. Correspondingly, all $^{1}H$, $^{1}H$ dipolar couplings are scaled by a factor of 6.5. Deuterated

samples in conjunction with fast MAS and very weak decoupling quench cross-relaxation from protons, show narrower resonances compared to their protonated analogues acquired with high power homonuclear decoupling  and allow recording of multi-dimensional spectra for assignment.

During this year a systematic study of the effect of the level of exchangeable protons on the observed amide proton line width obtained in perdeuterated proteins has been conducted in Oschkinat's research group. It has been shown that a decreased amount of amide deuterium in perdeuterated proteins from 90% to 0%, gave a 4-fold increase in line width for both $^1$H and $^{15}$N resonances.[9] At the same time, decreasing the D$_2$O concentration from 90 to 60% , a sensitivity gain of 3.8 has been obtained for the CP based experiments $^1$H detected. In conclusion, it has been demonstrated that the best compromise between the line width and the signal to noise ratio in CP experiments is represented by perdeuterated proteins 40% back exchanged with water.[9]

In 2006 Reif's group conducted a study on $^1$H-$^{15}$N proton detected correlation spectra in MAS ssNMR using deuterated proteins back exchanged with a crystallization buffer prepared in 1:9 H$_2$O:D$_2$O solvent.[10] The protein used for this study is the α-spectrin SH3 domain, a 62 amino acid protein whose ssNMR assignment is known,[11] that is largely used as standard protein for ssNMR methodological studies. A $^1$H line width on the order of 30 Hz has been observed for proton detected experiments at spinning frequencies of 13 kHz without homonuclear decoupling (Fig. 5.1A). While the line widths obtained in the same conditions with deuterated SH3, 100% $^1$H at labile proton positions, if protons are evolved in an indirect dimension ($^{15}$N detection) are of ca. 100 Hz. In addition, avoiding of decoupling reduces sample heating and is beneficial for sample stability (Fig. 5.1B).
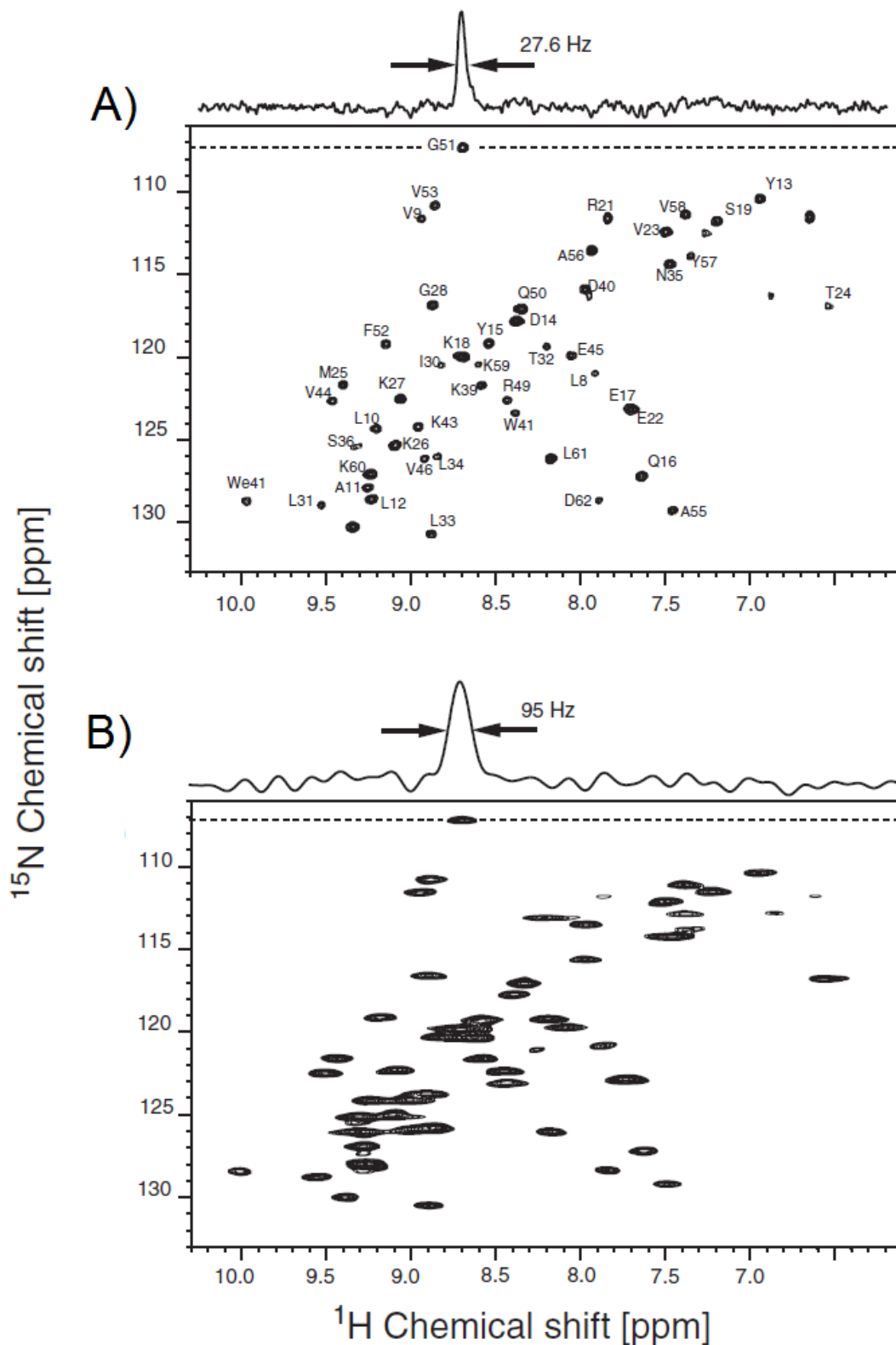
**Figure 5.1.** (A) $^1$H-detected 2D $^{15}$N–$^1$H correlation spectra of the deuterated SH3, 10% $^1$H at labile proton positions recorded at 600 MHz. (B) $^{15}$N detected 2D $^{15}$N–$^1$H correlation experiments for deuterated SH3, 100% $^1$H at labile proton positions recorded at 600 MHz.

*5.1.2. Deuterium spectroscopy in perdeuterated proteins*

Since in perdeuterated proteins protons are no longer available, a significant amount of information is compromised. In the last two years the use of deuterium CP has been proposed as an alternative to proton CP in perdeuterated proteins. The use of deuterium combined with $^{13}$C detection opens new fields for correlation and dynamic spectroscopy.

Reif's group shown that the deuterons can be used to probe chemical environment similarly to protons, and that the accessibility of deuterium quadrupolar tensors can reveal dynamic information. In particular it was demonstrated that the best way to obtain high-resolution deuterium ssNMR spectra is through Double-Quantum (DQ) deuterium-carbon experiments where the observed deuterium resonances are narrower than the ones observed in solution. Moreover, an sensitivity improvement by a factor of $2.46 \pm 0.82$ of the $^{2}$H,$^{13}$C correlation experiment compared to the $^{1}$H,$^{13}$C correlation was obtained.[12]

## 5.2. Aim of the project

Taking into account the sensitivity improvement of the $^{2}$H-$^{13}$C DQ CP versus the $^{1}$H-$^{13}$C CP and the better resolution of proton detected experiments in comparison with the carbon detected one, a novel strategy for protein assignment in ssNMR is in this work. The use of deuterium double quantum chemical shifts for separating carbon-carbon correlations into a third dimension, and thereby generating a starting point for a suite of pulse sequences for sequential assignment procedures that involve deuterium double quantum (equivalent to 2 times proton), and carbon chemical shifts is here reported. A solution-like concept for assigning side chain resonances was tested by 2D and 3D deuterium-carbon correlations of $^{2}$H,$^{13}$C,$^{15}$N-labelled ubiquitin employing deuterium double quantum generation prior to heteronuclear cross polarization. Moreover, considering the advantages derived from proton detection experiments, a DCH three dimensional correlation spectrum has been thought in order to achieve the sequential assignment.

### 5.3. Methods

*5.3.1. Deuterium*

Deuterium is a nucleus with I=1. A quadrupolar moment is associated with spins having a spin quantum number > 1/2. The quadrupolar moment interacts with the electric field surrounding it, created by the electron clouds around the nucleus in question. The large quadrupolar moment in combination with isotropic tumbling of molecules induces large quadrupolar relaxation and makes the direct observation of deuterons in solution disadvantageous. Typically, deuterium line widths in the order of 100-200 Hz are observed in solution.[13-15]. What about deuterium line width in ssNMR?

In the solid-state, narrow $^2$H resonance lines can in principle be achieved for two reasons: *i)* quadrupolar couplings inhomogeneous in nature can be completely averaged out under MAS; *ii)* the absence of molecular tumbling in solids eliminates the additional quadrupolar relaxation caused by tumbling.

*5.3.2. Deuterium-Carbon Cross-Polarization*

It is well known that deuterium gyromagnetic ratio is approximately 1.6 times smaller than the $^{13}$C one and that $^2$H has I=1 while $^{13}$C has I=1/2. Is the CP magnetization transfer from deuterium to carbon convenient? Comparing deuterium and carbon nuclear magnetic moments ($\mu_X$) is possible to calculate the gain in sensitivity:

$$\frac{\mu_D}{\mu_C} = \sqrt{\frac{I_D(I_D+1)}{I_C(I_C+1)}} \frac{\gamma_D}{\gamma_C} = 1.02 \qquad (5.1)$$

replacing $\frac{\gamma_D}{\gamma_C} = \frac{1}{1.6}$, $I_D = 1$ and $I_C = 1/2$. $I_D$ and $I_C$ are deuterium and carbon spin quantum numbers respectively and $\gamma_D$, $\gamma_C$ are the gyromagnetic ratios of deuterium and carbon respectively. It comes out that the magnetization transfer from deuterium to carbon is not inconvenient. In addition, deuterium being a quadrupolar nucleus, has an associated efficient quadrupolar relaxation mechanism which makes the spin-lattice relaxation times faster than proton[16,17] thus allowing fast repetition rates and therefore

the acquisition of a large number of scans for the same unit time. Moreover, deuterium is more abundant than protons in perdeuterated proteins.

In JACS 2009 Agarwal showed a resolution increase in the $^2$H-$^{13}$C DQ CP experiment with respect to the $^2$H-$^{13}$C SQ CP experiment.

He also found that $^2$H-$^{13}$C CP matching conditions for deuterons having different quadrupolar couplings (methyl groups and the Dα-Cα/ Dβ-Cβ) is different. During a CP spin lock pulse, deuterium spins are simultaneously exposed to two orthogonal fields: the constant Zeeman field created by *r.f.* and the electric gradient field of the quadrupolar nuclei.[18] The electric gradient field is continuously modulated in the z-direction parallel to the static magnetic field by spinning of the rotor. Thus, the modulation of size and direction of quadrupolar coupling is larger for nuclei having larger quadrupolar tensors. The time required for optimum $^2$H-DQ excitation is inversely proportional to the size of the quadrupolar coupling. Short DQ excitation and reconversion times (1 μs) yield exclusively correlations for C-Dα deuterons associated with large quadrupolar couplings ($e^2qQ/h$ =165 kHz), whereas a longer delay (9 μs) is necessary to observe correlations involving methyl deuterons associated with a smaller quadrupolar coupling ($e^2qQ/h$ = 55 kHz). An improvement by a factor of 2.46±0.82 of the $^2$H-$^{13}$C correlation experiment compared to the $^1$H-$^{13}$C correlation was observed. The $^2$H line widths were on the order of (38±14) Hz for methyls and of (30±9) Hz for carbon alpha.[19]

### 5.3.3. Double Quantum (DQ) $^2$H-$^{13}$C correlation experiment

Two hard 90° pulses separated by a delay τ constitute the simplest pulse sequence for generation the DQ coherence. The first 90° creates a transverse magnetization that evolves to an antiphase state under the influence of quadrupolar couplings during the delay τ. With the second 90° pulse this antyphase state is converted to a DQ coherence. The first and the second pulse should be both applied with the same phase. The DQ coherence evolves in the $t_1$ dimension. After evolution of DQ coherence in $t_1$, a time reversal approach is adopted to convert the DQ coherence back to SQ transverse magnetization. A fourth 90° pulse flips the deuterium magnetization along the z-axis. A z-filter of duration Δ of 2 μs was used to remove any residual transverse magnetization. Further $^2$H 90° pulse combined with $^2$H-$^{13}$C CP transfers magnetization to carbons for

detection. Proton decoupling was employed in the indirect dimension to remove the effect of long range heteronuclear dipolar coupling to deuterons.

For maximum efficiency of the sequence double-quantum excitation-evolution-reconversion period for the first $t_1$-increment is equal to one rotor period: $\tau_R = 3*\tau_{90°} + 2*\tau + t_1(0)$, where $\tau_{90°} = {}^2H$ 90° pulses, $\tau_{180°} = {}^{13}C$ 180° pulse, $\tau$ = free evolution period and $t_1(0)$ = deuterium evolution period for the first increment. All increments in the indirect dimension were adjusted as multiple of the rotor period: $t_1 = 2n* t_R$ with $n$ as integer number to avoid Multiple Quantum spinning side bands.
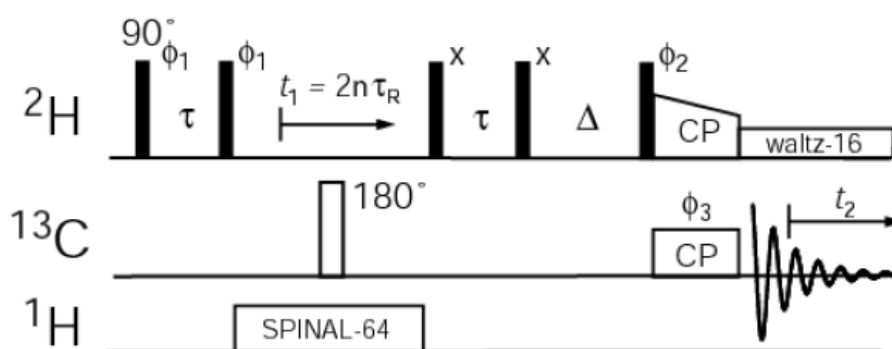


**Figure 5.2.** Pulse sequence for ${}^2H$, ${}^{13}C$ Double Quantum (DQ) correlation experiment. The double quantum excitation-evolution-reconversion period for the first t1 point was equal to a rotor period. Phase cycling: $\phi_1 = (x, y, -x, -y)$; $\phi_2 = (-y)$; $\phi_3 = (x)$; $\phi_4 = (4(x), 4(-y))$; $\phi_5 = (y)$; $\phi_6 = (8(x), 8(-x))$; $\phi_7 = (16(x), 16(-x))$; $\phi_8 = (2(x), 4(-x), 2(x))$; $\phi_{REC} = (2(x, -x), 4(-x, x), 2(x, -x))$. Phase sensitive detection in $t_1$ is achieved using TPPI (on phase $\phi_1$ and $\phi_4$).

### 5.3.4. Double Quantum (DQ) $^2H$-$^{13}C$-$^{13}C$ correlation experiment

Adding a $^{13}C$-$^{13}C$ magnetization transfer after the $^2H$-$^{13}C$ DQ-CP, a $^2H$-$^{13}C$-$^{13}C$ three dimensional correlation experiment has been performed. We used the DREAM adiabatic homonuclear recoupling magnetization transfer[20] which is demonstrated to bring the general advantages of adiabatic pulse schemes such as a complete polarization transfer and robustness with respect to chemical shift differences and experimental imperfections like RF-field inhomogeneity.
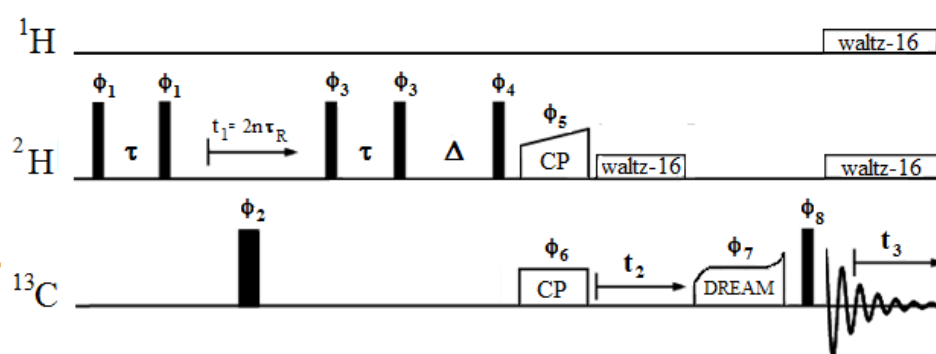


**Figure 5.3.** Pulse sequence for $^2H$-$^{13}C$-$^{13}C$ Double Quantum (DQ) correlation experiment. Phase cycling: $\phi_1 = (x, y, -x, -y)$; $\phi_2 = (-y)$; $\phi_3 = (x)$; $\phi_4 = (4(x), 4(-x))$; $\phi_5 = (y)$; $\phi_6 = (8(x), 8(-x))$; $\phi_7 = (16(x), 16(-x))$; $\phi_8 = (2(x), 4(-x), 2(x))$; $\phi_{REC} = (2(x, -x), 4(-x, x), 2(x, -x))$. Phase sensitive detection in $t_1$ was achieved using TPPI.

*5.3.5. Double Quantum (DQ) $^2H$-$^{13}C$-$^1H$ correlation experiment*

Adding a $^{13}C$-$^1H$ cross polarization transfer magnetization transfer after the $^2H$-$^{13}C$ DQ-CP, a $^2H$-$^{13}C$-$^1H$ three dimensional correlation experiment has been performed. Before the last CP pulses a z filter 90° pulse is introduced and immediately after four water suppression pulses. With a second 90° $^{13}C$ pulse, magnetization is recovered on the xy plane.
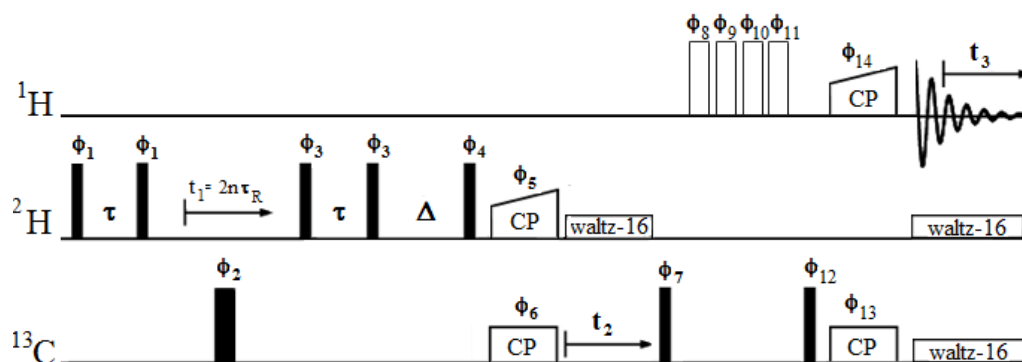


**Figure 5.4.** Pulse sequence for $^2H$-$^{13}C$-$^1H$ Double Quantum (DQ) correlation experiment. Phase cycling: $\phi_1$ = (x, y, -x, -y); $\phi_2$ = (-y); $\phi_3$ = (x); $\phi_4$ = (4(x), 4(-x)); $\phi_5$ = (y); $\phi_6$ = (8(x), 8(-x)); $\phi_7$ = (32(-y), 32(y)); $\phi_8$ = (x); $\phi_9$ = (y); $\phi_{10}$ = (-x); $\phi_{11}$ = (-y); $\phi_{12}$ =(16(y), 16(-y)); $\phi_{13}$ = (x); $\phi_{14}$ = (x); $\phi_{REC}$ = (2(x, -x), 4(-x, x), 2(x, -x), 2(-x, x), 4(x, -x), 4(-x, x), 4(x, -x), 2(-x, x), 2(x, -x), 4(-x, x), 2(x, -x)). Phase sensitive detection in $t_1$ was achieved using TPPI.

## 5.4. Assignment strategy

The aim of the novel experiments proposed in this section is to identify the amino acid spin systems trough $^{13}C$ and $D\alpha$ resonances and link the sequential spin systems via amide proton. The new three-dimensional chemical shift correlation experiments proposed for this purpose are the DCC and the DCH spectra. In Figure 5.4 is reported a scheme of the connectivities detected in the 3D DCC (blue) and DCH (green) experiments used for the sequential assignment. A peak in the DCC resonates at the frequencies of $D\alpha_i$-$C\alpha_i$-$CX_i$ and $D\beta_i$-$C\beta_i$-$CX_i$ where $CX_i$ can be one of the $^{13}C$ nuclear side chain spins of residue *i* ($C\alpha_i$, $C\beta_i$ etc.). Therefore, a strip at a given $^2H$ frequency in the DCC allows the identification of intraresidue (residue *i*) $C\alpha$, $C\beta$ etc.. A peak in the DCH resonates at the frequencies of $D\alpha_i$-$C\alpha_i$-$H_{i,i+1}$. Hence, a strip at a given $D\alpha$ frequency in the DCH allows the identification of the amide proton of residue *i* (intraresidue) and *i+1* (following residue). When the strips are visualized as in Figure 5.4 C it is possible to identify the intraresidue connectivities (residue *i*) $C\alpha$, $C\beta$ etc. and the sequential connectivities via amide protons.
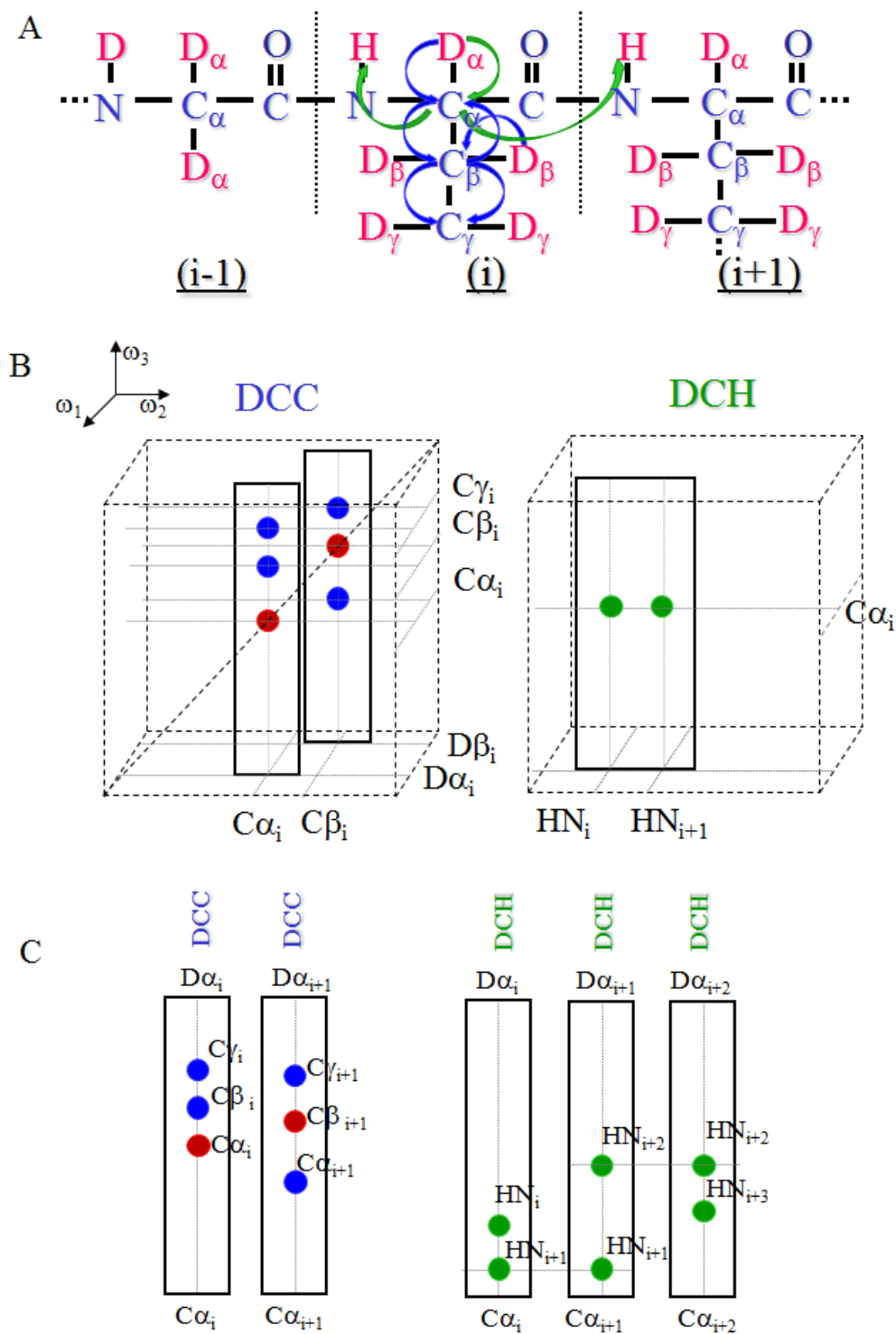
**Figure 5.4.** Schematic representation of the possible connectivities detected in the 3D experiments DCC (blue) and DCH (green) experiments. In the DCC cube representation the positive diagonal peaks are shown in red and the negative cross peaks in blue.

### 5.5. Results & Discussion

A $^2$H-$^{13}$C double quantum correlation experiment of uniformly $^2$H-$^{13}$C-$^{15}$N labeled ubiquitin 20% back exchanged with water is reported in Fig. 5.5.
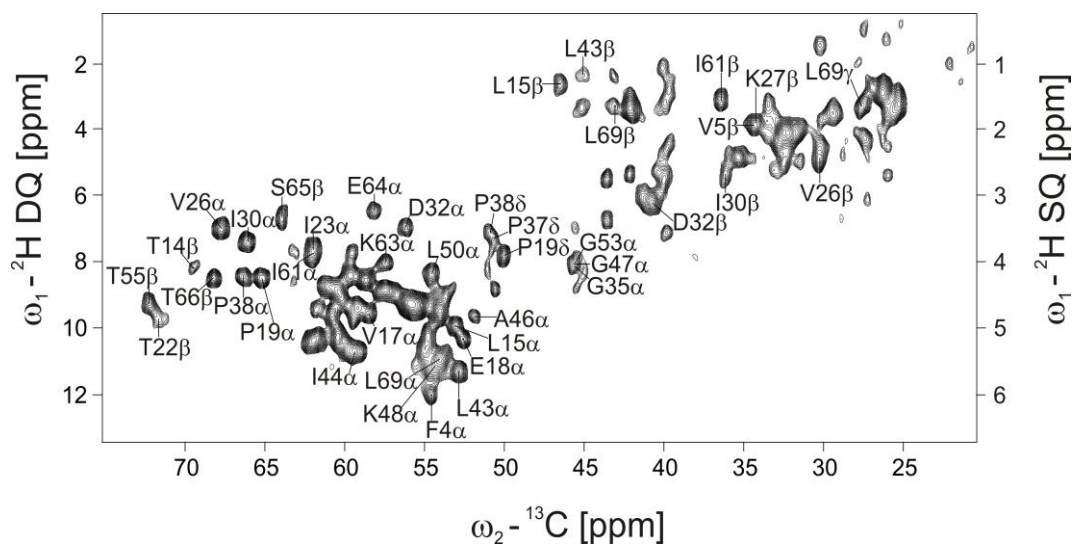


**Figure 5.5.** DQ $^2$H-$^{13}$C correlation spectrum. The spectrum was processed with a squared sine function with a sine bell shift of 5 in the indirect dimension while gauss-lorentz function in the direct dimension was carried out using exponential line broadening of -40 Hz and a gaussian max of 0.1.

The narrowest unscaled $^2$H DQ linewidth of approximately 36-37 Hz is observed for the Dα of A46. A list of DQ deuterium line width for different Dα groups is represented in Table 5.1.

| Amino acids | $^2$H DQ LW (Hz) |
|:---:|:---:|
| A46α | 36.5 |
| E64α | 40.3 |
| I30α | 40.9 |
| P38α | 38.6 |
| P19α | 38.7 |
| V26α | 42.0 |
| D32α | 47.0 |
| K63α | 39.9 |
| F4α | 47.5 |

**Table 5.1.** List of DQ deuterium line width for different Dα and Dβ groups.

A cube representation of the DQ $^2$H-$^{13}$C-$^{13}$C correlation spectrum of uniformly $^2$H-$^{13}$C-$^{15}$N labeled ubiquitin 20% back exchanged with water is reported in Fig. 5.6. The spectrum was processed as the DC one. The typical $^2$H line widths for CD$\alpha$ peaks are about 170 Hz.
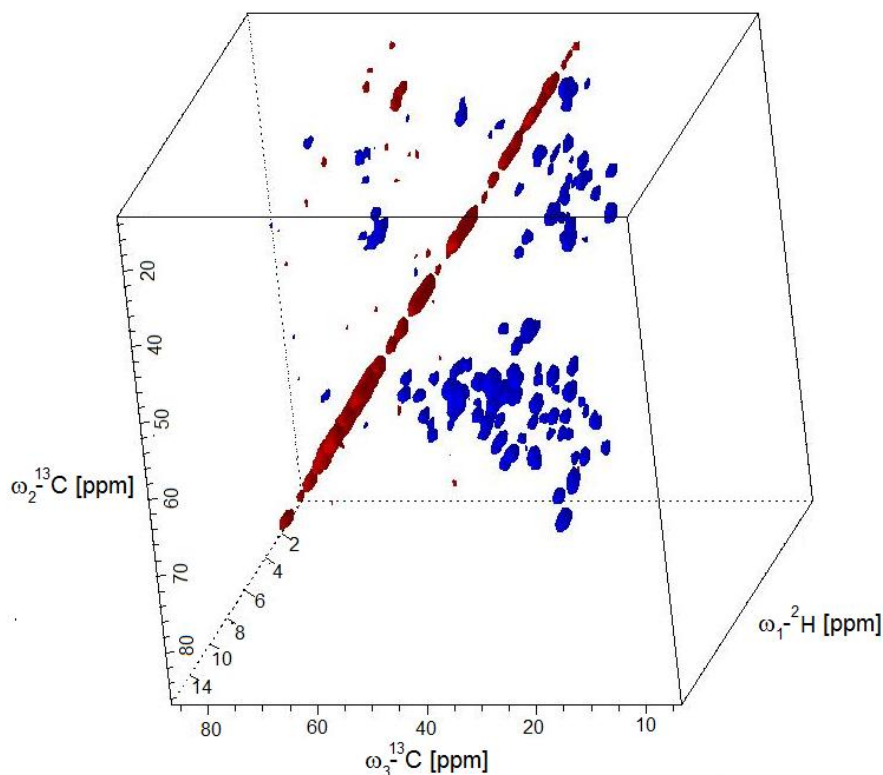


**Figure 5.6.** Cube representation of the three dimensional DQ $^2$H-$^{13}$C-$^{13}$C correlation.

The assignment was performed starting from carbon chemical shifts of crystal ubiquitin (BMRB 7111) that allowed the identification of the amino acid spin systems in the DREAM spectrum.[21] Linking the side chains of each amino acid in the DCC spectrum, was possible to assign almost all the D$\alpha$ cross peaks, 63% of the D$\beta$ and only a few C$\gamma$ (Fig 5.4 and 5.7). This is due to the different DC CP matching conditions for deuterons having different quadrupolar couplings (methyl groups, C$\alpha$-D$\alpha$ and C$\beta$-D$\beta$).[12] Recently, Reif and coworkers showed that the time required for optimum DC DQ excitation is inversely proportional to the size of the quadrupolar coupling constant: for C$\alpha$-D$\alpha$ having a large quadrupolar coupling constant (e$^2$qQ/h $\approx$ 170 kHz), a DQ excitation and reconversion time of 1 $\mu$s is needed; for methyl deuterons with a smaller quadrupolar coupling (e$^2$qQ/h $\approx$ 57 kHz), is necessary a delay 9 $\mu$s.[12]
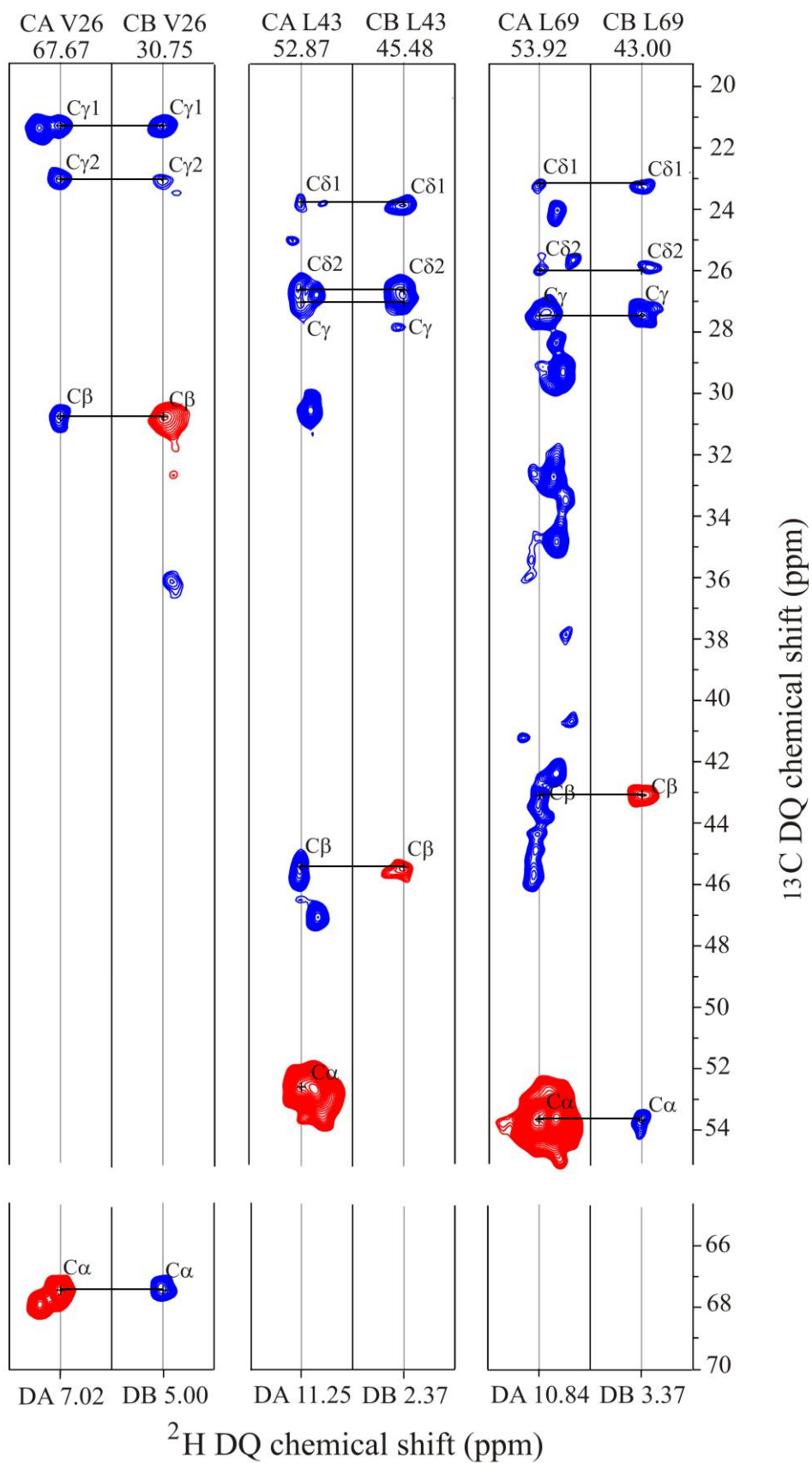
**Figure 5.7.** Representative strip plots of the 3D DCC spectrum, showing intraresidues connectivities for residues V26, L43 and L69. Blue and red peaks are

negative cross peaks and positive diagonal peaks respectively. Each strip is extracted at the $^{13}$C frequency ($\omega_3$) reported on the right. The $^2$H DQ frequency of the plane from which strips are extracted is reported on the bottom.

All the D$\alpha$–C$\alpha$ pairs could be identified, with the exception of the Met1. The 63% of the D$\beta$ was here assigned. For the prolines it was possible to recognize the D$\alpha$-C$\alpha$ and the C$\delta$-D$\delta$ pairs but not the C$\gamma$-D$\gamma$ and the C$\beta$-D$\beta$ pairs. On the other hand C$\beta$ and C$\gamma$ could be identified as connectivities from the D$\alpha$-C$\alpha$ and D$\delta$-C$\delta$ strips. For all the Gln and Glu (Gln2, Glu16, Glu18, Glu24, Gln31, Glu34, Gln40, Gln41, Gln49, Glu51, Gln58, Gln62, Glu64) the D$\beta$ was not detectable. The C$\beta$ was found from the D$\alpha$-C$\alpha$ strip. Beside these, the D$\beta$ of Ile3, Lys48, Ser57, Tyr59 were not identified, most probably due to low S/N. In the DCC spectrum all the correlations D$\alpha$–CX and D$\beta$–CX were recognized, provided the nuclei resonate between 75 an 15 ppm. In particular, for Leucines, Valines, Prolines, Threonines and Arginines all the cross peaks of the full pattern were observed, as well for the Isolucines, with the only exception of the C$\delta$ (that resonates around 14 ppm) and Lysines with the exception of the C$\varepsilon$ (low S/N). For the aromatic amino acids (Phe, Tyr, His), only the C$\alpha$ and C$\beta$ correlations are present in the spectrum.

### 5.6. Experimental section

#### 5.6.1. Protein samples

Uniformely $^2$H-$^{13}$C-$^{15}$N labeled human ubiquitin was expressed and purified as reported in literature.[22] Protein microcrystal were grown at 4°C with the sitting drop method with a crystallization buffer containing 80% of D$_2$O, following the procedure described by Igumenova.[23] In this form, about 5 mg of protein were transferred in a 2.5 mm rotor using an ultracentrifuge device.[24]

*5.6.2. NMR spectroscopy*

NMR experiments on ubiquitin were performed in Zurich on a Bruker Avance 600 wide-bore spectrometer operating at 14.09 T equipped of a Varian 2.5 Chemagnetic probe. The MAS frequency was set to 25 kHz. The sample was cooled at -15 °C.

For the DQ $^2$H-$^{13}$C correlation experiments the RF field amplitude for all hard pulses was adjusted to 100 kHz setting the RF carrier to 1.2 ppm and 100 ppm on $^2$H and $^{13}$C channels respectively. For $^2$H $^{13}$C CP the contact time was set to 1.4 ms. The $^2$H field during CP was ramped from 95 kHz to 105 kHz the $^{13}$C field was matched to the (-1) Hartmann-Hahn condition (approximately 75 kHz). The delay $\tau$ was set to 1 μs to yield efficient DQ excitation for CαDα. The delay $\Delta$ was set to 2 μs. The spectral width was 50 kHz and 3 kHz for the direct and indirect dimension. 4096 and 160 points corresponding to acquisition times of 41.0 ms and 25.6 ms were acquired in $t_1$ and $t_2$ respectively. 1024 transient were acquired per $t_1$ increment. The recycle delay was set to 150 ms leading to an acquisition time of about 9 h. WALTZ-16 decoupling on the $^2$H and $^1$H channels was applied using a RF field strength of 3 kHz.

The DQ $^2$H-$^{13}$C-$^{13}$C correlation experiments were recorded with the pulse sequence shown in Figure 5.3.. For $^2$H $^{13}$C CP was used a contact time of 1.5 ms. The $^2$H field during CP was ramped from 91 kHz to 101 kHz and the $^{13}$C field was matched to the (-1) Hartmann-Hahn condition (approximately 71 kHz). The WALTZ-16 decoupling on the $^1$H and $^2$H channels during the evolution and the detection time was set to a field strength of 3 kHz. The RF amplitude during the 5 ms DREAM recoupling period was tangentially swept from 11.50 to 10.45 kHz. The spectral width was 50 kHz, 18 kHz and 3 kHz for $t_3$ ($^{13}$C), $t_2$ ($^{13}$C) and $t_1$ ($^2$H) dimensions, respectively. 2048, 250 and 32 points corresponding to acquisition times of 20.5 ms, 6.9 ms and 5.1 ms were acquired in $t_1$, $t_2$ and $t_3$ respectively. 160 transient were acquired per $t_1$ increment. The recycle delay was set to 150 ms leading to an acquisition time of about 3 days.

Spectra were processed with the program TopSpin 2.0 (Bruker). The 2D and 3D maps were analyzed with the CARA 1.8.4.2 program.[25]

*5.6.3. NMR assignment of human ubiquitin*

Deuterium resonance assignments of ubiquitin was carried out starting from the reported assignment on BMRB entry 7111.[21]

## 5.7. Conclusion

Well-resolved $^1$H-$^{13}$C-equivalent correlations of deuterated ubiquitin are obtained by exploiting the long life time of $^2$H double quantum states. Sufficient signal-to-noise was obtained due to short deuterium $T_1$, allowing for high repetition rates and enabling 3D experiments within a reasonable time. 3D DQ D-C-C correlations of ubiquitin were recorded within 3 days each, allowing the sequence-specific assignment of all Dα and of 63% of Dβ. This approach provides the basis for a general heteronuclear 3D MAS NMR assignment concept by including $^1$H-like chemical shifts.

Reference List

1. Castellani, F. *et al.* Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420**, 98-102 (2002).

2. Zech, S. G., Wand, A. J. & McDermott, A. Protein structure determination by high-resolution solid-state NMR spectroscopy: application to microcrystalline ubiquitin. *J. Am. Chem. Soc.* **127**, 8618-8626 (2005).

3. Lange, A. *et al.* A concept for rapid protein-structure determination by solid-state NMR spectroscopy. *Angew. Chem. Int. Ed.* **44**, 2089-2092 (2005).

4. Ernst, M., Meier, M. A., Tuherm, T., Samoson, A. & Meier, B. H. Low-power high-resolution solid-state NMR of peptides and proteins. *J. Am. Chem Soc.* **126**, 4764-4765 (2004).

5. Zhou, D. H. *et al.* Proton-detected solid-state NMR spectroscopy of fully protonated proteins at 40 kHz magic-angle spinning. *J. Am. Chem. Soc.* **129**, 11791-11801 (2007).

6. Bielecki, A., Kolbert, A. C. & Levitt, M. H. Frequency-switched pulse sequences: Homonuclear decoupling and dilute spin NMR in solids. *Chem. Phys. Lett.* **155**, 341-346 (1998).

7. Vinogradov, E., Madhu, P. K. & Vega, S. High-resolution proton solid-state NMR spectroscopy by phase-modulated Lee–Goldburg experiment. *Chem Phys Lett.* **314**, 443-450 (1999).

8.  Lesage, A. *et al.* Experimental aspects of proton NMR spectroscopy in solids using phase-modulated homonuclear dipolar decoupling. *J. Magn Reson.* **163**, 105-113 (2003).

9.  Akbey, U. *et al.* Optimum levels of exchangeable protons in perdeuterated proteins for proton detection in MAS solid-state NMR spectroscopy. *J. Biomol. NMR* **46**, 67-73 (2010).

10. Chevelkov, V., Rehbein, K., Diehl, A. & Reif, B. Ultrahigh resolution in proton solid-state NMR spectroscopy at high levels of deuteration. *Angew. Chem Int. Ed Engl.* **45**, 3878-3881 (2006).

11. van Rossum, B. J., Castellani, F., Rehbein, K., Pauli, J. & Oschkinat, H. Assignment of the nonexchanging protons of the alpha-spectrin SH3 domain by two- and three-dimensional 1H-13C solid-state magic-angle spinning NMR and comparison of solution and solid-state proton chemical shifts. *Chembiochem.* **2**, 906-914 (2001).

12. Agarwal, V., Faelber, K., Schmieder, P. & Reif, B. High-resolution double-quantum deuterium magic angle spinning solid-state NMR spectroscopy of perdeuterated proteins. *J. Am. Chem Soc.* **131**, 2-3 (2009).

13. Skrynnikov, N. R., Millet, O. & Kay, L. E. Deuterium spin probes of side-chain dynamics in proteins. 2. Spectral density mapping and identification of nanosecond time-scale side-chain motions. *J. Am. Chem Soc.* **124**, 6449-6460 (2002).

14. Millet, O., Muhandiram, D. R., Skrynnikov, N. R. & Kay, L. E. Deuterium spin probes of side-chain dynamics in proteins. 1. Measurement of five relaxation rates per deuteron in (13)C-labeled and fractionally (2)H-enriched proteins in solution. *J. Am. Chem Soc.* **124**, 6439-6448 (2002).

15. Ishima, R., Louis, J. M. & Torchia, D. A. Optimized labeling of 13CHD2 methyl isotopomers in perdeuterated proteins: potential advantages for 13C relaxation studies of methyl dynamics of larger proteins. *J. Biomol. NMR* **21**, 167-171 (2001).

16. Hologne, M., Faelber, K., Diehl, A. & Reif, B. Characterization of Dynamics of perdeuterated proteins by MAS Solid-State NMR. *J. Am. Chem. Soc.* **127**, 11208-11209 (2005).

17. Hologne, M., Chevelkov, V. & Reif, B. Deuteration of peptides and proteins in MAS solid-state NMR. *Prog. NMR Spectrosc.* **48**, 211-232 (2006).

18. Muller, L. Proton--Deuterium polarization transfer in magic angle spinning polycrystalline solids in the rotating frame. *Chem. Phys.* **61**, 235-248 (1981).

19. Fung, B. M., Khitrin, A. K. & Ermolaev, K. An improved broadband decoupling sequence for liquid crystals and solids. *J. Magn. Reson.* **142**, 97-101 (2000).

20. Verel, R., Ernst, M. & Meier, B. H. Adiabatic dipolar recoupling in solid-state NMR: the DREAM scheme. *J. Magn Reson.* **150**, 81-99 (2001).

21. Schubert, M., Manolikas, T., Rogowski, M. & Meier, B. H. Solid-state NMR spectroscopy of 10% 13C labeled ubiquitin: spectral simplification and stereospecific assignment of isopropyl groups. *J. Biomol. NMR* **35**, 167-173 (2006).

22. Schanda, P., Meier, B. H. & Ernst, M. Quantitative analysis of protein backbone dynamics in microcrystalline ubiquitin by solid-state NMR spectroscopy. *J. Am. Chem Soc.* **132**, 15957-15967 (2010).

23. Igumenova, T. I., Wand, A. J. & McDermott, A. E. Assignment of the backbone resonances for microcrystalline ubiquitin. *J. Am. Chem. Soc.* **126**, 5323-5331 (2004).

24. Bockmann, A. *et al.* Characterization of different water pools in solid-state NMR protein samples. *J. Biomol. NMR* **45**, 319-327 (2009).

25. Keller, R. *The CARA/Lua Programmers Manual*. DATONAL AG, (2003).

# Summary and conclusions

### 6. Summary and conclusions

The present work has focused on some important biological processes involving protein intermolecular interactions.

Protein-small ligand studies have helped identifying new binders targeting the S100B-p53 interaction areas, that show potential for further optimization in a rational drug design process. Protein-protein interaction studies have provided the first structural model of cytochrome c binding to a Bcl-like protein, providing structural insights in cytochrome c dependent apoptosis. Both these activities relied on the use of well established NMR methodologies for ligand-screening and chemical shift perturbation mapping in the former case, and chemical shift perturbation mapping and data-driven docking in the latter.

On the other hand, the characterization of the interaction between iron and ferritin required the preliminary development of an original approach based on the combined used of solid state NMR and solution $^{13}$C-direct detection NMR, which partially overcame the intrinsic difficulties associated to the assignment of a such large protein by NMR. Analysis of the paramagnetic line-broadening in solution spectra of ferritin permitted to trace the iron path from the catalytic site to the nanocavity: this represent the first structural, albeit low resolution, information on the pre-biomineralization path of the iron inside the protein. High-field revisiting of the old Evans measurements gave indications on the nuclearity of the precursors of the biomineral.

Complementation between NMR and X-ray crystallography was used in some cases and resulted successful.

# List of Publications

1. Agamennone, M.; Cesari, L.; Lalli, D.; Turlizzi, E.; Del Conte, R.; Turano, P.; Mangani, S.; Padova, A.; Fragmenting the S100B-p53 interaction - Combined virtual/biophysical screening approaches to identify ligands; Chem. Med. Chem. 2010, 5(3), 428-435. (See Chapter 2).

2. Turano, P.; Lalli, D.; Felli, I. C.; Theil, E. C.; Bertini, I.; NMR reveals a pathway for iron mineral precursors to the central cavity of ferritin; PNAS 2010, 107(2), 545-550. (See Chapter 4).

3. Bertini, I.; Chevance, S.; Del Conte R.; Lalli, D.; Turano, P.; The anti-apoptotic Bcl-$x_L$ protein, a new piece in the puzzle of cytochrome c interactome.; PLoS ONE, submitted. (See Chapter 3).

4. Bertini, I.; Lalli, D.; Mangani, S. Pozzi, C.; Rosa, C.; Turano, P.; Snapshots of iron biomineralization in ferritin; In preparation. (See Chapter 4).

5. Lalli, D.; Schanda, P.; Ackbey, U.; Chowdhury, A.; Oschkinat, H.; A structure determination concept for large biological systems in solid; In preparation. (See Chapter 5).