UNIVERSITÀ DI SASSARI
DIPARTIMENTO DI CHIMICA

Dottorato in Scienze Chimiche
Ciclo XVIII (2002-2005)

MULTIDIMENSIONAL NUCLEAR MAGNETIC RESONANCE FOR METALLOPROTEINS CHARACTERIZATION

RISONANZA MAGNETICA NUCLEARE MULTIDIMENSIONALE PER LA CARATTERIZZAZIONE DI METALLOPROTEINE

Tesi di Dottorato

MASSIMILIANO FRANCESCO PEANA

TUTOR
Prof.ssa Maria Antonietta Zoroddu

COORDINATORE
Prof. Stefano Enzo

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AS	Alpha-Synuclein
CaBP	Calcium Binding Protein
CaM	CalModulin
Cap43	Calcium protein 43 kDa
CBD	Calmodulin Binding Domain
CD	Circular Dichroism spectroscopy
DLB	Dementia with Lewy Bodies
H4	Histone H4 protein
HSQC	Heteronuclear Single Quantum Coherence
LB	Lewy Bodies
LBVAD	Lewy Body Variant of Alzheimer's Disease
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect SpectroscopY
PCS	Pseudo-Contact Shift
PD	Parkinson's Disease
PDB	Protein Data Bank
RDC	Residual Dipolar Coupling
RMSD	Root Mean Square Deviation
ROESY	Rotational Overhauser Effect SpectroscopY
TOCSY	TOtal Correlation SpectroscopY
WT	Wild Type

CHAPTER 1

INTRODUCTION

1.1 The Post-Genomic Research

The explosive growth of genetic sequence information has offered us a comprehensive collection of the protein sequences found in many living organisms.

One of the great challenges of the *post-genomic era* is to be able to rapidly attribute specific biological functions to each gene sequence. The challenge of understanding these gene products has led to the development of functional genomics methods, which collectively aim to imbue the raw sequence with biological understanding.

At present (October 2005), genome sequences for 68 archea, 964 bacteria, 1007 eukaryota and 1488 viruses are available (Genome Projects Database: www.ncbi.nlm.nih.gov/Genomes/index.html). This wealth of information becomes particularly attractive because it can contribute to unraveling the basic mechanisms of life. However, this potential can be fully exploited only through a detailed study of

biomolecular structure and dynamics of the gene products, of the interactions between macromolecules (proteins, DNA and RNA) and of the complex regulation mechanisms.

Protein structure represents a powerful mean of discovering function, because structure is well conserved over evolutionary time, and it therefore provides the opportunity to recognize homology that is undetectable by sequence comparison.
This became apparent with the first two protein structures that were determined, because their common ancestry was clear from the three-dimensional fold [1], although their sequences did not contain recognizable similarity [2].

Structural genomic and *structural biology* represent two aspects of research in the *post-genomic era*. The former is the high throughput structural characterization of proteins, with complete coverage relative to genomic information. This can be achieved through computation of three-dimensional structural models of the molecules with a quite high accuracy in comparison to the top level obtainable with the presently available technologies. The second aspect consists of the refinement and the detailed analysis at atomic level of the structure of biological macromolecules, with the ultimate goal of investigating the relationship between structure and function.

1.2 NMR in Structural Biology

NMR has a central role within structural biology in the post-genomic challenge providing structures of macromolecules at atomic resolution.

Playing in concert with other techniques of structural analysis, mainly X-ray crystallography but also cryo-electron-microscopy and molecular modeling, together, they have unique strengths, providing a complementarity of approaches to completely describe the relationship between sequence and structure in molecular biology.
With the detailed three-dimensional structure of protein in solution, NMR plays, consequently, a pivotal role not only in the biological field, but also in pharmaceutical research as a basis for a detailed understanding of molecular functions and as a foundation for protein design.

Whereas x-ray crystallography requires single crystals, high field NMR spectroscopy is the only existing method for structure determination of soluble proteins and protein domains in solution.

Structural analysis by NMR does not require protein crystals. Most (~75%) of the NMR structures in the Protein Data Bank (www.rcsb.org/pdb/) do not have corresponding crystal structures, and many of these simply do not provide diffraction quality crystals. Furthermore, while most crystal structures are determined under physiologically relevant conditions, in many cases somewhat exotic solution conditions are required for crystallization. NMR studies can be carried out in aqueous solution under conditions quite similar to the physiological conditions under which the protein normally functions. This feature allows comparisons to be made between subtly different solution conditions that may modulate structure-function relationships.

For example, pH titration data can be used to determine pKa values of specific ionizable groups in the protein and to characterize the corresponding structure-function relationships.

The characterization in solution can go beyond the static picture of a single structure, as it can provide information on the internal mobility of the protein on various time scales. It has already been pointed out that internal mobility can play a determinant role in the biological functions.

NMR has special value in structural genomics efforts for rapidly characterizing the 'foldedness' of specific protein. The dispersion and lineshapes of resonances measured in 1D ^{1}H-NMR and 2D ^{15}N-^{1}H or ^{13}C-^{1}H correlation spectra provide 'foldedness' criteria with which to define constructs and solution conditions that provide folded protein samples (fig. 1.1). The 2D ^{15}N-^{1}H correlation spectra can be recorded in tens of minutes with conventional NMR systems and moreover, it offer us the advantage of being able to monitor the interaction of the proteins with other biomolecules (fig. 1.2).

With NMR spectroscopy, therefore, a complete description of the molecule can be obtained, ranging from the 3D detailed structure, to the folding process to the internal motions of the protein, all aspects which are essential for a deep understanding of its biological function.

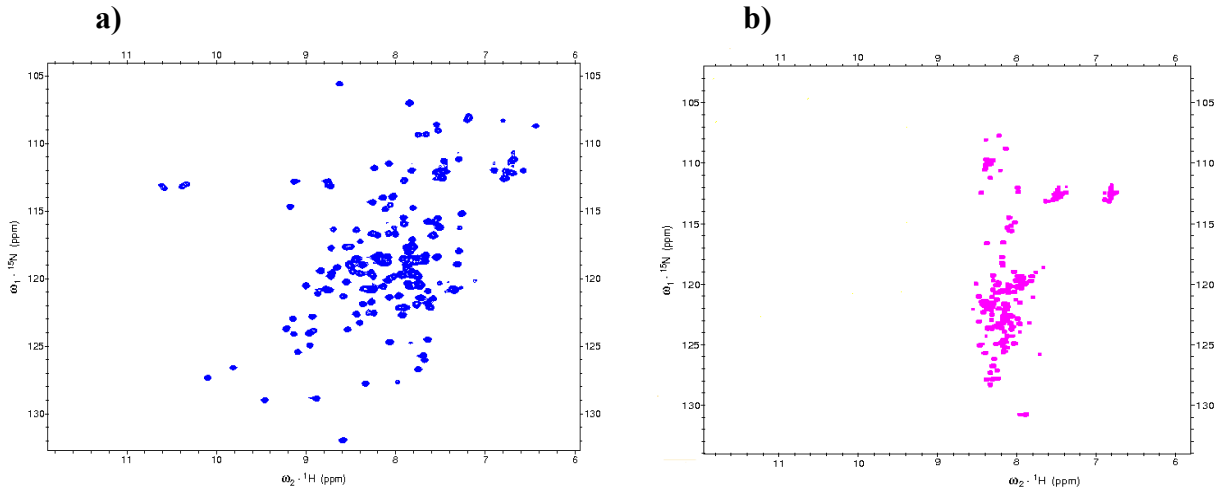


Fig. 1.1 Folded and unfolded protein: a) 2D ^{15}N-^{1}H HSQC spectra of Calmodulin (CaM) display a broad distribution of NMR frequencies resulting in a good spread-out of signals dispersion of folded protein, **b)** 2D ^{15}N-^{1}H HSQC for the unfolded alpha-Synuclein (AS) protein share similar NMR frequencies resulting in heavy overlap of signals (chapters 3-6)

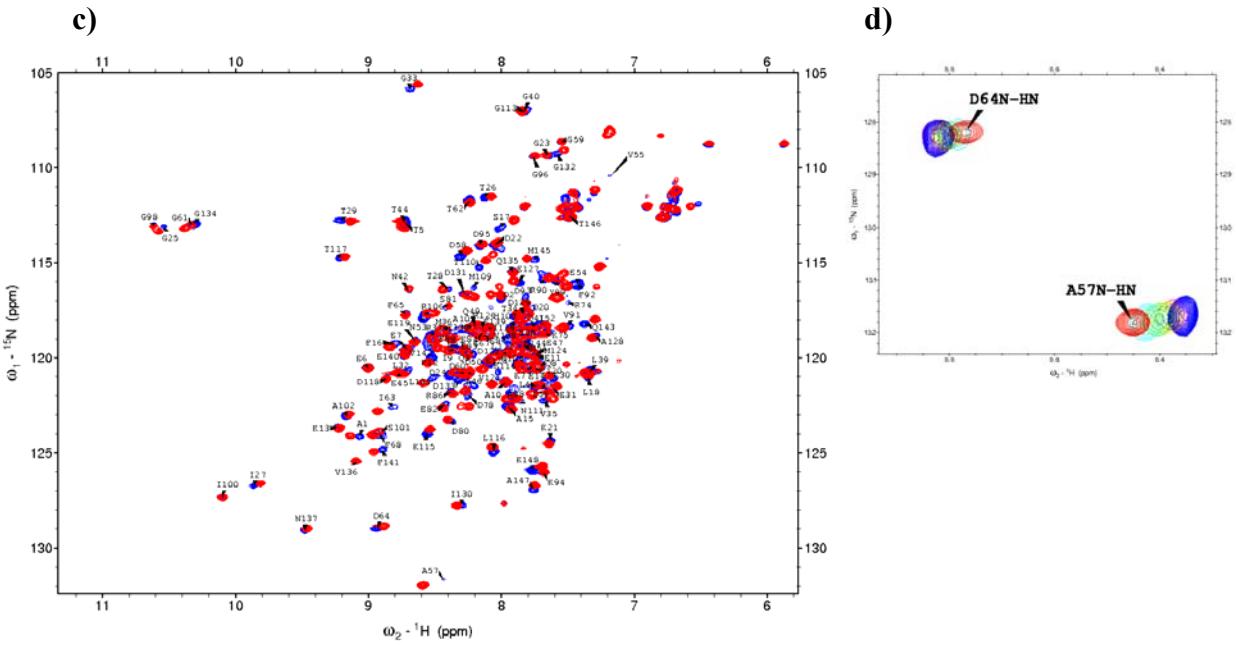


Fig. 1.2 Protein–protein interaction analysis by nuclear magnetic resonance spectroscopy: **c)** Superimposed ^{1}H-^{15}N HSQC spectra of Human Ca^{2+}-CaM (blue) and ^{1}H-^{15}N HSQC spectra of Human Ca^{2+}-CaM upon complexation (red) with unlabelled AS recorded at 700 MHz proton frequency. ^{15}N enriched CaM is titrated with increasing amounts of AS until 1:1 molar ratio, **d)** part of the spectrum contains a some peaks that change their position during the titration. The change in the amide proton chemical shifts of some residues of CaM in presence of non-labeled AS indicates that the proteins interact in solution. (chapter 6)

1.3 Protein Folding and Misfolding

The study of protein folding is an extraordinarily complex process, because it involves an extremely heterogeneous series of molecular transitions taking place in a solution environment often on a time-scale of seconds or less. Until recently, with the significant advances in theoretical, technical and experimental approaches, it is now possible to deal with a problem which was considered before as intractable.

NMR spectroscopy has been central in the studies of *protein folding problem* because of its capability to define the structures of proteins in solution, to design proteins with new functions and to characterize the dynamic properties that are inherent in the folding process. NMR 3D structural determination has led to the characterization of several proteins containing functionally relevant disordered regions. Additionally, NMR has surpassed other techniques, such as crystallography, in characterizing proteins that lack a rigid tertiary fold, i.e. natively unfolded proteins.

Many gene sequences in eukaryotic genomes encode entire proteins or large segments of proteins that lack a well-structured three-dimensional fold.
In contrast to the traditional view that function equates with a stable three-dimensional structure or depends on structural alternations between different conformations, examples of non-structured segments of proteins, which may play important roles in protein function, have been discovered. The literature now contains numerous reports of disordered regions, highly conserved between species in both composition and sequence, that are essential for function. These disordered regions include molecular recognition domains, protein folding inhibitors and flexible linkers.

For example, it is well known that locally structural disorder is necessary for Calmodulin binding activity. The flexibility within its interdomain linker, combined with the inherent deformability of the hydrophobic pockets, is thought to account for the ability of CaM to interact with a variety of different targets in a sequence-independent fashion [3,4].

The functional role of intrinsically disordered proteins in crucial process of regulatory functions in the cell, such as transcriptional regulation, translation and cellular signal transduction has only recently been recognized. The intrinsic lack of structure or partial disorder can confer functional advantages on a protein, such as their ability to alter their local and global structure in order to bind to several different

targets. In other words, flexible chains can adopt different conformations to fit to different ligands. Indeed, a significant advantage of intrinsic disorder is to allow one regulatory region or protein to bind to many different partners, including both proteins and nucleic acids.

An important regulatory mechanism that has emerged in recent years involves the post-traslational "marking" of regulatory regions at numerous sites to form a "code" that determines the biological response. The most well-know example of this case is the histone code, in which histone tails are subject to modifications by acetylation, phosphorylation, methylation and ubiquitylation, and which has a fundamental role in regulating access to DNA [5,6].

The N-terminal tails of histone are disordered in isolated histone proteins [7] and in the crystal structure of the nucleosome core particle [8]. The intrinsic disorder and the inherent flexibility in these regions are probably essential for their function: they leave the side chains exposed for modification by several different enzymes, for example acetyltransferase and deacetylases, methylases, and ubiquitin ligases, and provides the flexibility that is needed to allow adaptation to the varying conformational requirements of the active sites of these enzymes.

A number of examples of domains that are unstructured in solution but which become structured upon binding to their target have been noted. For example, several target proteins of β-catenin are intrinsically unstructured and fold upon binding to the extended scaffold presented by the armadillo repeat region of β-catenin. The cytoplasmic domain of E-cadherin is unstructured in solution [9], but becomes structured on binding to β-catenin [10] .

Many natively unfolded proteins require a metal ion or metal-containing cofactor. Signaling events that rely on Ca^{2+} are dependent on changes in protein structure induced by the binding of Ca^{2+}. Calcineurin, a protein with essential disorder, is a calcium/Calmodulin-activated serine/threonine phosphatase involved in both calcium and phosphorylation signaling pathways [11]. Often metalloproteins are unfolded in the apo state, i.e. their folding is linked to the binding of metal ions or cofactors. Examples include c-type proteins [12] and Ferredoxins [13]. While *in vitro* such apo proteins can be reconstituted, it is believed that under cellular conditions accessory proteins are required for the molecule to attain its tertiary fold. It should be

noted, however, that there are proteins that retain their tertiary structure in the absence of their metal cofactor, such as Cupredoxins [14] and Rubredoxins [15].

Several databases and bioinformatic tools have been developed to predict naturally unstructured regions using amino acid sequence information [16,17]. The sequence–structure relationships indicate that disorder is an encoded property, and the predictions strongly suggest that proteins in nature are much richer in intrinsic disorder than are those in the Protein Data Bank. Recent predictions on 29 genomes indicate that proteins from eucaryotes apparently have more intrinsic disorder than those from either bacteria or archaea, with typically >30% of eucaryotic proteins having disordered regions of length ≥ 50 consecutive residues. The high proportion of these sequences in the genomes of all organisms studied to date argues for important, as yet unknown functions, since there could be no other reason for their persistence throughout evolution.

Disorder in protein structures can be either local or global. Locally disordered regions are common and have been observed in numerous X-ray and NMR structures of proteins. In some of these cases their disordered regions have been linked to biological function.

Interestingly, several partially unfolded intermediates between the ordered state and the random coil have been observed. These folding intermediates exhibit side-chains with motional characteristics like those of the random coil but with backbone secondary structure like that of the ordered state. Ptitsyn *et al.* proposed a model to accommodate these observations [18]. In this model the protein converts from an ordered state into a form having characteristics similar to liquids. For example, the protein's side-chains go from rigid to non-rigid packing, while its secondary structure remains almost unchanged and the overall shape of the molecule remains compact. Furthermore, Ptitsyn and Crane- Robinson coined the term "*molten globule*" [19] to describe this liquid-like, partially folded state. There has been considerable uncertainty regarding the molten globule hypothesis. However, many of the critics now largely accept the term molten globule, if not all of the details of the original proposal.

Clearly the assumption that a folded three-dimensional structure is necessary for function needs to be re-examined [20].

Dunker and co-workers proposed "*The Protein Trinity Hypothesis*" [21]. In this view, in terms of their structure, native proteins can be defined as being in one of three

states: the solid-like ordered state (folded), the liquid-like collapsed disordered state (molten globule), or the gas-like extended-disordered state (random coil). Function is then viewed to arise from any one of the three states or from transitions among them. Clearly the role of disorder in determining protein activity in organisms has to be evaluated when studying protein function. Therefore, in the frame of obtaining insight into the structure-function relationship of proteins, it is necessary for structural biologists to give increased attention to the importance of intrinsic disorder and functional flexibility in proteins.

On the other hand, there are many cases in which flexibility may be a disadvantage to the cell and in which misfolding events lead to malfunctioning of living systems. There is increasing evidence that neurodegenerative conditions are correlated with misfolding of key cellular proteins and these must be structurally dissected to reveal potential therapeutic avenues. For example, amyloid diseases, such as Parkinson's disease, Alzheimer's disease and the prion diseases (BSE or ``mad cow'' disease), are characterized by the deposition of protein plaques that appear to arise from the misfolding of conformationally flexible proteins normally found in the brain and other tissues [22,23].

The pathological hallmark of Parkinson's disease (PD) is the presence of intracellular inclusions, Lewy bodies (LB), and Lewy neurities, in the dopaminergic neurons of the substantia nigra and several other brain regions. Insoluble alpha-synuclein (AS) constitutes the main component of LB. AS is a neuronal protein localized in the presynaptic terminals in mammalian brain that is natively unstructured [24] but adopts a β-stranded conformation in the fibrillar aggregated of the LB of PD [25] and in other neurodegenerative synucleinopathies. The misfolding and dysfunction of alpha-synuclein, is self-assembly and the resulting amyloid formation is believed to play a significant role for neuronal cell death [26].

1.4 Metalloproteins

It is well established that a number of metal ions are essential to life, while other metal ions are poisonous to living organisms, even when present at very low concentrations. Sometimes the same element may be beneficial or deleterious, depending on its speciation [27].

Metal-ions concentration in cells must be maintained within proper ranges. If the concentration of a given essential metal is too low, processes that need to use that ion will be adversely affected and the organism can suffer from metal ion deficiency. On the other hand, once the concentration of a given metal ion is higher than a minimal threshold, there will be enough of that ion to fulfill its biological functions. The concentration cannot be increased indefinitely without adverse consequences, however. Above an upper threshold the effects of metal ion toxicity will arise. For example, when a metal ion binds to an inappropriate site, it might compete with other beneficial metal ion for that site; or there might be undesirable reactivity of the metal ion when it is not properly controlled in its normal binding sites. On the other hand, some metals have no known or presumed biological function. When present in cells, they may be both innocuous or toxic.

Homeostasis, the maintenance of the concentration of beneficial metal ions in the correct range, and detoxification, the removal of toxic concentrations of non-beneficial metal ions, require balance among the processes of metal ion uptake, utilization, storage and excretion. As noted above, metal ions not utilized in biological systems can be toxic, often because they tend to bind non-specifically, but with high affinity, to certain types of sites. Because of this tight binding, which is often a consequence of kinetic inertness, these metals may bind to sites where they inhibit some normal processes in such a manner that they are not easily removed and excreted. Other possible causes of metal ion toxicity include the formation of insoluble salts in biological fluids, participation in hydrolytic reactions that degrade biopolymers or redox chemistry that produces damaging by products, such as hydroxyl radicals.

Cell activity maintains the concentration of essential metal ions by a specific metabolic pathways. There exist specific pathways involving proteins (transporters, metallochaperones) which protect and guide the metal ions through the cytoplasm, ultimately transferring the ions to specific partners proteins. It is becoming clear that

both the transporters and the metallochaperone proteins employ atypical coordination chemistry relative to the enzyme that ultimately incorporate the metal as a cofactor. Intracellular sequestration of metals is based on a sequence of cellular events involving a cascade of different ligands with increasing metal binding strengths. In addition, as already mentioned, some metal ions are toxic, therefore detoxification and resistance systems that employ a variety of mechanisms to rid the cell of these potentially lethal toxins have evolved. In most cases, the expression of such resistance system is controlled at the level of transcription by metal sensor proteins that "sense" specific metal ions via their direct coordination. It is the direct binding of inducing metal ions that allows for a change in the regulatory function of the metal sensor protein.

As a conclusion, cells manage metal-proteins acquiring ions that are in deficiency, while exporting or sequestering those that are in surplus or toxics. The beneficial intracellular concentration of metals is maintained by strictly regulation in the expression of metalloproteins involved in specific metal uptake, export or storage.

Metalloproteins perform a variety of functions, which can be classified into five basic types, according to the model proposed by Holm *et al.* [28]:

I. *structural* – configuration (in part) of protein tertiary and/or quaternary structure;
II. *storage* – uptake, binding and release of metals in soluble form;
III. *electron transfer* – uptake, release and storage of electrons;
IV. *dioxygen binding* – uptake, release and storage of oxygen;
V. *catalytic* – substrate binding, activation and conversion;

where V is an extensive class subdivided by the type of reaction catalyzed (dismutases, oxidases and oxygenases, nitrogenases and hydrogenases, oxotransferases, hydrolases, and others).

Recent discoveries have identified proteins that do not fit within these categories. For example, biochemical and structural data have revealed a new class of metal-binding proteins called, metallochaperones, as mentioned above, which function as intracellular shuttle service for metal ions [29-30].

1.3 Aims and Topics of the Research

The work carried out over three years of doctoral studies has been devoted to the structural and dynamic characterization by NMR spectroscopy of metalloproteins or their specific metal-binding domains in solution.

Part of these studies were focused on the conformational and dynamic characterization of the calcium-protein Calmodulin and on its functional implications.

Disorder plays an important role in this protein. CaM has two globular domains, each with 2 EF-hands that bind calcium. The connecting linker was crystallized as a rigid helical structure [31], while the solution structure revealed that part of this interdomain helical rod melts into a flexible linker [32], thus enabling CaM to wrap around its target. This limited disorder may account for the significant plasticity observed within the globular domains when they bind to different targets. The conformation and functional significance of this linker region has been the subject of an intensive debate. Indeed, the flexibility of this linker allows for the reciprocal reorientation of the two domains of CaM.

We decided to investigate the conformational space sampled by CaM.

To this purpose we developed a novel approach exploiting the use of long-range constraints. Particularly, we searched favored orientations of the two domains with respect to each other by using advanced NMR techniques supplied by a paramagnetism based strategy. The use of lanthanides as substitutes for the calcium centers, allowed us to exploit long-range restraints, such as pseudocontact shifts (PCS) and residual dipolar coupling (RDC) for structural and dynamical analysis.

The results found, by the use of a novel approach, try to address a long-standing puzzle and provide new structural information on the characterization of the conformational space sampled by the domain reorientation of CaM (chapter 3,4).

CaM is known to bind a number of target peptides and proteins. We became interested in the binding of a non-typical target: an unfolded protein. This polypeptide was alpha-Synuclein, a pre-synaptic protein that is implicated as a possible causative agent in the pathogenesis of Parkinson's disease [22].

Disorder in this protein may be functional. Free wild-type AS is largely unfolded in solution, but exhibits a region with a preference for helical conformations that may be important in the aggregation of AS into fibrils [33]. Previously, it was

shown by crosslinking experiments that CaM forms a complex with AS in the presence and in the absence of calcium [34]. Therefore, we set out to analyze this interaction by NMR and we tackled the task of understanding the nature of the CaM-AS interaction at the molecular level, and tried to relate it with its possible physiopathological effects (chapter 5,6).

A second fascinating part on my PhD work is focused on the structural characterization by multidimensional NMR and other spectroscopic techniques as well as potentiometric studies of specific binding sites of nickel ion Ni(II) to the Cap43 protein and to histone H4 protein .

Cap43, a novel gene induced by a rise in free intracellular Ca^{2+} following nickel exposure, encodes its protein namesake with molecular weight of 43.000 Da [35,36].

The role of this protein, still unknown, is perhaps related, like other stress responsive proteins, with a detoxification and protection of cell against oxidative stress induced by metals [37,38]. In this view, Cap43 protein can be examined to determine its behavior in binding nickel ions to neutralize their potential toxicity (chapter 7,8).

In the same way, the study of the interactions of histone H4 or histone model peptides with nickel ions, and their subsequent effects, i.e. the dramatic change in the conformation upon metal ion complexation, is carried out in order to contribute to the elucidation of the mechanism of toxicity and carcinogenicity of nickel compounds (chapter 9).

In fact, it is clear now that histones, basic proteins that play an important role in compression and organization of DNA, are also fundamental, integral and dynamic components of the machinery responsible for regulating gene transcription [39], and the impairment of their functions by interaction with nickel ions can lead to cancer promotion and progression.

LIST OF REFERENCES FOR CHAPTERS 1

1. Perutz M.F,. Rossmann M.G, Cullis A.F., Muirhead H., Will G. & A.C.T. North (1960) *Nature* 185, 416.
2. Kendrew J. C. & Watson H. C. (1961) *Nature* 190, 666.
3. Meador WE, Means AR, Quiocho FA. (1993) *Science* 262, 1718.
4. O'Neil KT, DeGrado WF. (1990) *Trends Biochem Sci.* 15, 59.
5. Jenuwein T, Allis CD. (2001) *Science* 293, 1074.
6. Berger S.L.(2002) *Curr Opin Genet* 12, 142.
7. Hartman P.G., Chapman G.E., Moss T., Bradbury E.M. (1977) *Eur J Biochem* b77,45.
8. Luger K., Mader A.W., Richmond R.K., Sargent D.F., Richmond T.J. (1997) *Nature* 389, 251.
9. Huber A. H., Stewart, D. B., Laurents D. V., Nelson W. J., and Weis W. I. (2001) *J. Biol. Chem.* 276, 12301.
10. Huber A. H. and Weis W. I. (2001) *Cell* 105, 391.
11. Kissinger, C. R., Parge, H. E., Knighton, D. R., Lewis, C. T., Pelletier, L. A., Tempczyk, A., Kalish, V. J., Tucker, K. D., Showalter, R. E., Moomaw, E. W., Gastinel, L. N., Habuka, N., Chen, X., Maldanado, F., Barker, J. E., Bacquet, R., and Villafranca, J. E. (1995) *Nature* 378, 641.
12. Pettigrew G. and Moore G. R. (1987) *Cytochromes c: Biological aspects*. SpringerVerlag, Berlin - Heidelberg - New York
13. Busch J. L. H., Breton J. L., Davy S. L., James R., Moore G. R., Armstrong, F. A., and Thomson, A. J. (2000) *Biochem. J.* 346, 375.
14. Durley R., Chen L., Lim L. W., Mathews F. S., and Davidson V. L. (1993) *Protein Sci*. 5, 739.
15. Zartler E. R., Jenney F. E., Terrell M. Jr., Eidsness M. K., Adams M. W. W., and Prestegard J. H. (2001) *Biochemistry* 40, 7279.
16. Li X., Romero P., Rani M., Dunker A. K., and Obradovic Z. (1999) *Menome Informatics* 10, 30.
17. Vucetic S., Obradovic Z., Radivojac P., Peng K., Iakoucheva L. M., Cortese M. S., Lawson J. D., Brown C. J., Sikes J. G., Newton C. D., and Dunker A. K. (2005) *Bioinformatics* 21, 137.
18. Dolgikh D. A., Gilmanshin R. I., Brazhnikov E. V., Bychkova V. E., Semisotnov G. V., Venyaminov S., and Ptitsyn O. B. (1981) *FEBS Lett*. 136, 311.
19. Ohgushi M. and Wada A. (1983) *FEBS Lett*. 164, 21.
20. Wright P. E. and Dyson H. J. (1999) *J. Mol. Biol*. 293, 321.
21. Dunker A. K. and Obradovic Z. (2001) *Nat. Biotenol*. 19, 805.

22. Koo E.H., Lansbury P.T.Jr, and Kelly J.W. (1999) *Proc Natl Acad Sci USA* 96, 9989.

23. Horwich A. L. and Weissman J. S. (1997) *Cell* 89, 499.

24. Weinreb PH, Zhen W Poon AW, Conway KA, Lansbury PT Jr. (1996) *Biochemistry, 35,* 13709.

25. Der-Sarkissian A, Jao CC, Chen J, Langen R. (2003) *J Biol Chem.* 278, 37530.

26. Goedert M, Spillantini MG, Davies SW. (1998) *Curr Opin Neurobiol.* 8,619.

27. Finney L. A. and O'Halloran T. V. (2003) *Science* 300, 931.

28. Holm R. H., Kennepohl P., and Solomon E. I. (1996) *Chem. Rev.* 96, 2239.

29. Pufahl R., Singer C. P., Peariso K. L., Lin S.-J., Schmidt P. J., Fahrni C. J., Culotta V. C., Penner-Hahn J. E., and O'Halloran T. V. (1997) *Science* 278, 853.

30. O'Halloran T. V. and Culotta V. C. (2000) J. *Biol. Chem*. 275, 25057.

31. Babu Y., Bugg C. E., and Cook W. J. (1988) *J. Mol. Biol.* 204, 191.

32. Wriggers W., Mehler E., Pitici F., Weinstein H., and Schulten K. (1998) *Biophys. J.* 74, 622.

33. Eliezer D., Kutluay E., Bussell R. J., and Browne G. (2004*) J. Mol. Biol.* 307, 1061.

34. Lee D., Lee S. Y., Lee E. N., Chang C. S., and Paik S. R. (2002) *Neurchem.* 82, 1007.

35. Zhou D, Salnikow K, Costa M. (1998) *Cancer Res*, 58,2182.

36. Salnikow K, Blagosklonny MV, Ryan H, Johnson R, Costa M. (2000) *Cancer Res*, 60, 38.

37. Agrowal A., Bhattacharya S., (1989) *Experientia* 45, 567.

38. Suerbaum S., Thiberge J.M., Kansau I.,. Ferrero R.L, Labigne A., (1994) *Mol. Microbiol.* 14, 959.

39. Grunstein M., (1997) *Nature* 389,349.

CHAPTER 2

PROTEIN STRUCTURE DETERMINATION BY SOLUTION NMR

An overview on a general assignment strategy for structure determination and the NMR derived constrains used in the present PhD work are presented in this chapter.

2.1 NMR for High Resolution Structure Determination

The increasing importance of NMR as a method for structure determination of biological macromolecules is manifested in the steadily rising number of NMR structures that are deposited in the Protein Data Bank [1]. In October 2005, there were a total of 4825 files available from the PDB with Cartesian coordinates of proteins, nucleic acids and macromolecular complexes that have been obtained by NMR techniques.

Until 1984 structural information at atomic resolution could only be determined by X-ray diffraction techniques with protein single crystals [2]. The introduction of nuclear magnetic resonance spectroscopy as a technique for protein structure determination [3] has made it possible to obtain structures with comparable accuracy but in a solution environment.

In fact the highest quality NMR structures have accuracies comparable to 2.0–2.5 Å X-ray crystal structures. Although atomic positions in high-resolution crystal structures are more precisely determined than in the corresponding NMR structures, the crystallization process may select for a subset of conformers present under solution conditions. For example, while high-quality NMR structures typically exhibit root mean square (RMSD) deviations of backbone and heavy atoms (excluding those of surface side chains) of 0.3–0.6 Å and 0.5–0.8 Å, respectively, analysis of a set of high-resolution X-ray crystal structures of bovine pancreatic trypsin inhibitor determined in different crystal forms [4] indicates similar variations of 0.2–0.6 Å in backbone atom positions due to preferential selection of distinct low energy conformers in the crystallization process.

The NMR method for the study of molecular structures depends on the sensitive variation of the resonance frequency of a nuclear spin in an external magnetic field with the chemical structure, the conformation of the molecule, and the solvent environment [5]. The dispersion of these chemical shifts ensures the necessary spectral resolution, although it usually does not provide direct structural information.

Different chemical shifts arise because nuclei are shielded from the externally applied magnetic field to different extent depending on their local environment. Three of the four most abundant elements in biological materials, hydrogen H, carbon C and nitrogen N, have naturally occurring isotopes with nuclear spin ½ , and are therefore suitable for high-resolution NMR experiments in solution. The proton (^{1}H) has the highest natural abundance (99±98%) and the highest sensitivity (due to its large gyromagnetic ratio) among these isotopes, and hence plays a central role in NMR experiments with biopolymers. Because of the low natural abundance and low relative sensitivity of ^{13}C and ^{15}N (1±11% and 0±37%, respectively) NMR spectroscopy with these nuclei normally requires isotope enrichment. This is routinely achieved by overexpression of proteins in isotope-labelled media. Structures of small proteins with molecular weight up to 10 kDa can be determined by homonuclear ^{1}H NMR.

Heteronuclear NMR experiments with ^{1}H, ^{13}C and ^{15}N [6] are indispensable for the structure determination of larger systems [7-8].

However, NMR is not a microscope with atomic resolution that would directly produce an image of a protein. Rather, it is able to yield a wealth of indirect structural information from which the three-dimensional structure can only be uncovered by extensive calculations. The pioneering first structure determinations of peptides and proteins in solution [9-13] were year-long struggles, both fascinating and tedious because of the lack of established NMR techniques and numerical methods for structure calculation, and hampered by limitations of the spectrometers and computers of the time. Recent experimental, theoretical and technological advances and the dissemination of the methodological knowledge, have changed this situation decisively. Today, with a sufficient amount of a purified, water-soluble, monomeric protein with less than about 200 amino acid residues, its three-dimensional structure in solution can be determined routinely by the NMR method, following the procedure described in the classical textbook of Wüthrich [3] and outlined in Fig. 2.1

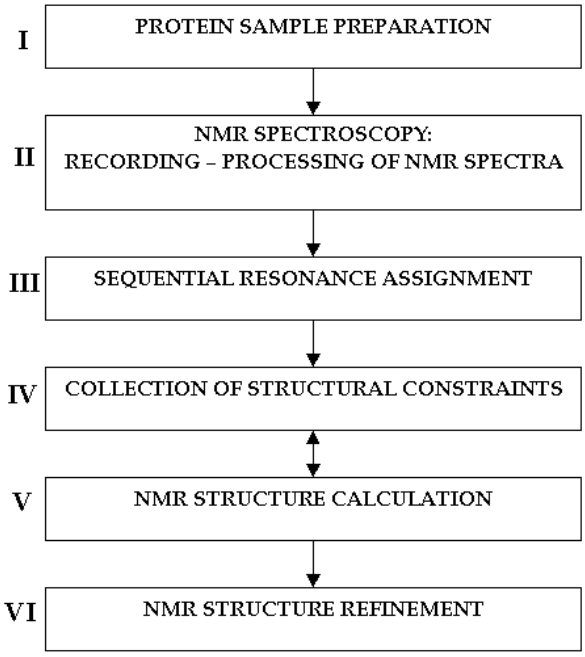


Fig. 2.1 Outline of the procedure for protein structure determination by NMR

The determination of a NMR solution structure may be dissected into six major parts:

(I) At the outset of the NMR study, a suitable sample, usually about 500 μL of a 1 mM protein solution is prepared. If the molecular weight of a protein exceeds ~10 kDa, enrichment with ^{13}C and ^{15}N isotopes is required in order to resolve spectral overlap in ^{1}H-NMR spectroscopy. Due to the availability of high-yield over-expression systems, stable isotope labeling has become routine.

(II) Subsequently, this sample is used to record a set of multidimensional NMR experiments, typically at temperatures around 30 ºC, which provide, after suitable data processing, the NMR spectra.

(III) These allow determination of (nearly) complete sequential NMR assignments, the measurement of resonance frequencies (chemical shifts) of the NMR active spins in the protein.

(IV) The resulting conformation-dependent dispersion of the chemical shifts is a prerequisite for deriving experimental constraints from various NMR experiments (such as NOE, scalar coupling, and dipolar coupling data) for the NMR structure calculation. The double arrow between steps (IV) and (V) indicate that the analysis of structural constraints and the calculation of NMR structures is generally pursued in an iterative fashion.

(V) Iterations involving structure calculations and identification of new constraints are carried out until the overwhelming majority of experimentally derived constraints is in agreement with a bundle of protein conformations representing the NMR solution structure. Conformational variations in the bundle of structures reflect the precision of the NMR structure determination.

(VI) Finally, the NMR structure can be refined using conformational energy force fields, which in essence reflect our current knowledge about conformational preferences of proteins.

2.2 Heteronuclear Multidimensional NMR Experiments

The introduction of the three and four-dimensional NMR experiments [14-18] and the availability of ^{13}C, ^{15}N-labelled proteins [19-29] allow one to assign the proton, nitrogen and carbon chemical shifts of proteins and protein-complexes with molecular weights above all 25 kDa and to determine their structures in solution [30-33]. The resonance assignment of singly (^{15}N or ^{13}C) labelled proteins using 3D experiments [34] is basically an extension of Wüthrich's strategy which exclusively relies on homonuclear ^{1}H NMR experiments.

The assignment strategy for proteins that are not isotopically enriched [35] makes use of a combination of COSY (COrrelation SpectroscopY), TOCSY (Total Correlation SpectroscopY) and NOESY (Nuclear Overhauser Effect SpectroscopY) or ROESY (Rotational Overhauser Effect Spectroscopy) spectra.

The spin systems attached to each H^N are identified in the COSY and TOCSY spectra. NOESY or ROESY are used for the sequential assignment of the individual spin systems. The establishing of the sequential connectivity relies on the occurrence of the resonance frequency of the Hα(i) proton of amino acid (i), which is observed in the H^N(i), Hα(i) cross peak in the COSY (TOCSY) spectrum and in the H^N(i+1), Ha(i) cross peak in the NOESY spectrum. Especially in α-helical proteins the sequential connectivity is also supported from H^N(i), H^N(i+1) cross peaks in the NOESY spectrum. Resonance overlap becomes so severe in the 2D spectra of proteins with more than 80 amino acids, that it is often impossible to identify adjacent spin systems by utilizing only a single common resonance frequency of the connecting cross peaks.

In 3D assignment experiments, each cross peak is labelled by three frequencies rather than by two frequencies as in 2D experiments. Using ^{15}N labelling it is then possible to resolve the resonance frequencies of the amide protons according to the ^{15}N chemical shift of the attached amide nitrogens. This is done in 3D $^{15}N,^{1}H$-HSQC-TOCSY [36-39] and 3D $^{15}N,^{1}H$-HSQC-NOESY [34, 40-44] experiments. While this facilitates the identification as well as the sequential assignment of the spin systems, the sequential connectivity still relies mainly on the

common chemical shift of the Hα(i) proton, which can introduce ambiguities. In order to resolve overlapping Hα chemical shifts, 3D HNHA spectra can be recorded on a ^{15}N-labelled sample. However, the use of the conformationally dependent NOE effect for the sequential assignment is a principal drawback of these strategies.

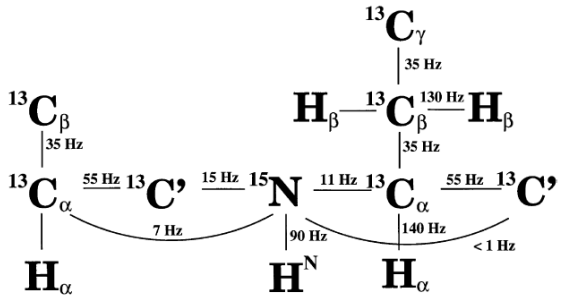


Fig. 2.2 Spin system of the peptide backbone and the size of the ^{1}J and ^{2}J coupling constants that are used for magnetization transfer in ^{13}C-, ^{15}N-labelled proteins.

More recently, new methodological developments have been introduced. They are a combinations of 3D experiments that including the chemical shifts of side-chain carbon and proton spins to achieve the sequential assignment.

The nomenclature for these triple resonance experiments reflects the magnetization transfer pathway of the experiments. Nuclei that are involved in magnetization transfers form the name of an experiment. Spins, whose chemical shifts are not evolved are put in parentheses. For an out-and-back type experiment [45], where magnetization of a spin is transferred to a remote spin and then brought back the same way, only the first half of the magnetization transfer is used for its name. For example, the out-and-back experiment that transfers magnetization from the amide proton (H^N) via the amide nitrogen (N) to the carbonyl C′(CO) of the previous residue is called HNCO. If another magnetization transfer step to the Cα (CA) of the previous residue is included and the corresponding Cα chemical shift is recorded, the experiment is referred to as H(N)COCA. The parentheses indicate that

magnetization is only transferred via the nitrogen spin, without chemical shift evolution taking place.

The strategy for resonance assignment originally proposed by Bax's group employs the 3D experiments [46-50] which exclusively correlate the resonances of the peptide backbone: H^N(i), N(i), Cα(i), Hα(i), Cα(i-1) Hα(i-1), C′ (i) and C′ (i-1).

Further developments involved including the chemical shifts of side-chain carbon and proton spins to achieve the sequential assignment. In these experiments chemical shifts of sidechain resonances are correlated to the amide proton by combinations of the 3D experiments CBCA(CO)NH [51] and CBCANH [52] (fig. 2.3). Information about the chemical shifts of the Cα and Cβ carbons is especially valuable for the assignment process, since they are characteristic of the different types of amino acids and can therefore help to position a sequentially connected stretch of amino acids within the known primary sequence of the protein [53]. From the combination of CBCA(CO)NH and CBCANH experiments the backbone resonance assignments and the sequential connectivities can be obtained. These experiments will be sensitive enough for proteins < 130 residues and provide the Cα and Cβ chemical shifts to establish the sequential link between neighboring residues. Further, the Cα and Cβ chemical shifts provide important information about the amino acid type and the secondary structure. In addition to sequential assignments, these two experiments can, in principle, provide the secondary structure of the protein.

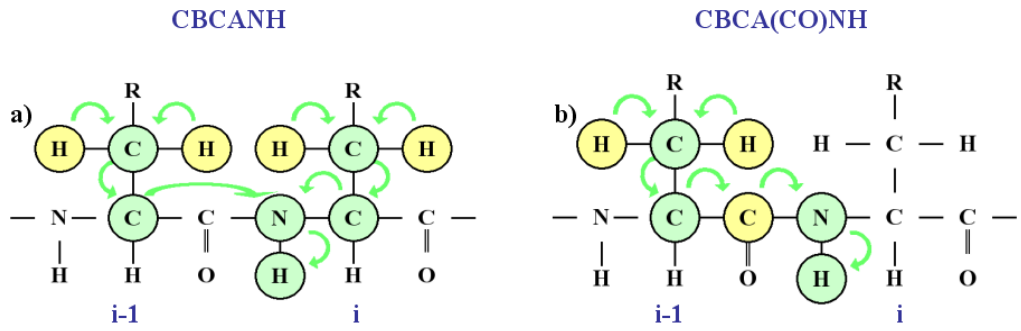


Fig. 2.3 **a)** The CBCANH experiment shows the correlations of N_i and H^N_i to both intra-residual $C\alpha_i$ and $C\beta_i$, as well as sequential correlations to the $C\alpha_{i-1}$ and $C\beta_{i-1}$ of the previous residue; **b)** the CBCA(CO)NH experiment only shows the correlations of N_i and H^N_i with $C\alpha_{i-1}$ and $C\beta_{i-1}$ of the residue i-1.

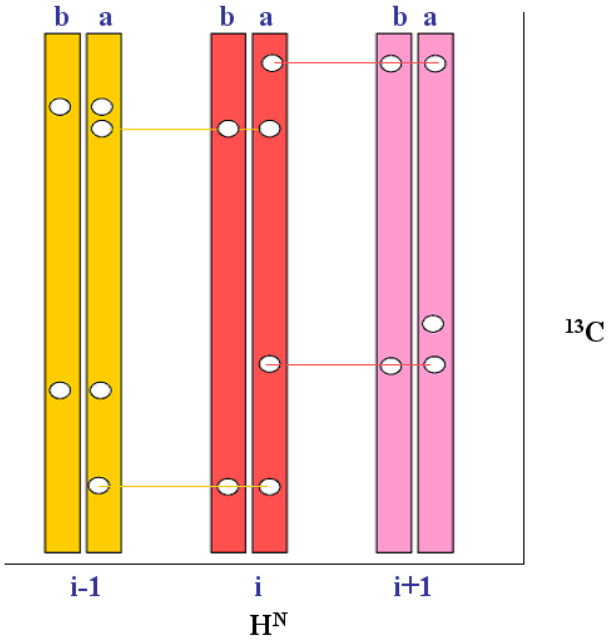


Fig. 2.4 Sequential assignment using CBCANH (a) and CBCA(CO)NH (b) spectra.

The Hα/β and Cα/β chemical shifts of the side chains are then used to link the side chain spin systems to the backbone assignments using 3D HCCH-TOCSY experiments.

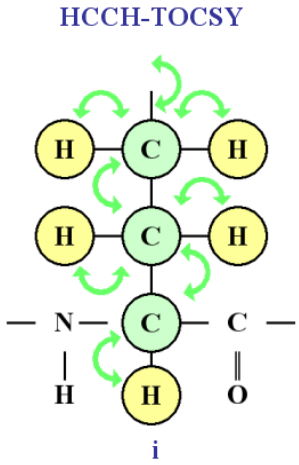


Fig. 2.5 The HCCH-TOCSY experiment provides nearly complete assignments of all aliphatic ^{1}H and ^{13}C resonances. In this experiment the magnetization is transferred via direct ^{1}J coupling from a sidechain (or backbone) proton to the direct attached carbon atom, to the neighboring carbon atoms and finally to their attached protons.

After the assignment of all or nearly all resonances of a protein, experiments for the extraction of structural parameters are analyzed. The most important parameter for NMR-based structure determination are ^{1}H, ^{1}H distances which are derived from NOE intensities. With increasing size of the molecules, overlap of cross peaks becomes a problem also for the extraction of NOE intensities and internuclear distances from 2D NOESY spectra. This overlap can be removed by the introduction of additional frequency dimensions [34]. Resolving the NOESY spectrum along a heteronuclear dimension yields a large gain in resolution, while the number of peaks remains constant. The available experiments for this purpose are ^{13}C or ^{15}N-HSQC NOESY.

2.3 Constraints for Protein Solution Structure Determination

The solution structure is obtained by measuring a sufficient number of structural constraints. These are structural parameters that can be either derived from NMR processed data or acquired directly. Various types of structural constraints can be obtained by exploiting interactions between nuclei of the molecule. The structural information depends on the constraint type. It is possible to extract nuclear-nuclear distances (generally between hydrogen atoms), the reciprocal orientation of chemical bonds and internuclear vectors (dihedral angles) or the orientation relative to a Cartesian axis system within the protein frame.

2.3.1 Nuclear Overhauser Effects

The NMR method for protein structure determination relies on a dense network of distance restraints derived from nuclear Overhauser effects (NOEs) between nearby hydrogen atoms in the protein [3]. NOEs are the essential NMR data

to define the secondary and tertiary structure of a protein because they connect pairs of hydrogen atoms separated by less than about 6 Å in amino acid residues that may be far away along the protein sequence but close together in space.

The NOE reflects the transfer of magnetization between spins coupled by the dipole-dipole interaction in a molecule that undergoes Brownian motion in a liquid [54-56]. The intensity of a NOE, i.e. the volume V of the corresponding cross peak in a NOESY spectrum [55,57], is related to the distance *r* between the two interacting spins by

(1) $V = \langle r^{-6} \rangle f(\tau_c)$

The averaging indicates that in molecules with inherent flexibility the distance *r* may vary and thus has to be averaged appropriately. The remaining dependence of the magnetization transfer on the motion enters through the function $f(\tau_c)$ that includes effects of global and internal motions of the molecule. Since, with the exceptions of the protein surface and disordered segments of the polypeptide chain, globular proteins are relatively rigid, it is generally assumed that there exists a single rigid conformation that is compatible with all NOE data simultaneously, provided that the NOE data are interpreted in a conservative, semi-quantitative manner [3].

More sophisticated treatments that take into account that the result of a NOESY experiment represents an average over time and space are usually deferred to the structure refinement stage [58].

In principle, all hydrogen atoms of a protein form a single network of spins, coupled by the dipole-dipole interaction. Magnetization can be transferred from one spin to another not only directly but also by 'spin diffusion', i.e. indirectly *via* other spins in the vicinity [55,59]. The approximation of isolated spin pairs is valid only for very short mixing times in the NOESY experiment. However, the mixing time cannot be made arbitrarily short because (in the limit of short mixing times) the intensity of a NOE is proportional to the mixing time. In practice, a compromise has to be made between the suppression of spin diffusion and sufficient cross peak

intensities, usually with mixing times in the range of 60-120 ms for high-quality structures.

Spin diffusion effects can be included in the structure calculation by complete relaxation matrix refinement [60-61]. Because also parameters about internal and overall motions that are difficult to measure experimentally enter into the relaxation matrix refinement, care has to be taken not to bias the structure determination by overinterpretation of the data. Since the line-widths can vary appreciably for different resonances, cross peak volumes should be determined by integration over the peak area rather than by measuring peak heights. While the reliable quantification of NOEs is important to obtain a high-quality protein structure, one should also keep in mind that, according to equation (1), the relative error of the distance estimate is only one sixth of the relative error of the volume determination. On the basis of equation (1), NOEs are usually treated as upper bounds on interatomic distances rather than as precise distance restraints because the presence of internal motions and, possibly, chemical exchange may diminish the strength of an NOE [62]. In fact, much of the robustness of the NMR structure determination method is due to the use of upper distance bounds instead of exact distance restraints in conjunction with the observation that internal motions and exchange effects usually reduce rather than increase the NOEs [3]. For the same reason, the absence of a NOE is in general not interpreted as a lower bound on the distance between the two interacting spins. Upper bounds u on the distance between two hydrogen atoms are derived from the corresponding NOESY cross peak volumes V according to 'calibration curves', $V=f(u)$. Assuming a rigid molecule, the calibration curve is

$$V = \frac{k}{u^6} \qquad (2)$$

with a constant k that depends on the arbitrary scaling of the NOESY spectrum. The value u obtained from equation (2) may either be used directly as an upper distance bound, or NOEs may be calibrated into different classes according to their volume, using the same upper bound u for all NOEs in a given class.

The constant *k* in equation 2 can be determined on the basis of known distances belonging to regular secondary structure elements, or by reference to a preliminary structure.

NOEs that involve groups of protons with degenerate chemical shifts, in particular methyl groups, are commonly referred to pseudoatoms located in the centre of the protons that they represent, and the upper bound is increased by a pseudoatom correction equal to the proton-pseudoatom distance. Sometimes, especially in the case of nucleic acid structure determination, where the standard, conservative interpretation of NOEs might not be sufficient to obtain a well-defined structure, also lower distance limits have been attributed to NOEs, either based on the intensity of the NOE or to reflect the absence of a corresponding cross peak in the NOESY spectrum.
But this was not the case in proteins.

2.3.2 Scalar Coupling Constants

Structural information about torsion angles can be obtained from scalar coupling three-bond $^3J_{HX}$ constants. The magnitude of these constants is related to dihedral angles subtended by the covalent bonds that connect the coupled nuclei H and X, where X can be ^1H, ^{15}N or ^{13}C. According to Karplus [63], the dependence of the three-bond J coupling constant on the dihedral angle θ subtended by the three successive covalent bonds that connect the coupled nuclei is embodied in the relation:

$$^3J(\theta) = A\cos^2(\theta) + B\cos(\theta) + C \qquad (3)$$

where θ = torsion angle (ϕ, φ, χ etc); *A*, *B* and *C* are Karplus constants depending on the type of the torsion angle. $^3J_{HNH\alpha}$ coupling constants (ϕ torsion angle) are obtained from the ratio between the intensity of the diagonal peak and that

of the cross-peak of the HNHA map. Through the analysis of the HNHB spectrum the $^{3}J_{HNH\alpha(i-1)}$ and $^{3}J_{HNH\beta}$ coupling constants (φ and χ torsion angles, respectively) can be derived.

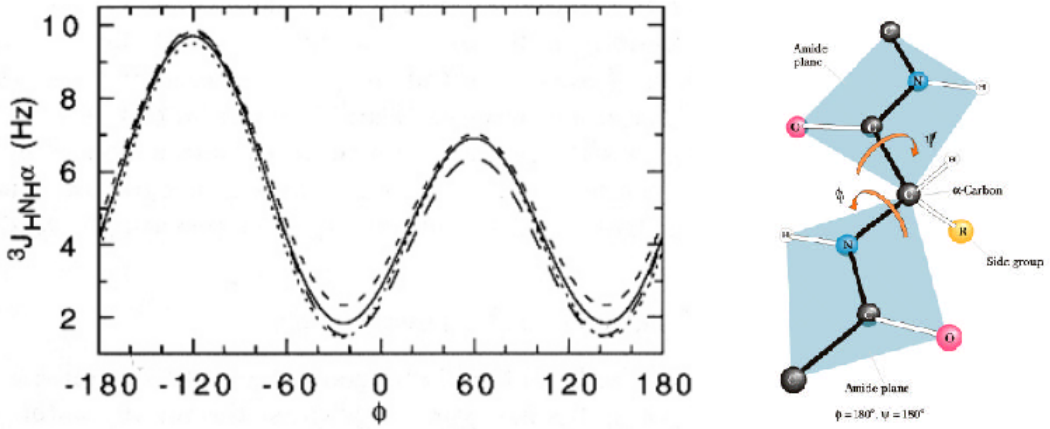


Fig. 2.3 Karplus relation describing the dependence of coupling constants $^{3}J_{HNHa}$ for the ϕ torsion angle.

2.3.2 Hydrogen Bonds

When a protein is expressed and purified in protic solvents (water-based buffers) the backbone amides HN are all protonated. When the protein is then quickly transferred to D_2O-based buffer, the NH groups begin to exchange to ND.

Under typical NMR conditions, exchange occurs within seconds for exposed NH groups. However, exchange may be as slow as hours or days if the NH groups are protected in hydrogen bonds or buried within the protein structure. Typically, an H-D exchange data set is collected and the amide protons remaining after a few hours are identified. In early stages of structure calculations, it becomes apparent what the hydrogen-bonding partners of these groups are.

The hydrogen bonds are then input as restraints for subsequent rounds of calculations.

2.3.3 Chemical Shifts

Chemical shifts are very sensitive probes of the molecular environment of a spin.

However, in many cases their dependence on the structure is complicated and either not fully understood or too intricate to allow the derivation of reliable conformational restraints [64,65]. An exception in this respect are the deviations of $^{13}C\alpha$ (and, to some extent, $^{13}C\beta$) chemical shifts from their random coil values that are correlated with the local backbone conformation [66,67].

$^{13}C\alpha$ chemical shifts larger than the random coil values tend to occur for amino acid residues in α-helical conformation, whereas deviations towards smaller values are observed for residues in β-sheet conformation. Such information can be included in a structure calculation by restricting the local conformation of a residue to the α–helical or β-sheet region of the Ramachandran plot, either through torsion angle restraints [68] or by a special potential [69] although care should be applied because the correlation between chemical shift deviation and structure is not perfect.

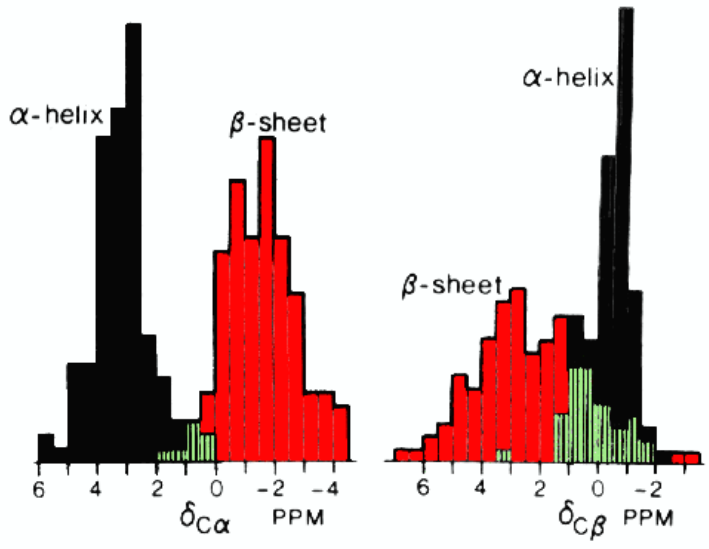


Fig. 2.4. Relation between the secondary structure and chemical shift deviations from random coil values of Cα and Cβ.

2.3.4 PCS and RDC

In the case of systems containing paramagnetic centers (like metalloproteins), it is possible to use the electron-nucleus interaction as a further source of structural information. The through-space contribution to nuclear relaxation depends on the distance between the metal ion (M) and the resonating nucleus (H), as the reciprocal of the sixth power of the distance. The *contact shifts* are used as constraints as they depend on the metal-donor-carbon-proton (M-D-C-H) dihedral angle, in the case of protons attached to an sp^3 carbon of a ligand of the metal ion.

The occurrence of magnetic anisotropy causes the dipolar coupling between the electron and the resonating nucleus not to average to zero in solution. This is because the electron magnetic moment changes with the orientation of the molecule with respect to the external magnetic field. This residual dipolar coupling provides a contribution to the nuclear chemical shift which is called *pseudocontact shift* (PCS) [70]. It is easily determined by measuring the chemical shift of the nuclei close to the metal but separated from it by enough chemical bonds that there is no unpaired electron spin density delocalized from the metal ion. To obtain the PCS values from these shifts it is necessary to subtract the corresponding shifts measured in an otherwise identical diamagnetic analog.

PCS are precious long-range constraints. In fact, the contribution to the shift due to the through-space coupling depends on the distance between the metal ion and the resonating nucleus (as the reciprocal of the third power of the distance) as well as on the magnetic susceptibility anisotropy of the metal ion and on the polar coordinates of the resonating nucleus with respect to the axis system formed by the principal directions of the magnetic susceptibility tensor.

PCS have been used from the early steps of structural calculations as well as for structure refinements. Furthermore, they provide a unique tool for the location of the metal ion in the protein frame, without any assumptions. As a result of the application of PCS, information on the magnetic susceptibility tensor of the metal (χ^{metal}) is obtained.

The occurrence of magnetic anisotropy implies that the molecular tumbling in solution is not random at high magnetic fields. Instead, a partial orientation occurs. This provides *residual dipolar couplings* (RDC) [71] between nuclei, having values different from zero. They can be conveniently detected by measuring ^{15}N-^{1}H ^{1}J (as well as ^{13}C-^{1}H ^{1}J) at variable magnetic fields. Such contributions can yield powerful structural constraints, in particular for paramagnetic molecules with large magnetic anisotropy.

2.4 Structure Calculation

Using NMR (distances and angles) constrains, calculation programs fold a random generated 3D structure in order to maximize the agreement between the structure and the structural constrains. As input for the programs the primary structure of the protein is taken, and all bond lengths and angles are taken fixed to their ideal values or allowed to deviate only slightly from these values. As the structural constrains used for calculations are not strict values and in some regions of protein the experimentally number constrains obtained is not sufficient to determine the local conformation, several different conformations may be calculated to satisfy the imposed constrains.

In general, it is preferred to have a large number of constraints of different types and to use a relatively large tolerance in order to minimize effects arising from miscalibration of the constraints. Indeed, if protocols are developed for applying new, non-classical constraints to structure calculations, improvement in the structure accuracy and precision can be achieved. In particular, inclusion of non-classical constraints derived from paramagnetism produces a more precise structure in the region surrounding the metal ion, and increases the accuracy.

The structural indetermination is a physical feature of the protein.
The NMR constraints only reflect the average structure. A NMR structure is thus represented by a family of conformers, usually composed by 20 to 50 members,

which are in good agreement with the structural constrains imposed. The precision of the structure is measured by the *root-mean-square-deviation* (RMSD) of the coordinates of the protein atoms for each conformer of the family from the mean structure.

The local RMSD value varies along the polypeptide chain. High local RMSD values can be due to the lack of experimental constraints, to actual disorder of the structure in solution or both. The accuracy of the structure is measured by an average *target function* of the family.

LIST OF REFERENCES FOR CHAPTERS 2

1. Bernstein F.C., Koetzle T.F., Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. (1977) *Eur J Biochem*. 80, 319.

2. Drenth J., (1994) *Principles of Protein X-ray Crystallography*. New York: Springer-Verlag.

3. Wüthrich K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York

4. Kossiakoff A.A., Randal M., Guenot M. & Eigenbrot C. (1992) *Proteins Struct. Funct.Genet.* 14, 65.

5. Ernst R.R., Bodenhausen G. & Wokaun A. (1987) *The Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Oxford: Clarendon Press.

6. Cavanagh J., Fairbrother W. J., Palmer III A. G. & Skelton N. (1996) *Protein NMR spectroscopy. Principles and Practice*. San Diego: Academic Press.

7. Clore G.M. & Gronenborn A.M. (1990) *Prog. NMR Spectrosc*. 23, 43.

8. Edison A. S., Abildgaard F., Westler W. M., Mooberry E. S. &Markley J.L. (1994) *Meth. Enzymol*. 239, 3.

9. Arseniev A. S., Kondakov V. I., Maiorov V.N. & Bystrov V.F. (1984) *FEBS Lett*. 165, 57.

10. Braun W., Bosch C., Brown L. R., Go N. &Wuthrich K. (1981). *Biochim. Biophys. Acta* 667, 377.

11. Clore G. M., Nilges M., Sukumaran D. K., Brunger A. T., Karplus M. & Gronenborn A.M. (1986) *EMBO J.* 5, 2729.

12. Williamson M. P., Havel T. F. & Wuthrich K. (1985) *J. Mol. Biol*. 182, 295.

13. Zuiderweg E. R. P., Billeter M., Boelens R., Scheek R. M., Wuthrich K. & Kaptein R. (1984) *FEBS Lett*. 174, 243.

14. Griesinger C.,. Sørensen O.W, Ernst R.R., (1987) *J. Magn. Reson*.73, 574.

15. Vuister G.W., Boelens R., Kaptein R., (1988) *J. Magn. Reson*. 80, 176.

16. Griesinger C., Sørensen O.W., Ernst R.R., (1989) *J. Magn. Reson.*84, 14.

17. Kay L.E., Clore G.M., Bax A., Gronenborn A.M., (1990) *Science* 249, 411.

18. Ernst R.R., (1992) *Angew. Chem*. 104, 817.

19. McIntosh L.P., Griffey R.H., Muchmore D.C., Nielson C.P., Redfield A.G., Dahlquist F.W., (1984) *Proc. Natl. Acad. Sci.* 84, 1244.

20. Griffey R.H., Loomis R.E., Redfield A.G., McIntosh L.P., Oas T.G., Dahlquist F.W., (1986) *J. Am. Chem. Soc*. 108, 6816.

21. McIntosh L.P., Dahlquist F.W., Redfield A.G., (1987) *J. Biomol. Struct. Dyn*. 5 21.

22. Oh B.H., Westler W.M., Derba P., Markley J.L., (1988) *Science* 240, 821.

23. Westler W.M., Stockman B.J., Markley J.L., Hosoya Y., Miyake Y., Kainosho M., (1988) *J. Am. Chem.* 110, 6256.

24. Stockman B.J., Reilly M.D., Westler W.M., Ulrich E.L., Markley J.L., (1989) *Biochemistry* 28, 230.

25. Muchmore D.C., McIntosh L.P., Russell C.B., Anderson D.E., Dahlquist F.W., (1989) *Methods Enzymol*. 177, 44.

26. Venters R.A., Calderone T.L., Spicer L.D., Fierke C.A., (1991) *Biochemistry* 30, 4491.

27. Hansen A.P., Petros A.M.,. Mazar A.P, Pederson T.M., Rueter A., Fesik S.W., (1992) *Biochemistry* 31,12713.

28. Kigawa T., Muto Y., Yokoyama S., (1995) *J. Biomol. NMR* 6 129.

29. Lustbader J.W., Birken S., Pollak S., Pound A.,. Chait B.T, Mirza U.A., Ramnarain S.,. Canfield R.E, Brown J.M., (1996) *J. Biomol. NMR* 7, 295.

30. Clore G.M., Gronenborn A.M., (1991) *Science* 252, 1390.

31. Weber C., Wider G., von Freyberg B., Traber R., Braun W., Widmer H., Wuthrich K., (1991) *Biochemistry* 30, 6563.

32. Fesik S.W., Gampe R.T.J., Eaton H L., Gemmecker G., Olejniczak E.T., Neri P., Holzman T.F., Egan D.A., Edalji R., Simmer R., Helfrich R., Hochlowski J., Jackson M., (1991) *Biochemistry* 30, 6574.

33. Ikura M., Clore G.M., Gronenborn A.M., Zhu G., Klee C.B., Bax A., (1992) *Science* 256, 633.

34. Fesik S.W., Zuiderweg E.R.P., (1988) *J. Magn Reson.* 78, 588.

35. Wuthrich K., *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York, 1986.

36. Marion D., Driscoll P.C., Kay L.E., Wingfield P.T., Bax A., Gronenborn A.M., Clore G. M., (1989) *Biochemistry* 28, 6150.

37. Wijmenga S.S., Hallenga K., Hilbers C.W., (1989) *J. Magn. Reson.* 84, 634.

38. Driscoll P.C., Clore G.M., Marion D., Wingfield P.T., Gronenborn A.M., (1990) *Biochemistry* 29, 3542.

39. Driscoll P.C., Gronenborn A.M., Wingfield P.T., Clore G.M., (1990) *Biochemistry* 29, 4668.

40. Shon K., Opella S.J., (1989) *J. Magn. Reson.* 82, 193.

41 Marion D., Kay L.E., Sparks S.W., Torchia D.A., Bax A., (1989) *J. Am. Chem. Soc*. 111, 1515.

42. Zuiderweg E.R.P., Fesik S.W., (1989) *Biochemistry* 28, 2387.

43. Ikura M., Kay L.E., Tschudin R., Bax A., (1990) *J. Magn. Reson*. 86, 204.

44. Zuiderweg E.R.P., McIntosh L.P., Dahlquist F.W., Fesik S. W., (1990) *J. Magn. Reson.* 86, 210.

45. Kay L.E., Ikura M., Tschudin R., Bax A., (1990) *J. Magn. Reson*. 89, 496.

46. Ikura M., Kay L.E., Bax A., (1990) *Biochemistry* 29, 4659.

47. Kay L.E., Ikura M., Tschudin R., Bax A., (1990) *J. Magn. Reson*. 89, 496.

48. Kay L.E., Ikura M., Zhu G., Bax A., (1991) *J. Magn. Reson*. 91, 422.

49. Powers R., Gronenborn A.M., Clore G.M., Bax A., (1991) *J. Magn. Reson*. 94, 209.

50. Grzesiek S., Bax A., (1992) *J. Magn. Reson*. 96, 432.

51. Grzesiek S., Bax A., (1992) *J. Am. Chem. Soc*. 114, 6291.

52. Grzesiek S., Bax A., (1992) *J. Magn. Reson*. 99, 201.

53. Grzesiek S., Bax A., (1993) *J. Biomol. NMR* 3, 185.

54. Solomon I. (1955) *Phys.Rev*. 99, 559.

55. Macura S. and Ernst R.R. (1980) *Mol.Phys*. 41, 95.

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

57. Jeener J., Meier B.H., Bachmann P., and Ernst R.R. (1979) *J.Chem.Phys.* 71, 4546.

58. Torda A.E., Scheek R.M., and Van Gunsteren W.F. (1990) *J.Mol.Biol.* 214, 223.

59. Kalk A. and Berendsen H.J.C. (1976) *J.Magn.Reson.* 24, 343.

60. Keepers J.W. and James T.L. (1984) *J.Magn.Reson.* 57, 404.

61. Yip P. and Case D.A. (1989) *J.Magn.Reson.* 83, 643.

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

63. Karplus M. (1963) *J.Am.Chem.Soc.* 85, 2870.

64. Oldfield E. (1995) *Protein Sci.* 5, 217.

65. Williamson M. P. & Asakura T. (1997) Protein chemical shifts. In *Protein NMR techniques* (ed. D. G. Reid), Totowa, NJ: Humana Press.

66. Spera S. & Bax A. (1991) *J. Amer. Chem. Soc*., 113, 5490.

67 de Dios A. C., Pearson J.G. & Oldfield E. (1993) *Science* 260, 1491.

68 Luginbuhl P., Szyperski T. & Wuthrich K. (1995) *J. Magn. Reson*. B109, 229.

69. Kuszewski J., Qin J., Gronenborn A.M. & Clore G.M. (1995) *J. Magn. Reson*. B106, 92.

70. Kurland R.J., McGarvey B.R. (1970), *J. Magn. Reson*., 2, 286.

71. Tjandra N. & Bax A. (1997) *Science* 278, 1111.

CHAPTER 3

STRUCTURAL AND DYNAMIC NMR CHARACTERIZATION OF CALMODULIN

3.1 Ca^{2+}-binding Proteins

Calcium (Ca^{2+}) is a intracellular secondary messenger responsible for controlling and regulating a wide variety of cellular processes, such as nucleotide metabolism, muscle contraction, cellcycle control, differentiation and signal transduction.

The cellular free Ca^{2+} concentration of living cells is tightly controlled within the range of 10^{-7} – 10^{-5} M via intricate Ca^{2+} influx and efflux mechanisms [1,2]. Upon external stimulation, the intracellular Ca^{2+} concentration transiently increases to ~ 10^{-5} M, resulting in the calcification of an array of Ca^{2+}-binding proteins (CBPs).

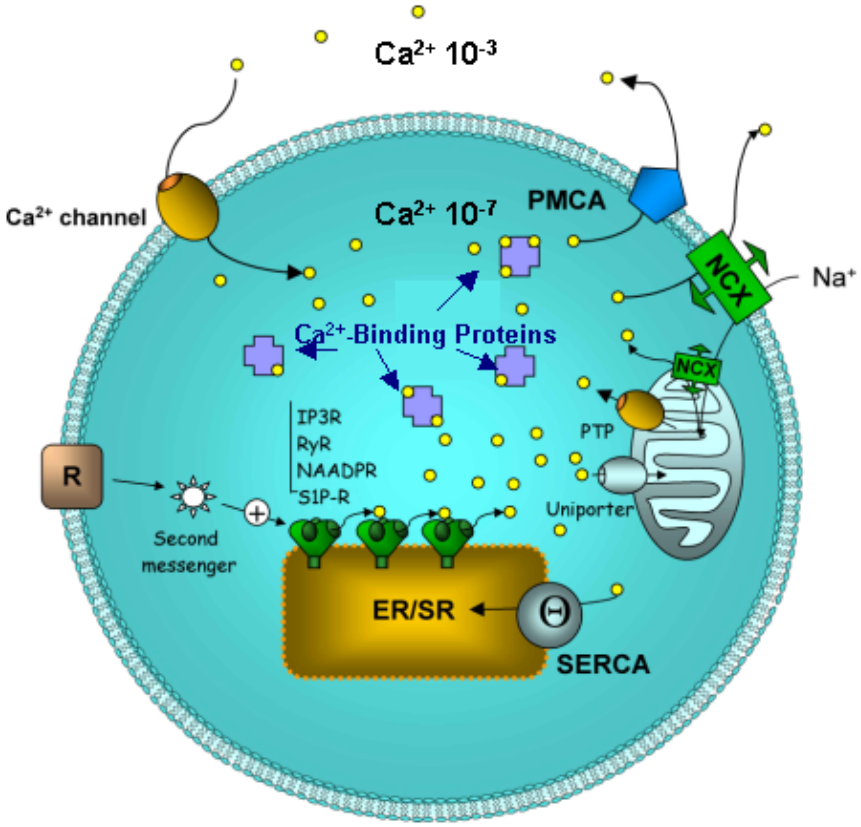


Fig. 3.1 Mechanisms of calcium signalling

Cell stimulation by agonists activate the formation of second messengers that induce the release of Ca^{2+} stored in the endoplasmic/sarcoplasmic reticulum (ER/SR) through the IP_3 receptor (IP_3R), the ryanodine receptor (RyR), the NAADP receptor (NAADPR) or the sphingosine 1-phosphate binding site (S1P-R). In addition, agonists stimulate the entry of extracellular Ca^{2+} through plasma membrane channels. Most of the cytosolic Ca^{2+} is bound to buffers, and only a small percentage binds to effectors and activates cellular functions. Ca^{2+} removal is mediated by various pumps and exchangers, including the Na^{+}/Ca^{2+} exchanger (NCX) and the plasma membrane Ca^{2+} ATPase (PMCA) that extrude Ca^{2+} from the cytosol and the sarcoendoplasmic reticulum Ca^{2+}-ATPase (SERCA) that pumps Ca^{2+} back to the ER/SR. During the Ca^{2+} signal mitochondria sequester Ca^{2+} through an uniporter that might be then released slowly into the cytoplasm through the NCX or the permeability transition pore (PTP).

The Ca^{2+} signaling pathway involves numerous CBPs, which can be divided into Ca^{2+} buffers and Ca^{2+} sensors, on the basis of their main functions. Cytosolic buffers are involved in shaping both the amplitude and duration of Ca^{2+} signals . Buffers also limit the spatial spreading of Ca^{2+} signals. This is particularly important

in neurons that contain high concentrations of buffers such as Parvalbumin and Calbindin, which ensure that Ca^{2+} signals are largely confined to synapses.

The second class, Ca^{2+} sensors proteins, comprises proteins that bind to Ca^{2+} for the specific purpose to decode its information, and doing so their real task is that of transmitting suitably processed Ca^{2+} signals to the desired targets.

The more common sensors are Troponin C and Calmodulin that respond to an increase in Ca^{2+} by activating diverse processes.

Troponin C has a limited function to control the interaction of actin and myosin during the contraction cycle of cardiac and skeletal muscle. Calmodulin (CaM) is used more generally to regulate many processes such as the contraction of smooth muscle, gene transcription, ion channel modulation and metabolism. The same cell can use different sensors to regulate separate processes. In skeletal muscle, for example, Troponin C regulates contraction whereas CaM stimulates the phosphorylase kinase to ensure a parallel increase in ATP production.

The binding of Ca^{2+} results in conformational changes of these CBPs, thus enabling these proteins to interact with their respective partner proteins and enzymes. The formation of complexes between CBPs and their targets induces further conformational changes in both components of the complexes, thereby allowing functional regulation of the targets. Therefore, CBPs play pivotal roles in Ca^{2+}-mediated signal transduction pathways.

3.2 EF-hand Motif

Calcium interacts with a very large number of proteins, but the variety of configurations of calcium binding sites is rather limited. The most common among these is the EF-hand motif.

The name EF-hand was devised by Kretsinger and Nockolds as a graphical description of the calcium-binding motif observed in parvalbumin [3] (fig. 3.2).

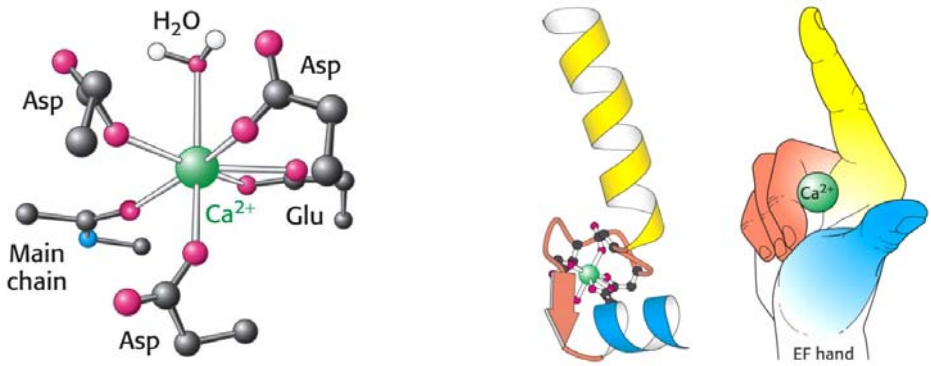


Fig. 3.2 The EF-hand Ca^{2+}-binding motif
In each EF-hand motif, two alpha-helices, oriented at a approx. 90° angle, flank a 12 amino acid-containing loop. The Ca^{2+} ion is coordinated by seven oxygen ligands in a pentagonal bipyramid arrangement; five amino acid residues at positions X,Y,Z,-X and -Y provide one oxygen ligand, whereas the residues at position -Z provide two.

The canonical sequence of the EF-hand motif is rather short (about 30 residues) and consists of a helix-loop-helix motif characterized by a binding loop of 12 residues with the pattern +X•+Y•+Z•-Y•-X••-Z, where ±X; ±Y; and ±Z are the ligands that participate in the metal coordination and the dots represent intervening residues. At positions X and Y, we usually find the sidechains of aspartic acid or asparagine; the sidechains of aspartic acid, asparagine or serine are found at Z and a peptide carbonyl oxygen lies at –Y. –X is usually a water molecule and –Z is a conserved bidentate ligand, glutamic acid or aspartic acid.

The two helices are arranged similar to the extended thumb and index finger of a hand.

This relatively simple structural motif can accommodate calcium and magnesium with differences in the geometry of the coordination and in the specific affinity [4]. Magnesium is usually bound by six ligands in an octahedron, whereas seven ligands at the vertices of a pentagonal bipyramid coordinate calcium.

The function of some EF-hand proteins, such as Parvalbumin in muscle physiology and CaM in the cytosol, depends on the difference in affinity for calcium and magnesium [5]. The Ca^{2+}-binding affinities of EF-hand proteins vary substantially (K_d = 10^{-4}-10^{-9} M) and are amino acid sequence dependent, especially with regard to the 12-residue consensus loop sequence that provides all the amino

acids that directly ligate a Ca^{2+} ion [6]. In many proteins, EF-hands exist as a pair, with a short antiparallel β-sheet interaction between the two Ca^{2+}-binding loops. *Zhang et al.* showed that concerted movement of the helices in these EF-hand pair is crucial for the cooperativity of Ca^{2+}-binding [7]. The EF-hand motif has been detected in a number of small proteins (e.g. CaM or S100) and within domains of much larger complex proteins (i.e. Myosin or Calpain).

In most known cases, EF-hand motifs occur in adjacent pairs: Parvalbumin and S100 proteins represent the minimal motif. The Eps15 homology (EH) domain is also formed by a pair of EF-hands and the structures of several EH domains of Eps15 show a fold similar to that found in S100 proteins. Proteins containing four EF-hand motifs usually have two domains, each formed by a pair of EF-hands, separated by a flexible linker, which can be extended in the classical dumbbell structures of CaM and Troponin C. One exception to the rule of EF-hand pairing is Calpain, which has five EF-hand motifs in the C-terminal part of its large subunit. However, Calpain forms a heterodimer with a small subunit composed of another five EF-hand motifs.

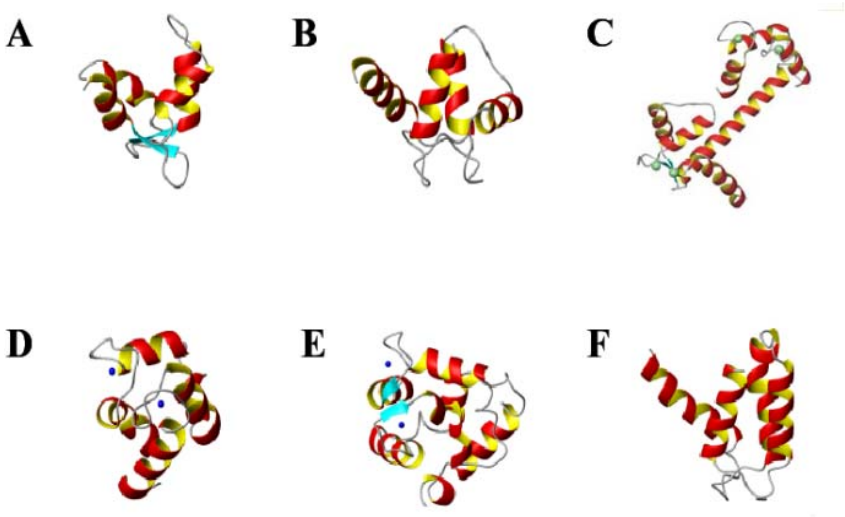


Fig. 3.3 Ribbon representation of the structures of a number of proteins of the EF-hand family: **A)** Myosin, **B)** Troponin C, **C)** Calmodulin, **D)** Calbindin D9K , **E)** Parvalbumin, **F)** S100B

3.3 Domain Organization and Functional Roles

The concept of regulatory and structural domains was first introduced for Troponin C, in which the high-affinity carboxy-terminal domain (structural) always binds Ca^{2+} in the muscle cell, and the low-affinity amino-terminal domain (regulatory) triggers the Ca^{2+} signal leading to muscle contraction [8]. EF-hand proteins with regulatory roles are often termed Ca^{2+}-sensor proteins, whereas those involved in Ca^{2+} buffer and transport functions are termed Ca^{2+}-buffer proteins [9]. Most EF-hand proteins consist of single, or multiple, pairs of the helix-loop-helix motif. Typically, a pair of EF-hand motifs forms a globular domain, while a protein containing four motifs is composed by two domains that can be either structurally independent, as observed in the structures of CaM and Troponin C, or well packed, resulting in a single oval shape, as seen in the structure of Recoverin. Some domains have a single, functional Ca^{2+}-binding site, for example, the amino- and carboxy-terminal domains of Recoverin and Myosin light chains.

EF-hand proteins undergo conformational changes upon binding of Ca^{2+}. This event is a crucial step in many Ca^{2+}-dependent cellular processes.

The high number of three-dimensional structures of various EF-hand proteins have led much insights into the structure and functional relationship of this large family of Ca^{2+}-binding proteins and have provided valuable information about this Ca^{2+}-induced conformational change. Indeed, both the carboxy- and amino-terminal domains of CaM undergo large structural rearrangements upon binding of Ca^{2+} from the ‘closed’ conformation (the two helices of each EF-hand are almost anti-parallel) to the ‘open’ conformation (the two helices are more perpendicular) [7]. By contrast, Calbindin D9K, retains the Ca^{2+}-free ‘closed’ conformation even after binding Ca^{2+} [10]. Indeed, the single domain of Calbindin D9K exhibits small conformational changes in response to binding of Ca^{2+}, which may be responsible for the structural stability and function in buffering intracellular Ca^{2+}. Thus, this type of domain can be referred to as a ‘structural’ or ‘buffer’ domain.

3.4 Calmodulin : Ca^{2+}-modulated Protein

Calmodulin is a primary Ca^{2+} -receptor protein. It is a small, 148 amino acid residues, highly conserved protein molecule found in all eukaryotes. The protein is able to bind hundred of target proteins and enzymes including various protein kinases and phosphatases, receptors, ion-channel proteins, phosphodiesterases, and nitric oxide synthases [11]. These CaM targets play essential roles in a wide variety of cellular events such as cytoskeletal dynamics, metabolism, cell proliferation, and development.

CaM consists of two similar domains, each containing two Ca^{2+}-binding sites. Each metal-ion binding site is comprised of a highly conserved, continuous stretch of 12 residues, the EF-hand helix-loop-helix motif. The residues at positions +X, +Y, +Z, -Y, and -X of this 12-residue loop each provide one ligand, and the side chain of the -Z residue supplies two oxygen ligands to each Ca^{2+} ion. Hence, Ca^{2+} forms a pseudo-bipyramidal coordination sphere in Ca^{2+}-saturated CaM [12].

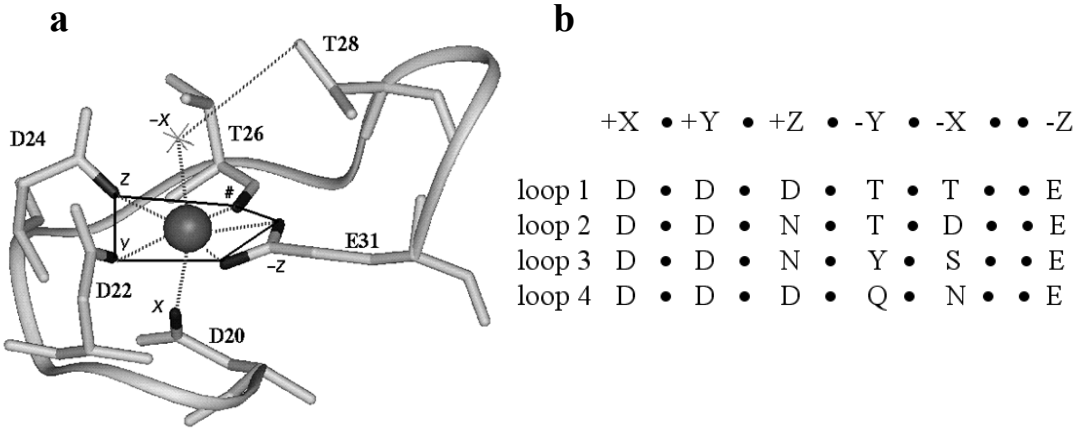


Fig. 3.4 **Calcium coordination in Calmodulin**

a) The first EF-hand in CaM is used to illustrate the pentagonal bipyramidal ligation of the calcium ion. Oxygen atoms ligating calcium are colored black. The water ligand is shown as a star. **b)** The primary structure of the four Ca^{2+}-binding loops of CaM.

The C-terminal domain of CaM cooperatively binds to two Ca^{2+} ions with a K_d of ~10^{-6} M, and the N-terminal pair of EF-hands also cooperatively binds to Ca^{2+} ions, albeit with an approximately 10-fold lower affinity [13-16]. The Ca^{2+}-binding affinity of CaM falls exactly in the range of the intracellular Ca^{2+} concentration fluctuation, ensuring that the protein is capable of responding to cellular Ca^{2+} signaling. In addition, the kinetic parameters of Ca^{2+} binding (k_{on} ~108 s^{-1} and k_{off} ~103 s^{-1}), a diffusion controlled on-rate and a millisecond off-rate) allow the protein to respond in a timely fashion to transient cellular Ca^{2+} concentration fluctuations. Therefore, CaM is ideally suited as a cellular Ca^{2+} signal trigger both thermodynamically and kinetically. The binding of Ca^{2+} to CaM induces large conformational changes in both domains of the protein. The Ca^{2+}-bound form of CaM is competent in binding with high affinity (with a K_d of ~10^{-8} to 10^{-11} M) to its targets, and thereby activating these proteins and enzymes.

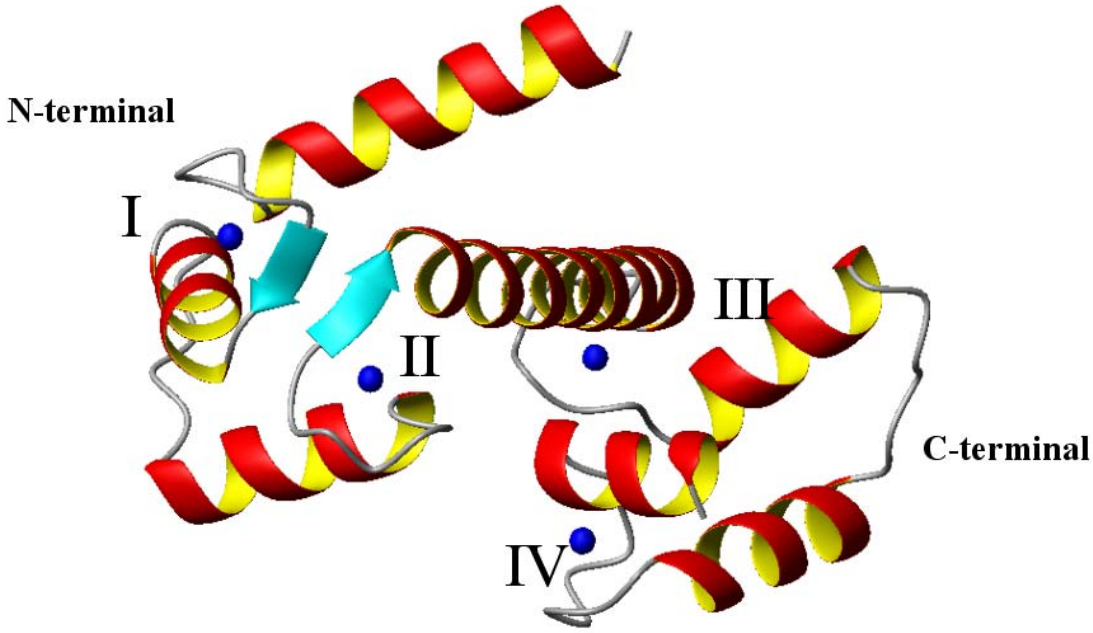


Fig. 3.4 Ribbon representation of the X-ray three-dimensional structure of Ca^{2+}-CaM refined at 2.2 Å resolution [12] (PDB ID code 3CLN).
The numbers are referred to the binding sites of calcium ions: I ,II in the N-terminal domain and III, IV in the C-terminal domain

3.5 Calmodulin is a Highly Flexible System

The crystal structure of Ca^{2+}- CaM has shown that the protein has a dumbbell shape comprising an N- and a C-terminal globular domain, which are connected by a long (27-residues), continuous, solvent-exposed α-helix, the so-called central helix [12] (fig. 3.5a). The two helices in each of the four helix–loop–helix (EF-hand) Ca^{2+}-binding sites are almost perpendicular to each other.

Residues 7–9 of each Ca^{2+}-binding loop form a mini β-strand, and the two β-strands in each domain form an antiparallel sheet structure. Other than forming the antiparallel β-sheet, the helices of two EF-hands pack with each other to form a compact, globular domain structure. In particular, extensive contacts between the first and fourth (helices A–D for the N-terminal domain and E–H for the C-terminal domain) and between the second and third helices (helices B–C and F–G) of each domain are observed.

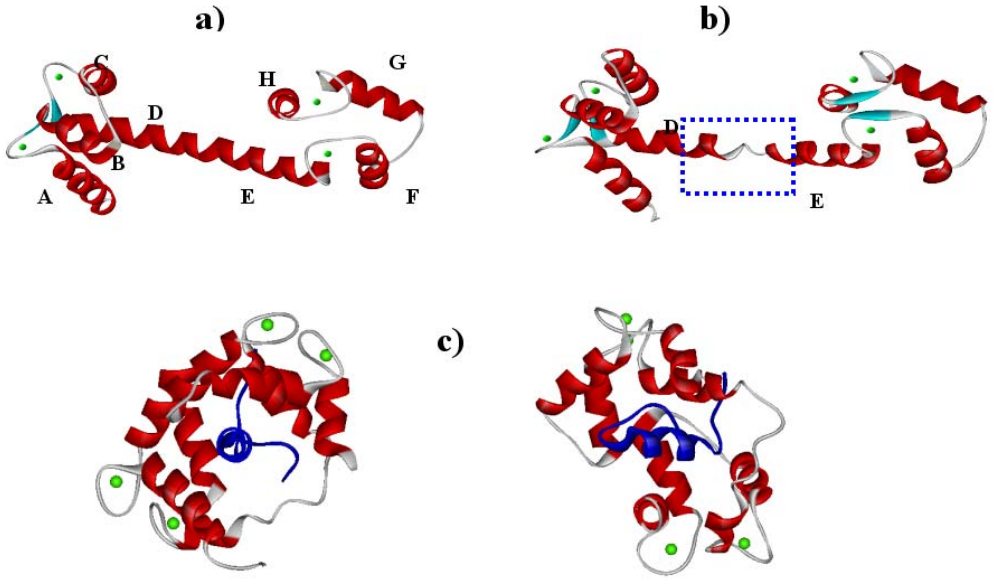


Fig 3.5 a) Ribbon representation of the early X-ray three-dimensional structure (extended conformation) of Ca^{2+}-CaM refined at 2.2 Å resolution (PDB code ID 3CLN); **b)** it shows the disorder in the central linker, between helix D and E, determined from small angle X-ray scattering and fluorescence energy transfer measurements and recently by NMR as well as by X-ray diffraction; **c)** NMR structure (closed conformation) of Ca^{2+}-CaM complex with Calmodulin Dependent Protein Kinase Fragment (shown in blue colour) (PDB code ID 1CKK)

The structure of Ca^{2+}/CaM, determined by X-ray crystallography showed that the two EF-hand pairs forming the C- and N-terminal lobes are connected by a rigid α-helix of 18 residues [17].

The length and expected rigidity of the central helix, as suggested from the crystal structure, does not agree well with data from small angle X-ray scattering and fluorescence energy transfer measurements in solution [18,19]. Additionally, it was shown by NMR as well as by X-ray diffraction data that the observed α-helical conformation of the central linker is a consequence of crystal packing [20,21].

The connecting helix resulted to be an artifact of crystallization. When CaM is in solution, part of the helical rod melts into a flexible disordered linker, allowing the protein to wrap around its target (fig. 3.6c).

Consequently, the N- and C-terminal lobes do not adopt a defined orientation relative to each other in solution, but display a tumbling motion, being held together by the central linker.

CaM undergoes an additional conformational change upon binding to target proteins in which the two hydrophobic pocket regions of CaM are brought near to one another and wrap around an amphiphilic α-helix in the target. In order to accommodate this large conformational change, the central helix unwinds in the middle and acts as a flexible linker that is capable of allowing the two lobes to adopt variable orientations when bound to the target helix [22,23]. This helical flexibility, combined with the inherent deformability of the hydrophobic pockets, is thought to account for the ability of CaM to interact with a variety of different targets in a sequence-independent fashion [24,25]. More generally, the prominent role that conformational dynamics play in both Ca^{2+} binding and target recognition illustrates the importance of structural plasticity in every aspect of CaM function.

It is therefore apparent that CaM has ample conformational freedom in solution, but that the extensive experimental data available do not allow us yet to make reasonable deductions on the nature and variety of the most favored conformers. Our goal during this work of thesis was to explore the conformational space sampled by CaM using a advanced NMR techniques supplied by a paramagnetism based strategy.

3.6 Lanthanide Substitution as a Tool for the Structural Investigation of Human Calmodulin

The conformational space sampled by the two-domain protein CaM has been explored by an approach based on four sets of NMR observables obtained on well-defined lanthanide-substituted proteins.

Lanthanides are known to be spectroscopic probes for calcium-binding proteins. They are widely employed as luminescent probes [26,27] as well as NMR probes. In the latter case, their paramagnetism is exploited to generate new constraints in solution structure determinations [28-34].

In view of these favorable properties, it would be desirable to have stoichiometrically definite and selectively substituted lanthanide-bound proteins [35]. Apparently, this is sometimes a difficult goal, especially in the presence of multiple calcium binding sites. The loop of each EF-hand contains a calcium binding site. Addition of lanthanides to calcium-loaded proteins sometimes provides selective lanthanide substitution [36], but this is not the case for vertebrate CaM [37].

Mutations in the sequence is therefore necessary in order to provide Ca_3Ln-derivatives with one lanthanide selectively substituted in one of the four calcium binding sites. The selective substitution of the calcium center in the second site of the N-terminal domain was obtained by using a N60D mutant, where asparagine N in position 60 was substituted by aspartate D, previously tested and found to increase the affinity of lanthanides for this site [38].

To explore favored orientations of the two domains with respect to each other we based our approach on the exploitation of the long-range restraints that are imposed on the nuclear coordinates (PCS) and internuclear vector orientations (RDC) in one domain (C-terminal domain) by the presence of paramagnetic metal ions with large and different magnetic susceptibility tensor anisotropies in the other domain (N-terminal domain). Four sets of paramagnetic restraints were obtained on Tb^{3+}- and Tm^{3+}- substituted proteins.

$^{15}N,^{13}C$-enriched vertebrate N60D $(Ca_2)_N(Ca_2)_C$CaM was subjected to the standard set of 2D and 3D experiments to assign backbone and side chains, and to obtain NOE. The relevant data are summarized in Table 3.1 e 3.2

	N-terminal domain	**C-terminal domain**
Meaningful NOEs	1495	1225
φ dihedral angles	50	43
ψ dihedral angles	50	43
PCS (Tb^{3+} and Tm^{3+})	125	165
RDC (Tb^{3+} and Tm^{3+})	-	67

Table 3.1 Type and Number of restraints used in the Dyana Calculation

Experiment	**Dimensions of acquired data**			**Spectral width (Hz)**		
	t_1	t_2	t_3	F_1	F_2	F_3
^1H-^{15}N-HSQC	256(^{15}N)	1024(^1H)		2500	11261	
[^1H-^1H]-NOESY	1024(^1H)	2048(^1H)		10416	10416	
CBCANH	128(^{13}C)	48(^{15}N)	2048(^1H)	13404	3548	10416
CBCA(CO)NH	128(^{13}C)	48(^{15}N)	2048(^1H)	13404	3548	10416
HNHA	128(^1H)	48(^{15}N)	1024(^1H)	10416	3548	10416
(H)CCH-TOCSY	128(^{13}C)	80(^{13}C)	2048(^1H)	12500	12500	10416
^1H-^{15}N-HSQC-NOESY	320(^1H)	48(^{15}N)	2048(^1H)	10416	2199	10416
^1H-^{13}C-HSQC-NOESY	128(^1H)	64(^{13}C)	1024(^1H)	10416	14082	10416
^1H-^{15}N-HSQC (PCS)	256(^{15}N)	1024(^1H)		2500	11261	
^1H-^{15}N-HSQC (J_{NH}-split)	256(^{15}N)	1024(^1H)		2500	11261	

Table 3.2 Acquisition parameters for NMR experiments

The structures obtained for the two domains are in good agreement with the latest solution structure refined with the help of RDC [39] resulting from orienting media.

N60D $(CaLn)_N(Ca_2)_C$CaM derivatives with Ln = Tm or Tb were prepared, and an extensive assignment of both N-terminal and C-terminal peptide NH nuclei was performed by using a semiautomated procedure [40,41].

Paramagnetic lanthanides are capable of orienting the protein in high magnetic fields to an extent similar to that obtained by using orienting devices, and each lanthanide orients according to its magnetic susceptibility tensor.

Thus, we use the PCS values also as additional constraints in the structure determination by the program PARAMAGNETIC DYANA [42] together with the obtained magnetic susceptibility anisotropy values. The resulting family of solution structures of the N-terminal domain of CaM containing the N60D mutation (fig. 3.6) is deposited in the Protein Data Bank (PDB ID code 1SW8).

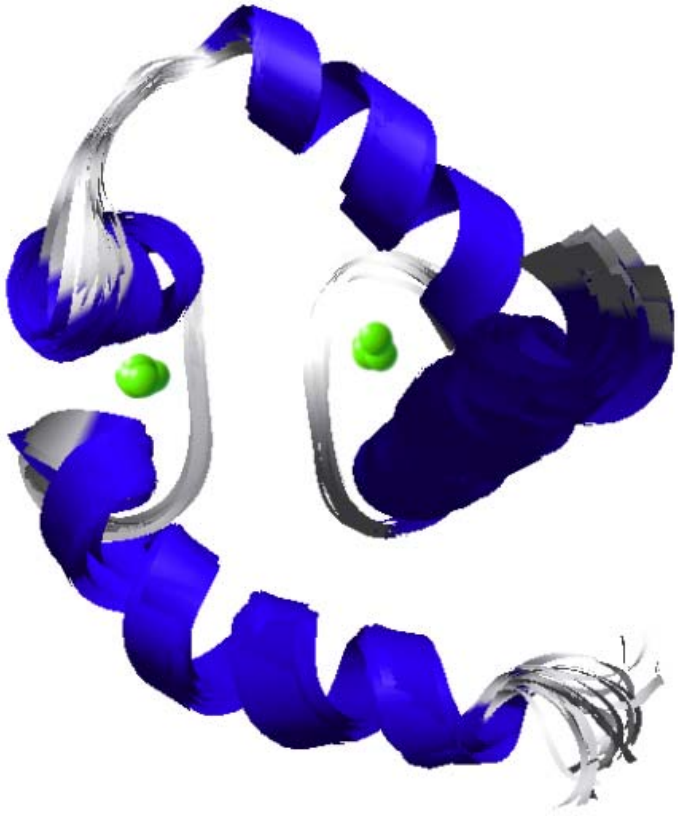


Fig. 3.6 Family of solution structures of the N-terminal domain of N60D CaM.

These magnetic susceptibility anisotropy tensors determine the frame to which the positions of the C-terminal domain are referred. Such positions are referred to as conformations of CaM, where the N-terminal domain is fixed and the C-terminal domain is not. The very same N-terminal tensor contributes to induce a partial orientation of the N-terminal domain in a magnetic field and tends to induce an orientation also on the C-terminal domain. Therefore, it is possible to extract information on the relative orientation of the C-terminal and N-terminal domains.

It was found that not all of the conformations are equally populated.

With respect to the N-terminal domain, the C-terminal domain resides in a region of space inscribed in an elliptical cone, in which its axis is tilted by about 30 degrees with respect to the direction of the N-terminal part of the interdomain helix (fig. 3.7).

Additionally, the RDC values show that the C-terminal domain undergoes rotation about the axis defined by the C-terminal part of the interdomain helix.

The methods and results obtained are more deeply discussed within this thesis (chapter 5).

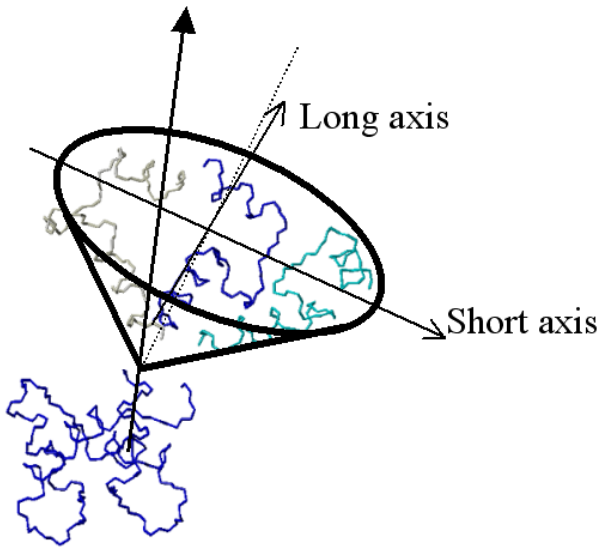


Fig. 3.7 Cone containing the three conformations of the C-terminal domain (only the first two helices are shown), which provided pseudo contact shifts and residual dipolar coupling with an average in good fit with the experimental data.

The applied procedure is of general application, as paramagnetic line broadenings decay with $1/r^6$ (r being the metal-nucleus distance), PCS decay with $1/r^3$ and RDC do not decay at all with distance. Therefore, no matter how strongly paramagnetic the metal center is, there will always be plenty of PCS and RDC to be detected outside the region of excessive broadening. For broadened lines, PCS can be measured more accurately than RDC. Therefore, if the metal sits in one domain, large and meaningful PCS can still be measured for that domain (and used to determine the tensor) while small but meaningful PCS and RDC can be measured for the other domain (and used to sample the conformational space).

These findings are unique, both in terms of structural information obtained on a biomolecule that samples multiple conformations and in terms of the approach developed to achieve the results. The same approach is in principle applicable to other multidomain proteins, as well as to multiple interaction modes between two macromolecular partners (chapter 6).

LIST OF REFERENCES FOR CHAPTERS 3

1. Berridge MJ, Lipp P, Bootman MD.. (2000) *Nat Rev Mol Cell Biol*. 1, 11.
2. Means AR, VanBerkum MFA, Bagchi I, Lu KP. and Rasmussen, C. D. *Intracellular Messengers* (1993) Pergamon Press, Oxford.
3. Kretsinger R. H. (1973) *J. Biol. Chem*. 248, 3313.
4. Malmendal A. (1999) *Biochemistry* 38, 11844.
5. Franchini P. L. (1999) *Arch. Biochme. Biophys* 372, 80.
6. Linse S, Forsen S. (1995). *Adv Second Messenger Phosphoprotein Res*. 30 ,89.
7. Zhang M, Tanaka T, Ikura M. (1995) *Nat Struct Biol.* 2, 758.
8. Potter J. D. and Gergely J. (2004) *J. Biol. Chem*. 250, 4628.
9. de Silva A. C. and Reinach F. C. (1991) *Trends In Biochem* 16, 53.
10. Finn B. E., Evenas J., Drakenberg T., Waltho J. P., Thulin E., and Forsen S. (1999) *Nat Struct. Biol* 22, 777.
11. Cohen P. & Klee C.B. *Calmodulin* (1988) Elsevier, New York.
12. Babu Y.S., Bugg C.E. & Cook W.J. (1988) *J. Mol. Biol.* 204, 191.
13. Ikura M., Hiraoki T., Hikichi K., Mikuni T., Yazawa M., and Yagi K. (1983) *Biochemistry*; 22, 2573.
14. Andersson A., Forsen S., Thulin E., and Vogel H.J. (1983) *Biochemistry*, 22, 2309.
15. Dalgarno D.C., Klevit R.E., Levine B.A., Williams R.J., Dobrowolski Z., and Drabikowski W. (1984) *Eur. J. Biochem.* 138, 281.
16. Thulin E., Andersson A., Drakenberg T., Forsen S., and Vogel H.J. (1984) *Biochemistry*, 23, 1860.
17. Babu Y., Sack J. S., Greenhough T. J., Bugg C. E., Means A. R., and Cook W. J. (1985) *Nature* 315, 37.
18. Seaton B. A., Head J. F., and Richards F. M. (1985) *Biochemistry* 24, 6740.
19. Small E. W. and Anderson S. R. (1988) *Biochemistry* 27, 419.
20. Barbato G., Ikura M., Kay L. E., Pastor R. W., and Bax A. (1992) *Biochemistry* 31, 5269.
21. Wall M. E., Clarage J. B., and Phillips G. N. (1997) *Structure* 5, 1599.
22. Ikura M., Clore G. M., Gronenborn A. M., Zhu G., Klee C. B. & Bax A. (1992) *Science*, 256, 632.

23. Meador W. E., Means A. R. & Quiocho F. A. (1992) *Scienc*e, 257, 1251.

24. Meador W. E., Means A. R. & Quiocho F. A. (1993). *Science*, 262, 1718.

25. Oneil K. T. & Degrado W. F. (1990). *Trends Biochem. Sci*. 15, 59.

26. Bruno J., Horrocks W. D., Jr., and Beckingham K. (1996) *Biophys. Chem.* 63, 1

27. Sudnick D. R., and Horrocks W. D. Jr. (1979) *Biochim. Biophys. Acta* 578, 135.

28. Lee L., and Sykes B. D. (1983). *Biochemistry* 22, 4366.

29. Geraldes C. F. (1993) *Methods Enzymol.* 227, 43.

30. Bertini I., Luchinat C., and Rosato A. (1996) *Prog. Biophys. Mol. Biol.* 66, 43.

31. Bentrop D., Bertini I., Cremonini M. A., Forsén S., Luchinat C., and Malmendal A. (1997) *Biochemistry* 36, 11605.

32. Bertini I., Donaire A., Jimenez B., Luchinat C., Parigi G., Piccioli M., and Poggi L. (2001) *J. Biomol. NMR* 21, 85.

33. Barbieri R., Bertini I., Cavallaro G., Lee Y. M., Luchinat C., and Rosato A. (2002) *J. Am. Chem. Soc.* 124, 5581.

34. Geraldes C. F. G. C., and Luchinat C. (2003) *Met. Ions Biol. Syst.* 40, 513.

35. Atreya H. S., Mukherjee S., Chary K. V. R., Lee Y.-M., and Luchinat C. (2003). *Protein Sci.* 12, 412.

36. Lee L., and Sykes B. D. (1981). *Biochemistry* 20, 1156.

37. Biekofsky R. R., Muskett F. W., Schmidt J. M., Martin S. R., Browne J. P., Bayley P. M., and Feeney J. (1999) *FEBS Lett.* 460, 519.

38. Bertini I., Gelis I., Katsaros N., Luchinat C., and Provenzani A. (2003) *Biochemistry* 42, 8011

39. Chou J. J., Li S., Klee C. B. & Bax A. (2001) *Nat. Struct. Biol.* 8, 990.

40. Allegrozzi M., Bertini I., Janik M. B. L., Lee Y.-M., Liu G. & Luchinat C. (2000) *J. Am. Chem. Soc.* 122, 4154.

41. Baig I., Bertini I., Del Bianco C., Gupta Y. K., Lee Y.-M., Luchinat C. & Quattrone A. (2004) *Biochemistry* 43, 5562.

42. Banci L., Bertini I., Cremonini M. A., Gori Savellini G., Luchinat C., Wuthrich K. & Guntert P. (1998) *J. Biomol. NMR* 12, 553.

CHAPTER 4

EXPERIMENTALLY EXPLORING THE CONFORMATIONAL SPACE SAMPLED BY DOMAIN REORIENTATION IN CALMODULIN

Ivano Bertini*, Cristina Del Bianco*, Ioannis Gelis‡, Nikolaus Katsaros‡, Claudio Luchinat§, Giacomo Parigi§, **Massimiliano Peana°**, Alessandro Provenzani*, and Maria Antonietta Zoroddu°

*Centre for Magnetic Resonance (CERM), and Department of Chemistry, University of Florence, Via Luigi Sacconi 6, I-50019 Sesto Fiorentino, Italy; ‡National Centre for Scientific Research Demokritos, Institute of Physical Chemistry, 15310 Agia Paraskevi Attikis, Greece; §Centre for Magnetic Resonance and Department of Agricultural Biotechnology, University of Florence, Piazzale delle Cascine 28, I-50144 Florence, Italy; and °Department of Chemistry, University of Sassari, Via Vienna 2, 07100 Sassari, Italy

Proc Natl Acad Sci USA, 2004 101 (18):6841-6846

Experimentally exploring the conformational space sampled by domain reorientation in calmodulin

Ivano Bertini*†, Cristina Del Bianco*, Ioannis Gelis‡, Nikolaus Katsaros‡, Claudio Luchinat§, Giacomo Parigi§, Massimiliano Peana¶, Alessandro Provenzani*, and Maria Antonietta Zoroddu¶

*Centre for Magnetic Resonance and Department of Chemistry, University of Florence, Via Luigi Sacconi 6, I-50019 Sesto Fiorentino, Italy; ‡National Centre for Scientific Research Demokritos, Institute of Physical Chemistry, 15310 Agia Paraskevi Attikis, Greece; §Centre for Magnetic Resonance and Department of Agricultural Biotechnology, University of Florence, Piazzale delle Cascine 28, I-50144 Florence, Italy; and ¶Department of Chemistry, University of Sassari, Via Vienna 2, 07100 Sassari, Italy

Edited by Harry B. Gray, California Institute of Technology, Pasadena, CA, and approved March 5, 2004 (received for review December 24, 2003)

The conformational space sampled by the two-domain protein calmodulin has been explored by an approach based on four sets of NMR observables obtained on Tb^{3+}- and Tm^{3+}-substituted proteins. The observables are the pseudocontact shifts and residual dipolar couplings of the C-terminal domain when lanthanide substitution is at the N-terminal domain. Each set of observables provides independent information on the conformations experienced by the molecule. It is found that not all sterically allowed conformations are equally populated. Taking the N-terminal domain as the reference, the C-terminal domain preferentially resides in a region of space inscribed in a wide elliptical cone. The axis of the cone is tilted by ≈30° with respect to the direction of the N-terminal part of the interdomain helix, which is known to have a flexible central part in solution. The C-terminal domain also undergoes rotation about the axis defined by the C-terminal part of the interdomain helix. Neither the extended helix conformation initially observed in the solid state for free calcium calmodulin nor the closed conformation(s) adopted by calcium calmodulin either alone or in its adduct(s) with target peptide(s) is among the most preferred ones. These findings are unique, both in terms of structural information obtained on a biomolecule that samples multiple conformations and in terms of the approach developed to achieve the results. The same approach is in principle applicable to other multidomain proteins, as well as to multiple interaction modes between two macromolecular partners.

Calmodulin (CaM) is a paradigm case in structural biology. The following brief survey of the history of the structural and dynamic studies on this protein serves the double purpose of putting the present findings in proper perspective and of acknowledging those pieces of previous information that were essential to allow the present approach to be developed and to yield novel structural information.

CaMs are two-domain proteins belonging to the large family of EF-hand proteins (1–3). They contain ≈150 amino acid residues, organized into two domains of ≈70 aa each and connected by a short linker. Each domain is made up of two special helix–loop–helix motifs (EF-hand motifs) that can bind a calcium ion in the loop. The two loops are held close to one another by two short antiparallel β-strands forming a three-hydrogen bond stretch of β-sheet. The function of CaM in cell cytoplasm is that of responding to sudden rises of calcium concentration by binding up to four calcium ions in the four EF-hand loops, by changing conformation because of metal binding, and by thus becoming able to recognize, bind to, and activate, a number of proteins and enzymes (1, 4–8). Early x-ray data (9) showed the four-calcium $(Ca_2)_N(Ca_2)_C$CaM form (subscripts N and C refer to the calcium atoms bound by the N- and C-terminal domains, respectively) to have a dumbbell shape, with helix 4, the last helix of the N-terminal domain, and helix 5, the first helix of the C-terminal domain, together with the interdomain linker, forming a long continuous helical structure (Fig. 1*A*). On the other hand, the protein takes up a closed conformation, with total loss of the helical character of the interdomain linker, when it binds to its target peptide(s) (Fig. 1*B*) (10–15). This binding involves the two domains getting closer to one another and clamping the recognition peptide in the target molecule between their exposed hydrophobic cores.

It was soon recognized that the NMR properties of $(Ca_2)_N(Ca_2)_C$CaM in solution were inconsistent with the rigid dumbbell shape observed in the early x-ray work, and that the central part of the helix loses its helical character and allows reciprocal reorientation of the two domains (16–20). Molecular dynamics simulations were performed (21, 22), confirming the flexibility of the two domains. An extended model-free analysis characterized the relative motions as occurring on a time scale of ≈3 ns, with a squared order parameter of 0.7 relative to the x-ray conformation, at room temperature (17). When temperature was raised to ≈40°C, a larger interdomain motion was observed, as a result of a doubling of the random coil residues in the central linker (23). Disorder in the central part of the interdomain helix was recently observed in an x-ray structure at 1.0-Å resolution (24). Finally, recent work has shown that native calcium-loaded CaM can also crystallize in the closed conformation (25).

It is therefore apparent that $(Ca_2)_N(Ca_2)_C$CaM has ample conformational freedom in solution, but that the extensive experimental data available do not allow us yet to make reasonable guesses on the nature and variety of the most favored conformers. Our approach is based on the exploitation of the long-range constraints that are imposed on the nuclear coordinates (pseudocontact shifts, pcs) and internuclear vector orientations (residual dipolar couplings, rdc) in one domain by the presence of paramagnetic metal ions with large and different magnetic susceptibility tensor anisotropies in the other domain. At least two metal ions are needed (26, 27) which in the present case are Tb^{3+} and Tm^{3+}. The approach capitalizes specifically on the results of recent work summarized below.

(*i*) It has been shown that lanthanide derivatives of *Drosophila melanogaster* CaM do induce appreciable pcs and rdc in the domain not bearing the lanthanide. It has been also observed that both pcs and rdc were sizably smaller than predicted by assuming that the x-ray conformation was maintained in solution, consistent with the presence of multiple conformations. The presence of a mixture of species deriving from lanthanide binding to different calcium sites apparently prevented a deeper analysis (28).

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: CaM, calmodulin; pcs, pseudocontact shifts; rdc, residual dipolar couplings; Ln, lanthanide; HSQC, heteronuclear sequential quantum correlation; NOE, nuclear Overhauser enhancement.

Data deposition: The solution structures have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 1SW8).

†To whom correspondence should be addressed at: Magnetic Resonance Center (CERM), University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino (FI), Italy. E-mail: bertini@cerm.unifi.it.

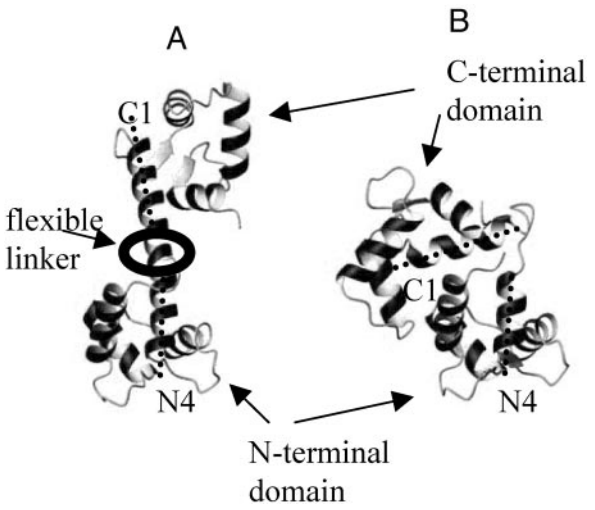


Fig. 1. Relative orientation of the N-terminal and C-terminal domains in CaM as early observed by x-ray in the absence of target peptides (*A*; extended conformation) and as observed in the presence of target peptides (*B*; closed conformation). Labels N4 and C1 indicate the fourth helix of the N-terminal domain and the first helix of the C-terminal domain, respectively.

(*ii*) An extensive series of mutants of vertebrate CaM aimed at altering the relative affinity of lanthanide (Ln) vs. calcium in the four metal binding sites was prepared and investigated by NMR. The N60D mutant was found to increase the relative affinity of calcium for the second site of the N-terminal domain to an extent that permitted the preparation of well defined $(CaLn)_N(Ca_2)_C$CaM derivatives (29).

(*iii*) The solution structure of CaM has been recently refined with an extensive use of rdc derived from an external orienting system. The structure showed significant differences in the helix orientations within the N-terminal domain with respect to the atomic resolution x-ray structure, and it provided reliable information on the orientation of internuclear vectors within both the N- and C-terminal domains (30). Therefore this structure is an essential starting point for the present study. We refer to this C-terminal domain structure as the "Bax structure." It should be noted that the information derived from the rdc obtained from an external orienting device is complicated by the fact that the orientation of each domain is the result of the contribution from the orientation of the other domain plus the direct and stronger orienting effect of the external device on the domain itself (31, 32).

Materials and Methods

Protein Expression. The plasmid pET16b-CaM was used to transform the BL21(DE3)gold *Escherichia coli* strain (Novagen) and transformants were used for the expression of unlabeled samples as previously described (29). ^{15}N-labeled N60D CaM was prepared as previously reported for unlabeled samples (29), except that immediately before induction the cells were gently pelleted and resuspended in M9 minimal medium supplemented with the appropriate isotopically labeled nitrogen (1.2 g of $^{15}NH_4Cl$ per liter) and carbon (3.5 g of [^{13}C]glucose per liter) sources. Bacterial cultures were induced with 0.5 mM isopropyl β-D-thiogalactoside, and the cells were further incubated at 37°C overnight. N60D $(Ca_2)_N(Ca_2)_C$CaM was purified as previously described (29).

Sample Preparation. After protein purification, NMR samples were prepared by buffer exchange by Centricon (Millipore) ultrafiltration with a membrane cutoff of 10,000 Da, as previously reported (29). Final conditions for NMR samples were 20 mM Mes/400 mM KCl, pH 6.5. Apo N60D CaM samples were carefully titrated up to 3 eq of Ca(II) by addition of $CaCl_2$. NMR samples (10% D_2O) were concentrated to about 1 mM protein solutions. N60D $(Ca_1)_N(Ca_2)_C$CaM samples were titrated with 50 mM solutions of analytical-grade $LnCl_3$ (Ln = Tb or Tm). The samples were kept at 4°C between measurements.

NMR Spectroscopy. NMR spectra were acquired on Bruker Avance 700 and 600 spectrometers operating at 700.13 and 600.13 MHz, respectively, equipped with triple-resonance inverse (TXI) gradient probes. Experiments were carried out at 300 K. Quadrature detection in the indirect dimensions was used and water suppression was achieved by means of WATERGATE (33). Experimental parameters are listed in Table 2, which is published as supporting information on the PNAS web site. All spectra were processed with the Bruker XWINNMR software package and analyzed by the program SPARKY (34).

pcs values of the Ln-substituted derivatives were obtained by recording ^{1}H–^{15}N heteronuclear sequential quantum correlation (HSQC) spectra (35) at 300 K and using a spectral width of 16 ppm and 32 ppm in the ^{1}H and ^{15}N dimensions, respectively. A total of 256 increments each with 1,024 complex data points and 16 transients were collected. The recycle delays were in the range between 0.7 s (paramagnetic derivative) and 1.5 s (diamagnetic derivative). pcs were calculated as the difference of the nuclear shifts between N60D $(CaLn)_N(Ca_2)_C$CaM and N60D $(Ca_1)_N(Ca_2)_C$CaM.

One-bond ^{1}H–^{15}N coupling constants were measured at 300 K and 700 MHz by fitting a series of $^1J_{NH}$-modulated HSQC spectra (36). rdc were calculated as the difference of the fitted $^1J_{NH}$ between N60D $(CaLn)_N(Ca_2)_C$CaM and N60D $(Ca_1)_N(Ca_2)_C$CaM.

Structure Calculation. Triple-resonance experiments were used to assign the backbone of N60D $(Ca_2)_N(Ca_2)_C$CaM. The backbone resonance assignment was obtained by the analysis of triple-resonance CBCANH and CBCA(CO)NH spectra (37, 38) performed at 700 MHz. Side-chain carbon and proton assignments were made by a (H)CCH-TOCSY experiment (39) at 700 MHz.

HNHA (40) and ^{1}H–^{15}N-NOESY-HSQC (41) spectra at 700 MHz allowed for torsion angle calculations. Backbone dihedral φ angles were independently derived from $^3J_{HNH\alpha}$ coupling constants through the appropriate Karplus equation. More specifically, $^3J_{HNH\alpha}$ values >7 Hz were constrained to φ angles between -155^o and -85^o and for those lower than 4.5 Hz the φ angles were constrained within -70^o and -30^o (40).

Backbone dihedral ψ angles for residue $i-1$ were also determined from the ratio of the intensities of the $d_{\alpha N}(i-1, i)$ and $d_{N\alpha}(i, i)$ nuclear Overhauser enhancements (NOEs) present on the $^{15}N(i)$ plane of residue i obtained from the ^{15}N-edited NOESY-HSQC. Ratio values of residue $i-1$ greater than one are indicative of β-sheets with ψ values ranging between 60^o and 180^o, whereas values smaller than one indicate a right-handed α-helix with ψ values between -60^o and -20^o (42).

NOEs were measured from 2D NOESY (43), 3D ^{1}H–^{13}C-HSQC-NOESY (41) (with a mixing time of 80 ms), and 3D ^{1}H–^{15}N-HSQC-NOESY (41) (with a mixing time of 100 ms) spectra collected at 700 MHz.

From the analysis of the 3D ^{15}N-edited and ^{13}C-edited NOESY-HSQC spectra and 2D NOESY spectrum, NOE cross-peaks were assigned and transformed into unique upper distance limits by using the program CALIBA (44). More specifically, for the N-terminal domain of N60D CaM, 2,360 NOE cross-peaks were assigned and transformed into 2,020 unique upper distance limits, of which 1,495 were found to be meaningful. For the C-terminal domain, 1,895 assigned NOEs were transformed into 1,606 unique upper distance limits and 1,225 were found to be meaningful.

The solution structure of the two domains was calculated with

Table 1. Experimental constraints collected for the N- and C-terminal domains of CaM

	No. of constraints	
	N-terminal domain	C-terminal domain
Meaningful NOEs	1,495	1,225
ϕ dihedral angles	50	43
ψ dihedral angles	50	43
pcs (Tb^{3+} and Tm^{3+})	125	165
rdc (Tb^{3+} and Tm^{3+})	—	67

the simulated annealing program DYANA (45) and PARAMAGNETIC DYANA (46).

Paramagnetism-Based Constraints. In paramagnetic metalloproteins the metal magnetic susceptibility tensor, χ^{para}, is usually anisotropic, owing to orbital contributions to the electron magnetic moment. In solution, this anisotropy produces pcs of the nuclei that are dipole–dipole coupled to the paramagnetic metal ion as well as rdc attributable to partial self-orientation (47–49). The pertinent equations are published as *Supporting Text* on the PNAS web site.

Results and Discussion

^{15}N,^{13}C-enriched vertebrate N60D $(Ca_2)_N(Ca_2)_C$CaM was subjected to the standard set of 2D and 3D experiments to assign backbone and side chains, and to obtain NOE. The relevant data are summarized in Table 1 and in Tables 3 and 4, which are published as supporting information on the PNAS web site. The structures obtained for the two domains are in good agreement with the latest solution structure refined with the help of rdc (30) resulting from orienting media (Bax structure). N60D $(CaLn)_N(Ca_2)_C$CaM derivatives with Ln = Tm or Tb were prepared, and an extensive assignment of both N-terminal and C-terminal peptide NH nuclei was performed by using a semi-automated procedure previously described (50, 51). The pcs and rdc values are summarized in Table 1 and in Tables 5 and 6, which are published as supporting information on the PNAS web site. Both pcs and rdc values of the C-terminal domain are rather small. However, they have been measured with good precision, the estimated uncertainty being ±0.05 ppm and ±0.3 Hz, respectively.

Orientations and Anisotropies of the Magnetic Susceptibility Tensors. The pcs values relative to the N-terminal domain were used together with the N-terminal atom coordinates from either the present or the Bax structure (30) to obtain the χ-tensor anisotropies and principal axes of both the Tb^{3+} and the Tm^{3+} derivatives according to a well established procedure (50, 52, 53). The best-fit values are reported in Fig. 2, where the satisfactory agreement between calculated and observed pcs is also shown (the correlation coefficient for Tb^{3+} and Tm^{3+} pcs being 0.995 and 0.987, respectively). Such good agreement can be seen as a validation of the available structures. pcs values were then provided as additional constraints to the program PARAMAGNETIC DYANA (46) together with the obtained magnetic susceptibility anisotropy values. The resulting family of solution structures of the N-terminal domain of CaM containing the N60D mutation is deposited in the Protein Data Bank (PDB ID code 1SW3).

These magnetic susceptibility anisotropy tensors determine the frame to which the positions of the C-terminal domain are referred. Such positions are referred to as *conformations* of CaM, where the N-terminal domain is fixed and the C-terminal domain is not. The very same N-terminal tensor contributes to induce a partial orientation of the N-terminal domain in a magnetic field

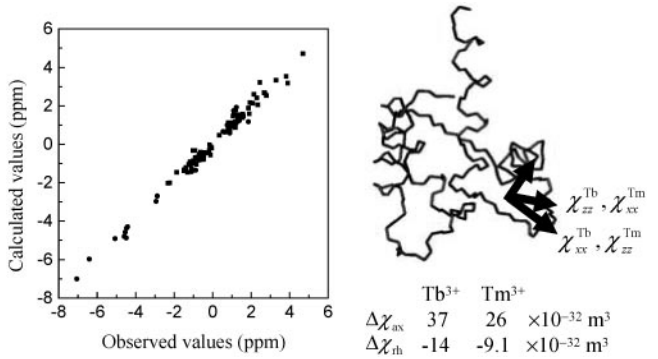


Fig. 2. Calculated vs. observed values of pcs of N-terminal nuclei for the terbium(III) (●) and thulium(III) (■) derivatives. The directions of the χ tensor axes are indicated, and the magnetic susceptibility anisotropies are reported.

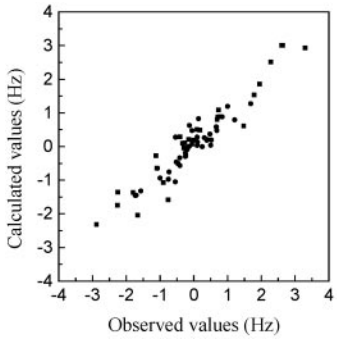
and tends to induce an orientation also on the C-terminal domain. Therefore, it is possible to extract information on the relative orientation of the C-terminal and N-terminal domains.

Testing Single and Uniformly Distributed Conformations. Simple minimization programs were constructed and used to search the conformational space of CaM. One of these programs allows movements of the C-terminal domain, assumed rigid (Bax structure), relative to the N-terminal domain, also assumed rigid, through a completely flexible linker of residues 78–81 (16, 17, 23, 25). For any given conformation of the C-terminal domain, the pcs and rdc values are calculated. This program was first used to check two extreme situations: (*i*) fixed single conformations, such as those observed in the solid state, i.e., the extended and the closed forms (the latter both with and without bound target peptide or peptides); (*ii*) a uniform distribution of sterically allowed conformations.

(*i*) A large number of starting conformations was generated and used in the minimization program. In all cases, the overall agreement was poor. In particular, for the conformations observed in the solid state a poor agreement was found for both pcs and rdc, either together or separately. The pcs and particularly the rdc calculated from the x-ray conformations were much larger in absolute value than the observed ones. As expected, there is not a single CaM conformation in solution.

(*ii*) At the other extreme, a uniform distribution of all sterically allowed conformations yields average pcs values that are in reasonable agreement with the experimental ones, but average rdc values that are in disagreement with the experimental ones and show two to six times smaller spreading of values. From the latter observation we conclude that the conformational space sampled by CaM may be ample, but *the distribution must be nonuniform*. In particular, within all possible *conformations*, there must be some with less favorable *orientations* of the C-terminal domain (note that, as usual in this field, the term *orientation* is not related to the position in the reference system). The fact that not all orientations are sampled equally is an unexpected finding, whose implications will be further analyzed below.

Analysis of the C-Terminal rdc. The values and the spreading of the rdc values for the two metal derivatives may contain information on the type of rotational average experienced by the C-terminal NH vectors. Before turning to a deeper analysis of these values, we need to assess the intrinsic quality of the two sets of experimental rdc. Any set of meaningful rdc is always describable by a single orientation tensor, independently of the fact that they originate from a weighted average of a number of conformations. The latter tensor can be obtained by a simple fit of the



	Tb^{3+}	Tm^{3+}	
$\Delta\chi_{ax}$	−1.6	−3.3	$\times 10^{-32}$ m^3
$\Delta\chi_{rh}$	0.8	2.2	$\times 10^{-32}$ m^3

Fig. 3. Calculated vs. observed values of rdc of C-terminal HN for the terbium(III) (●) and thulium(III) (■) derivatives. Each set of data was fit separately by using FANTAORIENT.

tensor parameters to the experimental rdc and to the existing C-terminal Bax structure, using the FANTAORIENT program (48, 52). The results of the fit are shown in Fig. 3. It is apparent that a satisfactory agreement is obtained. The tensor parameter values are much smaller than those obtained from the analysis of the pcs of the N-terminal domain NH nuclei. This reduction is, of course, a consequence of extensive orientation averaging. The good agreement is a proof only of the reliability of the two rdc sets.

It is instructive to back-calculate the rdc from the full susceptibility anisotropy tensors of the two metals (case A) and from the tensors obtained above from FANTAORIENT (case B). Fig. 4*A* shows the calculated distribution of the rdc values for case A. As expected, the spreading is somewhat larger for Tb^{3+} than for Tm^{3+}, because of the larger anisotropy of the former ion. Fig. 4*B* shows the distribution of the rdc calculated for case B. For comparison purposes, Fig. 4*C* shows the calculated distribution of rdc values when averaged over all conformations except those in steric clash with the N-terminal domain. These distributions are much narrower than in case B, indicating that in the latter there must be less favorable conformations besides those prohibited by steric clashes.

It is well known that the effect of fitting to a rigid structure the rdc arising from motional averaging is that of obtaining a motionally averaged alignment tensor (54, 55). In this case, a generalized order parameter (47, 56–58) can be defined qualitatively as the ratio of the spreading between the observed rdc distribution (Fig. 4*B*) and the rdc distribution calculated in the assumption of no motion (Fig. 4*A*). Such parameter is equal to 0.05 and 0.15 for Tb and Tm rdc values, respectively. Different order parameters for the different orienting metal ions indicate that the C-terminal motion causes different motional averaging because of the different directions of the principal χ^{para} axes of the two metals (Fig. 2). This observation is further evidence that not all sterically allowed conformations are equally probable.

Search for the Least-Favored Orientations by Using C-Terminal rdc. It is known that any set of meaningful rdc values can be reproduced by a given magnetic anisotropy tensor and a weighted average of three orientations. We grid-generated all C-terminal domain orientations and, for each one, we searched for two additional orientations (described by the Euler angles) which, combined with the given one, gave the best agreement with the experimental rdc data. In some cases, the sum of the squared residuals coincided with that from the fits reported in Fig. 3, i.e., the lowest possible value; in some other cases it did not. Details of the calculation can be found in the supporting text. Such analysis indicates that, even if all orientations could be sampled by the C-terminal domain, there are regions in the orientational space

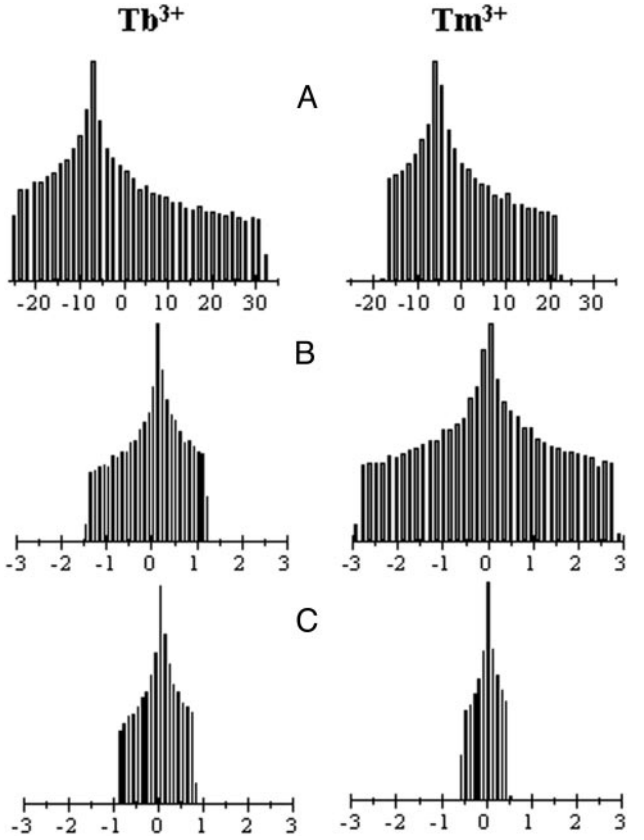


Fig. 4. Distribution of the rdc values calculated for all conformations by using the Bax structure and the magnetic susceptibility anisotropy values obtained from the experimental rdc of the N-terminal domain (*A*), from the experimental rdc of the C-terminal domain (*B*), and from the average rdc of the C-terminal domain obtained from sterically allowed uniformly sampled conformations (*C*).

less favored than others, and *these regions can be identified*. Now, we face the task of translating this information in Euler angle space into information on disfavored conformations in real Euclidean space.

Exploring the Conformational Space with C-Terminal pcs. We have mentioned that pcs calculated in either the extended or the closed conformations are not in agreement with the observed values. We then tried a six-parameter fit (three Euler angles and three translational parameters) to look for a single conformation providing a good agreement between calculated and observed values of pcs of both Tb(III) and Tm(III) simultaneously. A good fit was actually found corresponding to a conformation characterized by a tilt of the first helix of the C-terminal domain of ≈34°, with respect to the extended form, in the direction of the second metal binding site in the N-terminal domain. Such conformation has no physical meaning, as proved by the fact that the corresponding rdc are in strong disagreement with the observed values, but could be seen as a sort of "average" conformation.

If a conformation is chosen with the first helix of the C-terminal domain tilted by a large angle with respect to the extended form, as happens in the closed form where the tilt is ≈90°, a good fit is obtained for the pcs data *only* by considering other additional conformations with the interdomain helix again largely tilted, but in different directions with respect to the first one. This means that either the closed form is unfavored in

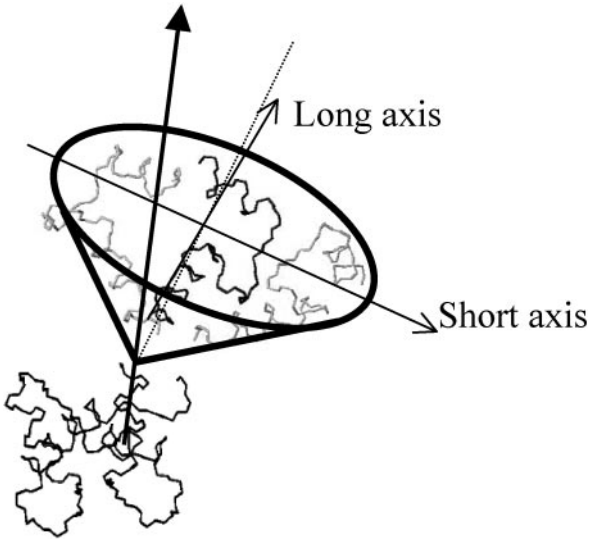


Fig. 5. Cone containing the three conformations of the C-terminal domain (only the first two helices are shown), which provide pcs and rdc with an average in good fit with the experimental data.

solution, or more than one widely different "closed" form should be present.

Minimal Sets of Conformations That Agree with Both C-Terminal rdc and pcs. pcs and rdc data were then fit simultaneously to search for the conformations providing the lowest value for the target function, keeping in mind that the lowest possible value for the target function is set by the sum of the target functions obtained from the fit performed on pcs (see *Exploring the Conformational Space with C-Terminal pcs*) and rdc (Fig. 3) (see *Analysis of the C-Terminal rdc* and *Search for the Least-Favored Orientations by Using C-Terminal rdc*) separately. The C- and N-terminal domains were held together by a completely flexible linker of residues 78–81 (16, 17, 23, 25) (such a linker was subsequently also extended to the range from residue 75 to residue 81, without significant deviations in the resulting features described below). The minimization program was then allowed to sample *triplets* of conformations with adjustable weights. The "successful" best-fit triplets provided target functions *coincident* with the lowest possible value defined above. We found a large "density" of such solutions for which the three conformations reside in an elliptical cone, whose axis is tilted by ≈30° on the direction of the second metal binding site of the N-terminal domain (Fig. 5). The major axis of the ellipsis is roughly parallel to the direction of the β-sheet located on the N-terminal domain. The fit also indicated that the C-terminal domain rotates about the axis defined by the C-terminal side of the interdomain helix. The amplitude of the semiangle of the cone ranges from 50° to 80°. Another set of calculations was performed allowing *four* conformations to be sampled. Often one of the four was with low weight, or the four were in the same region of space occupied by the conformation triplets.

To better visualize the results, we also restricted the Euler angles to assume only those values that provide orientations of the first helix of the C-terminal domain (see Fig. 1) within a fixed angle with respect to the orientation of the same helix in the extended form. This helix was thus restrained to move within a cone with fixed semiangle, and with symmetry axis along the direction of the helix in the extended form. The semiangle was set to increasing values from 10° in steps of 10°. We find that only a semiangle >40° provides a good fit. This result is due to the contribution by the rdc to the target function, as a good fit of pcs data could be obtained already for angles >10°. Another set of calculations was then performed, allowing *four* conformations to be sampled. Again, a semiangle of at least 40° was needed to provide a good fit. Such a model compares with that proposed by Baber *at al.* (17), according to which the rigid linking C-terminal helix is wobbling within a cone having a maximum semiangle of 54° with respect to the N-terminal helix.

The first helix of the C-terminal domain was then restrained to move within a cone with semiangle again variable in steps of 10°, but with the axis of the cone tilted with respect to the orientation of the same helix in the extended form. A grid search of solutions comprising three conformations was thus performed for different orientations and amplitudes of the cone. The results of the calculations show that there are indeed cones different from that depicted in Fig. 5, with semiangle larger than 40°, where it is possible to place three conformations that agree with the experimental data. In contrast, conformations inside cones with small semiangles are allowed only for the C-terminal domain bent in the same direction as that of the closed forms. We also tested the possibility that the extended form or one of the closed forms could be present among the conformations. This testing was done by searching for additional conformations that, combined with the given one, gave the best agreement with the experimental data. The result was that two additional conformations did not provide a good agreement in either case. Again, this result supports the idea that neither the extended forms nor the closed forms are particularly favored in solution. As expected, however, in both cases three additional conformations with similar weights provide a good fit.

Summary Considerations. The present research provides a further characterization of the conformational space sampled by domain reorientation of CaM by using pcs and rdc data.

The first conclusion to be drawn is that the available conformation space is quite ample and spans beyond the cone suggested by relaxation data (17). This conclusion is not in contradiction with the previous findings, as the analysis was limited to motions in the nano- to picosecond scale, whereas pcs and rdc are averaged by motions spanning a time scale that extends down to milliseconds. Apparently, slower motions than those affecting relaxation measurements may contribute to the sampling of conformational space in CaM.

The second finding is that C-terminal rdc clearly arise from averaging among very diverse orientations. Therefore, pendulum-like motions, no matter how ample, that do not imply also a rotation of the C-terminal domain about the axis of its first helix, are not sufficient in averaging the rdc down to the small values observed.

However, not all C-terminal orientations are equally probable (because the rdc would average zero), and exclusion of only those conformations in steric clash with the N-terminal domain is not enough to account for the magnitude of the observed rdc. Electrostatic repulsion between the negatively charged domains may contribute to the different probability of sterically accessible conformations.

pcs, taken alone, are seemingly less informative. It is easy to find two or more conformations, anywhere in the available conformational space, that agree with the experimental data, and even a single conformation is only in slightly worse agreement. However, there is information in this finding. The single conformation in reasonable agreement with the pcs is not far from the extended conformation (but distinct from it) and very far from the closed conformations (it is actually slightly tilted on the opposite side). As pcs depend on the reciprocal third power of the metal-nucleus distance, this finding by itself indicates that on average all nuclei of the C-terminal domain are rather far from the metal in the N-terminal domain. This in turn implies that, if a closed conformation is populated in solution, there must

be at least another similarly "closed" conformation on another side to compensate for the effect of the first.

Taken together, pcs and rdc indicate that, if motions within a cone are considered, with the exception of a few specific orientations of the axis of the cone in the direction of the closed forms, in all other orientations the cone semiangle must be at least 40° or wider. In light of the findings from relaxation data (17), we should exclude those few orientations with a too narrow cone, although they are consistent with the present data.

In conclusion, we point out that the applied procedure is of general application, as paramagnetic line broadenings decay with $1/r^6$ (r being the metal-nucleus distance), pcs decay with $1/r^3$ and rdc do not decay at all with distance. Therefore, no matter how strongly paramagnetic the metal center is, there will always be plenty of pcs and rdc to be detected outside the region of excessive broadening. For broadened lines, pcs can be measured more accurately than rdc. Therefore, if the metal sits in one domain, large and meaningful pcs can still be measured for that domain (and used to determine the tensor) while small but meaningful pcs and rdc can be measured for the other domain (and used to sample the conformational space).

Discussions with Luca Sgheri and Marco Longinetti are gratefully acknowledged. This work has been supported by the Ministero dell'Istruzione, dell'Università e della Ricerca, Italy ("Applicazioni innovative delle spettroscopie NMR ad alto campo nella ricerca industriale"), COFIN 2003 ("Il ruolo degli ioni metallici nei processi metabolici"), and by the European Commission, contract QLG2-CT-2002-00988.

1. Carafoli, E. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 1115–1122.
2. Kawasaki, K. & Kretsinger, R. H. (1994) *Protein* **1,** 343–517.
3. Biekofsky, R. R., Martin, S. R., Browne, J. P., Bayley, P. M. & Feeney, J. (1998) *Biochemistry* **37,** 7617–7629.
4. Chin, D. H. & Means, A. R. (2000) *Trends Cell Biol.* **10,** 322–328.
5. Cox, J. A., Comte, M., Malnoë, A., Burger, D. & Stein, E. A. (1984) in *Metal Ions in Biological Systems: Calcium and Its Role in Biology*, ed. Sigel, H. (Dekker, NY), pp. 215–273.
6. Crivici, A. & Ikura, M. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24,** 85–116.
7. Finn, B. E., Evenäs, J., Drakenberg, T., Waltho, J., Thulin, E. & Forsén, S. (1995) *Nat. Struct. Biol.* **2,** 777–783.
8. Peersen, O. B., Madsen, T. S. & Falke, J. J. (1997) *Protein Sci.* **6,** 794–807.
9. Babu, Y. S., Bugg, C. E. & Cook, W. J. (1988) *J. Mol. Biol.* **204,** 191–204.
10. Schumacher, R. T., Rivard, A. F., Bachinger, H. P. & Adelman, J. P. (2001) *Nature* **410,** 1120–1124.
11. Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Clee, C. & Bax, A. (1992) *Science* **256,** 632–638.
12. Ikura, M., Barbato, G., Klee, C. B. & Bax, A. (1992) *Cell Calcium* **13,** 391–400.
13. Meador, W. E., Means, A. R. & Quiocho, F. A. (1992) *Science* **257,** 1251–1255.
14. Osawa, M., Tokumitsu, H., Swindells, M., Kurihara, H., Orita, M., Shibanuma, T., Furuya, T. & Ikura, M. (1999) *Nat. Struct. Biol.* **6,** 819–826.
15. Elshorst, B., Hennig, M., Forsterling, H., Diener, A., Maurer, M., Schulte, P., Schwalbe, H., Griesinger, C., Krebs, J. F., Schmid, H. & Carafoli, E. (1999) *Biochemistry* **38,** 12320–12332.
16. Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W. & Bax, A. (1992) *Biochemistry* **31,** 5269–5278.
17. Baber, J. L., Szabo, A. & Tjandra, N. (2001) *J. Am. Chem. Soc.* **123,** 3953–3959.
18. Seaton, B. A., Head J. F. & Richardson, F. M. (1985) *Biochemistry* **24,** 6740–6743.
19. Heidorn, D. B. & Trewhella, J. (1988) *Biochemistry* **27,** 909–915.
20. Matsushima, N., Izumi, Y., Matsuo, T., Yoshino, H., Ueki, T. & Miyake, Y. (1989) *J. Biochem. (Tokyo)* **105,** 883–887.
21. Wriggers, W., Mehler, E., Pitici, F., Weinstein, H. & Schulten, K. (1998) *Biophys. J.* **74,** 1622–1639.
22. Barton, N. P., Verma, C. S. & Caves, L. S. D. (2002) *J. Phys. Chem. B* **106,** 11036–11040.
23. Chang, S.-G., Szabo, A. & Tjandra, N. (2003) *J. Am. Chem. Soc.* **125,** 11379–11384.
24. Wilson, M. A. & Brunger, A. T. (2000) *J. Mol. Biol.* **301,** 1237–1256.
25. Fallon, J. L. & Quiocho, F. A. (2003) *Structure* **11,** 1303–1307.
26. Barbieri, R., Bertini, I., Cavallaro, G., Lee, Y.-M., Luchinat, C. & Rosato, A. (2002) *J. Am. Chem. Soc.* **124,** 5581–5587.
27. Bertini, I., Longinetti, M., Luchinat, C., Parigi, G. & Sgheri, L. (2002) *J. Biomol. NMR* **22,** 123–136.
28. Biekofsky, R. R., Muskett, F. W., Schmidt, J. M., Martin, S. R., Browne, J. P., Bayley, P. M. & Feeney, J. (1999) *FEBS Lett.* **460,** 519–526.
29. Bertini, I., Gelis, I., Katsaros, N., Luchinat, C. & Provenzani, A. (2003) *Biochemistry* **42,** 8011–8021.
30. Chou, J. J., Li, S., Klee, C. B. & Bax, A. (2001) *Nat. Struct. Biol.* **8,** 990–997.
31. Goto, N. K., Skrynnikov, N. R., Dahlquist, F. W. & Kay, L. E. (2001) *J. Mol. Biol.* **308,** 745–764.
32. Skrynnikov, N. R., Goto, N. K., Yang, D., Choy, W.-Y., Tolman, J. R., Mueller, G. A. & Kay, L. E. (2000) *J. Mol. Biol.* **295,** 1265–1273.
33. Piotto, M., Saudek, V. & Sklenar, V. (1992) *J. Biomol. NMR* **2,** 661–666.
34. Goddard, T. D. & Kneller, D. G. (2000) SPARKY 3 (Univ. of California, San Francisco).
35. Schleucher, J., Schwendinger, M., Sattler, M., Schmidt, P., Schedletzky, O., Glaser, S. J., Sorensen, O. W. & Griesinger, C. (1994) *J. Biomol. NMR* **4,** 301–306.
36. Tjandra, N., Grzesiek, S. & Bax, A. (1996) *J. Am. Chem. Soc.* **118,** 6264–6272.
37. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. (1990) *J. Magn. Reson.* **89,** 496–514.
38. Muhandiram, D. R. & Kay, L. E. (1994) *J. Magn. Reson. B* **103,** 203–216.
39. Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R. & Forman-Kay, J. D. (1993) *J. Magn. Reson. B* **101,** 333–337.
40. Vuister, G. W. & Bax, A. (1993) *J. Am. Chem. Soc.* **115,** 7772–7777.
41. Wider, G., Neri, D., Otting, G. & Wüthrich, K. (1989) *J. Magn. Reson.* **85,** 426–431.
42. Gagné, S. M., Tsuda, S., Li, M. X., Chandra, M., Smillie, L. B. & Sykes, B. D. (1994) *Protein Sci.* **3,** 1961–1974.
43. Hwang, T.-J. & Shaka, A. J. (1995) *J. Magn. Reson. A* **112,** 275–279.
44. Güntert, P., Braun, W. & Wüthrich, K. (1991) *J. Mol. Biol.* **217,** 517–530.
45. Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997) *J. Mol. Biol.* **273,** 283–298.
46. Banci, L., Bertini, I., Cremonini, M. A., Gori Savellini, G., Luchinat, C., Wüthrich, K. & Güntert, P. (1998) *J. Biomol. NMR* **12,** 553–557.
47. Bertini, I., Luchinat, C. & Parigi, G. (2002) *Progr. NMR Spectrosc.* **40,** 249–273.
48. Banci, L., Bertini, I., Huber, J. G., Luchinat, C. & Rosato, A. (1998) *J. Am. Chem. Soc.* **120,** 12903–12909.
49. Bertini, I., Luchinat, C. & Parigi, G. (2001) *Solution NMR of Paramagnetic Molecules* (Elsevier, Amsterdam).
50. Allegrozzi, M., Bertini, I., Janik, M. B. L., Lee, Y.-M., Liu, G. & Luchinat, C. (2000) *J. Am. Chem. Soc.* **122,** 4154–4161.
51. Baig, I., Bertini, I., Del Bianco, C., Gupta, Y. K., Lee, Y.-M., Luchinat, C. & Quattrone, A. (2004) *Biochemistry* **43,** in press.
52. Bertini, I., Luchinat, C. & Parigi, G. (2002) *Concepts Magn. Reson.* **14,** 259–286.
53. Bertini, I., Janik, M. B. L., Lee, Y.-M., Luchinat, C. & Rosato, A. (2001) *J. Am. Chem. Soc.* **123,** 4181–4188.
54. Meiler, J., Prompers, J. J., Peti, W., Griesinger, C. & Brüschweiler, R. (2001) *J. Am. Chem. Soc.* **123,** 6098–6107.
55. Peti, W., Meiler, J., Brüschweiler, R. & Griesinger, C. (2002) *J. Am. Chem. Soc.* **124,** 5822–5833.
56. Tjandra, N. & Bax, A. (1997) *Science* **278,** 1111–1114.
57. Clore, G. M., Gronenborn, A. M. & Bax, A. (1998) *J. Magn. Reson.* **133,** 216–221.
58. Dosset, P., Hus, J. C., Marion, D. & Blackledge, M. (2001) *J. Biomol. NMR* **20,** 223–231.

CHAPTER 5

STRUCTURAL CHARACTERIZATION OF THE CALMODULIN-α-SYNUCLEIN COMPLEX BY NMR

5.1 Target Binding Versatility of Calmodulin

The major conformational change induced by the binding of Ca^{2+} ion to Calmodulin consists in the alteration of the relative orientation of the helices in each lobe. This rearrangement of the helices leads to the exposure of several hydrophobic residues to the solvent on the surface of each lobe, which are responsible for targets binding.

Biophysical and structural characterization of complexes between CaM and synthetic peptides indicated that the ternary complexes have a more compact shape than Ca^{2+}-CaM by itself and revealed different ways of binding. On one hand CaM can

bind target peptides forming a collapsed ellipsoid complex, where both N- and C-terminus domain are involved in the binding, as was observed in the structure of CaM-peptide complexes from skeletal Myosin light chain Kinase (fig. 5.1a).
On the other hand the NMR structure of the C20W peptide, representing the CaM-binding domain of the Ca^{2+} pump, showed that the peptide is bound only to the C-terminal domain and that the complex does not collapse into an ellipsoid shaped form, but remains in an extended conformation [1] (fig. 5.1b). In both cases, conformational rearrangements of the interdomain linker allow CaM to bind its target peptide. Interestingly, the peptide can be bound in a parallel or antiparallel orientation.

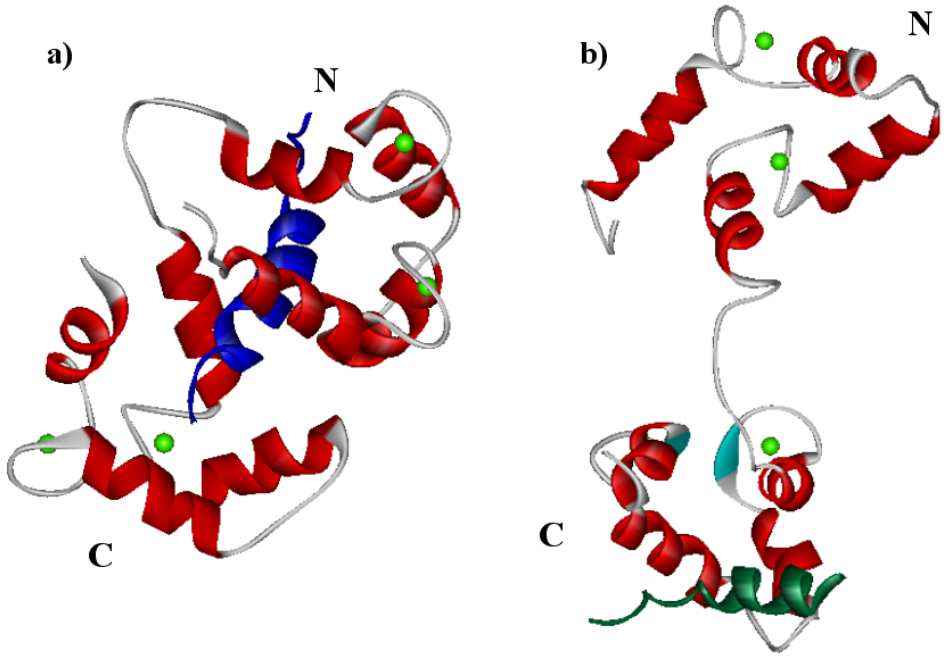


Fig. 5.1 a) Ribbon representation of the Ca_4CaM (red) complexed with Rabbit Skeletal Myosin Light Chain Kinase (blue) in which CaM adopt a collapsed ellipsoid shape (PDB ID code 2BBM). **b)** Ca_4CaM complexed with the CaM-binding domain of the Ca^{2+} pump, in which C20W peptide (green) is bound only to the C-terminal domain of CaM (PDB ID code 1CFF).

Analysis of CaM binding peptides revealed that they share no sequence homology, but they all have the potential to fold into a basic, amphiphilic α-helix. They usually display large hydrophobic residues in conserved positions and show an

α-helical conformation. The binding of CaM-binding peptides is largely driven by hydrophobic interactions between hydrophobic anchor residues of the peptide with the hydrophobic surface cavities of CaM.

Typically, CaM contains an abnormally high amount of methionine residues (9 Met out of a total of 148 amino acid residues), and 8 of the 9 Met residues are found in the hydrophobic surfaces of the Ca^{2+}-CaM, with 4 Met residues in each domain [2].

```
   1          11         21         31         41         51
   |          |          |          |          |          |
  1 ADQLTEEQIA EFKEAFSLFD KDGDGTITTK ELGTVMRSLG QNPTEAELQD MINEVDADGN    60
 61 GTIDFPEFLT MMARKMKDTD SEEEIREAFR VFDKDGNGYI SAAELRHVMT NLGEKLTDEE   120
121 VDEMIREADI DGDGQVNYEE FVQMMTAK
```

Fig. 5.2 Sequence of Calmodulin. The Met (M) residues are highlighted.

The four Met residues in each domain are located at the entrance of the hydrophobic-pocket made up of aromatic (Phe) and branched aliphatic (Ile, Leu, and Val) amino acid residues as well as short Ala residues. In fact, Met residues contribute to about 46% of the total exposed hydrophobic surface area of the protein[2] (fig. 5.3a).

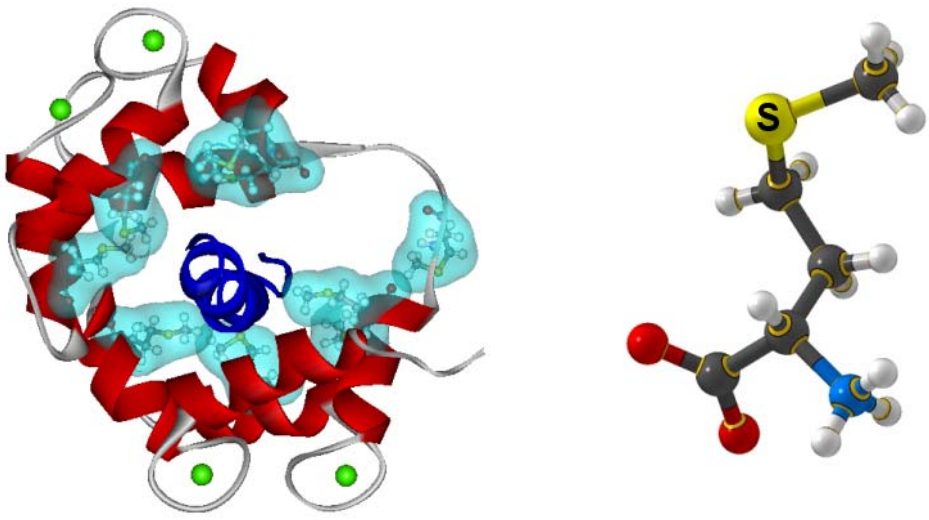


Fig. 5.3 **a)** Ribbon representation of Ca_4CaM complexed with Rabbit Skeletal Myosin Light Chain Kinase (2BBM). The methionines positioned in the entrance of hydrophobic-pocket are showed with their molecular surface, **b)** the methionine residue.

A surface rich in Met residues was proposed to allow CaM to bind many targets in a sequence-independent manner [3]. The existence of a sulphur atom in the unbranched side chain entails Met with unique flexibility and high polarizability. In fact the absence of branching permits methionine to assume many different shapes, thereby enabling this side chain to adopt many different conformations and so adapt to different hydrophobic surfaces (fig. 5.3 b).

In addition, high polarizability of the surfaces will not only allow the protein to interact with its targets via London dispersion forces but also stabilizes the large solvent-exposed hydrophobic surfaces in the absence of targets.

The majority of CaM binding motifs in proteins, which bind it in the canonical Ca^{2+}-dependent way, are un-structured in the absence of CaM, and become α-helical upon binding to it. The helix is amphipathic and its hydrophobic side interacts with hydrophobic pockets of CaM, whereas the hydrophilic face, containing four or more basic residues, interacts electrostatically with the acidic residues in CaM[4].

Recently, we have undertaken the characterization of the interaction in solution between human CaM and a novel target protein, α-Synuclein, which has unusual features with respect to other CaM target motifs.

5.2 Alpha-Synuclein and Parkinson's Disease

Parkinson's disease (PD), characterized by resting tremor, slowness of initial movement, rigidity, and general postural instability, is one of the most prevailing neurodegenerative disorders among elderly population. PD is characterized by the loss of dopaminergic neurons from the substantia nigra and the formation of fibrillar intraneuronal inclusions.

The importance of α-synuclein (AS) in the pathogenesis of PD is supported by a number of recent observations. Perhaps most suggestive are reports of two transgenic animal models for PD produced by introducing the gene for AS into mice

and into flies [5,6]. The resultant animals exhibit an age-dependent motor dysfunction, the appearance of neuronal AS deposits resembling Lewy bodies (LB), and a loss of dopaminergic neurons. This evidence follows on the heels of earlier work demonstrating that AS is the primary protein component of the LB deposits that are a diagnostic hallmark of PD [7], and that mutations associated with autosomal dominant early onset PD trace to the gene encoding AS [6,8].

Lewy bodies and dementia with Lewy bodies (DLB), constitute the second most common nerve cell pathology, after the neurofibrillary lesions of Alzheimer's disease. Apart from prevalent conditions such as PD and DLB, lately, AS pathology is reported to be associated with a number of neurodegenerative diseases. These include Lewy body variant of Alzheimer's disease (LBVAD), as well as rarer conditions including multiple systems atrophy (MSA), and neurodegeneration with brain iron accumulation type-1 (NBIA-1). Common in these diseases, sometimes referred to as α-synucleinopathies, are microscopic proteinaceous insoluble inclusions in neurons and glia that are composed largely of fibrillar aggregates of AS.

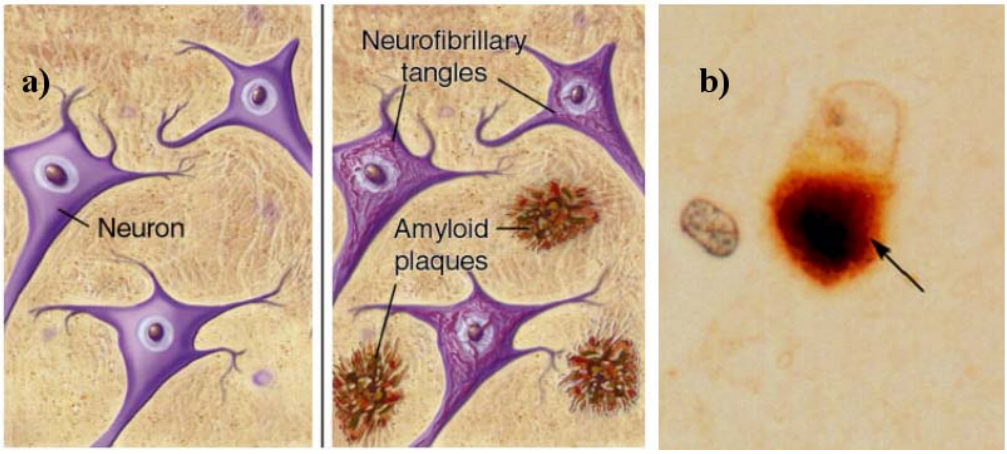


Fig. 5.4 a) Scheme of the formation of amyloid plaques and neurofibrillary tangles that are thought to contribute to the degradation of the neurons in the brain and the subsequent symptoms of Parkinson's disease, **b)** the arrow indicate an α-synuclein positive Lewy body in the cerebral cortex.

AS can form ordered fibrillar aggregates that are morphologically similar to the amyloid fibrils found in Alzheimer's disease neuritic plaques and in deposits associated with other amyloidogenic diseases [9-11]. As in other amyloidogenic diseases, it has been suggested that the aggregation of AS into fibril may play a causative role in the pathogenesis of PD. Despite the intense interest in AS generated by its potential role in PD, the structural properties of this protein have only been characterized at low resolution, using circular dichroism (CD) or other optical techniques. The reason for this is that AS does not appear to possess an intrinsic well-defined native structure [12], making it a member of a class of proteins referred to as intrinsically unstructured proteins [13].

5.3 Alpha-Synuclein: an Intrinsically Unstructured Protein

AS is a highly conserved presynaptic protein in vertebrates and only seven of its 140 amino acids differ between human and mouse.

```
    1          11         21         31         41         51
01  MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV GSKTKEGVVH GVATVAEKTK
61  EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA ATGFVKKDQL GKNEEGAPQE GILEDMPVDP
121 DNEAYEMPSE EGYQDYEPEA
```

Fig. 5.5 Sequence of Human alpha-synuclein.

AS contains a highly conserved amino-terminal repeat region, a hydrophobic central region (residues 61-95) and a less conserved negatively charged carboxy-terminal region (residues 96-140), rich in Pro and in acidic residues like Glu and Asp. The central part of AS (residues 61-95) is highly hydrophobic and is believed to be the mainly responsible for self-aggregation. The highly conserved amino-terminal of all synucleins is also disordered in solution, but can shift to an α-helical conformation that appears to consist of two α-helical regions (1-67) interrupted by a short break [14]. The α-helical N-terminal domain has an 11 amino acid repeat with

a highly conserved hexamer motif (KTKEGV) that contributes to its structural propensity for an amphipathic α-helix.

NMR spectroscopy studies of AS in solution have shown that its C-terminal part remains free and unfolded, and does not associate with vesicles or micelles [15]. AS, intrinsically unstructured in its native state, it can bind, however, a number of ligands and proteins, which likely alter its native state and lead to adopting a partially folded conformation [16].

5.3 Calmodulin interact with Alpha-Synuclein

AS experience dramatic structural alteration upon membrane interaction to form 80% of α-helicity from its natively unfolded structure [12,17]. Via the induction of amphipathic α-helices, AS has been suggested to be involved in membrane biogenesis. Since CaM has been known to recognize its target protein through the basic amphipathic α-helix [18-20], the physiological function of AS has been examined with respect to its CaM regulation and possible implications in cellular messenger system.

Recently, Martinez and co-workers [21] reported that AS interacts with the Ca^{2+}-CaM. Martinez and colleagues found this interaction in a screening test using a special form of AS, made by in-vitro protein translation that can be activated by light. Incorporating chemically modified lysine residues, this AS reacts to light by covalently binding to other proteins, but only if they are in close proximity, because the light-activation of the lysines lasts only a few nanoseconds. The authors spiked bovine brain extracts with this "smart" AS, irradiated the mixture with light, then affinity-purified the AS along with whatever had bound to it. Analysis of the latter showed that several proteins had made sufficient contact to be covalently captured, but subsequent experiments showed that one, about 17 kDa in size, was particularly abundant. After purification this protein was identified as CaM. When the authors tried similar experiments in the absence of calcium, they did not detect binding

between the two proteins, suggesting that their association is calcium dependent and thus may have a physiological role. What this could be is uncertain, but when the authors mixed AS, CaM, and calcium in vitro, they found that the formation of AS fibrils was accelerated. This led them to speculate that Ca^{2+}-CaM drives the assembly of AS-containing multimeric complexes, or perhaps regulates the oligomerization status of AS.

CaM and AS are abundant enough to make this a physiologically relevant interaction. Intracellular CaM concentrations have been estimated to be 1–10 mM, and CaM is estimated to be 0.1–1% of total protein in neurons [22]. Therefore CaM concentrations are high enough to allow for physiologically relevant interactions with AS, which shows the similar micromolar concentration range.

Therefore, we set out to analyze this interaction by NMR and we tackled the task of undestanding the nature of the CaM-AS interaction at the molecular level, and tried to relate it with its possible physiopathological effects.

The research are still in progress, however, preliminary results obtained are discussed in the chapter 6.

LIST OF REFERENCES FOR CHAPTERS 5

1. Elshorst B., Hennig M., Forsterling H., Diener A., Maurer M., Schulte P.,Schwalbe H., Griesinger C., Krebs J. F., Schmid H., and Carafoli E. (1999) *Biochemistry* 38, 12320.

2. Babu YS, Bugg CE, and Cook WJ. (1988) *J. Mol. Biol*. 204, 191.

3. Gellman SH. (1991) *Biochemistry* 30, 6633.

4. Crivici A and Ikura M. (1995) *Annu. Rev. Biophys. Biolmol. Struct.* 24, 85.

5. Feany MB, Bender WW (2000) *Nature* 404,394.

6. Masliah E, Rockenstein E, Veinbergs I, Mallory M, Hashimoto M, Takeda A, Sagara Y, Sisk A, Mucke L (2000) *Science* 287,1265.

7. Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M (1997) *Nature* 388,839.

8. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H, Rubenstein J, Boyer R, Stenroos ES, Chandrasekharappa S, Athanassiadou A, Papapetropoulos T, Johnson WG, Lazzarini AM, Duvoisin RC, Di Iorio G, Golbe LI, Nussbaum RL (1997) *Science* 276,2045.

9. Conway KA, Harper JD, Lansbury PT. (2000) *Biochemistry* 39,2552.

10. El-Agnaf OM, Jakes R, Curran MD, Wallace A (1998) *FEBS Lett* 440,67

11. Hashimoto M, Hsu LJ, Sisk A, Xia Y, Takeda A, Sundsmo M, Masliah E (1998) *Brain Res*. 799,301.

12. Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT Jr. (1996) *Biochemistry* 35,13709.

13. Wright PE, Dyson HJ. (1999) *J Mol Biol.* 293,321.

14. Chandra S., Chen X., Rizo J., Jahn R., Südhof T.C. (2003) *J. Biol. Chem.* 278, 15313.

15. Eliezer D., Kutluay, E., Bussell, R. and Browne, G. (2001) *J. Mol. Biol.* 307, 1061.

16. Weinreb P.H., Zhen, W., Poon, A.W., Conway, K.A. and Lansbury, P.T.J. (1996) *Biochemistry* 35, 13709.

17. Kim J (1997) *Mol Cells* 7,78.

18. Sanyal G., Richard L. M., Carraway K. L. and Puett D. (1988) *Biochemistry* 27, 6229.

19. O'Neil K. T., Erickson-Viitanen S. and DeGrado W. F. (1989) *J. Biol. Chem.* 264, 14571.

20. O'Neil K. T. and DeGrado W. F. (1990) *Trends Biochem. Sci.* 15, 59.

21. Martinez J, Moeller I, Erdjument-Bromage H, Tempst P, Lauring B. (2003) *J Biol. Chem.* 278,17379.

22. Gnegy M. E. (1993) *Annu Rev Pharmacol Toxicol* 33, 45.

CHAPTER 6

A STRUCTURAL MODEL OF THE CALMODULIN-α-SYNUCLEIN COMPLEX. IMPLICATIONS FOR PARKINSON'S DISEASE

Ivano Bertini[1,2], Cristina Del Bianco[1], Yogesh K. Gupta[1], Claudio Luchinat[1,3], Giacomo Parigi[1,3], **Massimiliano Peana[1,4]**, Maria Antonietta Zoroddu[4]

[1]Magnetic Resonance Center (CERM), University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy. [2]Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy. [3]Department of Agricultural Biotechnology, University of Florence, Via Gaetano Donizetti 6, 50144 Firenze, Italy. [4]Department of Chemistry, University of Sassari, Via Vienna 2, 07100 Sassari, Italy.

In preparation

6.1 Introduction

Parkinson's disease (PD) and other neurological disorders are related to the formation in neuronal cells of protein aggregates [1, 2]. In particular, in PD these aggregates (Lewy bodies) are constituted by fibrillar structures mainly containing α-synuclein (AS), a small cytoplasmic protein (15 kDa) that is essentially unfolded in its soluble, monomeric state [3, 4] and is abundant in the presynaptic space. AS is certainly involved in the dopamine signalling process [1]; however, the detailed mode of action of AS is still largely unknown. Silencing or knock-out experiments showed that AS is not essential for neuronal cell life, but has a role in modulating the dopamine signalling process, which would be impaired in the absence of AS especially under conditions of fast repeated signalling [5-7]. There are several ways by which this regulatory role can be exerted. Among them, a possible one is by regulating the activity of phospholipases [8, 9], possibly related to the recycling of presynaptic vesicles. Another possible way is direct binding of AS to presynaptic vesicles, which might contribute to regulate membrane plasticity. As a matter of fact, it has been shown that AS has a strong affinity for membrane surfaces; AS binds to membrane surfaces by adopting a α-helical conformation of its N-terminal domain [4, 10-13]. Increasing experimental evidence suggests that even the soluble, monomeric form of AS is not completely unstructured in solution [4, 14-17]. Soluble and membrane-bound forms of AS are reported to be present in similar amounts *in vivo* [1]. In virtually all the processes in which AS may be involved, calcium ions are also involved [18-23]. Interestingly, calcium ions interact with the carboxylate-rich C-terminal part of AS and modulate its ligand binding and oligomerization ability [24]. The nature and morphology of the initial fibrillar aggregates is also affected by the presence of calcium ions [25].

To further strengthen the links between calcium and AS function/misfunction, it has been shown that monomeric AS interacts strongly with calmodulin (CaM), with dissociation constants of the order of 10-100 nM [26, 27]. Whether this interaction has any physiological or pathological role is not clear, but its understanding may shed new light on the whole field of synucleinopathies.

With these premises, we tackled the task of understanding the nature of the CaM-AS interaction at the molecular level, and to try to relate it with its possible physiopathological effects. The interaction is strong, as previously reported. NMR measurements clearly show that the interaction occurs only with the calcium-loaded form of CaM, and involves the N-terminal part of AS. The lack of internuclear NOEs, between the C-terminal and N-terminal domains of CaM, evidences that the adduct is highly fluxional, possibly involving fast opening and closing of the two CaM domains. The large part of AS is not involved in the interaction with CaM and is still largely – but not completely – unstructured. Calcium-loaded CaM is found to slow down fibril formation by at least a factor two, while apo-CaM has no effect. The fibrils that form in the presence of CaM do not involve co-precipitation of appreciable amounts of CaM.

6.2 Methods

6.2.1 Protein Preparation

^{15}N and ^{13}C labeled wild type (WT) CaM and N60D CaM were prepared and purified as previously reported [28, 29]. Plasmid pDEST-17 (Invitrogen) able to express his-tagged AS was transformed into BL-21-AI *E. coli* strain. A primary culture was from a single bacterial colony diluted up to 1 L either with the LB media or M9 minimal media supplemented with the appropriate isotopically labeled nitrogen (1.2 g $^{15}NH_4Cl$/L) and carbon sources (3 g ^{13}C-glucose/L), and grown in a shaker. AS expression was induced at mid-log phase stage with 0.2% L-arabinose. The bacterial cells were harvested at 7 hrs of post-arabinose induction, centrifuged and resuspended in 15 mL of ice-chilled lysis buffer (0.05 M Tris, 0.5 M NaCl, 0.01 M Imidazole, pH 7.2) containing 0.05 mg/mL lysozyme and 1 mM PMSF and incubated on ice for 30 min to achieve the maximum activity of lysozyme. A lysate obtained by sonication was boiled for 15 min to inactivate all the bacterial proteases co-expressed with his-tagged AS and centrifuged. The supernatant containing his-

tagged AS was loaded on to a hi-trap chelating column charged with $NiSO_4$ solution. Other proteins were removed from the column by passing 25 mL of lysis buffer followed by 25 mL of washing buffer containing 0.02 M imidazole. The his-tagged AS was eluted with 15 mL of elution buffer, containing 0.3 M imidazole. The eluate fractions were pooled and the buffer of eluates was exchanged with the factor Xa protease buffer (0.02 M Tris, 0.1 M NaCl, 0.002 M $CaCl_2$, pH 8.0). His tag was cleaved off from the native AS by incubating the uncut protein with the factor Xa protease enzyme for 5 hrs at 37°C. Finally, the native un-tagged AS was purified by size exclusion chromatography using Superdex 75 column (Amersham pharmacia). The purity of the protein was checked by 1% SDS-15% PAGE electrophoresis.

6.2.2 NMR Sample Preparation and Measurements

NMR samples of CaM were prepared by buffer exchange by Centricon (MILLIPORE) ultrafiltration with membrane cutoff of 10000 Da, as previously reported [28]. NMR samples conditions were 200 mM KCl, 20 mM MES and 5% of D_2O. The pH was adjusted to 6.8 using 0.01 M NaOH or HCl solutions. Final concentrations of single and doubly-labelled samples were around 0.6 mM. Apo WT CaM and N60D CaM samples were titrated up to 4 equivalent of Ca(II) by addition of $CaCl_2$. Tb^{3+}-substituted N60D CaM samples, with Tb^{3+} replacing the Ca^{2+} ion in the second EF site of the N terminal domain, were prepared by titration up to a Tb^{3+}:Ca^{2+} ratio of 0.9:1 [29] to avoid occupancy of other sites.

Labeled WT CaM and N60D CaM were slowly titrated with unlabeled human AS up to final ratios of 1:1 (CaM-AS). No precipitation was observed under these conditions. The titration progress was followed by 2D 1H-15N HSQC spectra at 700 MHz and 298K. Titration of labeled human AS with unlabelled human CaM was performed under the same conditions.

The NMR spectra were acquired on Bruker AVANCE 600 and 700 spectrometers. Both spectrometers were equipped by a triple resonance (TXI) 5 mm

probe with a z-axis pulse field gradient. All spectra were taken at 298 K. The water signal was suppressed using presaturation during the relaxation delay and mixing time or by using the WATERGATE [30] method.

In order to obtain the pseudocontact shifts (PCS), 298 K ^{1}H-^{15}N HSQC spectra of $(Ca_2)_N(Ca_2)_C$CaM-AS and of $(CaTb)_N(Ca_2)_C$CaM-AS were recorded. 256 increments each with 1024 complex data points and 48 transients were collected. PCS were calculated as the difference in chemical shift between corresponding nuclei in the paramagnetic and diamagnetic samples. One bond ^{1}H-^{15}N coupling constants (RDC) were measured at 298 K and 700 MHz by using the IPAP method [31].

6.2.3 Resonance Assignments

Spin system identification of CaM when interacting with AS was made for a substantial number of amino acids trough the analysis of 2D ^{1}H-^{15}N HSQC. The analysis of a 3D ^{15}N-NOESY-HSQC allowed for the assignments of few missing residues. A total of 135 out of 148 HN an N backbone peaks was assigned. Assignment of side chains protons was obtained through the parallel analysis of ^{15}N and ^{13}C NOESY HSQC spectra.

For the AS when interacting with CaM, sequential backbone connectivities relied on 2D ^{1}H-^{15}N HSQC, CBCANH, and CBCA(CO)NH [32] spectra.

6.2.4 Structure Calculation

The program DYANA [33] employing torsion angle dynamics (TAD) combined with a simulated annealing algorithm was used to calculate a series of 200 structures in 10,000 annealing steps starting from randomly generated conformers.

The metal ions were introduced by a linker made of pseudoatoms to the C-terminal and were imposed to be between 2.0 and 2.8 Å from the donor atoms of the four calcium binding loops [34].

The 20 structures with the lowest target function were included in the final set for restrained energy minimization (REM) performed with the Sander module of Amber [35] by applying the NOE and dihedral angle restraints. The RMSD values from the mean for the backbone atoms and for all heavy atoms were determined. The program PROCHECK [36] was subsequently used for secondary structure analysis of the minimized average structures.

6.2.5 AS Fibril Formation

The effect of CaM on fibril forming propensity of AS was monitored by turbidity, and thioT-binding assays. 500 µl of freshly prepared AS (250 µM) was incubated with or without 1:1 amounts of CaM or apo-CaM at 37 °C in glass vials under constant stirring with micro stir bars. 250 µM of CaM without AS was also incubated under same experimental conditions. The buffer condition used for the incubation was as follows: 20 mM Na-MES, 0.1 M NaCl, pH 7.0.

Aliquots (1.5 µl) of AS were taken from each incubation vial at different time intervals and added to 2 mL of 5 µM thioT in 20 mM Na-MES, pH 7.0, according to published procedures [37]. Fluorescence measurements were carried out on a Cary Eclipse spectrofluorimeter using 3.5 mL quartz cuvette (Hellma) with a 1 cm light-path. Fluorescence intensity of thioT was measured at 480 nm, using excitation at 446 nm, an integration time of 0.1 second, and both excitation and emission bandwidths of 10 nm.

Aliquots (9 µl) of AS from the incubation vials were added to 441 µl of the same buffer used for incubation. The absorbance at 360 nm, an indicatore of turbidity [37], was recorded using a 1 cm path-length quartz cuvette (Hellma) in a Cary 50 Scan spectrophotometer (Varian).

6.3 Results

6.3.1 Binding Experiments and Structure Determination

The binding of human CaM to human AS was tested by following the changes in the ^{1}H-^{15}N HSQC spectrum of ^{15}N-labelled CaM upon addition of an increasing amount of unlabelled AS. A number of CaM HSQC signals were affected by the interaction. Along the titration, many signals shifted and some disappeared (fig. 6.1) while the average signal linewidth increased slightly. Further additions of AS did not cause further appreciable changes.

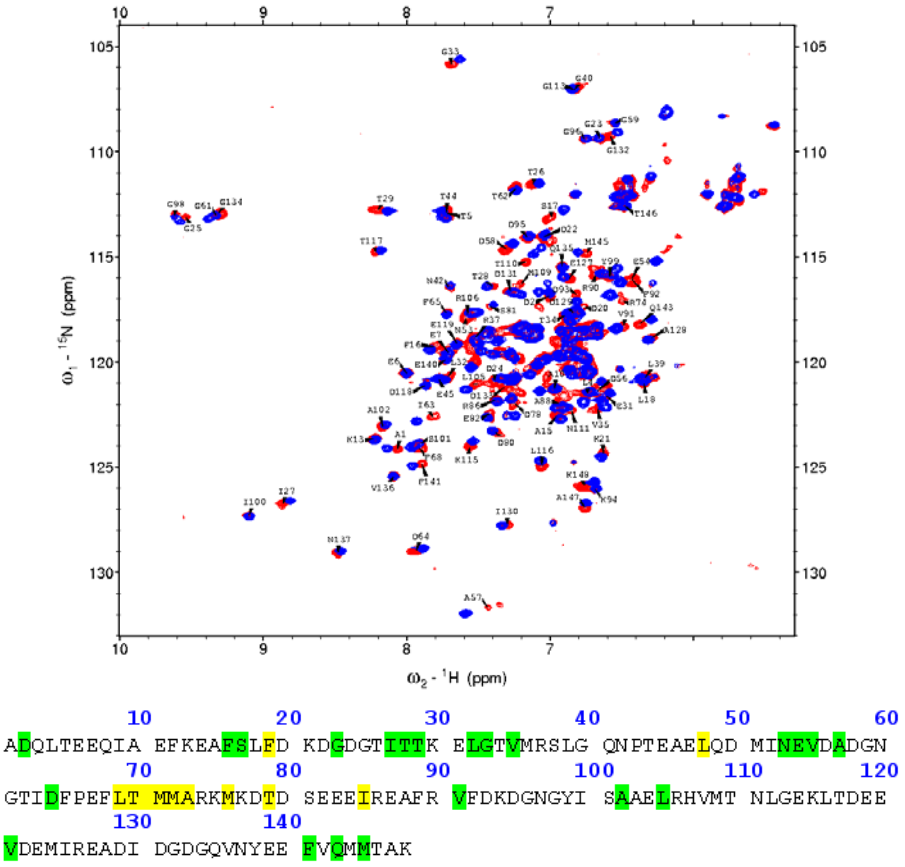


Fig. 6.1 Superposition of ^{1}H-^{15}N HSQC of Human CaM (blue) and ^{1}H-^{15}N HSQC of Human CaM upon complexation with AS (red). Assigned peaks are labeled with their corresponding residue number. The location in the sequence of the residues that shift (green) or disappear (yellow) in the CaM-AS adduct is shown.

The affected signals could be already identified from pre-existing CaM assignment [29] under similar conditions. The magnitude of the shifts is on average sensibly smaller than observed in other cases [38-40]; nevertheless, they clearly point to an interaction of canonical type, in the sense that it extends to both CaM domains as well as to the interdomain linker, similarly to what observed for several other adducts of CaM with target peptides [39, 40] in which CaM wraps around its target with formation of a channel between the two N and C-terminal domains.

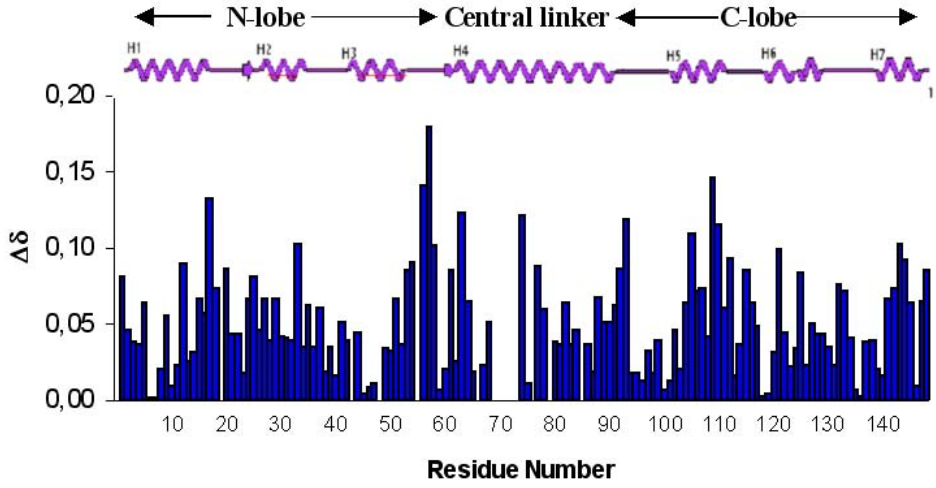


Fig. 6.2 Plot of the change in chemical shift observed for the Human CaM HN and N backbone atoms in the 1:1 ^{15}N CaM-AS complex. Shifts are reported as a weighted average using the formula: $\Delta\delta = \left[(\Delta\delta_{\mathrm{H}})^2 + (\Delta\delta_{\mathrm{N}} \times (\gamma_{\mathrm{N}}/\gamma_{\mathrm{H}}))^2\right]^{0.5}$

Shift changes during titration sets an upper limit for the dissociation constant around 10^{-5} M. Such experiments were performed with both the recombinant WT protein and the protein engineered with a mutation at one of the calcium binding site (N60D) to exploit the possibility provided by paramagnetism, and showed identical behaviour. All experiments described hereafter were then performed with the mutant protein. Analogous titration experiments performed with calcium-free CaM did not show any appreciable alteration of the HSQC spectrum upon addition of AS, pointing to a calcium-dependent binding mode.

Difference CD spectra were recorded by titrating CaM into AS and viceversa, at micromolar concentrations. The binding was essentially stoichiometric even at these much lower concentrations with respect to those of the NMR experiments, indicating a dissociation constant sizably below micromolar, consistent with the 10-100 nanomolar values previously reported [26, 27].

A 1:1 CaM-AS sample containing ^{13}C-^{15}N-doubly labelled CaM and unlabelled AS was thus prepared under the same solution conditions as the ^{1}H-^{15}N HSQC titration described above, and subjected to a complete set of 2D and 3D NMR experiments for structure determination. The acquisition parameters and the obtained constraints are reported in Tables 6.1-6.2.

	Dimension of acquired data			**Spectral width**		
Experiments	(Nucleus)			(Hz)		
	t_1	t_2	t_3	F_1	F_2	F_3
^{1}H-^{15}N HSQC	256 (^{15}N)	2048 (^{1}H)	-	2909	10000	-
^{1}H-^{13}C HSQC	256 (^{13}C)	2048 (^{1}H)	-	26455	9615	-
[^{1}H-^{1}H] NOESY	1024 (^{1}H)	4096 (^{1}H)	-	10000	10000	-
CBCA(CO)NH	96 (^{13}C)	48 (^{15}N)	1024 (^{1}H)	13888	2403	9765
CBCANH	96 (^{13}C)	48 (^{15}N)	1024 (^{1}H)	13888	2403	9765
^{1}H-^{13}C-HSQC-NOESY	240 (^{1}H)	60 (^{13}C)	1024 (^{1}H)	9765	13158	9765
^{1}H-^{15}N-HSQC-NOESY	340 (^{1}H)	48 (^{15}N)	1024 (^{1}H)	9765	2415	9765
^{1}H-^{15}N HSQC (PCS)	256 (^{15}N)	2048 (^{1}H)	-	2484	11261	-
^{1}H-^{15}N HSQC (J_{NH}-split)	256 (^{15}N)	2048 (^{1}H)	-	2484	11261	-
Isotope Filtered [^{1}H-^{1}H] NOESY	1024 (^{1}H)	4096 (^{1}H)	-	10000	10000	-

Table 6.1 Acquisition parameters for NMR experiments.

NMR restraints		
Distance Restraints	**N terminal**	**C terminal**
Total	1686	1285
Intraresidue	204	185
Sequential (i,i+1)	554	408
Medium range from (i,i+2) to (i,i+5)	619	444
Long range from (i,i+5)	309	248
$M^{2+/3+}$ Restraints		
Total	12	12
I site	6	0
II site	6	0
III site	0	6
IV site	0	6
Dihedral Angles		
Total	79	77
ϕ	41	39
ψ	38	38
STRUCTURAL STATISTICS		
Most favoured region	81.2%	83.9%
Additional allowed region	14.5%	12.9%
Generously allowed region	4.3%	3.2%
Disallowed region	0.0%	0.0%

Table 6.2 Type and Number of restraints used in the Dyana Calculation and Structural Statistics

These data, together with the full table of upper distance limits, are submitted to BMRB. The structure calculations, performed with the program DYANA and subsequent restrained energy minimization (REM), yielded well resolved structure families for both CaM domains (Table 6.2). The two structure families, shown in

figure 6.3 and 6.4, have been deposited in PDB (ID code 1YOQ, 1YOS) together with the minimized average structures (1YOR, 1YOT).

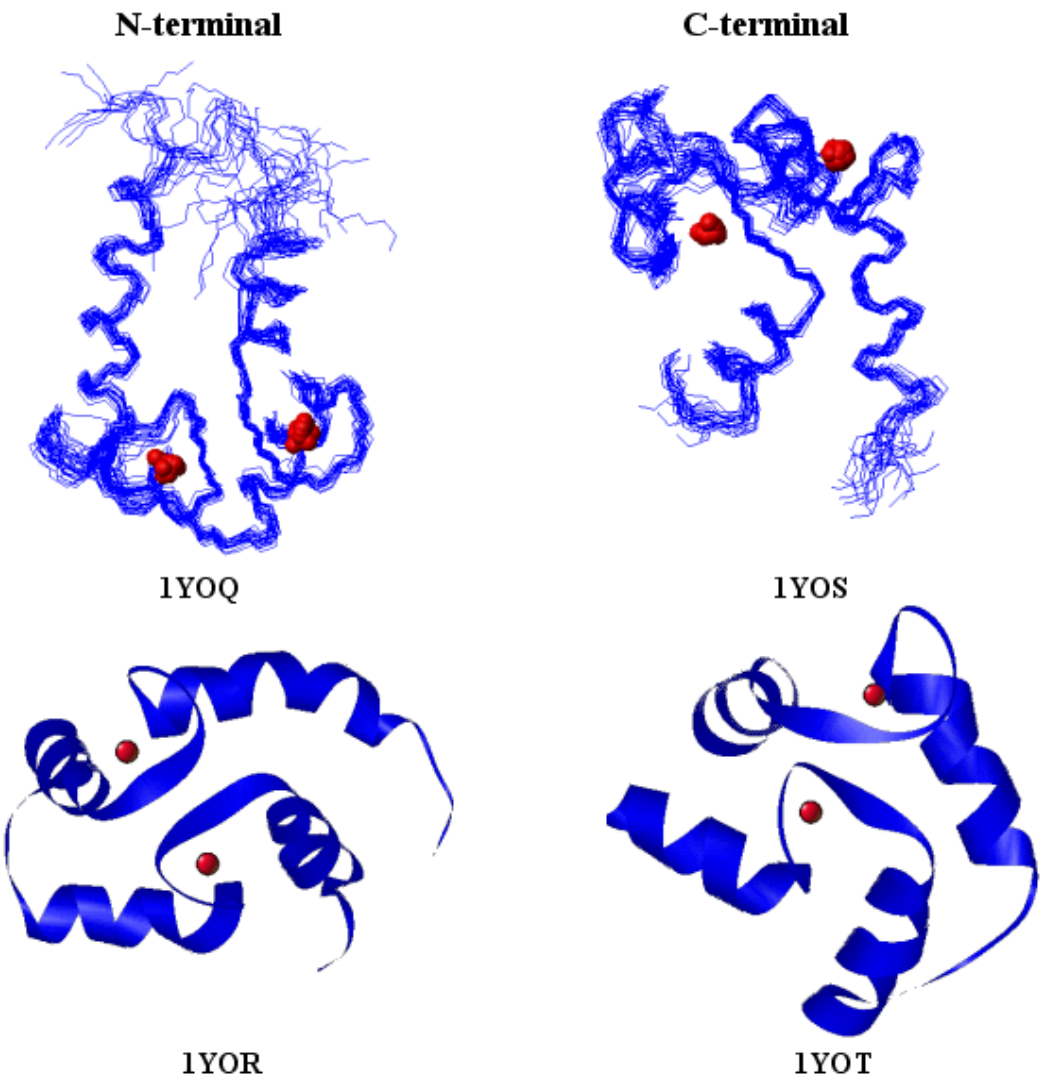


Fig. 6.2 NMR structure families (1YOQ) and minimized average structures (1YOR) of the N terminal, NMR structure families (1YOS) and minimized average structures (1YOT) of the C terminal of the two domains of Human CaM complexed with Human AS.

No NOEs could be unambiguously assigned to CaM-AS intermolecular contacts. Edited experiments either on the CaM or on the AS side also failed to reveal intermolecular NOEs. Possibly due to the unfolded nature of AS, the lack of intermolecular NOEs, poses the structural characterization of adduct between CaM and AS elusive, or it may be indicative of a true fluxional behavior of the complex.

The lack of intermolecular NOEs poses the problem of correctly placing AS within the channel. ^{13}C-^{15}N doubly labelled AS was thus prepared, and the backbone signals assigned using again standard 2D and 3D experiments.

Sidechain signals were also recorded. Very little shifts of either backbone or side chain signals were observed upon addition of CaM. However, several signals disappeared upon binding to CaM. Careful inspection of the spectra showed that these small perturbations tend to cluster in the first third of the AS sequence (fig. 6.6 and fig. 6.7). The N-terminal domain of AS is reported to be the part of the protein with larger helical propensity [4], and therefore to be the most likely region for interaction with CaM.

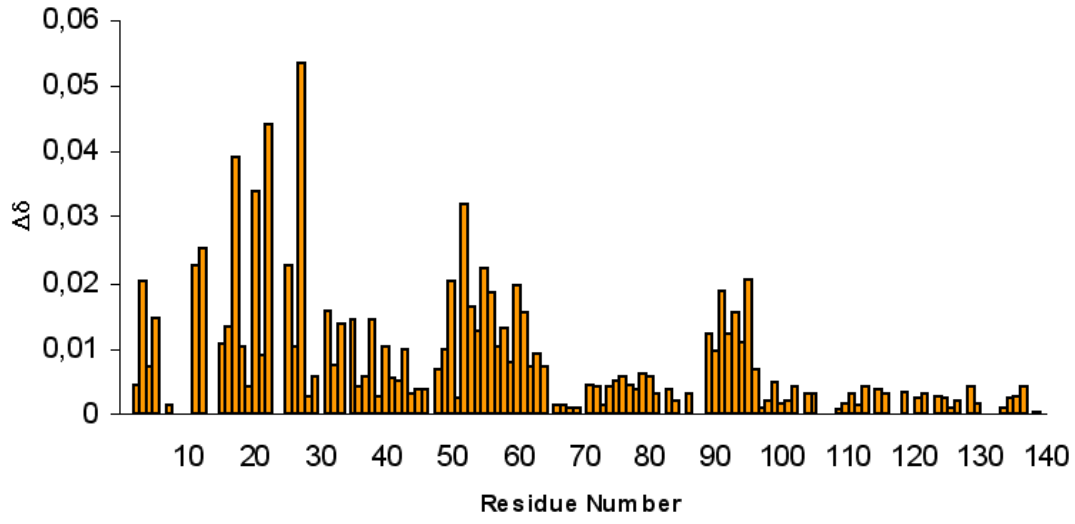


Fig. 6.6 Plot of the change in chemical shift observed for the Human AS HN and N backbone atoms in the 1:1 ^{15}N AS-CaM complex. Shifts are reported as a weighted average using the formula $\Delta\delta = \left[(\Delta\delta_H)^2 + (\Delta\delta_N \times (\gamma_N/\gamma_H))^2\right]^{0.5}$

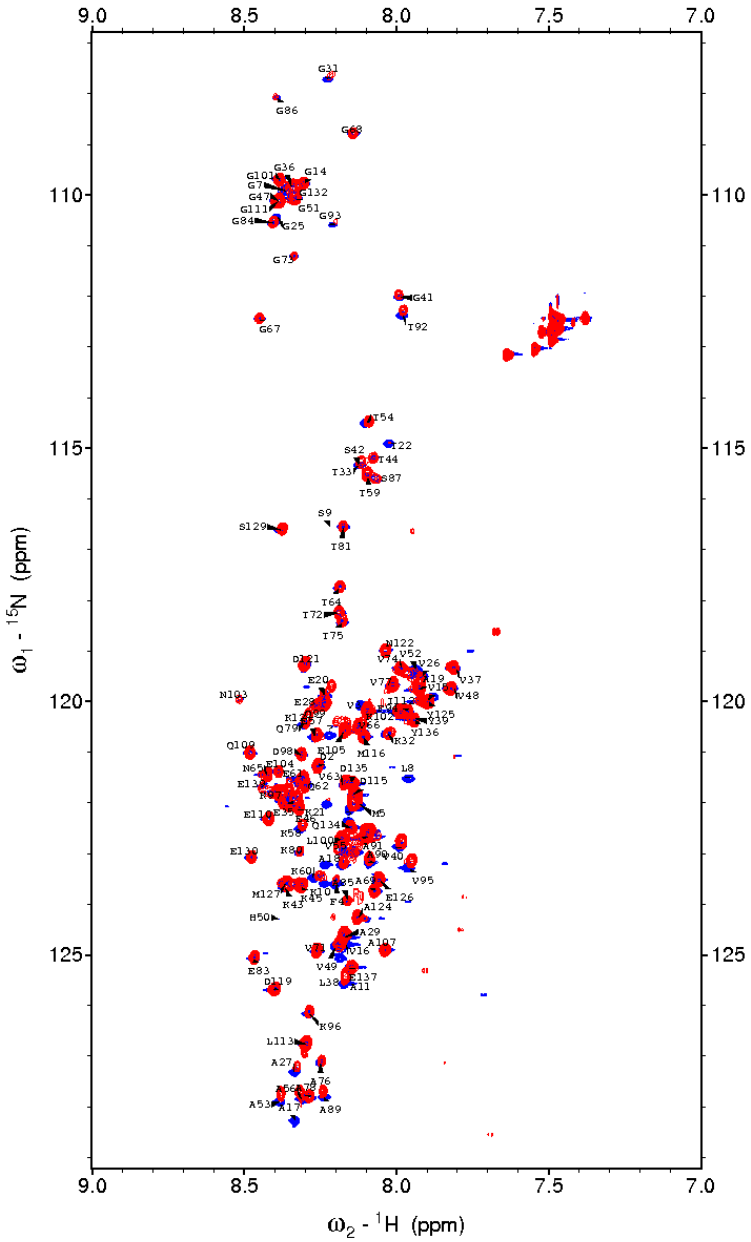


Fig. 6.7 Superposition of ^{1}H-^{15}N HSQC of Human AS (black) and ^{1}H-^{15}N HSQC of Human AS upon complexation with Human CaM (red).

6.3.2 The Fluxional Nature of the Adduct

The use of paramagnetic metal ions may help elucidating the long range spatial relationships and the dynamics in proteins and in protein-protein interactions [17, 34, 42]. Substitution of a paramagnetic lanthanide ion in one of the calcium binding sites

of CaM may provide information on the structure and conformational fluxionality of the system [29]. We have therefore prepared a CaM derivative where Tb^{3+} is selectively substituted at the second calcium binding site of the N-terminal domain of CaM. Selectivity is achieved by a N60D mutation at the second metal binding loop, as previously described [28]. Tb^{3+} provides large pseudocontact shifts (PCS) in the N-terminal domain of the protein and small PCS in the C-terminal domain. The small effect on the C-terminal domain is partly due to the much larger distance of the C-terminal nuclei from the paramagnetic center and partly to the very large fluxionality of the C-terminal domain in free calcium CaM in solution [29]. The presence of the paramagnetic lanthanide also induces residual dipolar couplings (RDC) that can be measured, for instance, in ^{1}H-^{15}N HSQC spectra [43]. RDC do not depend on distance, and therefore the spreading of their values should be approximately the same in both the N- and C-terminal domains, if there were no relative motion between the two. Indeed, in free CaM, due to the very large interdomain motions, the RDC spreading in the C-terminal domain was found to be only about 5% that of the N-terminal domain [29].

PCS and RDC were thus measured from the ^{1}H-^{15}N HSQC spectra of ^{15}N labelled Tb-CaM in its adduct with AS, and, conversely, in the ^{1}H-^{15}N HSQC spectra of ^{15}N labelled AS in its adduct with Tb-CaM. In order to interpret the results, the magnitude and orientation of the Tb-centered magnetic susceptibility tensor have to be established. This is accomplished by fitting the large PCS values of the N-terminal domain of CaM to its energy-minimised average structure, as obtained from the present experimental family (fig. 6.2 1YOR). The fit was good, the correlation coefficient being 0.988. No sensible improvement was obtained by using the PCS themselves in a few structure-refinement cycles using the program PARAMAGNETIC DYANA [44]. The tensor parameters (magnitude and orientations) are similar, but not identical, to those observed for free Tb-CaM [29], consistent with the slight but non negligible structural changes that usually occur from the free to the peptide bound state of the protein [45]. Using the magnetic susceptibility tensor parameters the spreading of the RDC values for the C-terminal domain, assumed rigidly anchored to the N-terminal domain, can be predicted and

compared with the experimental one, as well as with that observed in peptide-free Tb-CaM (fig. 6.8).

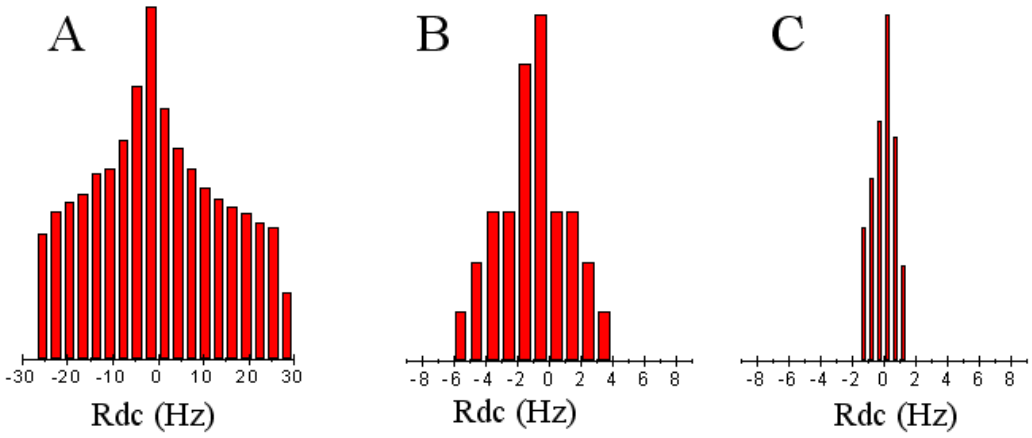


Fig. 6.8 Predicted **(A)** and observed **(B)** spreading of RDC values in the C terminal domain of Tb-CaM complexed with Human AS. The observed spreading in peptide free Tb-CaM **(C)** is also shown for comparison

It is apparent that the spreading is much smaller than predicted (about 15-20%), but still sizably larger than that observed in the peptide-free protein (ca. 5%). It can be concluded that indeed the CaM-AS adduct is highly fluxional.

The availability of the tensor parameters allows us also to predict the sign and magnitude of the PCS in the region of space around the N-terminal domain, and particularly i) in the region occupied by the C-terminal domain and ii) in the regions of space on either side of the channel, where the rest of the AS peptide should extend outside the binding region. The predicted PCS in the region occupied by the C-terminal domain are generally small, and span from slightly positive to slightly negative (fig. 6.9), the latter being predominant, consistently with the observed PCS values for this domain (Table 6.3).

Residue	N (ppm)	H_N(ppm)	Residue	N (ppm)	H_N(ppm)
L4	*0.64*	*0.73*	S101	*-0.10*	*-0.07*
T5	*0.97*	*0.99*	A102	*-0.28*	*-0.12*
E6	*1.22*	*1.08*	R106	*-0.13*	*-0.08*
E7	*1.12*	*0.99*	K115	*-0.09*	*-0.05*
A10	*1.04*	*1.45*	E119	*-0.12*	*-0.10*
E11	*1.30*	*1.25*	E127	*-0.05*	*-0.05*
E12	*1.55*	*1.54*	A128	*-0.04*	*-0.05*
E14	*0.65*	*1.30*	I130	*-0.04*	*-0.05*
A15	*0.68*	*1.71*	D131	*-0.05*	*-0.06*
L18	*0.70*	*0.83*	G132	*-0.02*	*-0.04*
K21	*-0.12*	*-0.12*	G134	*0.08*	*-0.07*
G23	*0.71*	*0.66*	Q135	*-0.03*	*-0.05*
D24	*1.42*	*1.05*	V136	*-0.05*	*-0.07*
N42	*-1.03*	*-0.89*	N137	*-0.06*	*-0.10*
E45	*-1.39*	*-1.04*	Y138	*-0.09*	*-0.10*
V91	*-0.15*	*-0.11*	E140	*-0.10*	*-0.11*
K94	*-0.05*	*-0.11*	F141	*-0.04*	*-0.10*
D95	*0.01*	*-0.05*	V142	*-0.04*	*-0.14*
G96	*0.01*	*-0.08*	Q143	*-0.40*	*-0.35*
N97	*-0.03*	*-0.07*	M145	*-0.21*	*-0.13*
G98	*0.02*	*-0.07*	T146	*-0.92*	*-0.11*
Y99	*-0.05*	*-0.08*	A147	*-0.07*	*-0.11*
I100	*-0.09*	*-0.07*	K148	*-0.11*	*-0.23*

Table 6.3 PCS values experimentally determined for Tb^{3+} substituted ^{15}N-Calmodulin: Synuclein 1:1

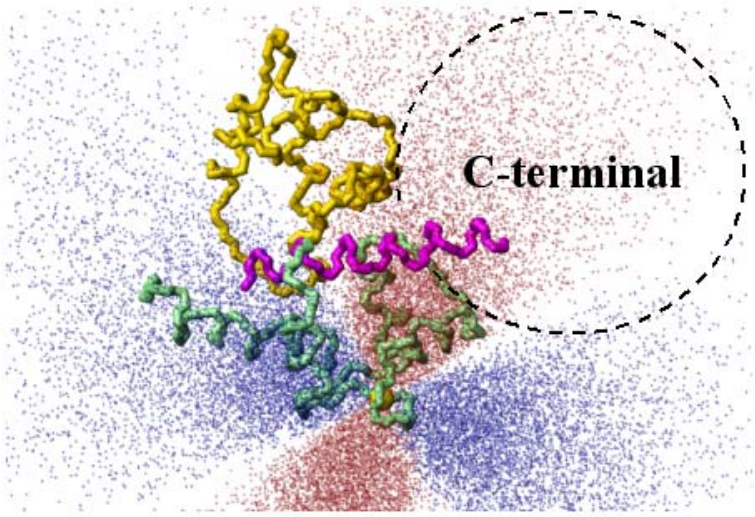


Fig. 6.9 Structural model of the CaM-AS adduct. Green: CaM N-terminal; orange: CaM C-terminal; magenta: N-terminal part of AS (residues 10-30), depicted as α-helical. The C-terminal part extends to the right. The blue dots indicate positive PCS regions, the red dots negative PCS regions, and the density of the dots represent the expected magnitude of the PCS. The PCS values are calculated from the magnetic susceptibility tensor centered on the Tb^{3+} ion in the second metal binding site of the N-teminal domain of CaM.

Experimentally derived PCS, although consistent with the structural model of CaM in a close conformation, have not been used for refinement of the C-terminal domain because of its large relative mobility. Whenever sizable mobility is present, the observed PCS are a weighted average of all sampled conformations [43]; refinement would imply the presence of a single conformation.

An interesting finding is that the sign of the PCS is opposite on opposite sides of the channel. This feature can be used to discriminate between the two possible orientations of the AS chain with respect to the channel. Despite AS is largely unstructured, its finite length poses intrinsic limits to the space that can be sampled by the C-terminal chain coming out of one of the two sides of the channel. In other words, it is expected that the PCS of the C-terminal part be predominantly positive if the C-terminal chain comes out on one side of the channel, and predominantly negative if it comes out on the other side. The measured ^{1}H and ^{15}N PCS on AS are small, and for the ^{15}N nuclei are within the indetermination of the measurements. A plot of the ^{1}H PCS as a function of residue number (fig. 6.10), however, shows a clear bias towards negative PCS values, especially around residues 70-100, i.e. well outside the CaM binding region.

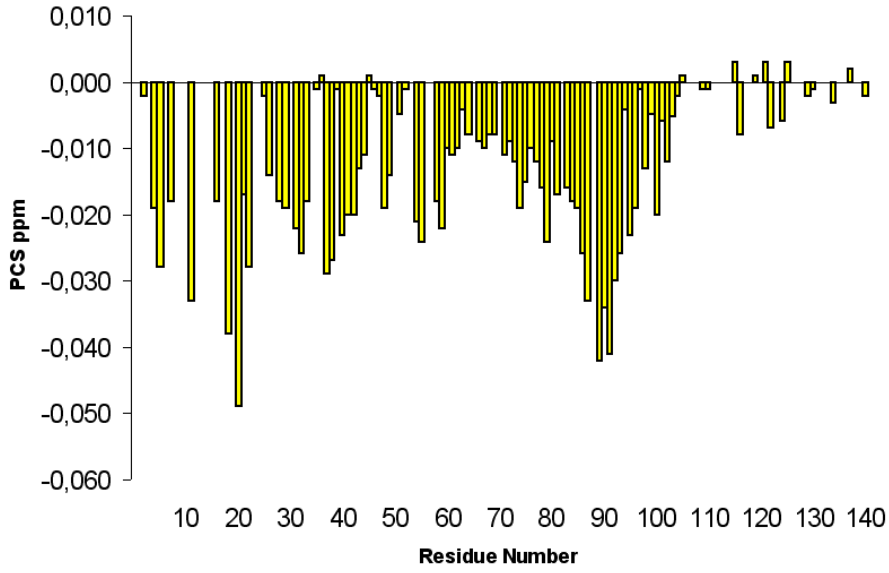


Fig. 6.10 Plot of PCS observed for the AS HN backbone atoms, for the 1:1 ^{15}N AS:CaM-Tb loaded complex.

This observation shows that the orientation depicted in figure 6.9 is the correct one. Adventitious binding of traces of Tb(III) to AS sites might in principle jeopardize the significance of the observed PCS, even if Tb(III) was used in substoichiometric amount with respect to CaM to minimize traces of free metal. However, metal ions are well known to have preferential binding sites in the carboxyl-rich C-terminal end (residues 109-140) of AS [24], and this preference is even more marked for lanthanides. Therefore, the clustering of negative PCS values in the region 70-100 of AS is unambiguously due to a true effect from the Tb(III) ion bound at the N-terminal of CaM.

In summary, the use of the paramagnetic metal has allowed us to i) verify the fluxionality of the adduct and ii) assess the orientation of the AS chain within the CaM channel. The question of the orientation is not a trivial one, as both orientations have been experimentally observed in other adducts of CaM with target peptides [46]. Figure 6.9 collects the structural information obtained in the present work: the two CaM domains are well structured; their reciprocal orientation has a higher uncertainly, but still allows one to derive a structural model consistent with a closed form of CaM; the interacting portion of AS is not known exactly but lies within the first 50-60 residues; finally, the orientation of the peptide within the channel is determined by the negative sign of most ^{1}H PCS in the C-terminal domain. The relative motion of AS with respect to CaM in the adduct could be visualized either as a gliding along the two domains or sticking to one of the two when CaM undergoes an opening/closing movement. In the latter case, AS would be predominantly bound to the C-terminal domain, as PCS and especially RDC would be larger than observed if it were bound to the N-terminal domain. The lack of any observed NOEs between AS and any CaM domain makes the gliding model preferable with respect to predominant binding to the C-terminal domain.

Another interesting finding is that not all RDC values of AS in the adduct with Tb-CaM are negligibly small. The RDC values of AS should be very close to zero if the peptide were fully unstructured and able to sample nearly all NH vector orientations. Instead, positive and negative values of up to 6-7 Hz in magnitude are observed, and they are scattered more or less homogeneously throughout the AS

sequence. It is of course impossible to derive any structural model from these data, but they clearly indicate that, although lacking any consistent three-dimensional organization, the central and C-terminal parts of AS may have a tendency to take up some local transient secondary structure. It is possible that this tendency is enhanced or at least altered by the entrapment of the N-terminal domain in the CaM channel. In turn, this tendency may not be innocent with respect to the tendency of AS to form fibrils.

6.3.4 Fibril Formation

The tendency of AS to form fibrils was estimated by turbidity measurements and thioflavin-T assay as previously reported [37]. AS alone and in the presence of apo-CaM showed an increase in both turbidity and thioflavin-T fluerescence around 5 hrs incubation time, while the onset of both phenomena occurred at times twice as long in the 1:1 AS-CaM sample. It can be concluded that CaM, but not apo-CaM, has a retarding effect on fibril formation under the present experimental conditions. It was also noted that the thioflavin-T response in the presence of CaM was more intense. As the thioflavin-T monitors exposed β-sheet structures, this observation suggests that the morphology of the resulting fibrillar material may be different. To assess whether this material were constituted by CaM-AS adducts instead of AS alone, calcium content was monitored both in the fibrils and in the supernatant. Only about 3% of CaM was found in the fibrils according to their calcium content, suggesting that the latter were essentially constituted by AS alone. These data are paralleled by UV measurements and gel electrophoresis on the supernatant, which contained mostly CaM and very little AS.

6.4 Discussion

Whatever the functional significance is of the CaM-AS interaction, and whatever can be its role in the pathogenesis of synucleinopaties, the present findings add solid knowledge to the mechanistic details of this interaction. Analyzing this knowledge in the frame of what is already known about these two proteins may actually provide hints and suggestions for future research. Two previous reports have shown that CaM interacts with AS both *in vitro* and *in vivo* [26, 27]. The binding is strong, as confirmed by the present research. CaM is essential for all cells and is relatively abundant, with concentrations in the micromolar range [47]. AS is localized in the presynaptic space of neurons, and the concentrations of its soluble monomeric form in these compartments are also above the micromolar range [48]. CaM cycles between its apo form and its calcium-bound form during calcium transients, which bring the intracellular calcium concentrations above 1 μM. Calcium CaM, when formed, triggers a number of events, depending on the particular state of the cell. More than 100 different physiological partners of calcium CaM are known [49]. Usually the concentration of calmodulin is sizably higher than the individual concentration of its partners, although probably lower than their total concentration [47], thereby allowing for multiple events being triggered by a single calcium transient. Only the calcium bound form of CaM is able to bind AS. The dissociation constant ensures that at the physiological concentrations the adduct would be close to 100% formed, although competition with stronger binding partners is possible. AS could thus be a novel *bona-fide* CaM partner. As the total concentrations of CaM and AS are of the same order of magnitude, from the thermodynamic point of view a large fraction of soluble AS in the presynaptic space could be bound by CaM after each calcium transient. A crucial point is the frequency of occurrence and duration of calcium loaded CaM in the absence of any partner. Measurements performed in dopaminergic neurons of the substantia nigra indicate that the basal frequency of calcium transients is around 1 Hz and the duration is of the order of tenths of seconds [50-52]. CaM responds effectively to this transients [53]. During this time scale the CaCaM-AS adduct should form. This event depends on the k_{on} rate constant, which

can be estimated from the present data. A few CaM ^{1}H-^{15}N HSQC cross peaks disappear during titration of CaM with AS, indicating exchange broadening, while several are in fast exchange. From the separation in chemical shift of the latter resonances, a reasonable estimate of the adduct lifetime of about 10^{-2} s is obtained. With a dissociation constant of about 10^{-7} M, and assuming that formation of the adduct is a one-step process, the rate constant for the formation of the adduct would be 10^{9} $M^{-1}s^{-1}$. This order of magnitude estimate, although relatively high, is reasonable as the k_{on} of proteins of similar size like barnase/barstar was measured to be 6×10^{8} $M^{-1}s^{-1}$ [54]. Such k_{on} and k_{off} values warrant the prompt formation and dissociation of the adduct at each calcium transient, i.e. for a significant fraction of time. During faster calcium firing events the adduct could remain formed for even longer times.

The lifetimes of CaM adducts with some of its target proteins are similar to that of the CaM-AS adduct [55], while many others are somewhat longer, as judged from the completely slow exchange observed in NMR experiments [39, 40]. Many of its adducts with downstream effector proteins might actually survive throughout moderately fast trains of calcium transients, while that with AS might not, except for the fastest. Therefore, rather than responding to calcium signals to perform a function, AS might actually act as a modulator of CaM operating on its calcium form.

On the opposite side, another implication of AS binding to CaM might be that CaM modulates the ability of AS to bind membranes, either of the dopamine-carrying vesicles or of the presynaptic junction. Binding to CaM essentially shifts the equilibrium towards the membrane unbound state of monomeric synuclein [26], while low calcium transient frequencies, corresponding to the resting state of the cell, would favour binding of AS to membranes. It is interesting that the mode of binding to CaM elucidated in the present work implies the N-terminal part of AS, which is the part involved in membrane binding, suggesting that modulation occurs through direct competition.

Finally, the retardation of fibril formation by calcium-CaM, but not by apo-CaM, may have implications for the molecular mechanism of synucleinopaties. Fibril

formation, like the formation of any aggregate, either amorphous or crystalline, is a complex phenomenon that needs a seeding event. Indeed, ample evidence suggests that AS fibrils are thermodynamically much more stable than monomeric AS [17, 56]. The latter should therefore be regarded as a metastable form, which survives for kinetic rather than thermodynamic reasons. If free monomeric AS occasionally assumes a "dangerous" β-strand-like conformation of its central part, accidental interaction with another AS molecule may trigger the same transition in the latter, and act as a seed for β-sheet formation. It is very likely that such transient β-strand-like conformation in CaM-bound AS is disfavoured. Therefore, in the presence of CaM, a lower amount of free AS in solution would be available to seed fibril formation. In this respect it should be noticed that, under the conditions of the experiments typically used to induce and monitor fibril formation, and also adopted here, many such seeds may form at the same time, producing a rather amorphous fibrillar material. Interestingly, the fibrils that eventually form in the presence of CaM are of better quality, as judged from the thioflavin-T assay, suggesting that the number of seeds has been reduced. This in turn is consistent with the idea that the protective effect is given by the subtraction of free AS from solution.

Fibril formation occurs over time scales that are much longer than the lifetime of the CaM-AS adduct and longer than the calcium transient intervals. Therefore, one might envision the transient binding of free AS to CaM when calcium concentration in the presynaptic space increases, as a way the cell adopts to reduce the probability that a β-strand-like conformation is adopted by a given AS molecule in the time interval between calcium transients when it is free in solution. Calcium transients in the presynaptic space are related to neuronal firing [57]; AS is also related to neuronal firing either through binding to vesicle membranes, or through regulating the synthesis of membrane components, or both; calcium is known to bind to the C-terminal part of AS and to modulate its tendency to fibrillate (very high, non-physiological calcium concentrations are needed for this interaction: it is possible however that such concentrations are reached locally in some pathologies). Now we have shed more light on the relationship that may exist between calcium concentration changes and AS through its binding to calcium CaM. This relationship

is more subtle than the direct calcium-AS interaction at high calcium concentrations, but far richer in physiological and possibly pathological implications.

LIST OF REFERENCES FOR CHAPTERS 6

1. Lotharius J. & Brundin P. (2002) *Nature Reviews Neuroscience* 3, 932.

2. Frasier M. & Wolozin B. (2004) *Experimental Neurology* 187, 235.

3. Weinreb P. H., Zhen W. G., Poon A. W., Conway K. A. & Lansbury P. T. Jr. (1996) *Biochemistry* 35, 13709.

4. Eliezer D., Kutluay E., Bussell R. Jr. & Browne, G. (2001) *J. Mol. Biol.* 307, 1061.

5. Abeliovich A., Schmitz Y., Farinas I., Choi-Lundberg D., Ho W. H., Cast P. E., Shinsky N., Verdugo J. M., Armanini M., Ryan A. *et al.* (2000) *Neuron* 25, 239.

6. Murphy D. D., Reuter S. M., Trojanowski J. Q. & Lee V. M. (2000) *J. Neurosci.* 20, 3214.

7. Chandra S., Fornai F., Kwon H.-B., Yazdani U., Atasoy D., Liu X., Hammer R. E., Battaglia G., German D. C., Castillo P. E. *et al.* (2004) *Proc Natl Acad Sci U S A* 101, 14966.

8. Jenco J. M., Rawlingson A., Daniels B. & Morris A. J. (1998) *Biochemistry* 37, 4901.

9. Narayanan V., Guo Y. & Scarlata S. (2005) *Biochemistry* 44, 462.

10. Bussell R. Jr. & Eliezer D. (2003) *J. Mol. Biol.* 329, 763.

11. Chandra S., Chen X., Rizo J., Hahn R. & Südhof T. C. (2003) *J. Biol. Chem.* 278, 15313.

12. Ulmer T. S., Bax A., Cole N. B. & Nussbaum R. L. (2004) *J. Biol. Chem.* 11, 9595.

13. Lee J. C., Langen R., Hummel P. A., Gray H. B. & Winkler J. R. (2004) *Proc Natl Acad Sci U S A* 101, 16466.

14. Bussell R. Jr. & Eliezer, D. (2001) *J. Biol. Chem.* 276, 45996.

15. Fernandez C. O., Hoyer W., Zweckstetter M., Jares-Erijman E. A., Subramaniam V., Griesinger C. & Jovin T. M. (2004) *The EMBO Journal* 23, 2039.

16. Bertini I., Gupta Y. K., Luchinat C., Parigi G., Schlörb C. & Schwalbe H. (2005) *Angew. Chem. Int. Ed.* 44, 2.

17. Bertoncini C. W., Jung Y.-S., Fernández C. O., Hoyer W., Griesinger C., Jovin T. M. & Zweckstetter M. (2005) *Proc Natl Acad Sci U S A* 102, 1430.

18. Ales E., Tabares L., Poyato J. M., Valero V., Lindau M. & de Toledo G. A. (1999) *Nature Cell Biology* 1, 40.

19. Schneggenburger R. & Neher E. (2000) *Nature* 406, 889.

20. Goncalves P. P., Meireles, S. M., Neves, P. & Vale, M. G. P. (2000) *Brain Research Protocols* 5, 102.

21. Miras-Portugal M. T., Pintor J. & Gualrix J. (2003) *J. Membr. Biol.* 194, 1.

22. Turner T. J. (2004) *J. Neurosci.* 24, 11328.

23. Macleod G. T., Marin L., Charlton M. P. & Atwood H. L. (2004) *J. Neurosci.* 24, 2496.

24. Nielsen M. S., Vorum H., Liendersson E. & Jensen P. H. (2001) *J. Biol. Chem.* 276, 22680.

25. Lowe R., Pountney D. L., Jensen P. H., Gai W. P. & Voelcker N. H. (2004) *Protein Sci.* 13, 3245.

26. Lee D., Lee S.-Y., Lee E.-N., Chang C.-S. & Paik S. R. (2002) *Journal of Neurochemistry* 82, 1007.

27. Martinez J., Moeller I., Erdjument-Bromage H., Tempst P. & Lauring B. (2003) *J. Biol. Chem.* 278, 17379.

28. Bertini I., Gelis I., Katsaros N., Luchinat C. & Provenzani A. (2003) *Biochemistry* 42, 8011.

29. Bertini I., Del Bianco C., Gelis I., Katsaros N., Luchinat C., Parigi G., Peana M. Provenzani A. & Zoroddu M. A. (2004) *Proc. Natl. Acad. Sci. USA* 101, 6841.

30. Piotto M., Saudek V. & Sklenar V. (1992) *J. Biomol. NMR* 2, 661.

31. Ottiger M., Delaglio F. & Bax A. (1998) *J. Magn. Reson.* 131, 373.

32. Muhandiram D. R. & Kay L. E. (1994) *J. Magn. Reson. Ser. B* 103, 203.

33. Güntert P., Mumenthaler C. & Wüthrich K. (1997) *J. Mol. Biol.* 273, 283.

34. Allegrozzi M., Bertini I., Janik M. B. L., Lee Y.-M., Liu G. & Luchinat C. (2000) *J. Am. Chem. Soc.* 122, 4154.

35. Pearlman D. A., Case D. A., Caldwell J. W., Ross W. S., Cheatham T. E., Ferguson D. M., Seibel G. L., Singh U. C., Weiner P. K. & Kollman P. A. (1997) *AMBER 5.0* (University of California, San Francisco).

36. Laskowski R. A., MacArthur M. W., Moss D. S. & Thornton, J. M. (1993) *J. Appl. Crystallogr.* 26, 283.

37. Hoyer W., Antony T., Cherny D., Heim G., Jovin T. M. & Subramaniam V. (2002) *J. Mol. Biol.* 322, 383.

38. Elshorst B., Hennig M., Forsterling H., Diener A., Maurer M., Schulte P., Schwalbe H., Griesinger C., Krebs J. F., Schmid H. *et al.* (1999) *Biochemistry* 38, 12320.

39. Ikura M., Kay L. E., Krinks M. & Bax A. (1991) *Biochemistry* 30, 5498.

40. Craescu C. T., Bouhss A., Mispelter J., Diesis E., Popescu A., Chiriac M. & Barzu O. (1995) *J. Biol. Chem.* 270, 7088.

41. Ikura M., Clore G. M., Gronenborn A. M., Zhu G., Clee C. & Bax A. (1992) *Science* 256, 632.

42. Pintacuda G., Keniry M. A., Huber T., Park A. Y., Dixon N. E. & Otting G. (2004) *J.Am.Chem.Soc.* 126, 2963.

43. Bertini I., Luchinat C. & Parigi G. (2002) *Progr. NMR Spectrosc.* 40, 249.

44. Banci L., Bertini I., Cremonini M. A., Gori Savellini G., Luchinat C., Wüthrich K. & Güntert P. (1998) *J. Biomol. NMR* 12, 553.

45. Kurokawa H., Osawa M., Kurihara H., Katayama N., Tokumitsu H., Swindells M. B., Kainosho M. & Ikura M. (2001) *J. Mol. Biol.* 312, 59.

46. Vetter S. W. & Leclerc, E. (2003) *Eur. J. Biochem.* 270, 404.

47. Persechini A. & Stemmer P. M. (2002) *Trends Cardiovasc. Med.* 12, 32.

48. Shtilerman M. D., Ding T. T. & Lansbury P. T. Jr. (2002) *Biochemistry* 41, 3855.

49. Shen X., Valencia C. A., Szostak J., Dong B. & Liu R. (2005) *Proc. Natl. Acad. Sci. U. S. A* 102, 5969.

50. Davis M. D. & Schmidt J. J. (2000) *J. Neuroscience Methods* 99, 9.

51. Wilson C. J. & Callaway J. C. (2000) *J. Neurophysiol.* 83, 3084.

52. Kerr J. N. D. & Plenz D. (2002) *J. Neurosci.* 22, 1499.

53. Martin S. R., Linse S., Johansson C., Bayley P. M. & Forsén S. (1990) *Biochemistry* 29, 4188.

54. Schreiber G. & Fersht A. R. (1993) *Biochemistry* 32, 5145.

55. Porumb T., Crivici A., Blackshear P. J. & Ikura M. (1997) *Eur. Biophys. J.* 25, 239.

56. Linding R., Schymkowitz J., Rousseau F., Diella F. & Serrano L. (2004) *J. Mol. Biol.* 342, 345.

57. Sharma G. & Vijayaraghavan S. (2003) *Neuron* 38, 929.

CHAPTER 7

SPECTROSCOPIC AND POTENTIOMETRIC STUDIES OF NICKEL BINDING TO THE C-TERMINAL SEQUENCE OF CAP43 PROTEIN

7.1 Cap43 Protein: a Nickel Carcinogenesis-related Marker

Cancer is a second, leading cause of death in humans worldwide. Exposure to a wide variety of natural and man-made substances in the environment accounts for at least two-thirds of all the cases of cancer. These environmental factors include lifestyle choices like cigarette smoking, excessive alcohol consumption, poor diet, lack of exercise, excessive sunlight exposure. Other factors include exposure to certain medical drugs, hormones, radiation, viruses, bacteria, and

environmental chemicals: solvents, dioxins, polycyclic aromatic hydrocarbons and several metal compounds that may be present in the air, water, food, and workplace.
In 1990, the International Agency for Research on Cancer (IARC) evaluated the carcinogenicity of nickel on the basis of well documented studies both *in vitro* and *in vivo* [1-7].
All Ni compounds except for metallic Ni were classified as carcinogenic to humans.
Human exposure to Ni is environmental and occurs primarily via inhalation and ingestion [8].
Although occupationally exposed nickel mining, smelting and refinery workers have an excess incidence of lung and nasal cancer [9], the utilization and disposal of nickel-containing products in modern industry lead to environmental pollution by both soluble and insoluble forms of Ni, resulting in a serious hazard for not typically work-related exposed targets (population). Additionally, combustion of fossil fuel contributes significantly to environmental burden, mostly by producing aerosols containing soluble Ni [10].
All these evidences make it a modern toxic and carcinogenic environmental contaminant.
The understanding of how nickel exerts its toxic and carcinogenic effects at a molecular level may be important in risk assessment, as well as in the treatment and prevention of occupational diseases.

The study on molecular level of Ni carcinogenesis has been challenging because carcinogenic nickel compounds were weakly mutagenic in most assay systems, even though they were able to produce oxidative DNA damage and inhibit DNA repair activity [8,11-15].
Current studies in the clinical cancer research area focus on either the early detection of this disease or on the development of new selective treatment tools. New tumour markers can provide aid in cancer early diagnosis and create novel treatment opportunities.

Cap43, also named Ndrg1, following the discovery of its gene *NDRG1*, is a stress responsive protein which shuttles between cytoplasm and nucleus upon certain noxious stimulations [16-20].

The Cap43 gene is induced by a rise in free intracellular Ca^{2+} following nickel exposure [16,21]. No other metal compound significantly induced expression of this gene, indicating that it is expressed with marked specificity to Ni(II) exposure [22].
Exploration of NDRG1 protein expression patterns in various tissues showed that Cap43 protein is overexpressed, compared to normal tissues, in a variety of cancers including brain, breast, lung, and melanoma cancers [16,22]. Because of its differential expression in cancer tissues and the high stability of the protein, Cap43 is proposed as a useful new tumor marker.
This finding makes the Cap43 protein an interesting candidate for studies of molecular mechanisms implicated in toxicity and carcinogenicity of nickel compounds, because a promising way to unveil these molecular events is to study the characteristics of the proteins expressed by the genes specifically induced by these carcinogens.

7.2 Nickel Binding Sites in Cap43 Protein

Recently it has been postulated that Cap43 is as a tumor suppressor gene because it causes the reversal of malignant phenotype in colon cancer cells [23].
Acute stress, such as heat, and some metals such as cadmium, mercury, lead, zinc and Ni(II), induce a specific group of stress proteins referred to as heat-shock proteins. The heat-shock proteins contribute to the survival of cells following a variety of stresses. Similarly nickel, like other heavy metals, is able to increase the levels of metallothionein (MT) and acute phase reactant proteins (APRs), a metal-binding protein that has also been found to have a protective and detoxification role in the cellular response to acute stresses determined by heavy metals.
These proteins have metal-binding domains that exert a protective effect against the cellular toxicity of metal ions. For example, MTs have high affinity for metals due to their high cysteine content [24]; purified Limulus C-reactive protein, a member of the APR family, has sulphydryl groups that bind strongly to mercury [25], and HSPA

from Helicobacter pylori features a series of cysteine and histidine residues, resembling anchoring binding site for metal ions [26].

For this reason we focused our attention on the ability of nickel(II) to interact with Cap43 protein.

Cap43 has no cysteine or histidine-rich motifs for metal binding, but it possesses a mono-histidine fragment composed by 10 amino acids (Thr–Arg–Ser–Arg–Ser–His–Thr–Ser–Glu–Gly) whose sequence is repeated consecutively three times, suggesting a nickel binding motif at the C- terminus (fig. 7.1). It should be mentioned that such mono-histidine fragments, e.g. octapeptide repeated regions in prion proteins, play a critical role in metal metabolism using sets of His imidazoles as the binding sites for metal ions [27].

```
  1 MSREMQDVDLAEVKPLVEKGETITGLLQEFDVQEQDIETLHGSVHVTLCGTPKGNRPVIL
 61 TYHDIGMNHKTCYNPLFNYEDMQEITQHFAVCHVDAPGQQDGAASFPAGYMYPSMDQLAE
121 MLPGVLQQFGLKSIIGMGTGAGAYILTRFALNNPEMVEGLVLINVNPCAEGWMDWAASKI
181 SGWTQALPDMVVSHLFGKEEMQSNVEVVHTYRQHIVNDMNPGNLHLFINAYNSRRDLEIE
241 RPMPGTHTVTLQCPALLVVGDSSPAVDAVVECNSKLDPTKTTLLKMADCGGLPQISQPAK
301 LAEAFKYFVQGMGYMPSASMTRLMRSRTASGSSVTSLDGTRSRSHTSEGTRSRSHTSEGT
361 RSRSHTSEGAHLDITPNSGAAGNSAGPKSMEVSC
```

Fig. 7.1 Sequence of nickel-specifically induced protein Cap43. The three-time repeated 10-amino acid fragment (green) is shown.

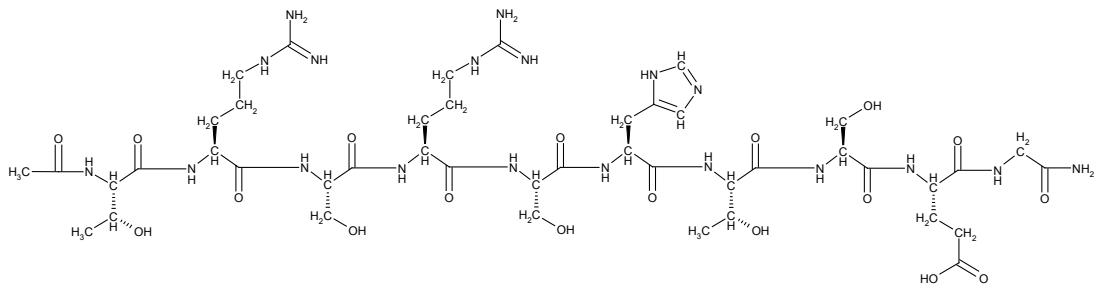


Fig. 7.2 Ten amino acids (Thr–Arg–Ser–Arg–Ser–His–Thr–Ser–Glu–Gly) sequence

During my PhD course I have analyzed, for Ni(II) binding, the 20-amino acid C-terminal sequence of the Cap43 protein -*TRSRSHTSEG-TRSRSHTSEG-,* and 30-amino acid sequence, *TRSRSHTSEG-TRSRSHTSEG-TRSRSHTSEG,* by a combined pH-metric and spectroscopic (UV-VIS, CD, NMR) study.

The results of a study on the interaction of Ni(II) ions with the fragments of Cap43 protein containing respectively two and three repeated –TRSRSHTSEG– amino acid sequences showed that each 10-amino acid fragment coordinates one metal ion. The coordination of the metal ion starts from the imidazole nitrogen atom of the histidine residue, and with increasing the pH, Ni(II) ions are able to deprotonate successive peptide nitrogen atoms, forming Ni(II)–N^- bonds, until a $NiH_{-3}L$ and $Ni_2H_{-6}L$ species for the 20- and $NiH_{-3}L$, $Ni_2H_{-6}L$ and $Ni_3H_{-9}L$ complexes for the 30-amino acid fragments, are formed (above pH 8). The formation of stable five-membered chelate rings by consecutive nitrogens is the driving force of the coordination process.

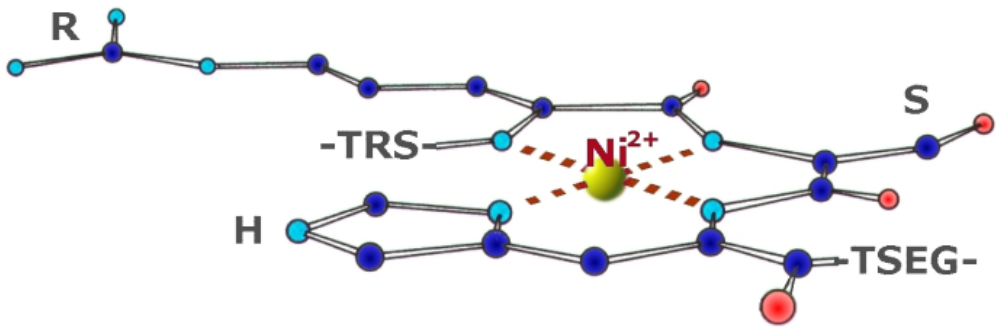


Fig. 7.3 Scheme of 4N coordination pattern

At physiological pH (7.4) and mM concentrations of nickel(II), dependently on the metal-to-ligand molar ratio, the 20-amino acid fragment forms the Ni_2L complex (2:1 molar ratio), while the 30-amino acid fragment forms the NiL (1:1), Ni_2L (2:1), and Ni_3L (3:1) complexes, where each metal ion is coordinated by the imidazole nitrogen atom of the histidine residue of each 10-amino acid sequence. The Ni_2L and Ni_3L complexes of these peptides are more stable by about 0.7–1.5 orders of magnitude in comparison to the stability constants evaluated for these

systems considering an independent coordination of each metal ion to a sing1e *TRSRSHTSEG* amino acid sequence [28]. The coordination of two or three metal ions with 4N {N_{Im}, $3N^-$} on this fragment ($Ni_2H_{-6}L$, $Ni_3H_{-9}L$ complexes respectively) is not cooperative.

The pH-metric and spectroscopic (UV-VIS, CD) results are deeply discussed in the chapter 8, while the data obtained from NMR studies are examined in the next paragraph.

7.3 NMR Study of Nickel Binding to C-terminal Sequence of Cap43 Protein

7.3.1 NMR Spectroscopy

NMR experiments were performed on a Bruker Avance 600 or 700 MHz spectrometer equipped with 5 mm TXI 1H-13 probe. Samples used for NMR experiments were 5 mM in concentration and dissolved in 90% H_2O/10% D_2O solutions. The acquisitions were performed at the temperature of 298 K. A series of 1D spectra of the free peptide was recorded at various pH values between 2.7 and 10.0 by step of 1.0. The titration experiments of Ni(II)-containing samples with metal-to-ligand molar ratios of 1:1 and 2:1 were performed at pH 9.0 and for samples with 3:1 molar ratio were performed at pH 10.0. The pH of the sample was adjusted to reach the final pH by addition of 1 N NaOH or 1 N HCl. Nuclear Overhauser Enhancement Spectroscopy (NOESY) with mixing times of 500 ms, Rotating Frame Overhauser Enhancement Spectroscopy (ROESY) with mixing times of 250 ms and Total Correlation Spectroscopy (TOCSY) with a mixing time of 50 ms were also performed. The combination of TOCSY, NOESY and ROESY experiments was used to assign the spectra of both free and Ni(II)-bound peptides at various pH. Solvent suppression for 1D, TOCSY, NOESY and ROESY experiments was achieved using WATERGATE pulse sequence or using excitation sculpting with gradients. All NMR data were processed using XWINNMR (Bruker Instruments) software on a Silicon Graphics Indigo workstation and analyzed using the Sparky 3.11 program.

7.3.2 NMR Characterization of C-terminal Sequence of Cap43

For the 30aa Cap43 free peptide 1D, 2D ^{1}H homonuclear TOCSY and NOESY spectra were performed at various pH. Table 7.1 presents the chemical shifts assignments at pH 2.7.

	Chemical Shift (ppm)													
Residue	HN	HA	QA	HB	HB1	HB2	QB	QG	HD1	HD2	QD	HE	HE1	HE2
T1	*8.047*	*4.246*		*4.155*				*1.131*						
R2	*8.396*	*4.316*			*1.816*	*1.700*		*1.579*			*3.130*	*7.097*		
S3	*8.255*	*4.355*					*3.792*							
R4	*8.393*	*4.316*			*1.817*	*1.697*		*1.578*			*3.130*	*7.097*		
S5	*8.215*	*4.342*					*3.760*							
H6	*8.566*	*4.738*			*3.249*	*3.134*			*7.382*	*7.237*			*8.532*	*6.991*
T7	*8.160*	*4.319*		*4.156*				*1.117*						
S8	*8.389*	*4.425*					*3.813*							
E9	*8.375*	*4.338*			*2.084*	*1.927*		*2.408*						
G10	*8.380*		*3.934*											
T11	*8.048*	*4.247*		*4.147*				*1.130*						
R12	*8.361*	*4.314*			*1.816*	*1.699*		*1.579*			*3.130*	*7.097*		
S13	*8.252*	*4.357*					*3.794*							
R14	*8.361*	*4.313*			*1.816*	*1.696*		*1.579*			*3.130*	*7.097*		
S15	*8.215*	*4.342*					*3.760*							
H16	*8.566*	*4.738*			*3.251*	*3.135*			*7.382*	*7.237*			*8.532*	*6.991*
T17	*8.155*	*4.320*		*4.157*				*1.118*						
S18	*8.373*	*4.423*					*3.815*							
E19	*8.373*	*4.336*			*2.083*	*1.927*		*2.409*						
G20	*8.378*		*3.934*											
T21	*8.049*	*4.248*		*4.143*				*1.130*						
R22	*8.359*	*4.314*			*1.815*	*1.699*		*1.579*			*3.130*	*7.097*		
S23	*8.254*	*4.355*					*3.793*							
R24	*8.363*	*4.314*			*1.817*	*1.679*		*1.579*			*3.130*	*7.097*		
S25	*8.215*	*4.339*					*3.760*							
H26	*8.566*	*4.738*			*3.250*	*3.134*			*7.382*	*7.237*			*8.532*	*6.991*
T27	*8.150*	*4.185*		*4.113*				*1.144*						
S28	*8.373*	*4.425*					*3.814*							
E29	*8.404*	*4.336*			*2.083*	*1.927*		*2.408*						
G30	*8.378*		*3.938*											

Table 7.1 ^{1}H NMR chemical shifts for Ac-*TRSRSHTSEG-TRSRSHTSEG-TRSRSHTSEG*-Am at pH 2.7. Degenerated protons are referred as Q.

2D NOESY spectra at various pH showed only short range (i, i+1) NOEs signals, as an indication of random coil secondary structure in agreement with the bioinformatic tools that were used to predict the structural conformation of the peptide using amino acid sequence information.

At pH 9.0 the resonances belonging to the 30 residues of the free peptide were assigned on the basis of 1D NMR spectra and 2D ^{1}H homonuclear TOCSY and NOESY experiments. In the region between 6.6 and 8.5 ppm, only the aromatic resonances of three histidine residues (HE1 and HD2, 7.607 ppm and 6.888 ppm respectively) were present, the three of them overlapping at the same chemical shift. All the amide resonances were in a fast exchange with water at this pH and their resonances were lost. In the aliphatic region, the HA of histidines appeared at 4.600 ppm under the water signal. Its assignment was based on the analysis of the TOCSY spectrum, where a correlation between histidinic HA and QB was clearly visible (QB 3.029 ppm). The TOCSY and NOESY spectra also allowed the assignment of the entire spin system of each amino acid. Table 7.2 presents the assignments at pH 9.0.

	Chemical Shift (ppm)									
Residue	**HA**	**QA**	**HB**	**HB1**	**HB2**	**QB**	**QG**	**HD2**	**QD**	**HE1**
T1	*4.228*		*4.156*				*1.121*			
R2	*4.307*			*1.794*	*1.687*		*1.557*		*3.111*	
S3	*4.357*					*3.758*				
R4	*4.307*			*1.795*	*1.687*		*1.557*		*3.111*	
S5	*4.361*					*3.755*				
H6	*4.600*					*3.026*		*7.603*		*6.887*
T7	*4.289*		*4.155*				*1.060*			
S8	*4.395*					*3.810*				
E9	*4.218*			*2.004*	*1.886*		*2.198*			
G10		*3.926*								
T11	*4.228*		*4.156*				*1.121*			
R12	*4.307*			*1.797*	*1.687*		*1.557*		*3.111*	
S13	*4.352*					*3.747*				
R14	*4.310*			*1.795*	*1.687*		*1.557*		*3.110*	
S15	*4.359*					*3.753*				
H16	*4.600*					*3.030*		*7.601*		*6.889*
T17	*4.289*		*4.155*				*1.057*			
S18	*4.396*					*3.812*				
E19	*4.218*			*2.006*	*1.886*		*2.197*			
G20		*3.926*								
T21	*4.228*		*4.156*				*1.121*			
R22	*4.307*			*1.795*	*1.687*		*1.557*		*3.111*	
S23	*4.349*					*3.750*				
R24	*4.310*			*1.795*	*1.687*		*1.557*		*3.110*	
S25	*4.359*					*3.758*				
H26	*4.600*					*3.033*		*7.601*		*6.884*
T27	*4.176*		*4.088*				*1.127*			
S28	*4.386*					*3.807*				
E29	*4.218*			*2.004*	*1.886*		*2.197*			
G30		*3.926*								

Table 7.2 ^{1}H NMR chemical shifts for Ac-*TRSRSHTSEG-TRSRSHTSEG-TRSRSHTSEG*-Am at pH 9.0. Degenerated protons are referred as Q.

7.3.3 NMR Characterization of Ni(II) Binding to C-terminal Sequence of Cap43

For the peptide-Ni(II) species, in order to test the coordination properties of individual mono-histidinic motif, the analysis was performed at different Ni(II)-to-ligand molar ratios (1:1, 2:1 and 3:1, respectively). The pH was chosen to approach maximum formation of the major planar diamagnetic species, as evidenced by potentiometric and spectroscopic measurements (pH=9 for 1:1 and 2:1 and pH=10 for 3:1) [30] (v. chapter 8).

The binding mode of Ni(II) to the three TRSRSHTSEG mono-histidinic motifs of Cap43 peptide, monitored by studing a series of 1D ^{1}H, 2D TOCSY and ROESY spectra at the different molar ratios (1:1, 1:2 and 1:3) for the peptide-metal system, is the same for the three molar ratios.

Addition of increasing amounts of Ni(II), at pH 9, causes the incremental loss in intensity of a number of resonances, commensurate with this is the appearance of a new set of peaks. These new resonances increase in intensity with increasing additions of Ni(II).

From the initial molar ratio 1:1 to the final 1:3, the two sets of the aromatic protons on the three histidines H_6, H_{16} and H_{26} showed exactly the same upfield shift with a $\Delta\delta$ = -0.172 ppm for HE1 and $\Delta\delta$= -0.015ppm for HD2.

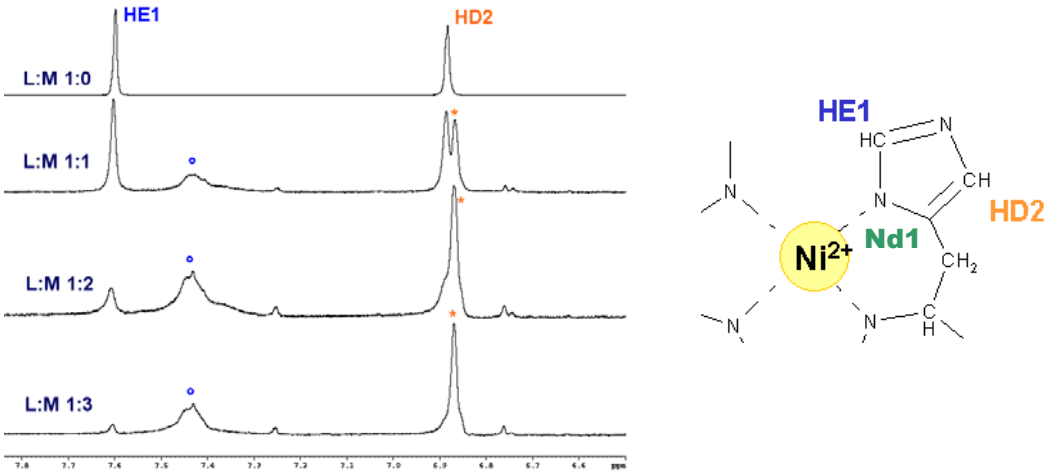


Fig. 7.4 Changes in intensity and chemical shift of the histidine aromatic protons (HE1 and HD2) by increasing nickel concentration. Comparison of the aromatic region of 1D ^{1}H NMR spectra of 30aa Cap43 peptide-Ni(II) in the molar ratios of 1:0, 1:1. 1:2, 1:3, respectively.

The new peaks increase in intensities with increasing of nickel concentration (fig. 7.4), suggesting the progressive involvement of the aromatic protons HE1 and HD2 for all three distinct histidine residues of the Ac-*TRSRSH$_6$TSEG-TRSRSH$_{16}$TSEG-TRSRSH$_{26}$TSEG*-Am peptide. The strong shift experienced by the HE1 protons respect to HD2 is due to their higher proximity to the metal ion.

HD2 is in the opposite side of the ring respect to the imidazolic nitrogen Nd1 involved in the complexation (fig. 7.4 b).

Strong shifts affect also the histidine HA and HB protons (HA 3.451 ppm, $\Delta\delta$ = -1.149; QB 2.883 ppm, $\Delta\delta$= -0.146). In the free peptide, the aliphatic protons of arginine R_2, R_4, R_{12}, R_{14}, R_{22} and R_{24}, experience equivalent chemical environments resulting in overlap in their resonances. On the contrary, the addition of Ni(II) causes a strong differentiation on their signals. In particular we can assign HA 4,024 ppm, HB_1 2,209 ppm, HB_2 2,047 ppm, QG 1,918 ppm, QD 3,290 ppm for R_4, R_{14}, R_{24} and HA 4,147 ppm, $HB_1$1,719 ppm, HB_2 1,619 ppm, QG 1,524 ppm, QD 3,097 ppm for R_2, R_{12}, R_{22} (fig. 7.5).

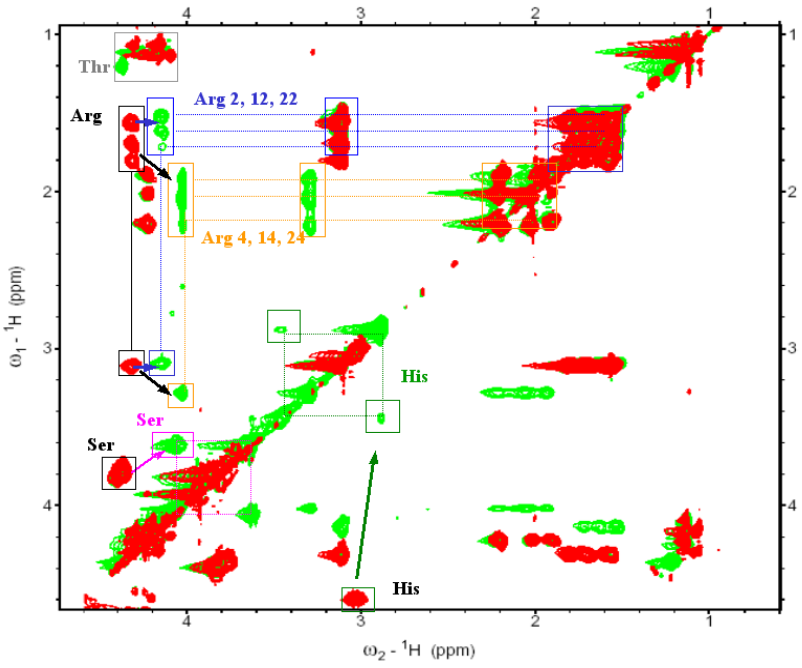


Fig. 7.5 Superimposition of 2D ^{1}H TOCSY spectra of free 30aa peptide (red) and 30aa peptide-Ni(II) complex (green) at 1:3 molar ratio.

This differentiation is due to the fact that the R_4, R_{14} and R_{24} residues are directly involved in the complex environment by the nitrogen amide H^N, and the metal-binding causes stronger shifts in their proton resonances than for the R_2, R_{12}, R_{22} residues, which are not directly involved in the coordination and remain, therefore, less perturbed. Consequently, the two different chemical environments for the arginine residues result in two new distinct spin systems that appear in the TOCSY and ROESY spectra.

In the same way, due to their direct involvement in the metal interaction, the proton resonances of serine residues S_5, S_{15}, S_{25} respect to S_3, S_8, S_{13}, S_{18}, S_{23} and S_{28}, appear at upfield chemical shift.

Threonine residues are located in position 1 and 7 in each repeated 10-amino acid fragment. From 1D and 2D spectra, small perturbations in the proton resonances after addition of nickel are clearly visible. Threonines in position 7, are closer to the coordination centre, respect to those in position 1. For this reason they are more sensitive to the chemical environment changes due to the metal. This fact permits to fix the correct assignments. Notably, spin systems that are unaffected by Ni(II) addition include glutammate E_{9+i} and glycine G_{10+i} (where i = 0, 10, 20) for every repeated fragment.

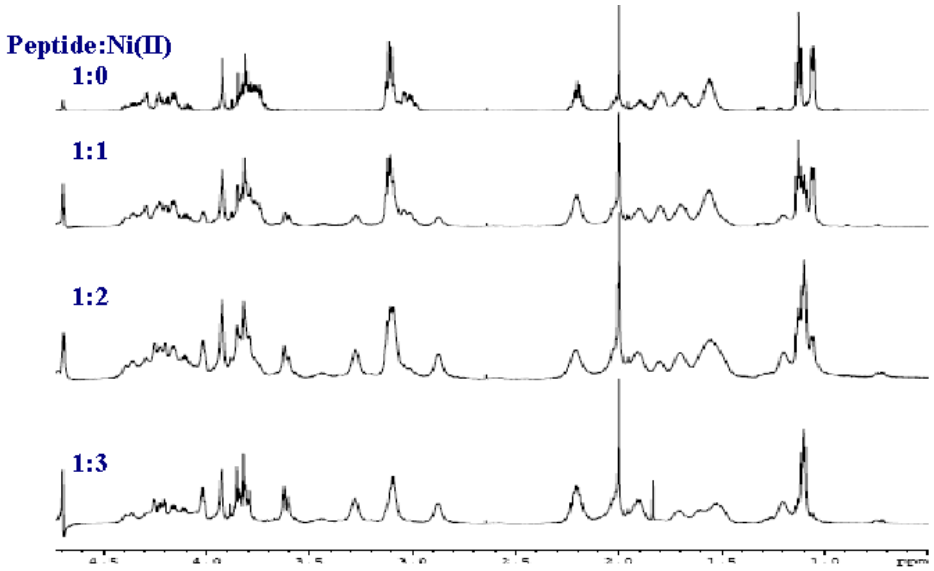


Fig. **7.6** Comparison of aliphatic region of 1D ^{1}H NMR spectra of 30aa Cap43 peptide -Ni(II) in the molar ratios 1:0, 1:1, 1:2, 1:3.

residue	free peptide	peptide-Ni(II) complex 1-3	chemical shift diff
R_2 HA	4.313	4.147	*-0.166*
R_2 HB1	1.803	1.719	*-0.084*
R_2 HB2	1.694	1.619	*-0.075*
R_2 QG	1.561	1.524	*-0.037*
R_2 QD	3.111	3.097	*-0.014*
R_4 HA	4.313	4.024	*-0.289*
R_4 HB1	1.803	2.209	*0.406*
R_4 HB2	1.694	2.047	*0.353*
R_4 QG	1.561	1.918	*0.357*
R_4 QD	3.111	3.290	*0.179*
S_5 HA	4.380	4.066	*-0.314*
S_5 QB	3.793	3.628	*-0.165*
H_6 HA	4.600	3.451	*-1.149*
H_6 QB	3.029	2.883	*-0.146*
H_6 HD2	6.888	6.873	*-0.015*
H_6 HE1	7.607	7.435	*-0.172*
T_7 HB	4.292	4.168	*-0.124*
T_7QG	1.125	1.206	*0.081*

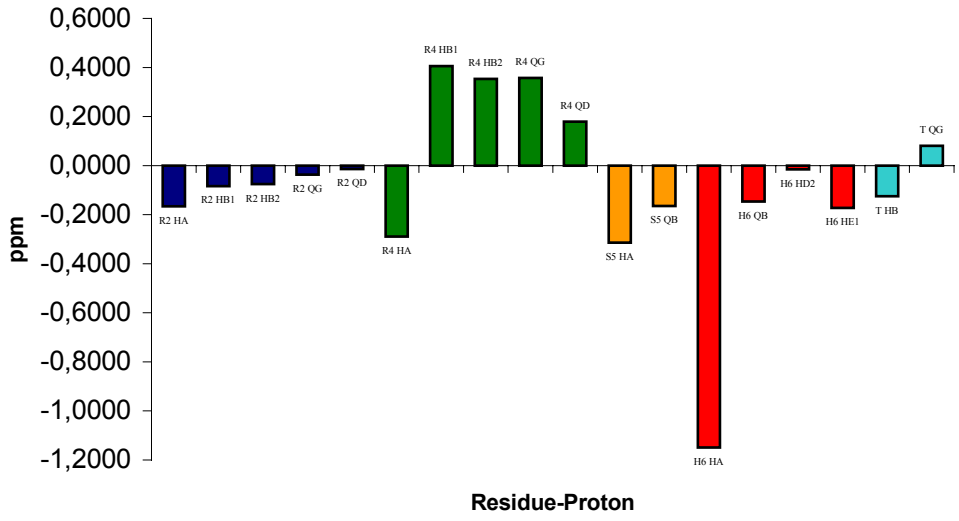


Fig. **7.7** Chemical shifts assignment, before and after nickel interaction, for the residues in the mono-histidine fragment -$T_1R_2S_3R_4S_5H_6T_7S_8E_9G_{10}$– involved in the complex formation. The chemical shift differences are also reported in the plot. The behavior appears to be the same for each of the three repeated region Ac-TRSRSHTSEG-TRSRSHTSEG-TRSRSHTSEG-Am under investigation.

7.4 Conclusions

It was not possible to obtain sequential main chain H^N ROE connectivities at pH 9, which is the pH required for Ni(II) binding, due to rapid exchange of the labile protons with the bulk water. In the absence of comprehensive ROESY data, large chemical shift changes upon Ni(II) coordination ($\Delta\delta$>0.3 ppm) can indicate the residues directly involved in coordination.

From NMR experiments, the shifts induced by Ni(II) were consistent with the binding of the metal ion in a square-planar site formed by four nitrogen atoms from H_{6+i} (Nd1,N_H) and from N_H of S_{5+i} and R_{4+i} (where i = 0, 10, 20), of each 10-aminoacid fragment. Strong shifts in the aliphatic proton resonances of arginine suggest an involvement of the side-chain in the complex formation and/or a conformational rearrangement.

In conclusion, NMR studies performed on the 30-amino acid peptide, support the existence of relatively effective metal binding site in the C-terminal region of Cap43 protein.

These results suggest that the entire Cap43 protein could be an interesting target for Ni(II) ions.

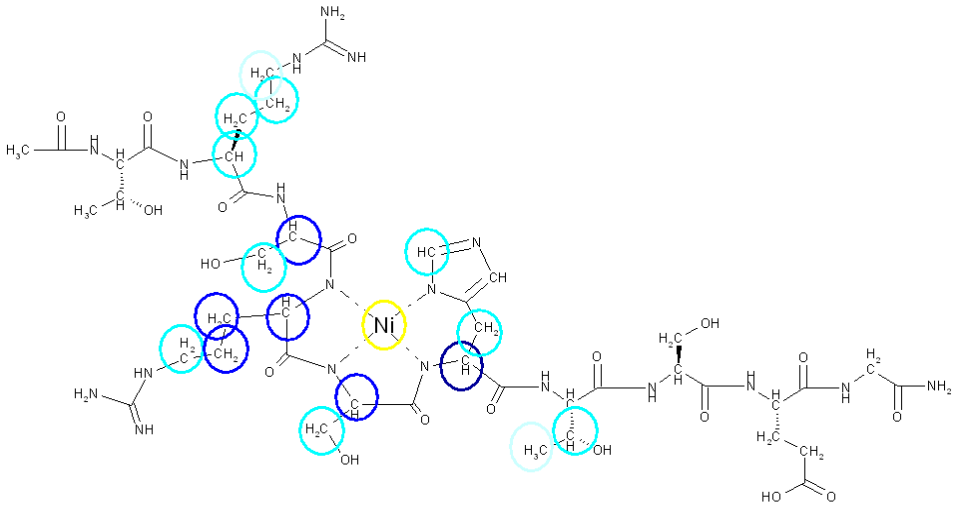


Fig. 7.8 Scheme of 4N coordination pattern of TRSRSHTSEG Cap43 fragment-Ni(II) complex. The circles indicate the most affected protons after nickel interaction. The hardness of colors indicate the strongness of changes.

LIST OF REFERENCES FOR CHAPTERS 7

1. Costa. M.. (1991) *Annu. Rev. Pharmacol.Toxicol*., 31, 321.
2. Biedermann. K. A., and Landolph. J. R. (1987) *Cancer Res*., 47,3815.
3. Oller. A. R., Costa. M., and Oberdorster. G. (1997) *Toxicol. Appl. Pharmacol*., 143: 152.
4. Knight. J. A., Plowman. M. R., Hopfer. S. M., and Sunderman. F. W. (1991) *Ann. Clin. Lab. Sci*., 21, 275.
5. Tveito. G., Hansteen. I. L., Dalen. H., and Haugen. A. (1989) *Cancer Res*., 49*,* 1829,.
6. Miura. T., Patierno. S. R., Sakuramoto. T., and Landolph. J. R. (1989) *Environ. Mol. Mutagen.*, 14*:* 65.
7. International Agency for Research on Cancer (IARC). (1990) IARC, Vol. 49, pp. 677. Lyon, France.
8. Denkhaus. E., and Salnikow. K. (2002) *Crit. Rev. Oncol. Hemat*., 42, 35.
9. Shen H.M., Zhang Q.F., (1994) *Environ. Health Perspect*. 102 (Suppl. 1) 275.
10. Biswas. P., and Wu. C. Y. J. (1998) *Air Waste Manag. Assoc*., 48, 113.
11. Biggart. N. W., and Costa. M. (1986) *Mutat. Res.,* 175, 209.
12. Mayer. C., Klein. R. G., Wesch. H., and Schmezer. P. (1998) *Mutat. Res*., 420, 85.
13. Kasprzak. K. S. Chem. Res. Toxicol., *4:* 604–615, 1991.
14. Hartwig. A., Mullenders. L. H. F., Schlepegrell. R., Kasten. U., and Beyersmann. D. (1994) *Cancer Res*., 54, 4045.
15. Klein. C. B., Frenkel. K., and Costa. M. (1991) *Chem. Res. Toxicol.,* 4, 592.
16. Zhou D, Salnikow K, Costa M. (1998) *Cancer Res*, 58,2182.
17. Salnikow K, Blagosklonny MV, Ryan H, Johnson R, Costa M. (2000) *Cancer Res*, 60,38.
18. Piquemal D, Joulia D, Balaguer P, Basset A, Marti J, Commes T (1999) *Biochim Biophys Acta*, 1450,364.
19. Kokame K, Kato H, Miyata T. (1996) *J Biol Chem*, 271,29659.
20. Pubmed OMIM Database, Gene #: 605262, Gene Function [http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=605262]
21. Salnikow K., Zhou D., Kluz T., Wang C., Costa M., in: A. Sarkar (Ed.), *Metal and Genetics*, Kluwer Academic, Plenum Publishers, New York, 1999, pp. 131–144.
22. Salnikow K., Kluz T., Costa M., (1999) *Toxicol. Appl. Pharmacol*. 160, 127.
23. Guan. R. J., Ford. H. L., Fu. Y., Li. Y., Shaw. L. M., and Pardee. A. B. (2000) *Cancer Res*., 60, 749.

24. Sigel H., in: *Metal Ions in Biological Systems*, vol. 25, Marcel Dekker, Inc., New York, 1989.

25. Agrowal A., Bhattacharya S., (1989) *Experientia* 45, 567.

26. Suerbaum S., Thiberge J.M., Kansau I., Ferrero R.L., Labigne A., (1994) *Mol. Microbiol.* 14, 959.

27. Viles J.H., Cohen F.E., Prusiner S.B., Goodin D.B., Wright P.E., Dyson H.J., (1999) *Proc. Natl. Acad. Sci*. USA 96, 2042.

28. Zoroddu M.A., Kowalik-Jankowska T., Kozłowski H., Salnikow K., Costa M., (2001) *J. Inorg. Biochem*. 85, 47.

CHAPTER 8

NICKEL(II) BINDING TO CAP43 PROTEIN FRAGMENTS

Maria Antonietta Zoroddu [a], **Massimiliano Peana** [a], Teresa Kowalik-Jankowska [b], Henryk Kozłowski [b], Max Costa [c]

[a] Department of Chemistry and Pharmacy Faculty, University of Sassari, Via Vienna 2, 07100 Sassari, Italy [b] Faculty of Chemistry, University of Wroclaw, Poland [c] Department of Environmental Medicine, N.Y. University, USA

Journal of Inorganic Biochemistry 98 (2004) 931–939

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Inorganic Biochemistry 98 (2004) 931–939
JOURNAL OF Inorganic Biochemistry
www.elsevier.com/locate/jinorgbio

Nickel(II) binding to Cap43 protein fragments

M.A. Zoroddu [a,*], M. Peana [a], T. Kowalik-Jankowska [b], H. Kozlowski [b], M. Costa [c]

[a] *Department of Chemistry and Pharmacy Faculty, University of Sassari, Via Vienna 2, 07100 Sassari, Italy*
[b] *Faculty of Chemistry, University of Wroclaw, Poland*
[c] *Department of Environmental Medicine, N.Y. University, USA*

Received 27 November 2003; received in revised form 10 March 2004; accepted 18 March 2004
Available online 6 May 2004

Abstract

Cap43 protein has been tested for metal binding domains. The protein, specifically induced by nickel compounds in cultured human cells, had a new mono-histidinic motif consisting of 10 amino acids repeated three times in the C-terminus.

The 20-Ac-TRSRSHTSEG–TRSRSHTSEG (Thr^{341}–Arg–Ser–Arg–Ser–His^{346}–Thr–Ser–Glu–Gly–Thr–Arg–Ser–Arg–Ser–His^{356}–Thr–Ser–Glu–Gly^{360} – peptide 1) and the 30–Ac-TRSRSHTSEG–TRSRSHTSEG–TRSRSHTSEG (Thr^{341}–Arg–Ser–Arg–Ser–His^{346}–Thr–Ser–Glu–Gly–Thr–Arg–Ser–Arg–Ser–His^{356}–Thr–Ser–Glu–Gly–Thr–Arg–Ser–Arg–Ser–His^{366}–Thr–Ser–Glu–Gly^{370} – peptide 2) amino acids sequence has been analyzed as a site for Ni(II) binding.

A combined pH-metric and spectroscopic (UV–visible, CD, NMR) studies of Ni(II) binding to both fragments were performed. The 20-amino acid peptide can bind one and two metal ions while the 30-amino acid fragment one, two and three metal ions. At physiological pH, depending on the metal to ligand molar ratio, peptide 1 forms the Ni_2L species while peptide 2 the NiL, Ni_2L and Ni_3L complexes where each metal ion is coordinated to the imidazole nitrogen atom of the histidine residue of the 10-amino acid fragment. Octahedral complexes at pH 8–9 and planar 4N complexes with (N_{Im}, $3N^-$) bonding mode at pH above 9, are formed.

This work supports the existence of an interesting binding site at the COOH-terminal domain of the Cap43 protein.

Keywords: Stability constants; Spectroscopic study; Nickel(II) complexes; Cap43 fragments

1. Introduction

Nickel is an essential element for many archea, bacteria and plants and may yet be found to play a role in the metabolism of higher organisms [1–3]. Although a deficiency disease for nickel in humans has not been identified, there is substantial evidence for the essential status of nickel [4]. However, under some circumstances nickel compounds are potent human carcinogens as reported from International Agency for Research on Cancer [5]. Humans are exposed to carcinogenic nickel compounds both occupationally and environmentally. Animals studies indicate that substantial variation in the carcinogenic potential of nickel compounds occurs, causing multiple types of cellular/nuclear damage. The majority of cellular damage may ultimately result in altered gene expression, rather than in direct DNA damage [6–9]. Ni(II) crosses cell membranes via calcium channel and probably competes with Ca^{2+} for specific intracellular receptors [10–12]. Ni(II) has been shown to block Ca^{2+} channels and release stores of free intracellular Ca^{2+} in a variety of systems, and some studies have reported that longer exposure to nickel compounds can elevate intracellular Ca^{2+}.

Cap43 is a novel gene induced by a rise in free intracellular Ca^{2+} following nickel exposure [13,14]. No other metal compounds significantly induced expression of this gene, indicating that it was expressed with marked specificity to Ni(II) exposure [15].

Cap43 expressed a 3.0-kb mRNA encoding a M_r 43,000 protein. The protein was not found in the nucleus, but was localized in the perinuclear region and cytoplasm. Although Ni(II) is known to induce heat shock proteins (HSPs), metallothioneins (MTs) and acute phase reactant proteins (APRs), none of these

* Corresponding author. Tel.: +39-079229529; fax: +39-079271305.
E-mail address: zoroddu@uniss.it (M.A. Zoroddu).

0162-0134/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jinorgbio.2004.03.005

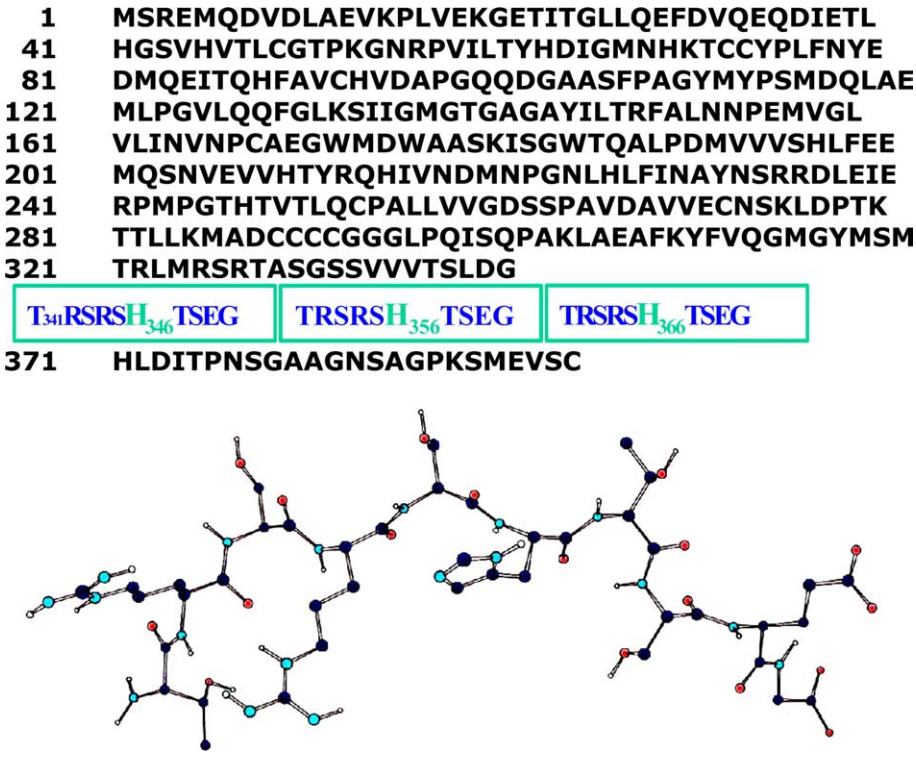


Fig. 1. Ten amino acids (Thr–Arg–Ser–Arg–Ser–His–Thr–Ser–Glu–Gly) sequence.

genes were specific to nickel and other heavy metals such as cadmium, mercury, lead and zinc also induced these genes. It has been suggested that the function of these three families of genes is in the detoxification and protection against oxidative stress induced by metals. Most of them have been proven to contain metal-binding domains. For example, MTs have high affinity for metals due to their high cysteine content [16]; purified Limulus C-reactive protein, a member of the APR family, has sulphydryl groups that bind strongly to mercury [17] and HSPA from *Helicobacter pylori*, features a series of cysteine and histidine residues, resembling anchoring binding site for metal ions [18]. Although, Cap43 has no cysteine or histidine-rich motifs for metal binding, it has a mono-histidine fragment, 10 amino acids (Thr–Arg–Ser–Arg–Ser–His–Thr–Ser–Glu–Gly) repeated three times, resembling a nickel binding motif at the COOH terminus (Fig. 1). It should be mentioned that such mono-histidine fragments, e.g. octapeptide repeated regions in prion proteins, play a critical role in metal metabolism using set of His imidazoles as the binding sites for metal ions [19]. The finding of the Cap43 protein makes it an interesting candidate for studies of molecular mechanisms of nickel carcinogesis.

The present paper reports the results of combined spectroscopic and potentiometric studies on the nickel(II) complexes of the 20-(TRSRSHTSEG–TRSRSHTSEG, peptide 1) and 30-(TRSRSHTSEG–TRSRSHTSEG–TRSRSHTSEG, peptide 2) amino acid sequences of C-terminal part of the Cap43 protein. The imidazole nitrogen atom of histidine residue is essential bonding site for nickel(II) ion, and the 20- and 30 amino acid sequences contain two and three histidine residues, respectively. Therefore, this study was also performed in order to examine the bonding abilities of these fragments to more than one metal ion.

2. Experimental

2.1. Peptide synthesis and purification

α-cyano-4-hydroxycinnamic acid was purchased from Sigma (Sigma Chemical Co., St. Louis, MO, USA). High performance liquid chromatography (HPLC)-grade trifluoroacetic acid (TFA) was purchased from Fluka (Buchs, Switzerland), HPLC-grade H_2O from Romil Ltd. (Amman Technik SA, Kölliken, Switzerland) and acetonitrile was purchased from Biosolve Ltd. (Chemie Brunschwig, Basel). All other chemicals were of highest purity and were used without further purification. All amino acid derivatives were obtained from Novabiochem (Läufelfingen, Switzerland) and other reagents from Fluka and Sigma Chemie (Buchs, Switzerland).

The peptides were chemically synthesized using solid phase Fmoc chemistry in an Applied Biosystems

Synthesizer [20]. The synthesis was performed using a fivefold excess of Fmoc amino acid derivatives, DCCI and HOBt as activating agents and a 60 min coupling time. Side chain protecting groups included: triphenylmethyl group for His; pentamethyl-chroman-sulfonyl group for Arg; *t*-butyl group for Glu, Ser and Thr. The peptides were deprotected and cleaved from the resin by treatment with 2.5% H_2O, 5% triethylsilan in TFA for 2 h at room temperature. After removal of the resin by filtration, the peptides were precipitated with *tert*-butyl-methyl-ether, centrifuged and the pellets resuspended in 50% acetic acid and lyophilized.

Crude peptides were reconstituted in 1 ml 50% acetic acid in H_2O and low molecular weight contaminants were removed by gel filtration on Sephadex G-25. The materials eluted in the void volume were lyophilised, reconstituted in 1 ml 50% acetic acid in H_2O and subjected to RP-HPLC on a Vydac column (250×22 mm, 10–15 μm). The column was eluted at a flow rate of 9 ml/min by a linear gradient of 0.1%TFA/acetonitrile on 0.1%TFA/H_2O, rising within 60 min from 10% to 100%. The optical density of the eluate was monitored at 220 or 280 nm. Fractions were collected and analysed by MALDI-TOF MS. Fractions containing the peptide of the expected molecular weight were pooled and lyophilised.

2.2. Potentiometric measurements

Stability constants for protons and Ni(II) complexes were calculated from pH-metric titrations carried out in argon atmosphere at 298 K using a total volume of 2 cm^3. Alkali (NaOH, 0.1 M) was added from a 0.250 cm^3 micrometer syringe which was calibrated by both weight titration and the titration of standard materials. The concentration of peptide 1 was 2×10^{-3} M and the ligand to metal molar ratio was 1.1:1 and 1.1:2. For peptide 2, the concentration was 1.3×10^{-3} M and the ligand to metal molar ratio was 1.3:1, 1.3:2 and 1.3:3. The pH-metric titrations were performed in ionic strength 0.10 M (KNO_3) on a MOLSPIN pH-metric system using a Russel CMAW 711 semi-micro combined electrode calibrated in hydrogen ion concentrations using HNO_3 [21]. The SUPERQUAD [22] and HYPERQUAD [23] computer programs were used for stability costant calculations. Standard deviations refer to random errors only. They are, however, a good indication of the importance of a particular species in the equilibrium. The sample were titrated in the pH region 2.5–10.5. In whole pH range for the solutions containing Ni(II) ions and the peptide 1 (2:1, metal to ligand molar ratio) or peptide 2 (2:1 and 3:1, metal to ligand molar ratio), the precipitations were not observed.

2.3. Spectroscopic measurements

Solutions for CD and UV–visible measurements were of similar concentrations to those used in the potentiometric studies. Absorption spectra were recorded on a Beckman DU 650 spectrophotometer. Circular dichroism (CD) spectra were recorded on a Jasco J-715 spectropolarimeter in the 750–250 nm range. The values of $\Delta\varepsilon$ (i.e., $\varepsilon_l - \varepsilon_r$) and ε were calculated at the maximum concentration of the particular species obtained from the potentiometric data. NMR experiments were performed on a Varian 300-VXR spectrometer. One-dimensional experiments were carried out in D_2O solution at a peptide concentration of 1.5×10^{-2} M and peptide: Ni(II) ratio of 1.1:1. Above pH 8 there was precipitation reaction under the condition used for NMR experiment.

3. Results and discussion

3.1. Protonation constants

The potentiometrically measured protonation costants for the ligands studied and for comparable peptides [24,25] are shown in Table 1 along with the calculated stepwise constants assigned to the respective peptidic functions. The 20-amino acid sequence (peptide 1) can be considered as a H_5L ligand; the deprotonation involves two histidine residues ($pK_a = 6.70$ and 5.89) [24,25] and two side chains of γ-carboxylic group of the Glu residues and a C-terminal COOH ($pK_a = 4.71$, 4.20, 3.56) [26].

The 30-amino acid sequence (peptide 2) is a H_7L ligand where the deprotonation involves three histidine residues ($pK_a = 6.88$, 6.26 and 5.71), three side chains of γ-carboxylic groups of the Glu residues and a C-terminal COOH ($pK_a = 4.76$, 4.34, 3.91, 3.39). As we can see from the values of the pK_a reported in Table 1, the first imidazole nitrogens of the histidine residues for both peptides investigated, peptides 1 and 2, are about 0.5 order of magnitude more acidic than the histidine in the monohistidinic 14-amino acid sequence of Cap43 C-terminal [24].

3.2. Ni(II) complexes containing one metal ion

Potentiometry detects a range of Ni(II) complexes in the pH range 2.5–10.5 with the formation costants reported in Tables 2 and 3. The values of $\log K^*$ ($\log K^* = \log \beta(NiH_jL) - \log \beta H_nL$), the protonation corrected stability constants which are useful to compare the ability of various ligands to bind a metal ion, are reported in Table 2 for 1:1 complexes [27]. Spectroscopic properties of the major complexes identified are given in Table 5.

Table 1
Protonation constants for Ac-TRSRSHTSEG–TRSRSHTSEG (peptide 1) and Ac-TRSRSHTSEG–TRSRSHTSEG–TRSRSHTSEG (peptide 2) and comparable peptides at 298 K, $I = 0.10$ M (KNO_3)

Peptide	log β/species						
	HL	H_2L	H_3L	H_4L	H_5L	H_6L	H_7L
Peptide 1	6.70 ± 0.01	12.59 ± 0.01	17.30 ± 0.01	21.50 ± 0.01	25.06 ± 0.01		
Peptide 2	6.88 ± 0.01	13.14 ± 0.01	18.85 ± 0.01	23.61 ± 0.01	27.61 ± 0.01	31.86 ± 0.01	35.25 ± 0.01
Cap43 (14 a.a. sequence)[a]	6.35 ± 0.01	10.36 ± 0.01					
Sequence 1*,[b]	11.79	22.54	32.62	42.65	51.85	57.84	
Stepwise protonation constants	log K						
	N_{Im}	N_{Im}	N_{Im}	COO^-	COO^-	COO^-	COO^-
Peptide1	6.70	5.89		4.71	4.20	3.56	
Peptide 2	6.88	6.26	5.71	4.76	4.34	3.91	3.39
Cap43 (14 a.a. sequence)[a]	6.35		4.01				
Sequence 1*,[b]	5.99						

[a] Ref. [24].
[b] Ref. [25].
* Sequence 1, Ac-SGRGKGGKGLGKGGAKRHRKVL-Am.

The histidine residues of both Cap43 C-terminal fragments can be the primary sites for the interaction of Ni(II) ions [24,25,28]. Ni(II) forms three monomeric species with peptide 1 and five monomeric species with peptide 2, respectively (Figs. 2 and 3). $NiH_{-1}L$ species ($\log K^* = -11.41$ and -10.89 for the peptides 1 and 2, respectively, Table 2) starts to form at pH about 7. It involves the coordination of imidazole nitrogen and a deprotonated amide nitrogen atom of a histidine residue. The stability of these complexes is comparable to each other and is one order of magnitude higher than that of the 22-amino acid (Sequence 1, Table 2). This stabilization may result from a different amino acid sequence and different structural organization of the peptides in Ni(II) complexes [29]. At pH 7.5 starts to form $NiH_{-3}L$ species ($\log K^* = -27.95$ and -27.14 for peptides 1 and 2, respectively, Table 2). It comes from the deprotonation of the other two amides from the backbone to give a 4N chromophore (N_{Im}, $3N^-$) as in the scheme in Fig. 4. The absorption and CD spectra of these complexes are typical for 4N planar complexes (Table 5) [24,25,28]. Above pH 9 $NiH_{-4}L$ species starts to form (Figs. 2 and 3); it likely comes from the deprotonation of a guanidine residue or the pyrrolic nitrogen atom of histidine residue ($pK_a = 11.54$ and 10.19 for peptides 1 and 2, respectively) [24,25].

NiHL and NiL species, were also found, from potentiometric data calculations, for peptide 2. The NiHL species, the maximum formation of which is obtained at pH 6.0–6.5 (Fig. 3) is the 1N $\{N_{Im}\}$ complex with one protonated imidazole nitrogen atom of the second histidine residue. The deprotonation constant of the NiHL ($NiHL \leftrightarrows NiL + H^+$) is -6.87; this value fits very well the deprotonation constant of imidazole nitrogen atom of histidine residue (-6.88, Tables 1 and 2). Therefore, it may suggest that the NiHL and NiL species are the

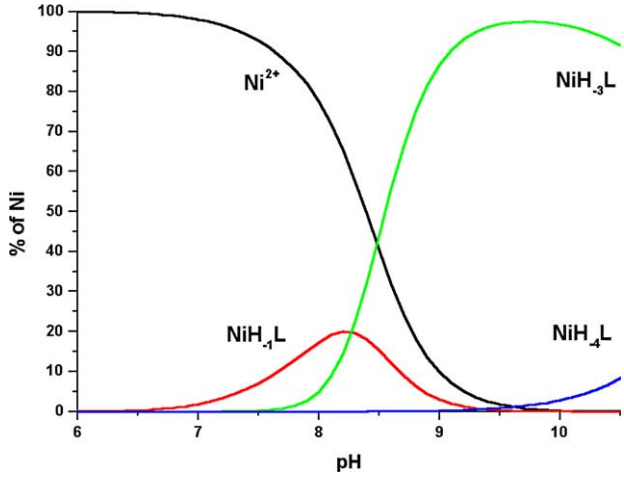


Fig. 2. Species distribution curves for Ni(II) – peptide 1, molar ratio 1:1.

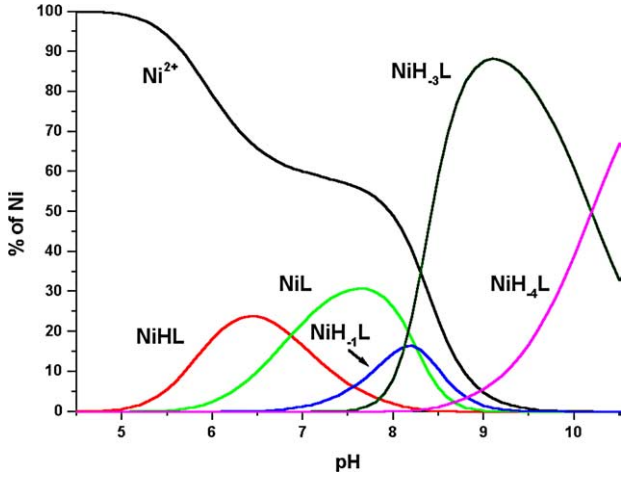


Fig. 3. Species distribution curves for Ni(II) – peptide 2, molar ratio 1:1.

Table 2
Stability constants of Ni(II) complexes containing one metal ion with Ac-TRSRSHTSEG–TRSRSHTSEG (peptide 1) and Ac-TRSRSHTSEG–TRSRSHTSEG–TRSRSHTSEG (peptide 2) and comparable peptides at 298 K, $I = 0.10$ M (KNO_3)

Peptide	log β/species								
	NiH_4L	NiH_3L	NiH_2L	NiHL	NiL	$NiH_{-1}L$	$NiH_{-2}L$	$NiH_{-3}L$	$NiH_{-4}L$
Peptide1						-5.52 ± 0.05		-22.06 ± 0.02	-33.60 ± 0.16
Peptide 2				9.94 ± 0.04	3.07 ± 0.01	-5.18 ± 0.06		-21.43 ± 0.02	-31.62 ± 0.04
Cap43 (14 a.a. sequence)[a]					2.73		-13.47	-21.81	-33.30
Sequence 1[b]	45.50	37.15	29.17	19.98	10.00	-0.16	-10.78	-21.98	
	log K^*								
	$1N\{N_{Im}\}$	$2N\{N_{Im}, N^-\}$	$3N\{N_{Im}, 2N^-\}$	$4N\{N_{Im}, 3N^-\}$					
Peptide 1		-11.41		-27.95					
Peptide 2	-2.64	-10.89		-27.14					
Cap43 (14 a.a. sequence)[a]	-3.62		-19.82	-28.16					
Sequence 1*,[b]		-12.34	-20.69	-28.67					

[a] Ref. [24].
[b] Ref. [25].
* Sequence 1, Ac-SGRGKGGKGLGKGGAKRHRKVL-Am.

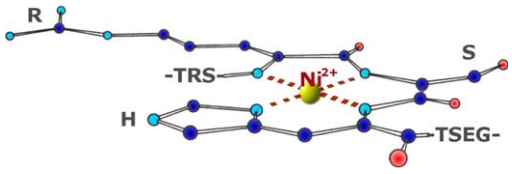


Fig. 4. Scheme 4N.

complexes with the same 1N $\{N_{Im}\}$ bonding mode. The absorption and CD spectra with changing the pH clearly support the presence in solution of the octahedral complexes till pH 8, for both peptides. In UV–visible spectra one d–d band at 390 nm (with ε 20–30 $M^{-1}\,cm^{-1}$) is observed (mM concentration of nickel(II)).

Variable pH NMR study in D_2O at 300 MHz from pH 3 to pH 8 was used to evaluate the pK_a of the glutammate carboxylic group from the upfield of the γ-CH_2 Glu, and the pK_a of the histidine residue from the shifts of the proton resonances of H(2) and H(4) for peptides 1 and 2. The resonances belonging to the 20-amino acid peptide were assigned on the basis of the 1D NMR spectra. The following assignment at pH 3 have been made: δ 8.58 (CH-2 of His); 7.27 (CH-4 of His); 3.94 (α-Glu); 3.82 (β-Ser); 3.2 (β-His); 3.15 (δ-Arg); 2.46 (γ-Glu); 2.03 (Thr); 1.8 (β-Arg); 1.6 (γ-Arg); 1.14 (γ-Thr). An overlap among the histidine α-proton appears as usually downfield of the other α-proton signals and it is partly obscured by the HDO signal. The amide protons appear to exchange rapidly with the solvent protons and are not observed in the NMR spectra. Interestingly, the multiplet centered at 1.14 ppm observed at pH 3 and assigned to γ-CH3–Thr, splits at pH 8 into two doublets centered at 1.2 and 1.1, showing that the threonine residues are not equivalent each other when the pH is raised to 8. The following assignments at pH 8 have been made: δ 7.77 (CH-2 of His); 6.9 (CH-4 of His) for peptide 1. Comparison of the 1D NMR spectra of TRSRSHTSEG–TRSRSHTSEG and Ni(II) species were performed a pH 7.2. Proton NMR spectra show significant chemical shifts changes, as a result of complexation, although the broadening of the lines made the interpretation of NMR spectra difficult. Nevertheless, some information on the binding mode of Ni(II) can be obtained. A minor shift of the two histidine aromatic protons δ at 8.43 (CH-2 of His) and δ at 7.23 and 7.199 (CH-4 of His) together with a visible changes involving the α-proton region of the spectrum were observed. Unfortunately, above pH 8 was not possible to have information from NMR spectra because there was precipitation reaction in the conditions used.

3.3. Ni(II) complexes containing more than one metal ion

The coordination study has been carried out also for molar ratio Ni(II) to peptides 2:1 (for peptides 1 and 2)

Table 3
Stability constants for Ni(II) complexes containing two or three metal ions with Ac-TRSRSHTSEG–TRSRSHTSEG-Am (peptide 1) and Ac-TRSRSHTSEG–TRSRSHTSEG–TRSRSHTSEG-Am (peptide 2) at 298 K, $I = 0.10$ M (KNO_3)

Peptide	log β/species					
	Ni_2L	$Ni_2H_{-1}L$	$Ni_2H_{-2}L$	$Ni_2H_{-4}L$	$Ni_2H_{-6}L$	$Ni_2H_{-7}L$
Peptide 1	6.11 ± 0.03	-1.95 ± 0.07	-10.27 ± 0.08	-26.70 ± 0.03	-44.31 ± 0.02	
Peptide 2	6.03 ± 0.07		-9.93 ± 0.07	-26.79 ± 0.09	-43.68 ± 0.03	-54.19 ± 0.05
	Ni_3L	$Ni_3H_{-1}L$	$Ni_3H_{-3}L$	$Ni_3H_{-6}L$	$Ni_3H_{-8}L$	$Ni_3H_{-9}L$
Peptide 2	9.51 ± 0.05	2.30 ± 0.03	-14.14 ± 0.04	-39.28 ± 0.04	-57.03 ± 0.04	-66.42 ± 0.04

and 3:1 (for peptide 2) in order to study the possible involvement of the set of histidines (His^6, His^{16} and His^{26}) in the coordination to the metal ion.

Five metal complex species can be fitted to the experimental titration curves obtained for Ni(II) – peptide 1 or peptide 2 systems with 2:1 metal to ligand molar ratio (Table 3). In Figs. 5 and 6 are reported the speciation curves for these complexes in the solution. The first species obtained for the peptides 1 and 2, the Ni_2L, starts to form at pH 5–5.5 and it reaches the maximum formation at pH 7–7.5. The stoichiometry of this complex suggests the $2 \times 1N$ $\{N_{Im}\}\{N_{Im}\}$ coordination mode (Table 4). The $\log K^*$ value for the Ni_2L complex is about 0.7 and 1.3 orders of magnitude higher (for peptides 1 and 2, respectively) in comparison to that (−7.24) obtained for the 14-amino acid sequence with the same bonding mode (Table 4, one metal ion coordinated to one histidine residue, −3.62). By raising the pH, consecutive deprotonations take place to give the $Ni_2H_{-1}L$ and $Ni_2H_{-2}L$ complexes. Above pH 8 the $Ni_2H_{-4}L$ species with maximum formation at pH 8.7, and the $Ni_2H_{-6}L$ with maximum formation above pH 9 can also be fitted to the experimental titration curves. The stoichiometry of the $Ni_2H_{-6}L$ complex is in agreement with the 4N $\{N_{Im}, 3N^-\}$ coordination of two

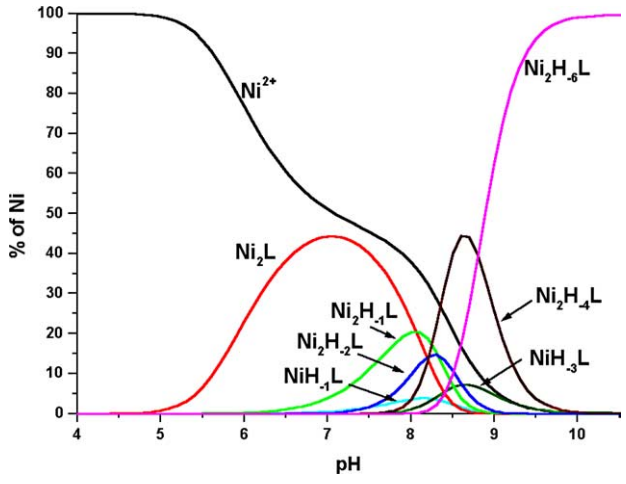


Fig. 5. Species distribution curves for Ni(II) – peptide 1, molar ratio 2:1.

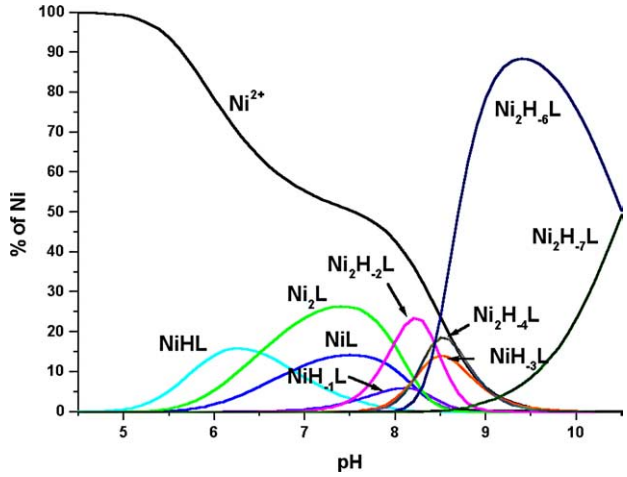


Fig. 6. Species distribution curves for Ni(II) – peptide 2, molar ratio 2:1.

metal ions to the peptides. This bonding mode of each metal ion is supported by spectroscopic data. The parameters of the absorption and CD spectra clearly indicate the formation of 4N planar diamagnetic complex with the $\{N_{Im}, 3N^-\}$ coordination mode (Table 5). The $\log K^*$ value of $Ni_2H_{-6}L$ complex, for peptide 1, is 0.6 order of magnitude lower compared to that evaluated when each metal ion is coordinated independently to the 14-amino acid fragment with the same coordination mode (Table 4, −56.32). It means that the coordination of two metal ions to the two repeated 10-amino acid sequence with 4N $\{N_{Im}, 3N^-\}$ bonding mode is not of cooperative type. The $Ni_2H_{-6}L$ complex of the peptide 2 is stabilized by about 0.7 order of magnitude compared to that obtained for the independently coordination of each metal ion to the 14-amino acid sequence (−56.32, Table 4).

Six metal complex species could be fitted to the experimental titration curves obtained for the Ni(II) – peptide 2 system with metal to ligand molar ratio 3:1, Ni_3L, $Ni_3H_{-1}L$, $Ni_3H_{-3}L$, $Ni_3H_{-6}L$, $Ni_3H_{-8}L$ and $Ni_3H_{-9}L$ (Table 3, Fig. 7). The coordination of metal ion starts at pH around 5 and Ni_3L complex (maximum concentration at pH ~6.6), is formed (Fig. 7). The absorpion and CD spectra indicate the formation in so-

Table 4
Log K^* values for the Ni(II) complexes with peptide 1 (20-amino acids) and peptide 2 (30-amino acids) containing more than one metal ion

$\log K^{*a}$	$1N\{N_{Im}\}$	$2N\{N_{Im},N^-\}$	$3N\{N_{Im},2N^-\}$	$4N\{N_{Im},3N^-\}$		
Cap43 (14 a.a. sequence)[b]	−3.62		−19.82	−28.16		
$\log K^*$ for the complexes containing two coordinated metal ions						
	$2\times 1N$ $\{N_{Im}\}\{N_{Im}\}$	$1N+2N$ $\{N_{Im}\}\{N_{Im},N^-\}$	$2\times 2N$ $2\times\{N_{Im},N^-\}$	$2\times 3N$ $2\times\{N_{Im},2N^-\}$	$2\times 4N$ $2\times\{N_{Im},3N^-\}$	
Peptide 1	−6.48 (−7.24)	−14.54	−22.86	−39.29 (−39.64)	−56.90 (−56.32)	
Peptide 2	−5.94		−21.90	−38.76	−55.65	
$\log K^*$ for the complexes containing three coordinated metal ions						
	$3\times 1N$ $\{N_{Im}\}\{N_{Im}\}\{N_{Im}\}$	$2\times 1N+2N$ $2\times\{N_{Im}\}\{N_{Im},N^-\}$	$3\times 2N$ $3\times\{N_{Im},N^-\}$	$3\times 3N$ $3\times\{N_{Im},2N^-\}$	$2\times 4N+1\times 3N$ $2\times\{N_{Im},3N^-\}\{N_{Im},2N^-\}$	$3\times 4N$ $3\times\{N_{Im},3N^-\}$
Peptide 2	−9.34 (−10.86)	−16.55 (−18.16)	−32.99	−58.13 (−59.46)	−75.88 (−76.14)	−85.27 (−84.48)

Log K^* values for the Cap43 (14-amino acids peptide) are reported for comparison.
[a] $\log K^* = \log\beta(NiH_jL) - \log\beta(H_nL)$.
[b] Ref. [24].

Table 5
Spectroscopic data for Ni(II) complexes of Ac-TRSRSHTSEG–TRSRSHTSEG (peptide 1) and Ac-TRSRSHTSEG–TRSRSHTSEG–TRSRSHTSEG (peptide 2)

Peptide, species coordination mode	Absorption		CD	
	λ (nm)	ε^a	λ (nm)	$\Delta\varepsilon^a$
Peptide 1				
$NiH_{-3}L$ (N_{Im}, $3N^-$)	437[b]	168	512[b]	+1.392
pH 9			420[b]	−3.091
			283[c]	+1.968
$Ni_2H_{-6}L$ $2\times(N_{Im}, 3N^-)$	439[b]	147	513[b]	+1.304
pH 10			420[b]	−2.969
Peptide 2				
$NiH_{-3}L$ (N_{Im}, $3N^-$)	437[b]	255	514[b]	+2.339
pH 9			421[b]	−5.344
			283[c]	+3.348
$Ni_2H_{-6}L$ $2\times(N_{Im}, 3N^-)$	437[b]	240	512[b]	+2.068
pH 9			420[b]	−4.452
			287[c]	+2.098
$Ni_3H_{-9}L$ $3\times(N_{Im}, 3N^-)$	438[b]	203	515[b]	+1.547
pH 10			421[b]	−3.696

[a] For ε and $\Delta\varepsilon$ results, $M^{-1}\,cm^{-1}$.
[b] d–d transition.
[c] Charge transfer transition N(amide) → Ni(II).

lution of octahedral complexes at pH 8.5 (data not shown). The stoichiometry of Ni_3L complex suggests $3\times 1N$ $\{N_{Im}\}\{N_{Im}\}\{N_{Im}\}$ coordination mode where each metal ion is coordinated to the imidazole nitrogen atom of the histidine residue of each 10-amino acid fragment of the peptide 2 (Table 4). The $\log K^*$ value for this complex is about 1.5 order of magnitude higher that that calculated for the same coordination mode of three systems containing one metal ion coordinated to the 14-amino acid sequence (3×-3.62, Table 4). It means that peptide 3 coordinated to three metal ions at His^6, His^{16} and His^{26} residues (one metal ion at one histidine residue) is additionally stabilized. With increasing the pH, sequential deprotonation and coordination to peptide nitrogen atoms occur and $Ni_3H_{-1}L$, $Ni_3H_{-3}L$, $Ni_3H_{-6}L$, $Ni_3H_{-8}L$ and $Ni_3H_{-9}L$ complexes are formed. The spectroscopic data are consistent with the results obtained from the potentiometric data calculations. In the solution at pH above 8.5, the formation of planar diamagnetic complexes is observed (Table 5). For $Ni_3H_{-9}L$ complex, each metal ion is coordinated to the 10-amino acid fragment with 4N $\{N_{Im}, 3N^-\}$ bonding mode. The $\log K^*$ value for this complex (−85.27) is about 0.8 order of magnitude lower compared to that evaluated for the system where each metal ion is coordinated to the 14-amino acid sequence in an independently way (3×-28.16, Table 4). The coordination of three metal ions to the three repeated 10-amino acid

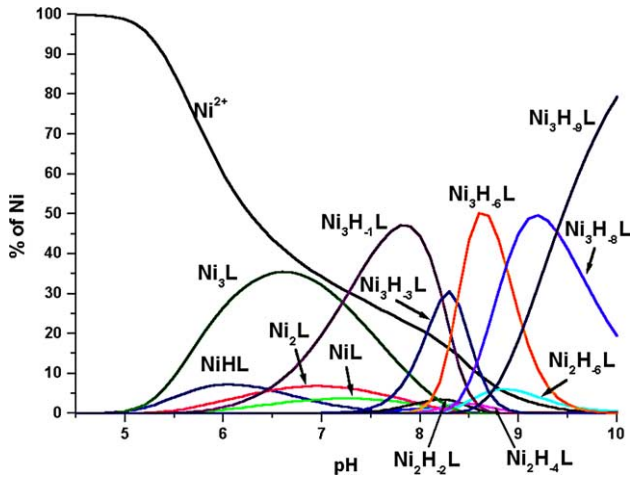


Fig. 7. Species distribution curves for Ni(II) – peptide 2, molar ratio 3:1.

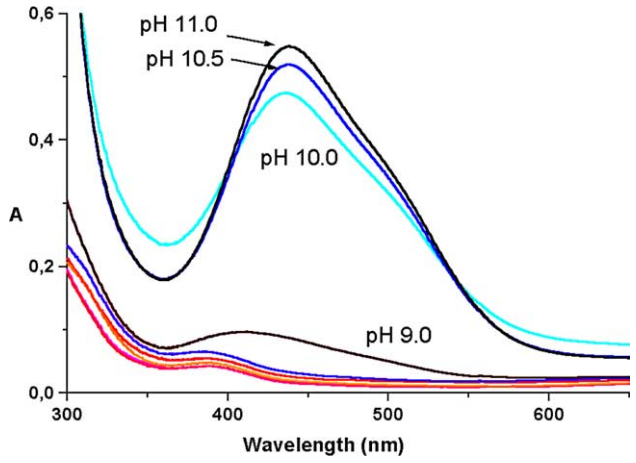


Fig. 8. UV–visible absorption spectra for Ni(II) – peptide 2 species, molar ratio 3:1, with changing the pH.

fragments of the peptide 2 in the $Ni_3H_{-9}L$ complex ($3 \times \{N_{Im}, 3N^-\}$ coordination mode) is also not cooperative (See Fig. 8).

4. Conclusions

The present paper reports the results of a study on the interaction of Ni(II) ions with the fragments of Cap43 protein containing two and three repeated –TRSRSHTSEG– amino acid sequence. The 10-amino acid fragment may coordinate one metal ion. The coordination of the metal ion starts from the imidazole nitrogen atom of the histidine residue, and with increasing the pH, Ni(II) ions are able to deprotonate successive peptide nitrogen atoms, forming Ni(II)–N^- bonds, until a $NiH_{-3}L$ and $Ni_2H_{-6}L$ species for 20- and $NiH_{-3}L$, $Ni_2H_{-6}L$ and $Ni_3H_{-9}L$ complexes for 30-amino acid fragments, are formed (above pH 8). The formation of stable five membered chelate rings by consecutive nitrogens is the driving force of the coordination process. At physiological pH (7.4) and mM concentrations of nickel(II), dependently from metal to ligand molar ratio, two repeated 10-amino acid fragment forms the Ni_2L complex (2:1 molar ratio), while 30-amino acid fragment forms the NiL (1:1), Ni_2L (2:1), and Ni_3L (3:1) complexes where each metal ion is coordinate by the imidazole nitrogen atom of the histidine residue of each 10-amino acid fragment. The Ni_2L and Ni_3L complexes of these fragments are more stable by about 0.7–1.5 order of magnitude in comparison to that evaluated for the systems with an independently coordination of each metal ion to the 14-amino acid sequence with the same coordination mode. The coordination of two or three metal ions with 4N $\{N_{Im}, 3N^-\}$ coordination mode of each metal ion ($Ni_2H_{-6}L$, $Ni_3H_{-9}L$ complexes), is not cooperative. In conclusion, this work supports the existence of relatively effective metal binding site in the C-terminal region of Cap43 protein. Our result suggest that the entire Cap43 protein could be one interesting target for Ni(II) ions.

References

[1] R.P. Hausinger, in: Biochemistry of Nickel, vol. 12, Plenum Press, New York, 1993.
[2] F.H. Nielsen, J. Nutr. 126 (suppl. 9) (1996) 2377S–22385S.
[3] G.I. Stangl, M. Kirchgessner, J. Nutr. 126 (1996) 2466–2473.
[4] F.H. Nielsen, T.J. Zimmerman, T.R. Shuler, B. Brossart, E.O. Uthus, J. Trace, Elem. Exp. Med. 2 (1989) 21–29.
[5] IARC, Lyon, France Monographs on the evaluation of carcinogenic risks to humans, vol. 49, Chromium, Nickel and Welding, 1990.
[6] A.R. Oller, M. Costa, G. Oberdorster, Toxicol. Appl. Pharmacol. 143 (1987) 152–166.
[7] P.J. Haley, D. Bice, B.A. Muggernburg, F.F. Hahn, S.A. Benjamin, Toxicol. Appl. Pharmacol. 88 (1987) 1–12.
[8] F.W. Sunderman Jr., Scan. J. Work Environ. Health 15 (1989) 1–12.
[9] Environmental Protection Agency. Project Summary Health Assessment Document for Nickel. Washington, DC: Office of Health and Environmental Assessment, 1990 EPA/600/S8-83/012.
[10] A.P. Fox, M.C. Nowycky, R.W. Tsien, J. Physiol. 394 (1987) 149–172.
[11] B. Mlinar, J.J. Enyeart, J. Physiol. 469 (1993) 379–386.
[12] G.W. Zamponi, E. Bourinet, T.P. Snutch, J. Membr. Biol. 151 (1996) 77–90.
[13] D. Zhou, K. Salnikow, M. Costa, Cancer Res. 58 (1998) 2182–2189.
[14] K. Salnikow, D. Zhou, T. Kluz, C. Wang, M. Costa, in: A. Sarkar (Ed.), Metal and Genetics, Kluwer Academic, Plenum Publishers, New York, 1999, pp. 131–144.
[15] K. Salnikow, T. Kluz, M. Costa, Toxicol. Appl. Pharmacol. 160 (1999) 127–132.
[16] H. Sigel, in: Metal Ions in Biological Systems, vol. 25, Marcel Dekker, Inc., New York, 1989.
[17] A. Agrowal, S. Bhattacharya, Experientia 45 (1989) 567–570.
[18] S. Suerbaum, J.M. Thiberge, I. Kansau, R.L. Ferrero, A. Labigne, Mol. Microbiol. 14 (1994) 959–974.

[19] J.H. Viles, F.E. Cohen, S.B. Prusiner, D.B. Goodin, P.E. Wright, H.J. Dyson, Proc. Natl. Acad. Sci. USA 96 (1999) 2042–2047.
[20] J. Meienhofer, M. Waki, E.P. Heimer, T.J. Lambros, R.C. Makofske, C.D. Chang, Int. J. Pept. Protein Res. 13 (1979) 35–42.
[21] H.M. Irving, M.H. Miles, L.D. Pettit, Anal. Chim. Acta 38 (1967) 475–488.
[22] G. Gans, A. Sabatini, A. Vacca, J. Chem. Soc., Dalton Trans. 6 (1985) 1195.
[23] G. Gans, A. Sabatini, A. Vacca, Talanta 43 (1996) 1739–1753.
[24] M.A. Zoroddu, T. Kowalik-Jankowska, H. Kozlowski, K. Salnikow, M. Costa, J. Inorg. Biochem. 85 (2001) 47–54.
[25] M.A. Zoroddu, M. Peana, T. Kowalik-Jankowska, H. Kozlowski, M. Costa, J. Chem. Soc. Dalton Trans. (2002) 458–465.
[26] J.F. Galey, B. Decock-Le Reverevd, A. Lebkiri, L.D. Pettit, S.I. Pyburn, H. Kozlowski, J. Chem. Soc. Dalton Trans. (1991) 2281–2287.
[27] H. Kozlowski, W. Bal, M. Dyba, T. Kowalik-Jankowska, Coord. Chem. Rev. 184 (1999) 319–346.
[28] M.A. Zoroddu, T. Kowalik-Jankowska, H. Kozlowski, H. Molinari, K. Salnikow, L. Broday, M. Costa, Biochem. Biophys. Acta 1475 (2000) 163–168.
[29] T. Kowalik-Jankowska, M. Ruta-Dolejsz, K. Wisniewska, L. Lankiewicz, J. Inorg. Biochem. 92 (2002) 1–10.

CHAPTER 9

NMR STUDY OF NICKEL BINDING TO N-TERMINAL SEQUENCE OF HISTONE H4

9.1 Nickel binding to Histone H4

Both water soluble and insoluble nickel compounds are implicated in the etiology of human lung and nasal cancers, although animal studies reveal that the insoluble nickel compounds, such as Ni_3S_2 are more potent carcinogens than their water soluble counterparts [1,2].

Water insoluble nickel compounds have been shown to enter cells by phagocytosis and to be stored inside cytoplasmic vacuoles, whose pH lowers to acidic values, thus accelerating the dissolution of insoluble nickel particles in a region of the cell very close to the nucleus.

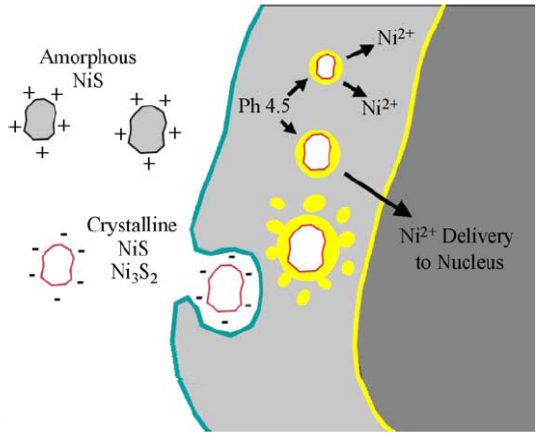


Fig. 9.1 Model of nickel uptake into cells. Cartoon illustrates the phagocytosis and intracellular dissolution of nickel sulphide particles inside a potential cancer target cell.

The carcinogenic potency of nickel compounds is consistently related to the ability of Ni(II) to access chromatin and cause multiple types of cellular nuclear damage via direct or indirect mechanisms. The mechanistic concepts proposed for nickel carcinogenesis include promutagenic DNA damage [3,4], epigenetic effects in chromatin [5-7], and impairment of DNA repair [8].

The binding of Ni(II) ions within the cell nucleus is a crucial element of all these concepts. Obviously, detection and chemical and structural description of specific Ni(II) binding sites in the nuclear compartment would provide molecular basis for a better understanding of the effects that eventually lead to cancer development.

The DNA polymer binds Ni(II) only weakly [9], leaving nuclear proteins as possible targets for Ni(II). Due to their abundance in chromatin, histones seem to be the most likely ligands. In fact, the core histone octamer (formed by two copies of histones H3, H4, H2A and H2B) together with the linker histone H1, package eukaryotic DNA into repeating nucleosomal units that are folded into higher order chromatin fibers.

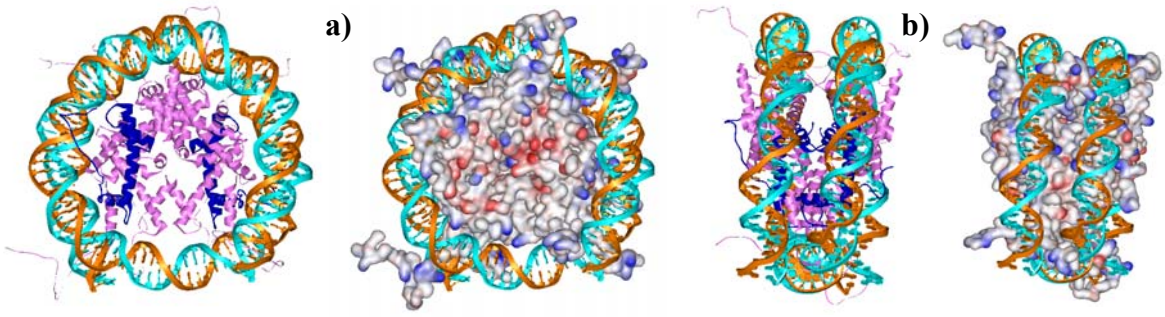


Fig. 9.2 Nucleosome core particle: ribbon traces for the 146-bp DNA phosphodiester backbones (brown and turquoise) and eight histone protein main chains (pink for H3, H2A, H2B and blue for H4). The views are down the DNA superhelix axis for **a)** and perpendicular to it for **b).** For histones molecular surface are also showed [19].

General chemical preferences of Ni(II) and, in particular, the binding modes of Ni(II) to proteins [10] and peptides [11,12] indicate that His imidazoles and Cys thiols are preferred by Ni(II) amongst the donor groups provided by the amino acid residues of proteins. The linker histone H1 does not contain any His or Cys residue, while H3 and H2A histones, for instance, contain two or more such residues capable of strong coordination to this metal.

In order to identify putative Ni(II) binding motifs on histones, we analyzed the data reported in the literature for amino acid sequences [13,14] as well as the structural data [15,18] of these core proteins, getting a better insight into the formation and physicochemical properties of Ni(II) complexes with selected motifs.

Once considered as static nonparticipating structural elements, histones are now clearly recognized as integral and dynamic components of the machinery responsible for the regulation of gene transcription. In particular, histone N-terminal tails are central to the processes that modulate nucleosome structure and function.

The N-terminal regions of the histone proteins are extensively and reversibly post-translationally modified by acetylation, phosphorylation, ubiquitination and methylation. Of these modifications, acetylation has generated more interest since gene expression was directly correlated with histone acetylation. The site of acetylation is the lysine residue of the positively charged amino-terminal tails where each acetyl group added to a histone reduces its net positive charge weakening and modulating the interactions between histones and the surface of the nucleosome.

The positive charge of the histone tail is neutralized by acetylation to the lysine residues increasing the access to transcription factors and making active chromatin more available for modifications and the gene expression [19].

It has been reported that nickel is a potent suppressor of histone H4 acetylation, in both yeast and mammalian cells [20].

Interestingly, an anchoring binding site for nickel ion on this protein, histidine H_{18}, is close to sites for post-translational modification involved in nickel toxicity.

All these informations point to the H4 tail as a candidate for a Ni(II) binding site in the histone octamer. Because of its structural position, this terminal part could be accessible for metal binding.

For this reason we analyzed, for Ni(II) binding, the sequence of the N-terminal tail of the histone H4.

```
   1          11         21         31         41         51
 1 SGRGKGGKGL GKGGAKRHRK VLRDNIQGIT KPAIRRLARR GGVKRISGLI YEETRGVLKV
61 FLENVIRDAV TYTEHAKRKT VTAMDVVYAL KRQGRTLYGF GG
```

Fig. 9.3 Sequence of Histone protein H4. The N-terminal tail are red highlighted. Histidine H_{18} are showed in green colour.

Previously we had reported, by potentiometric and spectroscopic (Uv-Vis, CD) studies, about the interaction of Ni(II) with minimal models of the H4 tail: the two peptides with 6 amino acids Ac-AKRHRK-Am and with 22 amino acids Ac-SGRGKGGKGLGKGGAKRHRKVL-Am, respectively[21-23].

From these studies it resulted that the histidine residue can be a basic binding site for Ni(II) ions in the histone core of the nucleosome. Therefore, the H4 tail peptide fragment can potentially be one of the biologically relevant sites for nickel genotoxicity.

Here we present our recent results on the coordination ability of Ni(II) to the N-terminal tail of histone H4, the 30-amino acid peptide Ac-SGRGKGGKGLGKGGAKRH_{18}RKVLRDNIQGIT-Am, achieved by the use of multidimensional NMR spectroscopy.

9.2 NMR Study of Nickel Binding to N-terminal Sequence of H4 Histone

9.2.1 NMR Spectroscopy

NMR experiments were performed on a Bruker Avance 600 or 700 MHz spectrometers equipped with 5 mm TXI 1H-13 probe. Samples used for NMR experiments were 5 mM in concentration and dissolved in 90% H_2O/10% D_2O solutions. All acquisitions were performed at the temperature of 298 K. The pH of the sample was adjusted to the final pH by addition of 1 N NaOH or 1 N HCl. The titration experiments on Ni(II)-containing samples with peptide-to-metal molar ratio of 1:1 were performed at pH 8.7. Nuclear Overhauser Enhancement Spectroscopy (NOESY) with mixing times of 500 ms and Total Correlation Spectroscopy (TOCSY) with a mixing time of 50 ms were also performed. The combination of TOCSY and NOESY experiments was used to assign the spectra of both free and Ni(II)-bound peptide. Solvent suppression for 1D, TOCSY and NOESY experiments was achieved using WATERGATE pulse sequence or using excitation sculpting with gradients. All NMR data were processed using XWINNMR (Bruker Instruments) software on a Silicon Graphics Indigo workstation and analyzed using the Sparky 3.11 program.

9.2.2 NMR Characterization of N-terminal Sequence of Histone H4

A comparison of the 1D, 2D ^{1}H homonuclear TOCSY and NOESY NMR spectra of H4 30aa free peptide and of peptide-Ni(II) species was performed at pH= 8.7. This pH was chosen to approach maximum formation of the major planar diamagnetic species, as evidenced by potentiometric and spectroscopic measurements [22].

The resonances belonging to the 30-residues free peptide Ac-SGRGKGGKGLGKGGAKRH$_{18}$RKVLRDNIQGIT-Am were assigned on the basis of 1D NMR spectra and 2D ^{1}H homonuclear TOCSY and NOESY experiments. The alkaline pH required for nickel binding results in exchange of the amide protons with bulk water. Therefore their resonances were lost. Only the aromatic resonances of histidine (HE1 and HD2 at 7.595 and 6.862 ppm, respectively) were present in the region between 6.6 and 8.5 ppm (fig. 9.4)

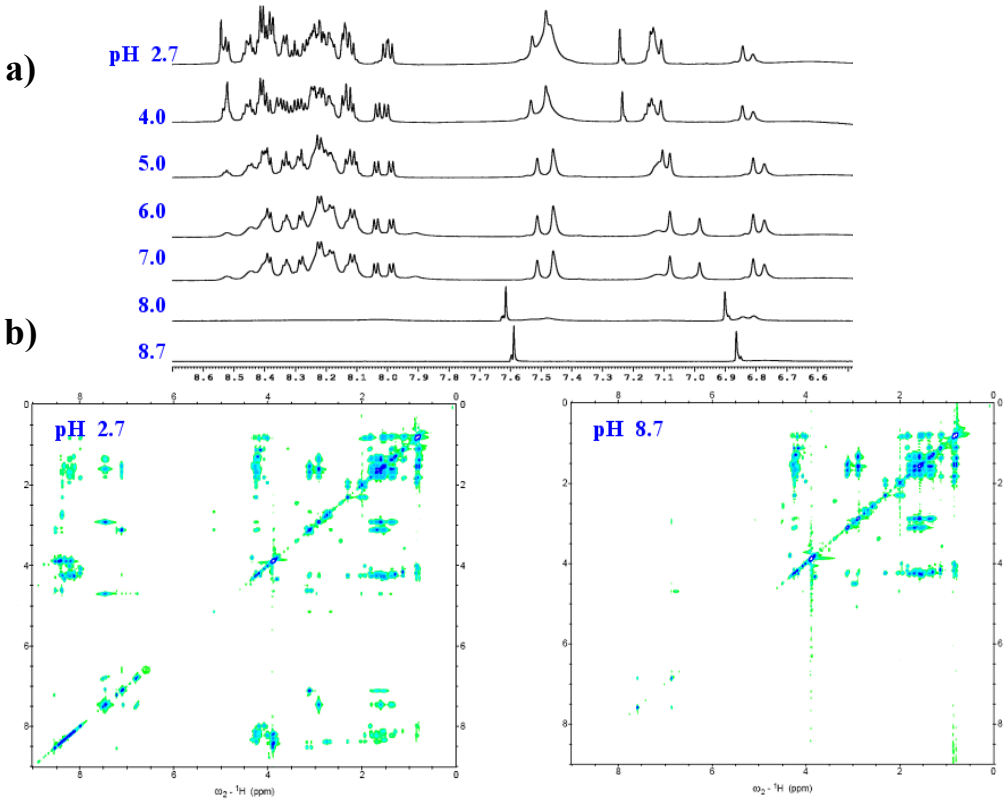


Fig. 9.4 a) ^{1}H 1D aromatic region of the free peptide N-terminal sequence of Histone H4 in dependence of pH, **b)** 2D ^{1}H Tocsy spectra of the peptide a pH 2.7 and 8.7 respectively.

The stepwise procedure applied to achieve detailed assignments by using ^{1}H 2D TOCSY and NOESY spectra made at pH 2.7 and at the successive points (pH 4.0-8.0) enabled unproblematic assignments at pH 8.7 where exchange with bulk water suppressed H^{N} amidic protons. Table 8.1 reports the chemical shift assignments at pH 8.7.

	Chemical shift ppm														
Residue	**HA**	**HB**	**HB1**	**HB2**	**HG**	**HG1**	**HG2**	**HN**	**QB**	**QD**	**QD1**	**QD2**	**QE**	**QG**	**QG2**
S1	4.340								3.804						
G2	3.872														
R3	4.257		1.683	1.793						3.116				1.541	
G4	3.898														
K5	4.270		1.687	1.793						1.580			2.905	1.355	
G6	3.898														
G7	3.898														
K8	4.277		1.687	1.793						1.580			2.901	1.355	
G9	3.898														
L10	4.264				1.541				1.570		0.832	0.837			
G11	3.898														
K12	4.251		1.687	1.793						1.580			2.900	1.355	
G13	3.898														
G14	3.898														
A15	4.215								1.291						
K16	4.218		1.687	1.793						1.580			2.896	1.355	
R17	4.260		1.683	1.793						3.119				1.541	
H18	4.513		3.042	2.970				8.522							
R19	4.257		1.683	1.793						3.116				1.541	
K20	4.193		1.687	1.793						1.580			2.894	1.355	
V21	4.007	1.953												0.838	
L22	4.306				1.531				1.572		0.827	0.838			
R23	4.260		1.683	1.793				8.128		3.117				1.557	
I26	4.076	1.817				1.356	1.104			0.802					0.823
Q27	4.252		1.934	2.042										2.309	
G28	3.898														
I29	4.186	1.839				1.384	1.126			0.810					0.841
T30	4.260	4.164												1.322	

Table 9.1 ^{1}H NMR chemical shifts for Ac-SGRGKGGKGLGKGGAKRH$_{18}$RKVLRDNIQGIT-Am at pH 8.7. Degenerated protons are referred as Q.

9.2.3 NMR Characterization of Ni(II) Binding to N-terminal Sequence of Histone H4

The binding mode of Ni(II) to the H4 sequence was studied at pH = 8.7 with increasing nickel concentrations up to the final peptide-to-Ni(II) molar ratio of 1:1.

Unfortunately, at this final molar ratio precipitation was observed; nevertheless, clear information on the binding mode of the metal can be obtained from a series of 1D ^{1}H, 2D TOCSY and NOESY complex spectra up to molar ratio of 1:0.8.

Both the histidine aromatic protons show an upfield shift of their signals upon metal coordination: HE1 proton appears now at 7.472 ppm, with $\Delta\delta$= -0.123 ppm respect to the free peptide, and HD_2 proton moves to 6.837 ppm, with a $\Delta\delta$= -0.025. The effect of incremental addition of Ni(II) on ^{1}H NMR signals for H_{18} ring protons has also been monitored. Figure 9.3(a) shows the ^{1}H mono-dimensional spectra collected for a Ni(II) titration of our peptide at pH 8.7. Figure 9.5(b) indicates the change in absolute peak height for resolved HE1 proton on H_{18} ring plotted versus equivalents of Ni(II) up to 1.0 eq (from 0.8 to 1.0 eq a linear prediction was used). The appearance of a new set of signals for H_{18} ring protons HE1 and HD2 is clearly visible in the mono-dimensional spectra. However, only the HE1 peak intensities have been plotted in Figure 9.5 (b) as the HD2 signals are overlapped with both apo-form and Ni(II)-form signals. The latter signals have reached about 90% of their maximal intensity after addition of 1 eq of Ni(II), suggesting that H_{18} binds Ni(II) with a high affinity, in a ratio close to 1:1.

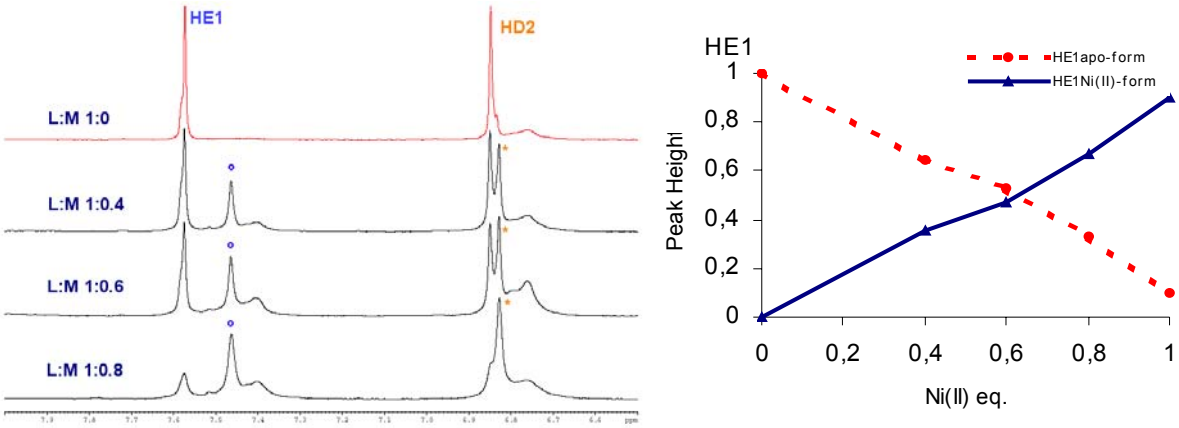


Fig. 9.5 a) Changes in intensity and chemical shift of the histidine aromatic protons (HE1 and HD2) by increasing nickel concentration. Comparison of aromatic regions of 1D ^{1}H NMR spectra of H4 30aa peptide-Ni(II) binding in the molar ratios of 1:0, 1:0.4, 1:0.6 and 1:0.8 respectively, **b)** Intensity, as absolute peak height, of selected peak HE1 as a function of added Ni(II).

Ni(II) has the ability to adopt a range of coordination geometries. Octahedral and tetrahedral geometries form high-spin paramagnetic complexes while a square-planar or a five-coordinated square-pyramidal geometry produces a low-spin diamagnetic complex. The line-width at half-height for the new Ni(II) bound signals is comparable to the apo signals, indicating that Ni(II) coordinates to H_{18} in a low spin diamagnetic complex. The estimated 10% of free high-spin octahedral paramagnetic Ni(II) aqua ion is responsible for the slight broadening found in all the signals of the spectra.

Comparison of the free peptide and peptide-Ni(II) complex 1D, 2D TOCSY spectra at pH 8.7 points out several additional peaks appearing upon coordination.

Strong shifts involve the HA and HB protons of H_{18} (HA 3.481 ppm, $\Delta\delta$= -1.032; HB_1 2.840 ppm, $\Delta\delta$= -0.202, HB_2 2.709 ppm, $\Delta\delta$= -0.261).

Moreover, new spin systems appear for R and K residues, attributable to K_{16} and R_{17}, in which the strongly shifted resonance of the HA protons of K_{16} (at 3.937 ppm, $\Delta\delta$= - 0.281) and R_{17} (at 3.565 ppm, $\Delta\delta$= -0.695), are identified.

.

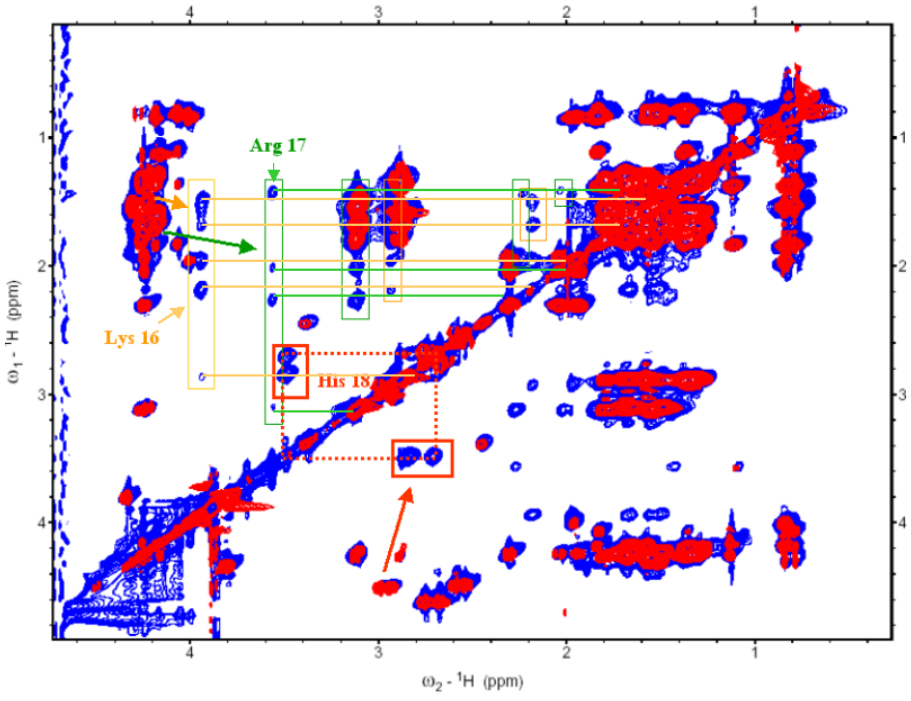


Fig. 9.6 Superimposition of TOCSY spectra of free 30aa peptide (red) and 30aa peptide-Ni(II) (blue) 1:0.8 molar ratio. The new spin systems for Lys_{16}, Arg_{17} and His_{18} after nickel interaction are also indicated.

The chemical shift values in the new system identified for K_{16} , R_{17} and H_{18}, and the Δδ differences with the resonances of the free peptide are reported in the next table.

.

assignment	free peptide	peptide :Ni(II) complex 1 :3	chemical shift diff
	ppm	ppm	ppm
K_{16} QG	1.355	1.499	*0.144*
K_{16} HB2	1.793	2.183	*0.39*
K_{16} HB1	1.687	1.969	*0.282*
K_{16} HA	4.218	3.937	*-0.281*
R_{17} QD	3.119	3.119	*0.000*
R_{17} QG	1.541	1.429	*-0.112*
R_{17} HB2	1.793	2.255	*0.462*
R_{17} HB1	1.683	2.019	*0.336*
R_{17} HA	4.26	3.565	*-0.695*
H_{18} HB1	3.042	2.84	*-0.202*
H_{18} HB2	2.97	2.709	*-0.261*
H_{18} HA	4.513	3.481	*-1.032*
H_{18} HE1	7.595	7.472	*-0.123*
H_{18} HD2	6.862	6.837	*-0.025*

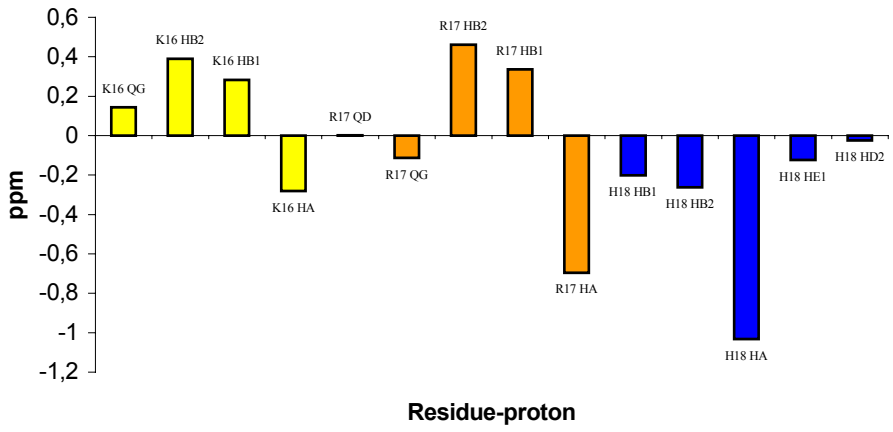


Fig. 9.7 Chemical shift assignment for the residues K_{16}, R_{17} and H_{18} in Ac-SGRGKGGKGLGKGGAKRH$_{18}$RKVLRDNIQGIT-Am directly involved in the complex formation, before and after nickel interaction; the chemical shifts differences are also reported in the plot.

The 2D NOESY spectra, respect to the TOCSY ones, showed additional meaningful cross-peaks. Although the alkaline pH required for Ni(II) binding, due to rapid exchange of the labile H^N protons with the bulk water, did not allow the observation of the potential H^{Nx}-H^{Ny} or H^{Nx}-H_i correlations, the aromatic histidine protons HE1 and HD2 clearly show a new set of NOE cross-peaks appearing after nickel binding.

Preliminary check of the data showed that some of the NOEs involve both HE1 and HD2 interactions with aliphatic protons HA, HB, HG of Lys K_{16} residue.

The figure 9.8 represent the aromatic region of 2D NOESY spectra showing the new spin systems.

Presently, these data are under analysis in order to collect a detailed set of overall structural constrains required for a full structural determination of the complex.

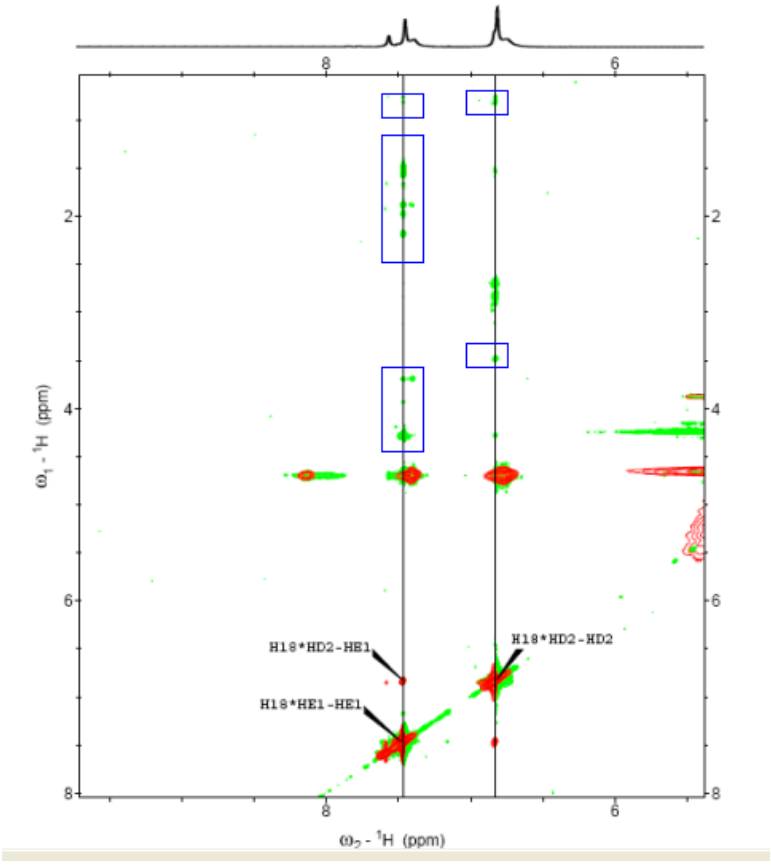


Fig. 9.8 Superimposition of aromatic region for TOCSY (red) and NOESY (green) spectra, of peptide-Ni(II) 1:0.8 molar ratio. Some of NOE cross-peaks appearing after nickel binding are already assigned and found to involved HE1 and HD2 Histidine H_{18} protons with HA, HB, HG of Lys K_{16}.

9.3 Conclusions

From NMR experiments, the chemical shifts induced by Ni(II) coordination to the peptide were consistent with the binding of the metal ion in a square-planar site by four nitrogen atoms from His_{18} ($Nd1,N_H$) and from Arg_{17} and Lys_{16} (N_H).

The proton NMR spectra showed significant downfield and upfield changes particularly with regard to the HA protons of the backbone as a result of complexation of Ni(II) to the peptide.

The HB1, HB2, HG protons of arginine Arg_{17} and the HB1, HB2 protons of lysine Lys_{16} exhibited a downfield shift, suggesting a change of hydrophobic packing of the side chains following nickel binding. The appearance of new sets of NOE indicates that when Ni(II) forms 4N-complexes with the peptide ligand, the peptide conformation is dramatically affected; this aspect is particularly important with donors able to bind metal ions through the backbone peptide nitrogens, since they will influence both the physical and biological properties of the peptide itself.

More detailed informations of the binding mode of Ni(II) to the N-terminal sequence of histone H4 and the relate conclusions will be available after the complete analysis of the data and the consequent structure calculations.

At the moment, the results from NMR studies together with our previous spectroscopic (UV-Vis, CD) and potentiometric studies [21-23], support the existence of relatively effective metal binding sites in the N-terminal region of H4 protein.

These results suggest that the entire H4 protein could be an interesting target for Ni(II) ions.

Although the complexation with Ni(II) at a physiological pH, under our experimental conditions, is not very effective, the formation of a rigid square planar complex may result, somehow paradoxically, in a higher specificity of Ni(II) to produce a particular conformation of the peptide [24]. The presence of positively charged residues close to the metal binding site in the H4 tail can result in a site-selectivity association of the Ni(II) complexed tail with the negatively charged DNA backbone [25,26]. In addition, the hydrophobic environment in the entire protein is expected to

enhance metal binding capabilities, due to the multiple non-bonding interactions available, as reported in the literature [27,29].

In conclusion, the histidine H_{18} residue can be a primary binding site for Ni(II) ions, therefore the H4 tail can potentially be one of the biologically relevant sites for nickel genotoxicity.

LIST OF REFERENCES FOR CHAPTERS 9

1. Salnikow K, Blagosklonny MV, Ryan H, Johnson R, Costa M, (2000) *Cancer Res*, 60,38.

2. Piquemal D, Joulia D, Balaguer P, Basset A, Marti J, Commes T, (1999) *Biochim Biophys Acta*, 1450,364.

3. Kasprzak, K. S. (1995) *Cancer Invest*. 13, 411.

4. Kasprzak, K. S. (1997)In *Cytotoxic, Mutagenic and Carcinogenic Potential of Heavy Metals Related to Human Environment* (Hadjiliadis, N., Ed.) 73, NATO ASI Ser. 2, Environment, Vol. 26, Kluwer,Dordrecht.

5. Costa, M. (1991) *Annu. Rev. Pharmacol. Toxicol.* 31, 321.

6. Salnikow, K., Cosentino, S., Klein, C., and Costa, M. (1994) *Mol. Cell. Biol.* 14, 851.

7. Lee, Y.-W., Klein, C. B., Kargacin, B., Salnikow, K., Kitahara, J., Dowjat, K., Zhitkovich, A., and Costa, M. (1995) *Mol. Cell. Biol.* 15, 2547.

8. Hartwig, A. (1995) Current aspects in metal genotoxicity. *Bio-Metals* 8, 3.

9. Lee, J. E., Ciccarelli, R. B., and Wetterhahn Jenette, K. (1982) *Biochemistry* 21, 771.

10. Halcrow, M. A., and Christou, G. (1994) *Chem. Rev.* 94, 2421.

11. Pettit, L. D., Gregor, J. E., and Kozłowski, H. (1991) In *Perspectives on Bioinorganic Chemistry* (Hay, R.W., Dillworth, J. R., and Nolan, K. B., Eds.) Vol. 1, 1, JAI Press, Greenwich.

12. Sigel, H., and Martin, R. B. (1982) *Chem. Rev.* 82, 385.

13. GenPept (GenBank Gene Products). For GenBank, cf. Burks, C., Cassidy, M., Cinkosky, M. J., Cumella, K. E., Gilna, P., Hayden, J. E.-D., Kelley, T. A., Kelly, M., Kristofferson, D., and Ryals, J. (1991) GenBank. *Nucleic Acids Res.* 19, 2221.

14. von Holt, C., Brandt, W. F., Greyling, H. J., Lindsey, G. G., Retief, J. D., Rodrigues, J. D. A., chwager, S., and Sewell, B. T. (1989). In *Methods in Enzymology, Vol. 170, Nucleosomes* (Wassarman, P. M., and Kornberg, R. D., Eds.) 503, Academic Press, San Diego.

15. Arents, G., Burlingame, R. W., Wang, B.-C., Love, W. E., and Moudrianakis, E. N. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88, 10148.

16. Arents, G., and Moudrianakis, E. N. (1994) *Structural Biology: The State of the Art, Proceedings of the Eighth Conversation, State University of New York, Albany, NY, 1993*(Sarma, R. H., and Sarma, M. H., Eds.) 93, Adenine Press, New York.

17. Arents, G., and Moudrianakis, E. N. (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92, 11170.

18. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) *Nature* 389, 251-260.

19. M. Grunstein, (1997) *Nature* 389, 349.

20. L. Broday, W. Peng, M.H. Kuo, K. Salnikow, M.A. Zoroddu, M. Costa (2000) *Cancer Res.* 60, 238.

21. Zoroddu MA, Kowalik-Jankowska T, Kozłowski H, Molinari H, Salnikow K, Broday L, Costa M. (2000) *Biochim Biophys Acta*. 1475,163.

22. M.A. Zoroddu, M. Peana, T. Kowalik-Jankowska, H. Kozłowski, M. Costa, (2002) *J. Chem. Soc., Dalton Trans*., 458.

23. M.A. Zoroddu , L. Schinocca, T. Kowalik-Jankowska, H. Kozłowski, K. Salnikow , M. Costa, (2002) *Environ Health Perspect*., 110 Suppl 5, 719

24. H. Kozłowski, W. Bal, M. Dyba and T. Kowalik-Jankowska, (1999) *Coord. Chem. Rev.*, 184, 319.

25. D. F. Shullenberg, P. Eason and E. C. Long, (1993) *J. Am. Chem. Soc.*, 115, 11038.

26. Q. Liang, P. D. Eason and E. C. Long, (1995) *J. Am. Chem. Soc.*, 117, 9625.

27. W. Bal, G. N. Chmurny, B. D. Hilton, P. J. Sadler and A. Tucker, (1996) *J. Am. Chem. Soc.*, 118, 4727.

28. M. M. Yamashita, L. Wesson, G. Eisenman and D. Eisenberg, (1990) *Proc. Natl. Acad. Sci. USA*, 87, 5648.

29. L. Regan, (1993) Annu. Rev. Biophys. Biomol. Struct., 22, 257.

ACNOWLEDGEMENTS

First and foremost, I am deeply indebted to my Tutor Professor Maria Antonietta Zoroddu, whose help, stimulating suggestions and encouragement supported me in all the time of research.
I express my profound sense of gratitude to Professor Ivano Bertini, Director of Magnetic Resonance Center (CERM) of University of Florence, for giving me the opportunity to conduct part of my doctoral work within an excellent research infrastructure in a very stimulating scientific environment.
I would like to thank my supervisor in CERM, Professor Claudio Luchinat for his expert guidance during my time spent there and for introducing me into the wonderful world of Calmodulin. I appreciate several enthusiastic scientific discussions with him and his timely critics and full support, which helped me to achieve my objectives.

I take this opportunity to sincerely thank all my colleagues and friends: Yogesh Gupta, Cristina Del Bianco, Giacomo Parigi, Moreno Lelli, Nicola d'Amelio, Marco Fragai, Yong-Min Lee, Massimiliano Maletta, Manuele Migliardi, Sara Albanese, Elisa Libralesso, Sara Grassi, Tilemakhos Karavelas, Ileana Leontari, Manolis Matzapetakis Serenella Medici, Serena Sirigu, Murugendra Vanarotti and Christos Chasapis for their invaluable suggestions, help, hints or simple friendship.

I wish also to thank Professor Stefano Enzo, Coordinator of the PhD course.

All NMR experiments were performed on a Bruker Avance spectrometers 600 and 700 MHz at the Magnetic Resonance Center (CERM) of University of Florence.