University of Florence

International Doctorate in Structural Biology

Cycle XX (2005-2007)

Genomics and structural proteomics studies on metalloproteins

Ph.D. thesis of

Eva Grassi

Tutor

Prof. Claudio Luchinat

Coordinator

Prof. Claudio Luchinat

S.S.D. CHIM/03

This thesis has been approved by the University of Florence, the University of Frankfurt and the Utrecht University

A me stessa

e a tutti coloro che mi hanno sostenuto

....it is our choices that show what we truly are,
far more than our abilities.

Albus P. W. B. Dumbledore

Acknowledgements

Lo so che potrà sembrare strano, ma sicuramente questa è la pagina più difficile da scrivere. Infatti sono innumerevoli le persone che sento di dover ringraziare, per cui temo di dimenticarne qualcuna.

Desidero ringraziare tutta la mia famiglia e gli amici che in questi tre anni mi hanno supportato e sopportato con tanto affetto, in particolar modo la mia mamma sempre pronta ad aiutarmi nella vita quotidiana e il mio papà, che sono sicura, continua a proteggermi.

Ringrazio il Prof. Luchinat e il Prof. Bertini per avermi dato la possibilità di conseguire questo dottorato.

Un sentito ringraziamento va a Marco, la persona che più di tutti ha contribuito a questo dottorato. Senza il suo prezioso aiuto e la sua infinita disponibilità probabilmente le cose sarebbero andate diversamente….. Grazie soprattutto per avermi insegnato durante le nottate insonni a lavoro, che nella ricerca, come nella vita, bisogna essere "tigre", non demordere lamentandosi di fronte ai risultati negativi, ma ingegnarsi e rimboccarsi le braccia.

Grazie a Elisa che in ogni momento si è dimostrata una vera amica, sempre pronta ad aiutarmi e incoraggiarmi.

Come non ringraziare Francesca e Antonio che mi hanno insegnato i rudimenti della spettroscopia NMR e mi hanno introdotto nel magico mondo del paramagnetico!…….quanta fatica!! ma alla fine ce l'abbiamo fatta!!!

Desidero ringraziare anche il Prof. Parigi per la sua disponibilità e il suo prezioso aiuto nell'analisi dei dati.

Un pensiero particolare va a Nunzia, Chiara, Adele, Sara Dragonetti/Albanese e Valentina le amiche delle stanze 2 e 3 sempre pronte a dividere gioie e dolori porgendoti una spalla o strappandoti un sorriso.

Ringrazio tutto lo staff di ProtEra s.r.l. e soprattutto le "galline proteriane" (Mercia, Eleonora, Cristina e Mariapina).

Uno speciale ringraziamento va a Beatriz, Ania, Manuele Martinelli/Migliardi, Stefano, Leonardo, Murugendra, Massimiliano Peana/Maletta, Serena, Serenella, Emanuele Chirivino/Marzocchi, Nicola, Maxime, Moreno, Antonella, Luisa, la parabola del Grillo, il Gallo e il Cacciatore, la banda dei greci e degli indiani… Tutte persone che ho avuto la fortuna di incontrare al CERM e che mi hanno trasmesso e insegnato qualcosa durante questo dottorato.

Contents

Chapter I

INTRODUCTION

1.1 STRUCTURAL BIOLOGY

Genome sequencing projects have made available the complete genome of more than 60 species in addition to thousands of viruses. This sudden increase in the amount of genomic information has provided a new starting point to understand our basic genetic machinery.

Proteins, encoded by genes, perform their functions by interacting with each other in coordinated networks. But, only a fraction of these networks has been identified and characterized through classical biochemistry. A much more comprehensive view of proteins interaction patterns within cells will become possible with sequence information from multiple species, including humans. The knowledge of the full complement of our genes, and the function of correspondent proteins, should enable us to identify all of the metabolic pathways in the human body.

In the latest years, besides the genome sequencing projects, the field of structural genomics has seen an explosive growth. The efforts of structural biology projects tend to obtain a full coverage of the fold space and to structurally characterize all the products of selected genomes.

Structural biology requires a large number of process steps to convert sequence information into a 3D structure; which includes: choosing proper expression constructs, setting up of right growth conditions and efficient purification strategy.

Usually, the three-dimensional structures of proteins, or protein domains, are obtained by X-ray crystallography or NMR spectroscopy. NMR is unique in that it allows researchers to study the conformation of these molecules in the liquid state, which more closely resembles the cellular environment. Moreover, it easily allows the characterization of biological molecules under several different experimental conditions, such as different pH and ionic strength values, as well as in the presence of denaturating agents for the study of the unfolding process.

Many enzymes exert their cellular functions in the context of larger assemblies and not as isolated molecules, since they often are part of more than one cellular complex. Understanding proteins behaviour in vivo, therefore, requires the study of their structure and function within the context of such complexes, since they are the mediators of biological function in the cell, and so the relevant targets for drug design.
Thus, the role of NMR will be crucial not only in the structure determination of proteins as stand-alone entities, but also in the dynamic characterization of their functional complexes and the investigation of ligand binding.

1.2 METALLOPROTEINS

Although traditionally regarded as organic, life is inorganic too. Indeed metals are commonly found as natural constituents of several biomolecules. At present, we are aware of 13 metals which play an essential role as trace elements in many biological systems[11] for plants and animals[11;46]. Four of these, sodium, potassium, magnesium and calcium, are present in large quantities and are known as bulk metals[46]. The remaining nine, which are present in small quantities, are the d-block elements vanadium, chromium, molybdenum, manganese, iron, cobalt, nickel, copper and zinc, and are known as the trace metals.
The importance of metals in living organisms is mainly due to the feasibility of their transfer in different cell compartments, where they can be employed in various biological processes such as: electron transfer reactions, oxygen transport, signals transductions, and in a large variety of catalytic processes.
The Lanthanide ions play a particular role between all metals. Indeed the striking similarity of lanthanide(III) ions to calcium(II) ions in size, coordination environment and ligand preferences, along with strong Lewis acidity, suggests that this crossover could reasonably be made. Although

not as rare as their name would imply, the low biological availability of the lanthanides no doubt precluded this evolutionary outcome. However, this presents an exciting opportunity for chemists to explore the properties, structure, selectivity and reactivity of Ln's in Ca^{2+} binding sites. Indeed calcium-binding proteins are characterized by a native metal-binding site, where the calcium can be naturally substituted with different lanthanide ions. It is generally accepted that the substitution of Ca^{2+} with Ln^{3+} does not significantly alter the conformation of the proteins[15;19;61], so this strategy can be used to perform structural proteomic studies through the analysis of paramagnetic restraints, as reported[14;25]. The trivalent Ln ions have an array of useful chemical and spectroscopic properties that include luminescence, paramagnetism (for all but La^{3+} and Lu^{3+}) and hydrolytic activity (due to strong Lewis acidity). These characteristics make them valuable probes to study structure, function and dynamic interactions in biomolecules.

Current estimates indicate that about one third of all structurally characterized proteins are capable of binding one or more metal ions (hence called metalloproteins) either as a structural component, as a triggering agent for their activities, or as a catalytic co-factor[30] thanks to their redox and acid/base properties.

In addition to the metal cofactor, an enzyme can have a prosthetic group, usually covalently bound to some specific residues, that often play an important role in its functions. Heme group is one prosthetic subunit extensively characterized, since the proteins with this cofactor (as well known as hemeproteins) are involved in several biological processes. For esample, one small heme protein, as cytochrome c[12], is an essential component of the electron transfer chain[66] and also an intermediate of apoptosis[40;41;47;57].

To complete the picture, metals are also very important for the structure and function of nucleic acids, specifically RNA.

Several metals are thus essential for organism survival, but toxic when in excess. Consequently, their intracellular concentration as well as the distribution among the various cell compartments, and their incorporation

into metalloproteins, must be tightly controlled[21;30]. A proper balance of all the involved processes is necessary for a healthy phenotype.
Therefore, the characterization of metalloproteins is an important aspect to understand several biological systems.
Genome sequencing projects have provided a huge number of protein primary sequences, but although several different elaborated analyses and annotations have been enabled by a rich and ever-increasing portfolio of bioinformatic tools, metal-binding properties remain difficult to predict as well as to investigate experimentally. Consequently, the present knowledge about metalloproteins is only partial[30].
An integrated and multidisciplinary approach, including bioinformatic, molecular biology and structural proteomic, must to be used to investigate the metalloprotein family.

1.3 AIMS AND TOPICS OF THE RESEARCH

During these three years of PhD, my research was focused on the investigation of metalloproteins using a multidisciplinary approach. In this frame, I tried to integrate different scientific fields for this common aim.
This thesis is divided into two main projects. The first one is focused on genomics studies dealing with the search of polymorphisms in a mitochondrial metalloprotein. The second one is a structural proteomics study, involving the exploitation of the calcium-lanthanide substitution to obtain structural information in solution.
The reason behind the choice of studying Single-Nucleotide-Polymorphisms (SNPs) lies in the fact that several genomic studies have evidenced the relationship between DNA sequence variations, as mutations or SNPs, and diseases[69].
Therefore I decided to perform a study in order to understand the degree of polymorphism of metalloproteins in possible relation with their

misfunctions.
During this analysis I noticed that the SNP density of the human Cyt c gene (CYCS) is 1 order of magnitude higher than the average SNP density. In particular I focused my attention on the coding region of Cyt *c*, since the existence of Cyt *c* variants (cSNP) affecting biological functions without being incompatible with life, could be hypothesized. From the evolutional point of view, such variability could be also the indication of an exceptional propensity of the human Cyt *c* protein to individual plasticity.
Since all these cSNPs were obtained by computational methods only[20], the aim of our research was the experimental validation of these putative cSNP.

In the second project I exploited a structural proteomic approach in order to investigate protein function through structural data and biophysical proprieties.
My attention was mainly devoted to the expression of proteins, which present a calcium-binding motif, since the calcium in their native metal-binding site can be naturally substituted with different lanthanide ions. The presence of these peculiar metals in a protein allows to exploit paramagnetic restraints (T1 enhancements, pcs and rdc) to obtain more accurate structural informations on protein-protein interaction or protein complexes [3;16], thus increasing the feasibility of NMR characterization.
My interest was focused on two proteins involved in the copper transport, HAH1 and first metal domain of ATP7A (MNK1)[48;49], as a model system to investigate the protein-protein interaction using the paramagnetic approach. In particular, I have expressed some constructs of human protein HAH1 that despite lacking a native Ca^{2+}-binding site, can acquire the ability to bind lantanide ions by inserting calcium-binding site as tags[29;50;82] in its constructs. In this frame, I exploited a peptide (LBT), already used in several research works[31;82], in order to determine using a paramagnetic approach the structure of the metal-mediated adduct, which was formed by HAH1 and MNK1 proteins. A significant potential shortcoming of the LBT strategy is that these relatively short polypeptides

may feature a significant conformational with respect to the protein or domain that they are fused to. This may harness the usefulness of the pcs measured for structure determination of proteins or protein complexes.
Also in this context, I developed a new paramagnetic tag proposing as candidate the entire CalbindinD_{9k} , since the flexibility of this tag respect to the tagged protein could be restricted taking advantage of the steric volume occupied by tag and protein. In this case the entire protein P43M Calbindin D_{9k} (CABP) was fused to the first metal domain of ATP7A (MNK1), in order to demonstrate the feasibility of this new tag.

Bertini I, Grassi E, Luchinat C, Quattrone A, Saccenti E. Genomics. 2006 Nov;88(5):669-72.

Chapter II

Monomorphism of human cytochrome *c*

INTRODUCTION

Single Nucleotide Polymorphism

The completion of the human genome project (HGP)[24] has allowed the identification of thousands of gene polymorphisms, so the scientific attention is now rapidly shifting towards the study of individual genetic variations. In fact the sequencing of the human genome would open up the possibility to systematically identify all possible gene variants in different human populations, associate their presence with individual phenotype, including disease susceptibility, and determine the functional impact of such variations. The most abundant source of genetic variations in the human genome comprises single nucleotide polymorphisms (SNPs)[23;71]. A polymorphism (a term that comes from the Greek words "poly," or "many," and "morphe," or "form") is a DNA variation. They occur, on average, once every 300 base pairs of sequence with a minor allele frequency (MAF) greater than 1% [43;64;73]. Although the majority of SNPs are probably functionally neutral, some clearly have functional effects, and thus are presumed to underlie the differences in human biological traits and individual response to therapeutic treatment or environmental exposure. The SNPs in gene coding regions (cSNPs), altering the amino-acid sequence, can lead to changes in the structure and biological properties of the encoded protein. For example, SNPs in the NAT2 gene significantly decrease the enzyme metabolism of aromatic amines[10], and individuals who are homozygous for low activity alleles, have an increased risk for smoking-induced bladder cancer[53].

Several studies about the peroxisome proliferator-activated receptor-gamma2 (PPAR_-gamma2) have reported conflicting results with regard to the association between a frequent polymorphism of this gene and complex traits, such as insulin sensitivity, Type 2 diabetes and obesity[76].

In the same way, cSNP in the MTHFR gene have been clearly associated with a decreased enzyme activity, so these polymorphisms could be used

as genomic predictors of clinical response to fluoropyrimidine-based chemotherapy[52].

In contrast, the SNPs in non coding gene regulatory regions, such as promoters, enhancers, silencers, and introns, may affect gene expression levels in an allele-specific manner[63], and these functional polymorphisms represent an important but relatively unexplored class of genetic variations[37]. There are several examples of a SNP in a regulatory region causing either complete elimination of the natural transcription factor binding site[18;78] or formation of a novel spurious site[42;62]. In addition, SNPs in transcription factor binding sites can lead to allele-specific binding of transcription factors and can modulate gene expression[35].

So, in the future, the characterization of these processes and the understanding of the impact of polymorphism will possibly prevent and control the environmentally induced diseases and allow progress in pharmacological research. Indeed, there are a lot of projects devoted to the determination and validation of the presence of SNPs in several genes. The NIEHS Environmental Genome Project (EGP) is a systematic attempt to identify genetic variation in environmentally responsive genes and up till now, over 44669 polymorphisms in 332 environmentally responsive genes have been discovered and catalogued in NIEHS Gene SNP database (http://www.genome.utah.edu/genesnps/). Another initiative, designed to characterize the polymorphisms, is SNP500Cancer. This project is a part of the National Cancer Institute's Cancer Genome Anatomy Project (CGAP) and represents a resource for the mapping of SNP variations in cancer-related genes (http://snp500cancer.nci.nihgov/)[59].

Although these initiatives are aimed at providing a reliable source of information regarding well characterized human genes, their degree of coverage of the about 25,000 annotated human genes is still negligible.

For this reason, the main bioinformatics resource to retrieve starting information on variations in human genes, remains the dbSNP database developed at NCBI[20;22;39].

Since metalloproteins are involved in numerous essentials process of life, the discovery and the validation of new metalloproteins SNPs are an

important advancement for the scientific research.

Search for SNPs of Cytochrome c

Cytochrome *c* (Cyt *c*) is an intensively studied metalloprotein because it has a relatively small size (104 amino acids in mammalians) and it is involved in numerous cellular processes essentials for life, such as aerobic and anaerobic respiration or apoptosis[40;41;57]. It is an ubiquitous protein, found in all prokariotes and to date in more than 100 eukariotes. Their alignment showed that the primary sequence is very highly conserved, with residue identity ranging between 45% and 100% across all eukariotes[9]. The Cyt *c* sequence and overall fold are thus well preserved during evolution. The amino acids coordinating the iron ion (His18 and Met80) are conserved especially in the region around the axial Met80, whereas the one around the axial His18 is slightly more variable. Several functions of cytochrome c are associated with the protein-protein interactions between the polypeptide itself and its interacting partners, so in this case the presence of SNPs could change the efficiency of these process. To support such hypothesis, it is well known that in anthropoid primates the genes, encoding components of the mitochondrial electron transport chain (ETC), have been major targets of darwinian positive selection[65;83].

Sequence comparison have provided evidence that marked increase of non-synonymous substitution rates occurred in anthropoid ETC genes encoding subunits of complex III and IV and cytochrome *c*; moreover this accelerated amino acid replacement rates of COX (cytochrome *c* oxidase) and CYC (cytochrome *c*) in the stem anthropoid and stem catarrhine lineages were coadaptive. These molecular evolution changes might have had functional consequences for the mechanism and the rate of aerobic energy production in anthropid cells, since this phenotypic changes could be associated with enlarged neocortex, prolonged fetal development and extended lifespan, all process supported by adaptations in aerobic energy

production.

While components of the COX and of the apoptosome complexes have been shown to undergo disease-causing mutations in humans and mice[36;60], evidence of Cyt c sequence variation in metazoan has never been reported. Prenatal lethality due to this variation is suggested by death at midgestation of embryos with targeted disruption of the Cyt c gene, even if cell lines established from early Cyt c null embryos were viable if compensated for defective oxidative phosphorylation[45]. Moreover extensive experimental mutagenesis has been performed in the yeast Cyt c ortholog, with the generation of several viable mutant strains[70], and some knock-in mice generated with a point-mutated Cyt c gene reached adulthood[34]. Therefore, the existence of Cyt c variants affecting biological functions without being incompatible with life, could be hypothesized.

MATERIAL and METHODS

SNP Analysis

Genome browsing

One of the main challenges that the molecular biology community has to face today, is to make sense of the wealth of data produced by the genome sequencing projects. In the past decade, bioinformatics have become an integral part of research and development in the biomedical sciences. Bioinformatics now has an essential role both in deciphering genomic, transcriptomic and proteomic data generated by high-throughput experimental technologies, and in organizing information gathered from traditional biology. It is an interdisciplinary research area at the interface between the biological and computational science. The ultimate goal of bioinformatics is to uncover the biological information hidden in the mass of data, in order to obtain a clearer insight into the fundamental biology of organisms.

A number of data banks are available and provide the scientific community with tools for gene banks searching protein sequences analysis and the prediction of a variety of protein properties. Primary databases contain information and annotations of DNA and protein sequences, DNA and protein structures and protein expression profiles.

Some available databases for genome browsing are:

- **NCBI** (www.ncbi.nlm.nih.gov) This web site integrates information from several databases (Swissprot, EMBL, all geneBank, etc...)
    1. **dbSNP** (www.ncbi.nlm.nih.gov/projects/SNP/). In collaboration with the National Human Genome Research Institute, the National Center for Biotechnology Information has established the dbSNP database to serve as a central repository for both

single base nucleotide substitutions and short deletion and insertion polymorphisms.

2. **BLAST** (www.ncbi.nlm.nih.gov/BLAST/): Standard BLAST (Basic Local Alignment Search Tool) is a set of similarity search programs, designed to explore all of the available sequence databases regardless of whether the query is protein or DNA.

- **Ensembl** (www.ensembl.org) is a joint project between the EMBL-EBI and the Wellcome Trust Sanger Institute, that aims at developing a system that maintains automatic annotation of large eukaryotic genomes. It is a comprehensive source of stable annotation with confirmed gene predictions that have been integrated from external data sources. Ensembl annotates known genes and predicts new ones, with functional annotation from InterPro, OMIM, SAGE and gene families.
- **PDB** (www.rcsb.org/pdb) is a 3-D biological macromolecular structure database.
- **CLUSTALW** (www.ebi.ac.uk/clustalw/) is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen through viewing Cladograms or Phylograms.

cSNP retrieval

Coding Single Nucleotide Polymorphisms have been retrieved in the dbSNP BUILD 125 (www.ncbi.nlm.nih.gov/sites/entrez?db=snp&cmd=search&term=) and the Uniprot/SwissProt databases (http://expasy.org/sprot/). The Ensembl (www.ensembl.org) has been also used for SNP annotation (Fig.1).

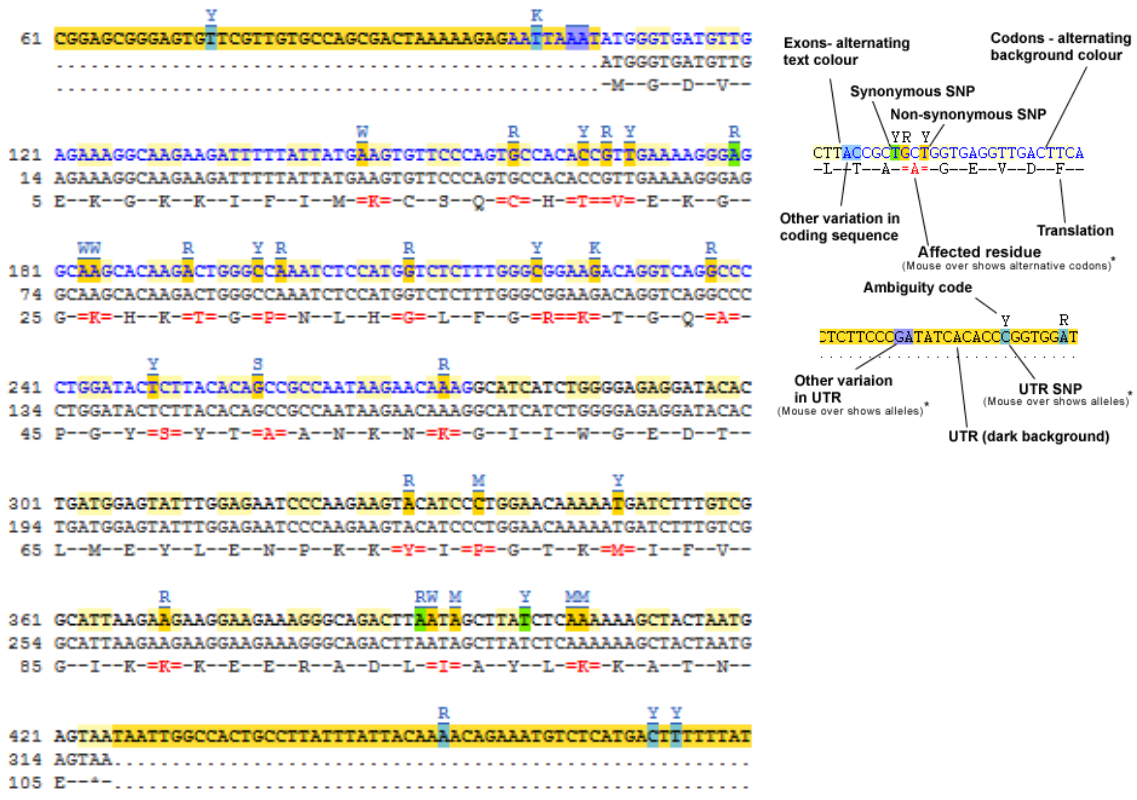

Figure 1 Alignment of the nucleotide and aminoacid sequences of Cyt c showing the SNP positions

Surveying the dbSNP database

The public SNPs database dbSNP (build 125) (http://www.ncbi.nlm.nih.gov/projects/SNP/) was queried using either the ENTREZ SNP filter mask or queries with Boolean operators. The success rate of candidate cSNPs derived from *in silico* analysis of public EST libraries was estimated by restricting the queries to the panel of 546 genes re-genotyped by the NIEHS Environmental Genome Project (EGP) (http://egp.gs.washington.edu). EGP-derived SNPs and SNPs, coming from the two EST–based computational methods published in *Nature Genetics,* were retrieved in dbSNP according to the criterion that they had the [HANDLE] tags "EGP_SNPS" for the NIEHS project dataset, the [HANDLE] "CGAP-GAI" for the Buetow *et al* dataset[20] (query restricted to "computed"[METHOD]) and the [HANDLE] "LEE" for the Irizarry *et al.* dataset[39].

Experimental cSNPs were retrieved in dbSNP according to the criterion that they had the following [METHOD] tags: “sequence”, “hybridize”, “rflp”, “sscp”, “dhplc”. Computational cSNPs were retrieved in dbSNP according to the criterion that they had the “computed” [METHOD] tag.

Resequencing of cytochrome c

Reliability of our experimental method

A mono-directional sequencing for the 3’ UTR flanking regions of CYCS gene for both SNP rs7810784 and rs12700584 was performed using standard methods. Genomic DNA, as template, was obtained form Coriell.
For SNP rs7810784 the forward primer 5’-TTTAACCCAGAAGTAATCAGCCCAGTAGTA-3’ was used to sequence of a PCR segment (410 bp). Said PCR fragment was amplified using the following forward primer 5’-TTTAACCCAGAAGTAATCAGCCCAGTAGTA-3’ and reverse 5’-CCAACACAGACCTTAATATAGGAGGCATAG-3’(30)
For SNP rs12700584 the primer 5’-AAACTAGATAACTGGGCGTCGTGGT-3’ was used to sequence PCR segment (530 bp). In this case, the primer forward 5’-AAACTAGATAACTGGGCGTCGTGGT-3’ and reverse 5’-ATACTGATGACGGATTGCCAAAAGA-3’ were used in order to amplify the DNA fragment, containing analysed SNP.

Experimental method of CYCS cSNP validation

The gene CYCS (NM_018947.4, gi 34328939) was amplified from the genomic DNA of 95 samples from Coriell panel using the primers 5’-AGTGGCTAGAGTGGTCATTCATTTACA-3' and 5’-TCATGATCTGAATTCTGGTGTATGAGA-3', spanning the coding part of exon 2, the short intervening intron 2-3, and the coding part of exon 3 for a PCR segment of 609 bp. The amplified DNA fragments were analyzed by

electrophoresis on agarose gel and purified by QIAquik PCR Purification Kit (Qiagen).
Bi-directional sequencing of the coding portion of the human cytochrome *c* was performed by standard methods

Human DNA Samples

Genomic DNA was obtained from the same panel of 95 individuals composing the Panel 2 used in the NIEHS Environmental Genome Project (http://egp.gs.washington.edu/).
This population is defined here as the sample of N=95 DNAs with different etnicities. The anonymized samples were obtained from the Coriell Cell Repositories (Coriell Institute for Medical Research, Camden, NJ, USA), and represent four ethnic groups: 27 Africans (15 African-American, 12 Africa Yoruba), 22 Caucasians (CEPH/ Utah), 23 Hispanics (Mexican-American community of Los Angeles) and 24 Asians (12 Chinese, 12 Japanese). The individuals comprised 45 male and 50 female. These individuals are not a random sample of any specific human population, and thus the predictive value of the sequence and genotype data provided will vary for different population samples. This panel is the same used by NIEHS SNPs Project.

RESULTS and DISCUSSION

During the analysis of the degree of polymorphism in metalloproteins, it emerged that an unusually high number of cSNPs for Cytochrome *c* were reported in the dbSNP build 125. For the coding region of human CYCS gene (Uniprot/SwissProt: P99999; Ensemble: ENS00000172115), 15 coding non-synonymous SNPs (nsSNPs) and 3 synonymous SNPs (sSNPs) have been retrieved in the dbSNP build 125 database (Table 1)**.**

N°Residue	SNP ID	SNP type	Alleles	Ambiguity code	Alternate residues	Handle|Submitter ID
14	Rs11548821	ns	T/A	W	K, STOP	CGAP-GAI|1494305
18	Rs17851278	ns	C/T	Y	C/Y	YMGC_GENOME_DIFF|BC015130 x37538470-C24457294T
20	Rs11548816	ns	G/A	R	T, I	CGAP-GAI|1494299
21	Rs11548815	ns	C/T	YR	I, V	CGAP-GAI|1494291
24	Rs11548797	s	T/C	Y	-	CGAP-GAI|1494238
26	Rs11548802	ns	T/A	WW	M, K, STOP, L	CGAP-GAI|1494260
29	Rs11548799	ns	T/C	Y	A, T	CGAP-GAI|1494240
31	Rs11548796	ns	G/A	RY	S, P	CGAP-GAI|1494237
35	Rs11548805	ns	C/T	Y	D, G	CGAP-GAI|1494265
39	Rs11548791	ns	G/A	R	W, R	CGAP-GAI|1494228
40	Rs11548812	ns	C/A	M	N, K	CGAP-GAI|1494287
44	rs1154880	ns	C/T	Y	A, T	CGAP-GAI|1494268
48	Rs11548783	ns	A/G	R	S, P	CGAP-GAI|1494195
51	Rs11548772	ns	C/G	S	A, P	CGAP-GAI|1494160
56	Rs11548795	ns	T/C	Y	R, K	CGAP-GAI|1494235
75	Rs11548820	ns	T/C	Y	C, Y	CGAP-GAI|1494304
77	Rs11548818	ns	G/T	K	H, P	CGAP-GAI|1494302
81	Rs11548778	ns	A/G	R	T, M	CGAP-GAI|1494175
88	Rs11548785	ns	T/C	Y	R, K	CGAP-GAI|1494210
95	rs3211448	s	T/C	Y	-	LEE|e2078449, SEQUENOM|sqnm209891
96	rs3211449	ns	T/A	WK	F, I, L	LEE|e2078452; SEQUENOM|sqnm209892
98	Rs11548776	s	A/G	R	-	CGAP-GAI|1494172
100	rs3211451	ns	T/G	KK	Q, T, P,K	LEE|e2078484

Table 1 cSNPs reported in dbSNP build 125 for CYCS

Our attention was focused on the coding region of Cyt *c*, since 19 cSNPs

appeared in an ORF length of 319 bases and a SNP occurs, on average, once every 300 base pairs of sequence. Such SNP density on the human Cyt c gene is 1 order of magnitude higher than the average SNP density.
That being so, we realized that the primary source of these SNP (15 out of 19) was a dataset deposited in dbSNP in 2003 by CGAP-GAI project, on the basis of EST-based computational method[20]. The other cSNPs derived from a similar approach, again based on EST data[39] (LEE, SEQUENOM). Such EST-based dbSNP data are generally used for genomic studies with the same level of confidence as well as SNPs obtained with experimental methods. Indeed, the papers [20;39], relating to describe such methods, claimed an overall success rates of 70-80% through sample validation experiments.
However, the internal check that we performed within the dbSNP database in order to determine the percentage of SNP retrieved by computational methods and verified by experimental approaches, showed that only 28.2% of the cSNPs deriving from EST-based computational methods have been detected also by experimental methods. This success ratio was already unexpectedly low, but we found that it decreased further (21.1%) when the comparison method was tightened by restricting it to the genes analysed in the frame of the NIEHS Environmental Genome Project (EGP) (http://egp.gs.washington.edu/) (Table 2). In this case, the LEE[39] cSNP dataset scored 32,6%, and the CGAP-GAI[20] cSNP dataset scored as low as 19,1%.

Therefore, the CYCS gene cSNPs were analysed in order to evaluate the potential structural and functional impact for such protein. Two of these reported cSNPs, namely the rs11548796 (Pro31Ser) and the rs11548778 (Met81Thr), represent quite dramatic variations, since they are involved in the correct binding of the iron ion of the heme group.
This prostethic group, covalently bound to Cyt c by thioether links between the vinyl groups of the heme and the sulphur atoms of two cysteins, contains an iron ion that lies in a six-coordinated geometry defined among the others by the Nε1 atom of the imidazole ring of His19

and the sulphur of Met81. The Hδ1 proton of said imidazole is hydrogen bound to the carbonyl oxygen of Pro31.

Both residues Pro31 and Met81 are absolutely conserved in 113 mitochondrial cytochrome *c* from different eukaryotes[9]; furthermore, residues surrounding Met81 are conserved at least in the 95% of all the organisms. On the contrary, in this region dbSNP reported two non synonymous variations, rs11548820 (Tyr75Cys) and rs11548818 (Pro77His), which could in principle impair the protein functionality.

Several studies[68;74] have evidenced that Mendelian diseases associated cSNPs tend to occur at highly conserved amino acid positions even in quite distantly related proteins suggesting that they generally have severe impact on protein function.

	Non Synonymous SNPs			Synonymous SNPs			Total coding SNPs		
	Deposited in dbSNP125	**Confirmed by EGP_SNPs**	**% Confirmed by EGP_SNPs**	**Deposited in dbSNP125**	**Confirmed by EGP_SNPs**	**% Confirmed by EGP_SNPs**	**Deposited in dbSNP125**	**Confirmed by EGP_SNPs**	**% Confirmed by EGP_SNPs**
Experimental(a)	1085	453	41.7%	981	523	53.3%	2066	976	**47.2**%
Computed(b)	848	131	15.4%	775	211	27.2%	1623	342	**21.1**%
CGAI-GAP(c)	293	36	12.3%	229	74	32.3%	522	100	**19.1**%
LEE(d)	194	48	24.7%	177	73	41,2%	371	121	**32.6**%

Table 2. *Number of cSNPs retrieved in the dbSNP build 125 restricted to the panel of 555 genes re-probed for variation by sequencing in the frame of the NIEHS Environmental Genome Project.*

cSNPs identified through experimental methods (sequencing, hybridization, RFLP, SSCP, dHPLC) but EGP_SNPS.

cSNPs identified through computational methods.

cSNPs identified by CGAP-GAI[20]

cSNPs identified by LEE[39].

These statistical considerations linked to the putative high structural and functional impact of some reported mutations led us to search experimental validation for these SNPs, since the existence of Cyt *c* variants (cSNP), affecting biological functions without being incompatible with life, could be hypothesized.
Since all these human CYCS cSNPs are still lacking validation or independent discovery by other methods, we PCR-amplified and bi-directionally resequenced the coding parts of two human Cyt *c* coding exons and the short inverting intron.

In order to assess the goodness and the reliability of our experimental method to determine the presence of SNPs in the coding region of human CYCS, we choose to re-validate, using the same method, two SNPs present in 3'-UTR region of same gene. Two SNPs were selected: rs7810784 and rs12700584. Both these SNPs have been previously tested and validated by non-computational methods, as reported in dbSNP database.
In order to re-validate the SNP rs7810784, we selected a sub-panel of 12 individuals (Coriell ID: NA18526, NA18524, NA18562, NA18537, NA18545, NA18572, NA18609, NA18552, NA18566, NA18621, NA18577, NA18635) from the CSHL-HAPMAP project (HAPMAP-HCB population) and re-sequenced the corresponding 3'-UTR flanking region. We obtained the following genotype data G/G 0.416, A/G 0.250, A/A 0.333 and allele frequencies A 0.458, G 0.542, fully consistent with the experimental data from CSHL-HAPMAP reported in dbSNP125 (genotypes: G/G 0.429, A/G 0.310, A/A 0.262, Alleles: A 0.417, G 0.583).
As for the SNP rs12700584 genotype and allele frequencies data are available for each individual of our panel in the CSHL-HAPMAP project (HAPMAP-HCB population), we choose to re-sequence three individuals (Coriell ID: NA18526, NA18609, NA18545).
Three individuals reported to be C/C homozygote (NA18526), C/T heterozygote (NA18609), and T/T homozygote (NA18545) were selected, and their 3'-UTR sequences showed the correct genotype for all three.
Again, the genotype data from our sequencing are consistent with data

from CSHL-HAPMAP project.

These results shown the full reliability of our sequencing method.

Since the accuracy of our experimental method was verified, we analysed the presence of CYCS cSNPs in a statistically meaningful sample of different individuals, and we selected the same pool of individuals used in the EGP project. The coding parts of the two CYCS coding exons and the short intervening intron were PCR-amplified from the selected panel of 95 individuals (Fig.2) and bi-directionally sequenced.

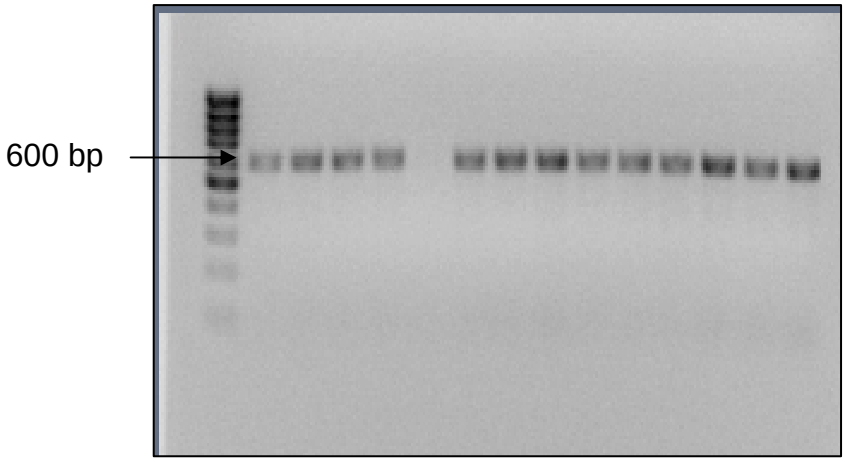

Figure 2 Agarose gel of gene CYCS amplified by PCR from several individuals of Coriell panel. Line 1 is marker of molecular weight (100 bp Fermentas)

We expected this quite high number of experimentally verified cSNPs to be able to provide a reasonable estimate of the reliability of the candidate cSNPs for this gene. Surprisingly, none *of the 19 cSNPs was found in 95-individuals panel* (190 chromosomes) *and no other new SNPs were found.*

It could be argued that 190 chromosomes may not be sufficient to discover rare SNPs, and therefore several, if not all, of the candidate cSNPs of CYCS will be eventually validated by experimental studies on samples of larger depth. However, it should be kept in mind that also the method by CGAI-GAP[20] uses a number of selected EST sequences of at most a few hundreds (194 in the case of CYCS), i.e. comparable to the number of chromosomes in the present analysis. Therefore, it is extremely

unlikely that a significant number of the 19 candidate cSNPs of CYCS be rare SNPs.

Taken together, the re-examination of the SNPs deposited in dbSNP by EST-based methods casts serious doubts on the validity of this computational approach. From the present statistics, we should conclude that on average three out of four candidate cSNPs are instead sequence errors. It is likely that it is not the computational approach itself that is faulty, but rather the reliability of the raw data, i.e. of the deposited EST sequences. This should not be surprising, as the depth of the EST sequences was dictated by the need to reliably identify a consensus expressed sequence for any given cluster, not to reliably identify outliers. In other words, the quality of the EST sequences was good enough to identify the major allele for each variant base in the genome, but not good enough to discriminate a real SNP from an error (in reverse transcription, DNA amplification or sequencing).

In conclusion, from our resequencing of the human CYCS gene in a 95-individuals panel we derived a likely monomorphism for the ORF of this gene in the human population, resulting in the absence of allelic variants of the Cyt *c* protein. Phylogenetically, an accelerated amino acid replacement rate, especially at the interface with COX, has been described recently for Cyt *c* orthologs in anthropoid primates but not in other mammalian lineages[65]. This finding has been interpreted as a possible adaptive response to the increase production of oxygen radicals by mitochondrial respiration in enlarged brains of longer-living individuals, two phenotypic features characteristic of the anthropoid evolution. If true, this hypothesis should be reconciled with our finding of Cyt *c* monomorphism in our species, possibly favoured by the structural constraints imposed to Cyt *c* by the interactions with two mitochondrial (Cytochrome bc1 and COX) and one cytosolic (apoptosome) complexes while performing its key cellular roles in respiration and programmed cell death.

Chapter III

The use of pseudo-contact shift induced by a Lanthanide-biding tag for the study of protein-protein interactions.

INTRODUCTION

NMR spectroscopy is a well-established technique for the determination of the three-dimensional structure of proteins, nucleic acids and their adducts in solution. The structural characterization of multi-domain proteins or of protein-protein complexes is often a challenging task due to the paucity of restraints that can be measured when the domain-domain interaction is relatively weak, and therefore interdomain dynamics is present, or when the affinity of the complex is low, and thus again dynamics can be present. Even when the extent of dynamics is modest, the determination of inter-domain or inter-protein NOEs to be used in structural calculations can be technically difficult.

In this context, the incorporation of lanthanide ions into proteins can be quite effective thanks to the effects that they can induce in the NMR spectra, which are a potential source of information on long-range interatomic distances, on the reciprocal orientation of domains, and on interdomain or inter-protein dynamics. This wealth of information is due to the paramagnetism of the trivalent ions of lanthanides (Ln^{3+}). Ln^{3+} are chemically stable and are paramagnetic (except La^{3+}, Lu^{3+}) due to the presence of unpaired electrons in their *f* orbitals. The paramagnetism causes pronounced changes in the chemical shifts of the nuclear spins located around the metal ion and causes partial orientation of the molecule at high magnetic fields. These effects are very well understood and can be described quantitatively. Other effects that are induced by these paramagnetic metals in NMR spectra are the enhancement of nuclear relaxation rates, and various cross-correlation effects.

The above-mentioned paramagnetic contribution to the measured chemical shift of nuclei located several chemical bonds away but relatively close in space to the Ln^{3+} ion is called pseudocontact shift (pcs). Pseudocontact shifts can be measured quite accurately and easily, and provide structural information at distances up to 30-40 Å. Therefore , they have long been used for structure determinations of paramagnetic

metalloproteins by NMR spectroscopy. More recently, the residual dipolar couplings (rdc's) that can be measured for protein bond vectors thanks to the partial alignment induced by the Ln^{3+} ion at high magnetic field have also been shown to be very useful for structural purposes.

Due to these favourable properties, and also to other useful properties such as luminescence or anomalous scattering of X-ray radiation, various methods for the incorporation of lanthanide ions into biomolecules have been explored. In various applications, the similarity of the Ln^{3+} ions to Ca^{2+} in ionic radius and oxophilicity has enabled their direct incorporation into calcium-binding proteins. The majority of proteins, however, lack native calcium-binding sites. Thus, one approach has been to incorporate lanthanide-chelating prosthetic groups e.g. through the chemical modification of cysteine side chains. These chelators can bind Ln^{3+} ions extremely tightly, but the process requires considerable manipulation. A straightforward approach is to integrate a lanthanide-binding sequence as a protein co-expression tag via molecular biology strategies. Recent design and engineering studies on calcium-binding structural motifs have resulted in the development of short polypeptides comprising 20 encoded amino acids or fewer that bind tightly and selectively to lanthanides. These peptides, dubbed "lanthanide-binding tags" (LBT's), show affinities up to 10^{9}-10^{10} M^{-1} and are selective for lanthanides over other common metal ions. Various applications of the LBT strategy are available from the literature. A significant potential shortcoming of the LBT strategy is that these relatively short polypeptides may feature a significant conformational with respect to the protein or domain that they are fused to. This may harness the usefulness of the pcs measured for structure determination of proteins or protein complexes. In this paper, we used the adduct formed by the copper(I)-chaperone HAH1 and the first metal-binding domain of the human ATPase ATP7A (MNK1) as a model system to investigate this matter, by fusing a LBT to the C-terminus of HAH1. It is shown that while the dynamics of the LBT prevents the use of pcs data for structure determination of HAH1 itself, the pcs induced on MNK1 still permit a low-resolution structural characterization of the complex

MATERIAL and METHODS

Cloning, expression and purification of LBT-HAH1 and MNK1

The QuikChange mutagenesis kit (Stratagene) was used for insertion of a large DNA fragment encoding the LBT directly at the 3' end of the gene encoding HAH1 in our previously developed plasmid for expression of the His_6-tagged protein. In this way, the LBT tag was fused to the C-terminus of HAH1. We chose to use the sequence YIDTNNDGWYEGDELLA (17 residues) for the LBT, which has been optimized for the present kind of applications[54;72;82]. The cloning protocol applied here represents an expansion of the classical QuikChange protocol, where PCR fragments are used as megaprimers for the insertion mutagenesis[33;80]. Three different constructs were prepared, where zero, one or two Gly residues were inserted between the protein and the tag (G_0LBT, G_1LBT, G_2LBT, respectively). In the remainder, HAH1-LBT will be used to indicate the samples used in procedures identically applied to all three constructs.

Protein expression and purification was carried out as previously reported[2]. The various HAH1-LBT constructs were expressed in *E. coli* strain BL21(DE3)Gold. Cells were grown at 37 °C in M9 minimal medium containing $(^{15}NH_4)_2SO_4$ as the sole nitrogen source. The medium was supplemented with a vitamin mix and a trace metal solution and contained ampicillin (100 µg/ mL). Protein expression was induced using 1 mM IPTG at an OD_{600} of 0.6, and the cells were harvested 5 hours later. Cells were then harvested by centrifugation at 7000 rpm and resuspended in the lysis buffer (20 mM Na_2HPO_4 pH 8, 0.5 M NaCl, 5 mM Imidazole). Cell lysis was accomplished with 10 cycles of sonication; the suspension was finally centrifuged at 30000 rpm for 30 min. Protein purification was carried out by affinity chromatography with a HiTrap chelating FF affinity column (Amersham GE Healthcare Life Science) previously charged with Zn^{2+}. The protein was eluted with an isocratic gradient up to 60 mM EDTA.

After purification, the His_6 tag was cut by overnight incubation with Factor Xa (40 µg of enzyme/20 mg of recombinant protein) under reducing atmosphere in a glove-box chamber. The protein without the His_6 tag was purified by passing the solution again through the HiTrap column charged with Zn^{2+}, under conditions where the protein retaining the tag bound to the column. The typical protein yield was 30- 40 mg of purified protein/L of culture.

The first metal-binding domain (5-77 A.A) of ATP7A (MNK1) had already been cloned in pET20b+. Said plasmid was transformed into *E. coli* Rosetta pLysS strain. M9 medium containing $(^{15}NH4)_2SO_4$, as the sole nitrogen source, 100 µg /ml of ampicillin and 34 µg /ml chloramphenicol was inoculated with 10 ml/l overnight culture *E. coli* Rosetta pLysS expressing the human MNK1 and grown at 37°C in shaker. At OD_{600} of 0.6 the expression of the recombinant protein was induced with 0.5 mM IPTG and the colture was incubated O/N at 17°C in order to prevent the formation of inclusion bodies.
The purification protocol and sample preparation were the same used for the HAH1-LBT.

NMR sample preparation

All manipulations of the purified proteins (both HAH1-LBT and MNK1) were performed under reducing atmosphere in a glove-box chamber. Apo-protein samples were reduced with excess dithiothreitol (DTT), and then washed with 100 mM Hepes, 100 mM NaCl, 2mM DTT buffer at pH 7.0 using an Amicon stirred cell. Copper(I)-containing samples were prepared by incubating the apo-protein with a slight excess of $[Cu(I)(CH_3CN)_4]PF_6$, as previously reported. NMR tubes were sealed before they were removed from the chamber. The final protein concentration in all samples was about 0.3 mM. NMR samples also contained 10% (v/v) ^2H_2O for NMR

spectrometer lock.

NMR samples of the HAH1-LBT:MNK1 adduct were prepared by mixing in the same NMR tube equimolar amounts of HAH1-LBT and copper(I)-loaded MNK1. Samples prepared by mixing copper(I)-HAH1-LBT and MNK1 gave identical NMR spectra.

Ln^{3+}-containing samples were prepared by adding aliquots from a stock 33 mM solution of the chloride salt, up to a 1:1 Ln^{3+}: HAH1-LBT ratio, either to HAH1-LBT alone or to the HAH1-LBT:MNK1 adduct. Binding of the Tb^{3+} ion was checked by fluorescence, as described in the literature[31;51].

^{1}H and 2D [^{1}H,^{15}N]-HSQC NMR spectra were acquired using a Bruker AVANCE spectrometer operating at a proton frequency of 600 MHz and at a temperature of 298 K.

Calculations

Pseudocontact shifts (pcs, defined as the isotropic rotational average of dipolar shifts) are observed in NMR spectra of paramagnetic molecules in solution when there is a metal-based magnetic anisotropy. They are given by[44;56]:

$$PCS = \delta^{pc} = \frac{1}{12\pi r_i^3}\left[\Delta\chi_{ax}\left(3\cos^2\theta - 1\right) + \frac{3}{2}\Delta\chi_{rh}\left(\sin^2\theta\cos 2\Omega\right)\right] \quad (1)$$

where $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and the rhombic anisotropies of the magnetic susceptibility tensor, r_i, θ and Ω are the polar coordinates of nucleus i with respect to the orthogonal reference system formed by the principal axes of the magnetic susceptibility tensor. The pcs is defined as the difference between the chemical shift of a nucleus observed in the paramagnetic compound and that of the same nucleus in a hypothetical molecule deprived of the unpaired electron(s) but with the same structure. In practice, the diamagnetic reference shift is often taken equal to the chemical shift of the same nucleus in an analogous diamagnetic

compound. Here, pcs are taken as the difference in chemical shift between the system loaded with a paramagnetic Ln^{3+} ion and with La^{3+}, which is diamagnetic.

Pcs simulations and fittings were performed with a suite of programs written in-house[8]. Structural calculations were run with the PARA-DYANA package[77]**.** For modelling, the available solution structures of HAH1[2], MNK1[26], Calbindin[13], and of the adduct of the yeast homologous proteins[7] (Atx1:Ccc2) were used.

In particular, the model of the HAH1:MNK1 was built from the Atx1:Ccc2 adduct by superimposing the structure of HAH1 to that of Atx1 and the structure of MNK1 to that of Ccc2.

In PARA-DYANA calculations, all dihedral angles of HAH1, MNK1 and LBT were kept fixed, with the exceptions described in the following. This procedure is comparable to a pcs-driven rigid-body docking of the various polypeptides. The position of the Ln^{3+} ion was always kept fixed with respect to the LBT. For calculations on HAH1-G_1LBT where the LBT was allowed to move with respect to the protein, the backbone dihedral angles of the last amino acid of HAH1, of the linker Gly, of the first amino acid of the LBT were set as variable. This procedure was adopted to define the position of the LNT, and consequently of the Ln^{3+} ion, with respect to the protein. For calculations on HAH1-G_1LBT:MNK1, the two proteins were connected through a 226-residue linker. The LBT was blocked in the best-fit position resulting from calculations on HAH1-G_1LBT by fixing the values of the backbone dihedral angles of the three above-mentioned residues. The atoms in the linker had null Van der Waals radius. This procedure allows free reorientation of one protein with respect to the other, driven only by experimental data[7].

RESULTS

Production of HAH1-LBT and of the HAH1-LBT:MNK1 adduct

The three different constructs of HAH1-LBT with zero, one or two Gly residues between the protein and the LBT could be all purified with good yield. In all cases, the HSQC spectrum was essentially superimposable to that of wild-type HAH1, with the exception of the two most C-terminal residues (Figure 1).

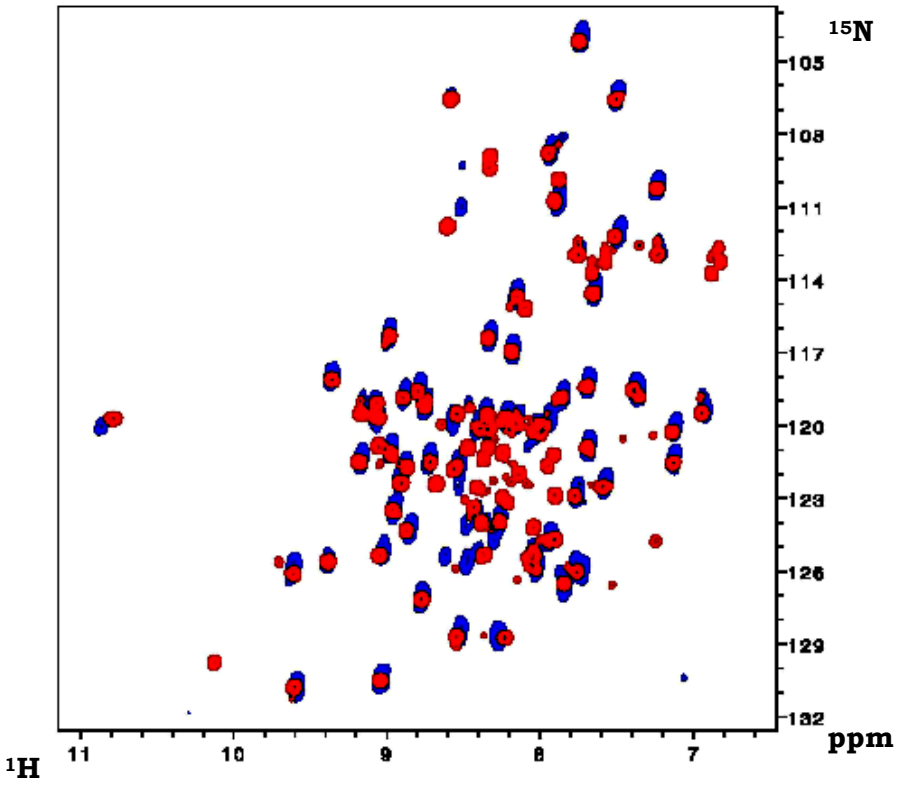

Figure 1 Superimposition of the [^{1}H,^{15}N]-HSQC spectra at 600 MHz, and 298 K in 100mM HEPES, 100mM NaCl, 1mM DTT pH 7 of
0.3 mM **apoHAH1-G$_1$LBT** (with one connecting Gly) and
0.3 mM **apo HAH1**

Copper(I)-binding capabilities were also unaffected. We found that all three constructs could bind Ln^{3+} ions. However, complete metallation of the tag in the HAH1 G_0LBT construct lacking any linking Gly residue

required an excess of Ln^{3+}. This was not the case for the other two constructs. Figure 2 shows the HSQC spectrum of apo- HAH1 G_1LBT in the presence of equimolar Tb^{3+}. Note that here and in the remainder of the paper apo- HAH1-G_1LBT refers to the protein devoid of copper(I).

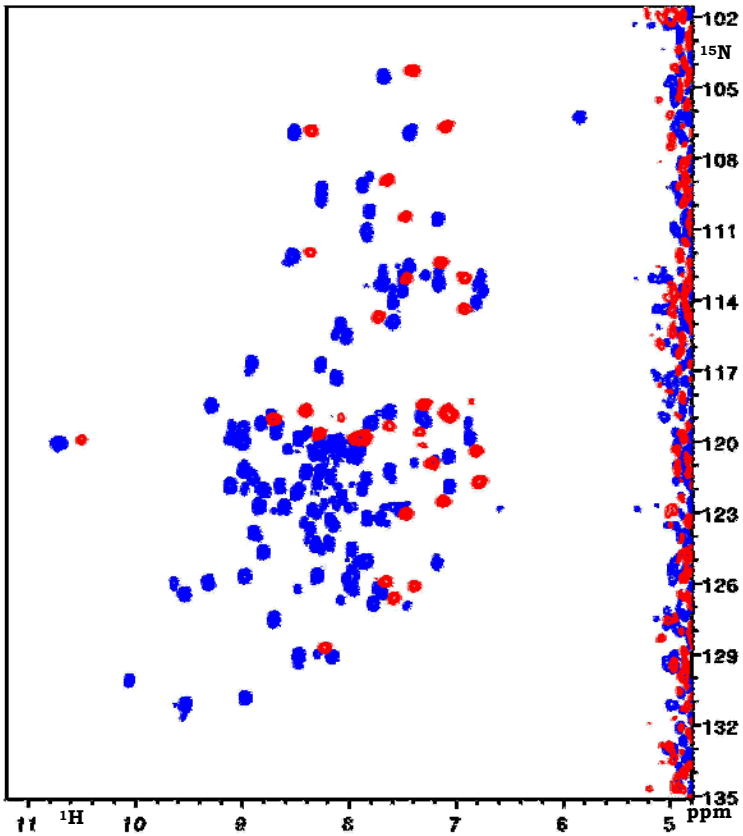

Figure 2 Superimposition of the [^{1}H,^{15}N]-HSQC spectra at 600 MHz, and 298 K in 100mM HEPES, 100mM NaCl, 1mM DTT pH 7 of
0.3 mM **apoHAH1-G_1LBT** (with one connecting Gly)
0.3 mM **apoHAH1-G_1LBT + Tb^{3+} 1:1** (in the presence of one equivalent of Tb^{3+})

As expected[5;82], binding of the higly paramagnetic Tb^{3+} ion induces broadening beyond detection of the majority of the signals from this small (68 amino acids protein). Indeed, this may be seen as a desirable feature when studying protein-protein complexes, due to the simplification of the spectra. On the basis of the above observations, we decided to use the construct with one Gly for all subsequent studies of the HAH1-LBT:MNK1 adduct, in order to keep the linker length minimal so to minimize the degrees of conformational freedom between HAH1 and the LBT. Note that the C-terminus of HAH1 is involved in the formation of the last strand of

the four-stranded β sheet of the protein, and thus its mobility is relatively restricted.

The HAH1-G_1LBT:MNK1 adduct was formed by mixing equimolar amounts of apo-HAH1-G_1LBT and copper(I)-MNK1. As already described for the yeast and a more complex human construct, the proteins do not form the adduct at amounts detectable by NMR in the absence of copper(I). Mixing the two proteins both enriched in ^{15}N allowed us to measure simultaneously pcs values for both systems, thereby avoiding potential pitfalls such as differences in ionic strength or pH in two different samples were one protein was enriched and the other not. This is made possible by the fact that most signals of HAH1-G_1LBT are not observable when the tag is loaded with the paramagnetic lanthanide ion. The lanthanides used were Tb^{3+}, Dy^{3+}, Er^{3+}, Tm^{3+}, whereas La^{3+} was taken as the diamagnetic reference. The various lanthanides induce pcs on the same nuclei that shift the signal left or right with respect to the position of the diamagnetic reference depending on the sign of the magnetic anisotropy of the metal (Figure 3). Magnetic anisotropies are known for the various lanthanide ions[82].

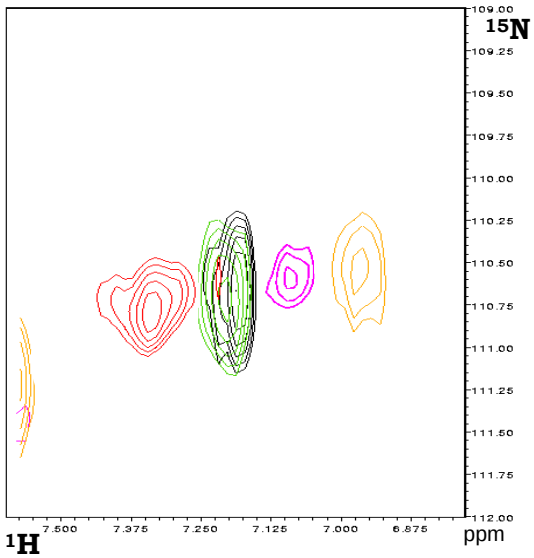

Figure 3 Variation in the position of the signal from the backbone amide group of Asn47 is HAH1-LBT:MNK1 samples loaded with different lanthanide ions (**La^{3+}, Tm^{3+}, Dy^{3+}, Er^{3+}, Tb^{3+}**)

Pseudocontact shift simulations for HAH1-G$_1$LBT

As mentioned, we collected pcs data for the HAH1-G$_1$LBT loaded with Tb^{3+}, Dy^{3+}, Er^{3+}, Tm^{3+} vs. the La^{3+} diamagnetic reference (Table 1).

Residue HAH1	**Dy^{3+}-La^{3+}**		**Tb^{3+}-La^{3+}**		**Tm^{3+}-La^{3+}**		**Er^{3+}-La^{3+}**	
	Chemical Shift		*Chemical Shift*		*Chemical Shift*		*Chemical Shift*	
	HN	***N***	***HN***	***N***	***HN***	***N***	***HN***	***N***
N 24	-		-0,128	-0,076	-		-	
L 26	-0,207	-0,145	-0,097	-0.041	-		0,025	0,029
G 27	-0,244	-0,061	-0,155	-0.040	-		0,017	0,051
G 28	-0,235	-0,221	-		0,005	0.067	0,015	0,011
V 29	-0,327	-0,091	-0,205	-0.001	0,080	0.161	0,077	0,294
L 35	-		-		-		0,007	0,143
K 39	0,056	0,003	-		-		-	
V 40	-		-		-		-0,079	-0,235
A 53	-		-0,142	-0.240	-0,008	-0.037	0,035	0,029
T 54	-0,275	-0,026	-0,091	-0.046	-		-	
T 61	0,037	0,089	-		-		-	
V 62	-		-		-		-0,046	-0,022
G 71	-		-		-		0,239	0,257

Table 1 The measured values of pcs for the resonaces of HAH1-G$_1$LBT protein loaded with Tb^{3+}, Dy^{3+}, Er^{3+}, Tm^{3+}.

The pcs data obtained could be fit to the HAH1 structure only assuming magnetic anisotropy values significantly smaller than those reported for Calbindin D_{9k}. For Dy^{3+}, the best-fit tensor values had $\Delta\chi_{ax}$ as small as $7.5 \cdot 10^{-32}$ m^3, vs. a reported value of $35 \cdot 10^{-32}$ m^3. Also noteworthy was the fact that the position of the Ln^{3+} ion was not uniquely defined. In fact two alternate positions could be identified, either when fitting one set of data at the time or simultaneously all sets for all paramagnetic ions introduced in the LBT. Letting PARA-DYANA optimise the position of the tag with fixed $\Delta\chi$ values also resulted in the same observation. It has to be noted however that the uncertainty on the tensor parameters and metal position determined is quite large, due to the fact that these parameters are all optimised simultaneously, at variance with what happens when a native, high-affinity binding site exists in the protein[8]**.** The present findings suggested to us the possibility that in solution there is relevant reorientation of the LBT as a whole with respect to HAH1. The pcs data

alone did not allow us to ascertain whether the LBT is in fast equilibrium between two different conformations or is exsploring a larger variety of conformations. Indeed, literature data already hinted at LBT's to be quite flexible with respect to the protein they are fused to. This is in agreement with the smaller than expected rdc values measured for ubiquitin bound to the same LBT used here[82] as well as from the measurement of order parameters reported recently for ubiquitin fused to the LBT or to a double LBT[54].

To obtain more insight into the problem, we simulated the effect of the reorientation of the tag as a rigid body with respect to the HAH1, again taken as a completely rigid structure. This was done by generating 2,000 structural models of HAH1-G_1LBT without any pcs constraint. This procedure in practice generates all possible conformations that are allowed given the length of the flexible part of the polypeptide chain and the steric hindrance of both HAH1 and the LBT. Pcs were calculated for all conformations assuming the orientation of the tensor with respect to the LBT was constant, and then averaged. The pcs values simulated for the HAH1 atoms in this dynamic situation were significantly reduced in magnitude with respect to what predicted if only a single conformation was populated in solution (with reduction factors up to a ten-fold or more for residues of HAH1 at intermediate distance from the LBT). Moreover, when looking at individual conformations one by one in several cases changes in the sign of the pcs could be observed, due to regions of HAH1 passing from the regions of space corresponding to positive pcs to the regions corresponding to negative pcs or viceversa.

We also evaluated the impact of this on our capability of safely fitting the pcs data to identify the position of the metal with respect to HAH1 and the orientation of the tensor, when assuming different constant values for the magnetic anisotropies. An example of these results is given in Figure 4.

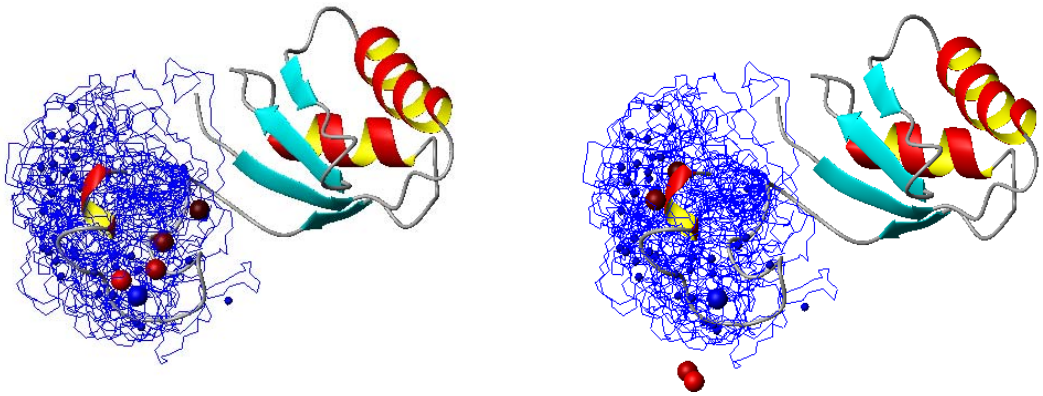

Figure 4 Fitting of the position of the lanthanide ion with respect to HAH1 assuming known magnetic anisotropies in the case of (left) no mobility of the LBT (only the first conformation was included in simulation, with the metal position indicated by the blue sphere) or of (right) large conformational span of the LBT (the first 50 conformations simulated are shown as a thin line). The pcs values were simulated with $\Delta\chi_{ax}$ = 28.3·10^{32} m^3 and $\Delta\chi_{rh}$ = 7.5·10^{32} m^3, i.e. an anisotropy in between that of Dy^{3+} and Ho^{3+}. Fittings were then run assuming $\Delta\chi_{ax}$ values of 28.3, 24.5, 14.1, 5.7·10^{32} m^3 (respectively: bright red, dark red, light brown, dark brown) and $\Delta\chi_{rh}$ values of 7.5. 1.5, 3.75, 0.75·10^{32} m^3 (respectively: bright red, dark red, light brown, dark brown).

It is readily observed that the fitting procedure, which allows some tolerance for deviations between best-fit and input (i.e. simulated) pcs values, readily allows the localization of the metal close to the real position (within 1 Å) when the correct magnetic anisotropies are assumed. The analysis of the residual deviations also permits the correct identification of the choice of magnetic anisotropy values. On the other hand, when the conformational variability is taken into account, an average position of the metal cannot be identified safely and the fitting procedure is very unstable. Magnetic anisotropy values significantly smaller than those used for the generation of the simulated data tend to yield the lowest residual deviations, although this depends in part on the orientation of the magnetic anisotropy tensor within the LBT, which is different for different Ln^{3+} ions[14;82].

Pseudocontact shift data for HAH1-G$_1$LBT:MNK1

Pcs data were measured for the HAH1-G$_1$LBT:MNK1 adduct loaded with Tb^{3+}, Dy^{3+}, Er^{3+}, Tm^{3+} (Table 2).

Residue MNK1	Chemical Shift HN			
	Dy^{3+} -La^{3+}	Tb^{3+} -La^{3+}	Tm^{3+}- La^{3+}	Er^{3+}- La^{3+}
V3	0,031	0,031	-0,015	-
S5	-0,015	-	-	-
V6	-0,032	-	-	-
T7	-	-	-	0,015
I8	-0,053	-	-	-
S9	-	-	0,054	0,054
V10	-	0,039	-0,025	-
E11	-	-	-0,09	-0,081
C15	-	-	-	0,018
T21	-	-	-	-0,019
I22	-	0,022	0,030	0,043
Q24	-	-	-0,014	-0,02
G27	-	-	-0,015	-
K28	-	0,022	-0,010	-
V29	-	0,010	-0,017	-
N30	-	0,026	-0,017	-
G31	0,015	0,015	-	-
V32	-	0,018	-0,013	-
H34	-0,009	-0,007	-0,009	-
I35	-	0,035	-	0,039
K36	-	-	-	0,006
V37	-	-0,041	-0,042	-0,087
L39	0,026	0,020	-	-
E40	-	0,012	-	-
E41	0,039	-	-0,019	-0,024
K42	0,064	0,027	-0,008	-
N43	-	-	-0,016	-
I46	-0,035	-	-	0,01
I47	-0,024	-0,016	-	-
Y48	-0,021	-0,006	-	-
L52	-	-	-0,015	-
T54	-0,013	0,023	0,012	-0,022
T57	-	0,010	-	-
K56	-0,066	-	-0,015	0,006
D64	-0,044	-	-	-
M65	-0,040	-	-	-
G66	-0,037	-	-	-
I71	-	-	-	0,073
H72	-0,014	0,005	-	-
N73	-	-	-	0,009

Table 2 The measured values of pcs for the resonaces of MNK1 for the HAH1-G$_1$LBT:MNK1 adduct loaded with Tb^{3+}, Dy^{3+}, Er^{3+}, Tm^{3+}.

Figure 5 shows a mapping of the measured pcs values onto the model structure of the HAH1-G_1LBT:MNK1 adduct for two different lanthanide ions.

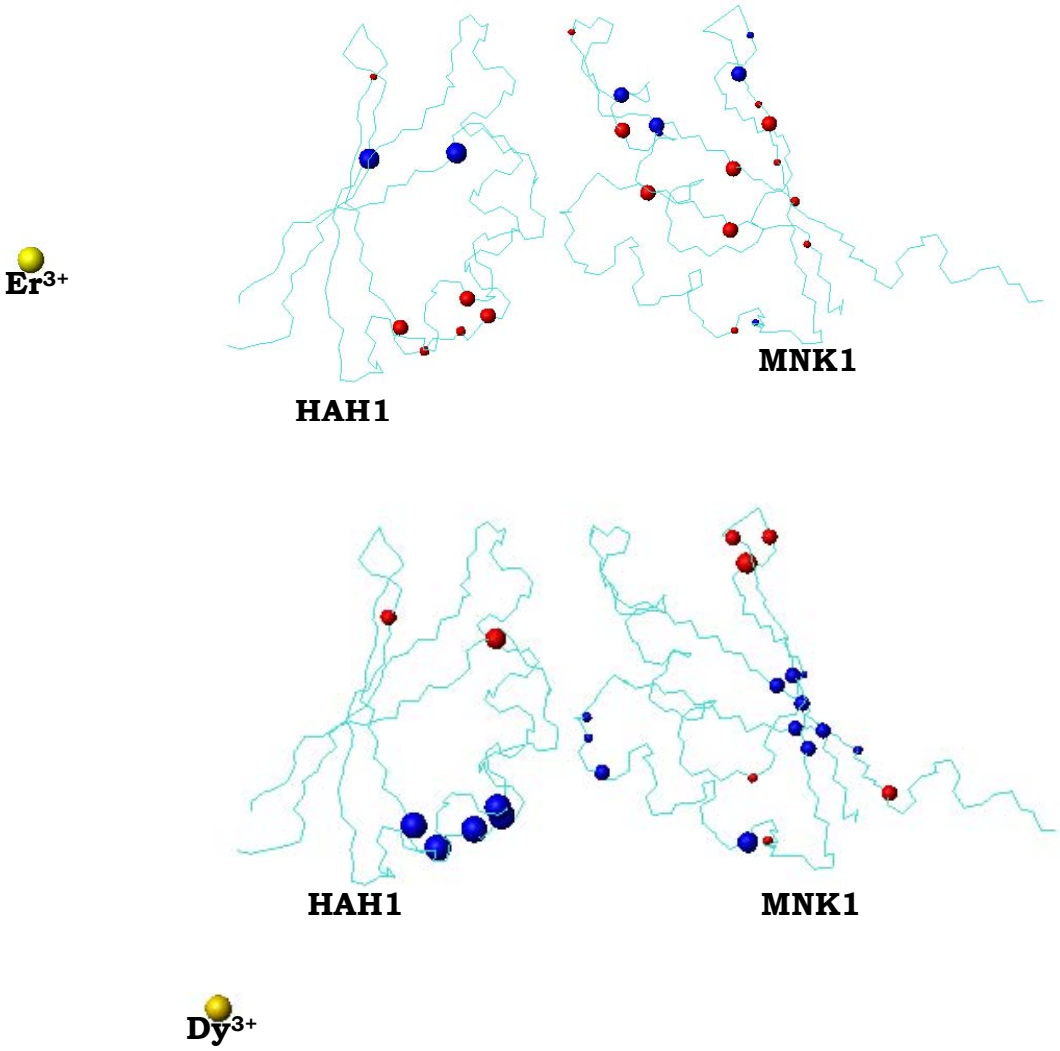

Figure 5. Pseudocontact shift values measured for the HAH1-LBT:MNK1 adduct mapped onto a structural model of the HAH1:MNK1 complex. Amide groups are indicated for which a pcs value could be measured are marked by spheres with a radius proportional to the average absolute value of the pcs measured. Red and blue spheres indicate positive or negative pcs, respectively.

While some residues at the interface could not be analysed due to line broadening induced by formation of the complex (presumably because of exchange averaging at the interface[6]), it is readily observed that residues with pcs values of same sign and similar magnitude tend to cluster close in space. After fixing the conformation of HAH1-G_1LBT to either one of

those detected through the fitting of the HAH1 data alone, rigid-body docking calculations were run using PARA-DYANA. In practice, we let the program optimize the position of MNK1 with respect to HAH1-G_1LBT based on pcs data alone, using also fixed $\Delta\chi$ values. The contribution to the target function of PARA-DYANA due to pcs data was strongly overweighed with respect to that of Van der Waals interactions, in order to compensate for the imperfect conformation of HAH1 and MNK1 side chains. We therefore obtained two different bundles of structures for the adduct, differing for the conformation of the LBT, i.e. for the position of the Ln^{3+} ion with respect to HAH1. Only calculations starting with one of the two HAH1-G_1LBT conformations resulted in a structure of the adduct very similar to that expected on the basis of homology modeling (Figure 6).

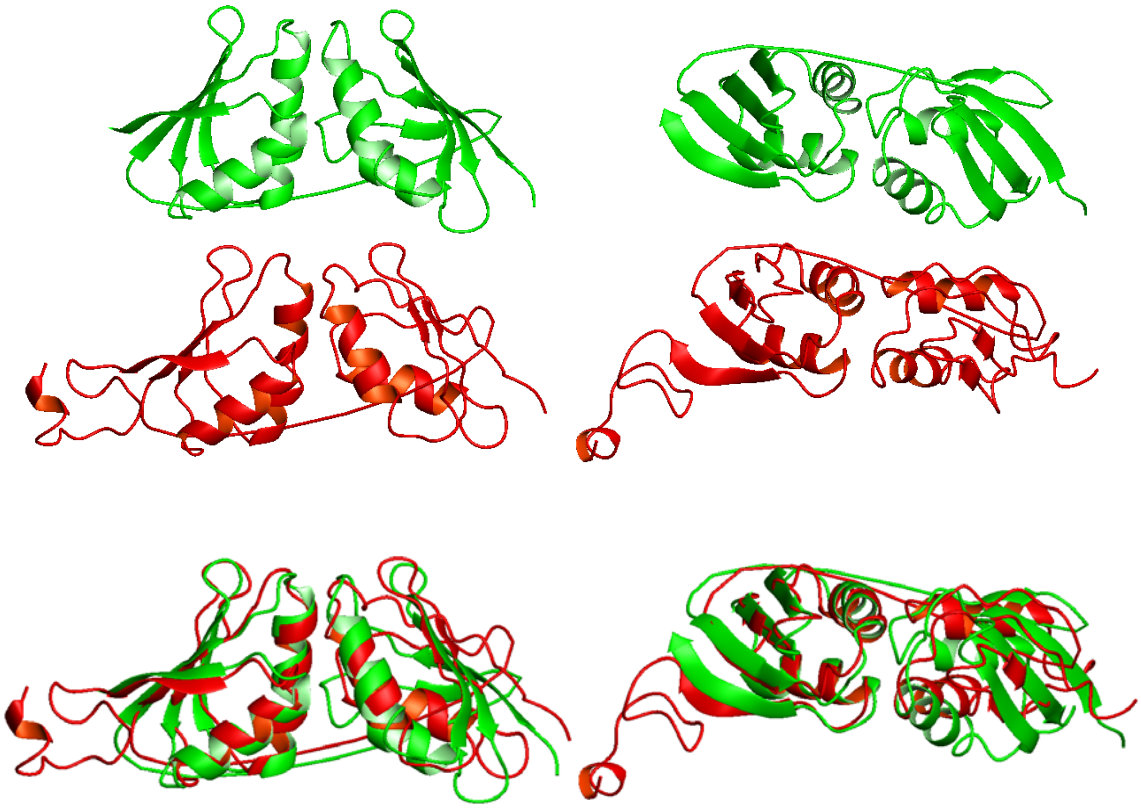

Figure 6. Comparison of (red) the mean structure of the HAH1-G_1LBT:MNK1 adduct and of (green) the structure of the HAH1:MNK1 adduct obtained from homology modelling. The structures have been superimposed on the backbone of HAH1, which is the protein on the left.

Notably, this adduct had a better agreement with the pcs data than the adduct resulting from calculations with the other conformation of HAH1-G_1LBT.

Simulations similar to those described in the preceding section were applied to the homology model of HAH1-G_1LBT:MNK1. Also in this case, we observed that the conformational averaging reduced, to an extent that is again dependent on the orientation of the magnetic anisotropy tensor, the magnitude of the induced pcs. Remarkably however, the interface residues and their neighbouring regions typically have chemical shift values of the same sign and close in value in the two proteins. This is because at some distance (more than 30 Å) from the LBT tag, which in the present system is fused at the C-terminus of the protein i.e. at the opposite side of the protein with respect to the protein-protein interface, the conformational flexibility of the LBT results in a similar averaging of the pcs for the nuclei. Thus, the simulations validate the concept that a highly flexible LBT still allows the use of pcs data to generate reliable structural models of protein:protein adducts.

DISCUSSION

The use of artificial lanthanide-binding sites to aid the determination of the solution structure of proteins and, particularly, of protein-protein and protein-ligand adducts has been actively pursued by several research teams [27;79;81]. This is due to the array of highly informative effects that can be induced by the paramagnetic Ln^{3+} ions in the system, which, in turn, can be captured with relative ease through NMR spectroscopy. These effects include mainly pseudocontact shifts (which result from the dipolar interaction of the nuclear spins with the magnetic moment of the lanthanide), residual dipolar couplings (which result from the partial alignment of the system in the magnetic field), and nuclear relaxation rate enhancement (also resulting from the dipolar interaction of the nuclear

spins with the magnetic moment of the lanthanide). All three effects can be exploited for structural purposes, as it has been extensively documented in the literature[4;17]. Various means of introducing the artificial lanthanide-binding site into the protein of interest have been developed and evaluated. One which is particularly straightforward is fusing a lanthanide-binding short (20 residues or less) polypeptide at one of the protein termini through protein engineering methods. These peptide stretches have been dubbed "lanthanide-binding tags" (LBT's)[31;32;54;55;58;67;82].

Here we have focused on the possible use of pcs resulting from the use of an LBT described in the literature for structure determination purposes. Heterologous expression and purification of the fusion protein were performed using essentially the same protocol as adopted for the wild-type protein. The introduction of the LBT is thus indeed very straightforward from the protein manipulation point of view. Lanthanide ions could also be titrated in the system with ease, as shown both by fluorescence measurements (thanks to the presence of the indole side chain of a Trp in the LBT[31;51]) and NMR (Figure 2). Insertion of the more highly paramagnetic Ln^{3+} ions, which is mandatory to measure pcs at the long-distances which are required for the characterization of a protein-protein interface while making sure that the LBT does not interfere with the inter-protein interaction, had also the advantage of providing spectra which are devoid of many signals of the fusion protein. This in turn allowed us to measure pcs values for both proteins in an adduct where both partners were enriched in ^{15}N, so that a complete set of data can be collected from an individual sample.

As shown by the present data, and in agreement with previous literature[82], the present fusion approach may allow significant conformational mobility of the LBT with respect to the protein of interest (HAH1 in our case). This makes the pcs values measured somewhat smaller than what one would expect in a rigid, but otherwise identical, system. Based on steric considerations alone, the conformational freedom of the LBT can be large enough to allow some nuclei of the protein to experience both negative and

positive contributions to the pcs, which thus tend to vanish. Under these conditions, fitting the observed pcs values to a magnetic anisotropy tensor, as normally done for proteins containing a native, rigid binding site for the paramagnetic ion (e.g. ferricytochromes, lanthanide-substituted calcium binding proteins), can yield to a significant underestimation of the magnetic anisotropy values and to an incorrect localization of the metal ion (Figure 4). The structural information provided by the pcs values thus becomes unreliable. Notably, this problem does not affect rdc values, which are simply scaled down by a factor related to the degree of freedom of the LBT with respect to the protein, and thus may be more difficult to measure precisely, but maintain their information content[82].

The mobility of the LBT influences also the pcs data measured for the partner in a protein:protein adduct, again reducing the magnitude of the effect. This makes their measurement more difficult and enhances the impact of the experimental error on the determination of the configuration of the adduct. Nevertheless, there is a clear indication that the data are still informative on the configuration of the adduct, i.e. on the relative position of the two proteins. This could be ascertained here as by using a homology model of the HAH1:MNK1 adduct as a "golden standard" to evaluate the results of docking calculations driven by pcs data. A chemical shift mapping of the HAH1:MNK1 interaction is already available, and confirms the correctness of the homology model[6]. Pcs-driven docking of MNK1 to HAH1-G_1LBT with PARA-DYANA required some variations with respect to the protocol followed e.g. for Calbindin D_{9k}. While in both protocols, the $\Delta\chi$ values are kept fixed, here to ensure convergence we also had to block the conformation of the LBT with respect to HAH1 in order to freeze the position of the paramagnetic Ln^{3+} ion. Because the fittings of the pcs data to the HAH1 structure showed that two different positions of the metal were in similar agreement with the data, we correspondingly run two parallel calculations. Simulated data suggested that when the tag is flexible, it may indeed by difficult to identify an optimal position for the metal ion (Figure 4). Importantly, this position does not correspond to a physically meaningful quantity, such as the average position of the

paramagnetic Ln^{3+} ion. Of the two families of structures that resulted from the two parallel calculations, one had better agreement with the pcs data, with all its members having a lowest target function than that the best member of the other. As shown in Figure 6, the average structure of this best family resembles closely the homology model of HAH1:MNK1. Also based on simulated data, we can thus conclude that notwithstanding the flexibility of the LBT, the pcs data measured for MNK1 are still informative enough to safely drive calculations and obtain with relative ease a structural model of a quality comparable to that obtainable from chemical shift mapping data[28].

As a concluding remark, we can highlight that the development of less flexible tags is a strong requirement to really boost the applicability of the technique. Some reports have already shown the usefulness of chemical tags that can be attached to pairs of protein side chains rather than to the side chain of an individual cysteine. Given the ease of use of fusion tags, as done here, another line of development would towards the use of bulkier tags, which can have more extensive contacts with the protein they are fused to and thus restricted flexibility. Initial results in this direction are also available in the literature[54] and are also being explored in our lab.

Acknowledgements

We thank Barbara Imperiali for the kind gift of a sample of a control protein fused to the LBT.

Chapter IV

Design and experimental validation of a lanthanide binding protein as a rigid tag

INTRODUCTION

Paramagnetism-based restraints, like pcs and rdc, have been demonstrated to represent a powerful tool to determine the relative position of two proteins for which such data have been obtained[4;17]. The advantages with respect to the information available with using external orienting devices to induce rdc are due to the fact that pcs are not available in the latter approach, and thus only information on the relative orientation, but not on the position, can be obtained in such cases. Furthermore, if the system experiences conformational variability, the paramagnetic pcs and rdc restraints can provide information on the preferred conformations sampled by the system, whereas such information is not achievable using external devices since the latter orient each protein or mobile protein domain independently. The paramagnetic restraints pcs and rdc are however available only if a paramagnetic metal is located inside the protein complex. This may be the case when one of the interacting proteins is a paramagnetic metalloprotein, or a metalloprotein where the diamagnetic metal ion has been substituted by a paramagnetic ion. In order to extend such strategy for the investigation of protein-protein complexes to systems without any metal ion inside, in the last years paramagnetic tags have been developed[29;38;54;75;82]. The most promising and investigated class of tags is probably constituted by the lanthanide binding tags, which are able to bind 11 different paramagnetic metal ions, corresponding to the paramagnetic lanthanide ions. The possibility to easily obtain pcs and rdc referred to more than one metal is quite important because allows for removal of the degeneracy of the solutions, intrinsic in single metal pcs and rdc measurements.

A possible drawback due to the use of paramagnetic tags is that the latter must be rigidly attached to the protein, without significant flexibility, to make possible to exploit pcs as a source of information on the position of a protein in a reference frame. A strategy to improve the rigidity of the tag is to increase the number of contacts with the protein to which the tag is

fused[38;75]. Another strategy is to restrict tag flexibility taking advantage of the steric volume occupied by tag and protein. Here we have developed a novel tag corresponding to the entire protein P43M CalbindinD$_{9k}$ (CABP), which we have fused to the first metal-binding domain of ATP7A protein (Menkes 1 or MNK1). CalbindinD$_{9k}$ is composed by two EF-hand motifs connected by a loop; each EF-hand motif can bind a calcium ion. The protein has been shown to have a rigid and well defined structure, and to be able to selectively bind any lanthanide ion in its second EF-hand motif[1;13;14].

Menkes1 has been fused to CalbindinD$_{9k}$ starting from residue Met 5 (CABP-MNK1); the first residues have been discarded due to their mobility. The structure of Menkes1 has already been solved. Our purpose in this work is to check whether the CABP tag is effective as a rigid tag, so that no significant mobility is present and pcs can be profitably used.

MATERIAL and METHODS

Cloning, expression and purification of CABP-MNK1

The chimeric protein (CABP-MNK1) was obtained from the fusion of the bovine CalbindinD$_{9k}$ P43M (4-79a.a.) (wild type CABP: NM174257, gi:110347578) with the human first metal-binding domain of ATP7A or MNK1 (a.a. 5-77) (NM_000052., gi:115529485).

The CABP cDNA sequence was fused by overlap extension PCR at the N-terminal region of the MNK1. A first PCR reaction was performed using the plasmid expressing bovine CalbindinD$_{9k}$ P43M as template, in order to obtain a PCR fragment where the 3'-region contained a flanking sequence homologous to the 5'-region of MNK1 (primer: Forward 5'-CACC-ATGAAATCTCCGGAAGAACTGAAAGGTATC-3', Reverse 5'-CAGAATTCACACCCATCTGAGAGATCTTTTTAACCAGAACCTGG-3').

Viceversa, the plasmid expressing human MNK1 was used as template, in order to obtain a PCR fragment, where the 5'-region contained a flanking sequence homologous to the 3'-region of CABP (primer: Forward 5'-

GTTAAAAAGATCTCTCAGATGGGTGTGAATTCTGTTACCATTTCTGTTG-3’ Reverse 5’-CTACTAATTATGGATAACAGCATCAAAGCCCATGTC-3’). In the second step, the fusion of the two proteins was obtained exploiting the complementary region and using the forward primer of CABP and the reverse primer of MNK1, and obviously, as template, both of the previous PCR fragments of the single gene (Fig.1). Finally, the chimeric gene was cloned in pDEST17 vector via Gateway system (Invitrogen).

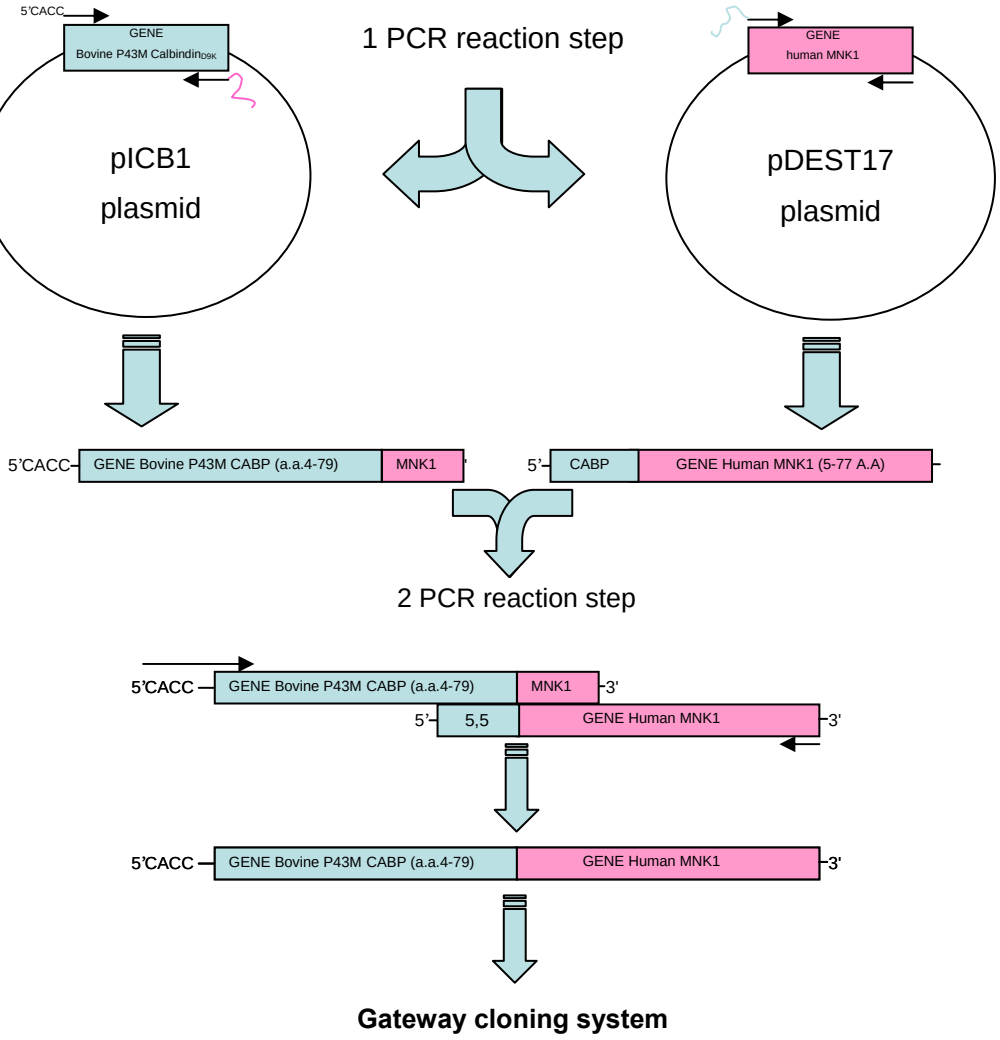

Figure1 Scheme of steps for CABP-MNK1 cloning (the line arrows rappresent primer)

The chimeric protein (CABP-MNK1) was expressed in ^{15}N labeled M9 minimal media in E. *coli* Rosetta pLysS (Novagen); the employed bacterial strain possesses a plasmid that encodes *E. coli* rare codons for arginine, isoleukine, leukine and proline

Cells were grown at 37°C until cell density measured at 600 nm reached the value of 0.6; protein expression was then induced with 0.5 mM IPTG and temperature changed to 17°C. After 14-16h cells were harvested by centrifugation and resuspended in lysis buffer (20mM Phosphate buffer pH7.2, 0,6M NaCl, 100µg/ml Lysozime, 2mM $MgSO_4$). Cell lysis was done by heat-shock treatment and 3 cycles of sonication; the suspension was finally centrifuged at 20000g for 30'. Protein solution was then loaded onto a HisTrap chelating FF column (Amersham GE Healthcare Life Science) equilibrated with 20mM Phosphate buffer pH7 2and 0,6M NaCl. Protein was eluted almost pure with a step-wise gradient in the fraction containing 200 mM imidazole. The His-tag was then cleaved with AcTEV protease at 25°C with over-night incubation. Uncut protein and cleaved His-tag were removed by a second step of purification using HisTrap chelating FF column.

NMR measurements

The protein samples in deionized water were prepared with buffer-exchange chromatography (H_2O) under reducing atmosphere in a glove-box chamber. The pH of the samples were adjusted to 6.0 by means of diluted solutions of NaOH and HCl.

Yb^{3+}, Er^{3+}, Tb^{3+} and La^{3+} ions have been added to the CalbindinD_{9k} tagged Menkes1 (CABP-MNK1) protein in order to replace the calcium ion present in the second metal binding motif of the CABP tag.

Diluted solutions of CABP-MNK1 in water at pH 6 were concentrated up to 80 µM in presence of D_2O (10%) for an NMR spectrometer lock.

Lanthanide-containing CABP-MNK1 samples were obtained by titrating the Ca^{2+} form of protein with stock solutions (30mM) of $LnCl_3$. The solutions were prepared by dissolving $LaCl_3$, $YbCl_3$, $ErCl_3$ and $TbCl_3$ in H_2O, adjusting the pH at 5 with diluted solutions of NaOH and HCl. Titrations were followed by 1D ^1H NMR and by 2D ^1H, ^{15}N-HSQC experiments using Bruker AVANCE spectrometer operating at proton frequencies of 500 MHz and at 298K. The NMR instrument was equipped with a cryo-probe.

RESULTS and DISCUSSION

Designer, cloning and expression of CABP-MNK1 construct

Several constructs were designed in order to reduce the linker length and to prevent any steric clash between the two proteins. The different constructs were then evaluated by performing energy minimization analysis of the relative structural models using the CYANA program. The models were generated starting from the PDB of the isolated CalbindinD$_{9k}$ (1KQV) and MNK1 (1KVJ) structures. In the calculation all the dihedral angles of CalbindinD$_{9k}$ and of MNK1 were keeped fixed and only the few aminoacids connecting the two proteins were left free to move. The model, costituited by a fusion of the CABP and the MNK1, showed that no sterical hindrance occurs between the two proteins by linking directly the C-terminal of CABP (a.a 4-79) to the N-terminal of MNK1(5-77a.a.). (Fig.2). The sequence of the selected construct is reported below.

MKSPEELKGIFEKYAAKEGDPNQLSKEELKLLLQTEFPSLLKGMSTLDELFEE
LDKNGDGEVSFEEFQVLVKKISQMGVNSVTISVEGMTCNSCVWTIEQQIGKV
NGVHHIKVSLEEKNATIIYDPKLQTPKTLQEAIDDMGFDAVIHN

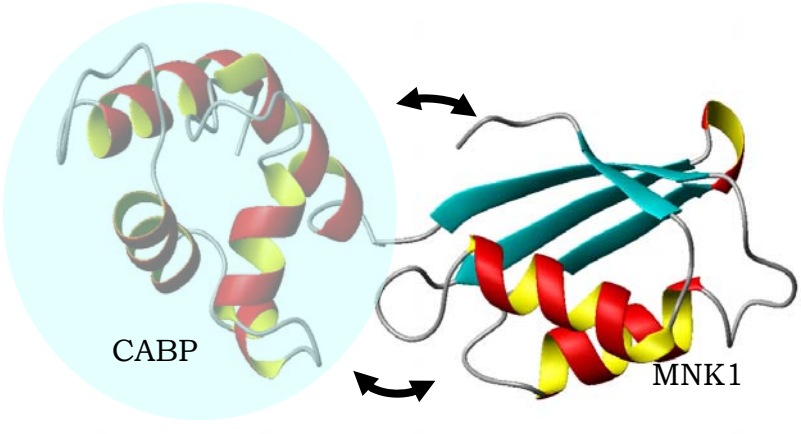

Figure 2 The CYANA model of the cloned CABP-MNK1 chimeric protein

CABP-MNK1 was expressed and purified using the same protocol of wt MNK1 protein both for the natural and the ^{15}N isotopically enriched samples, since the conjugation of the two polypeptides does not affect their biophysical and biochemical properties with respect to them as separate entities.
The final yield of purified protein was 15 mg/L.

NMR characterization of CABP-MNK1

The CABP-MNK1 yielded a $^{1}H,^{15}N$-HSQC spectrum of good quality corresponding to a well folded protein (Fig.3).

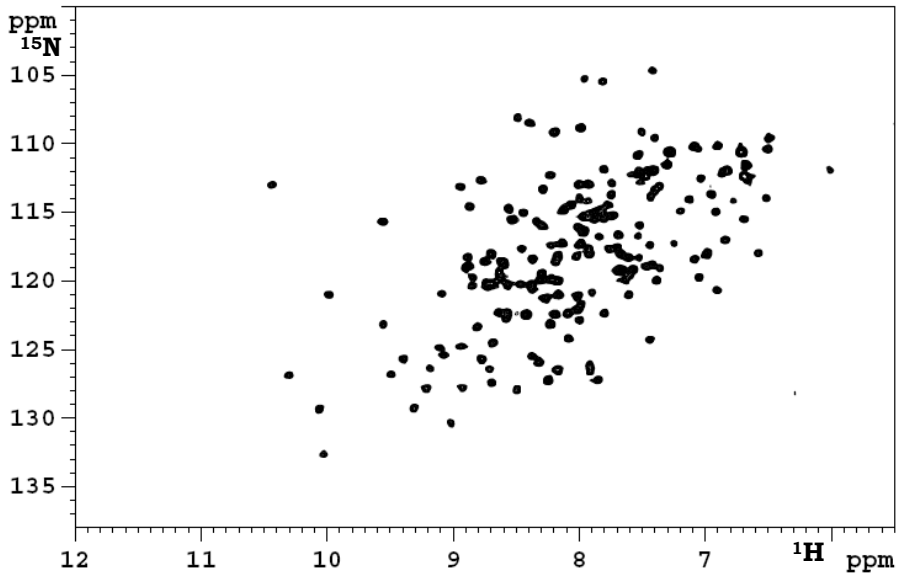

Figure 3 $^{1}H,^{15}N$-HSQC Spectrum 500 MHz, 298 K, H_2O pH6 of the CABP-MNK1 protein

The comparison of spectra with those of the isolated CABP and MNK1 acquired in the same experimental conditions, showed that only few peaks are shifted in the CABP-MNK1 adduct (Fig.4).

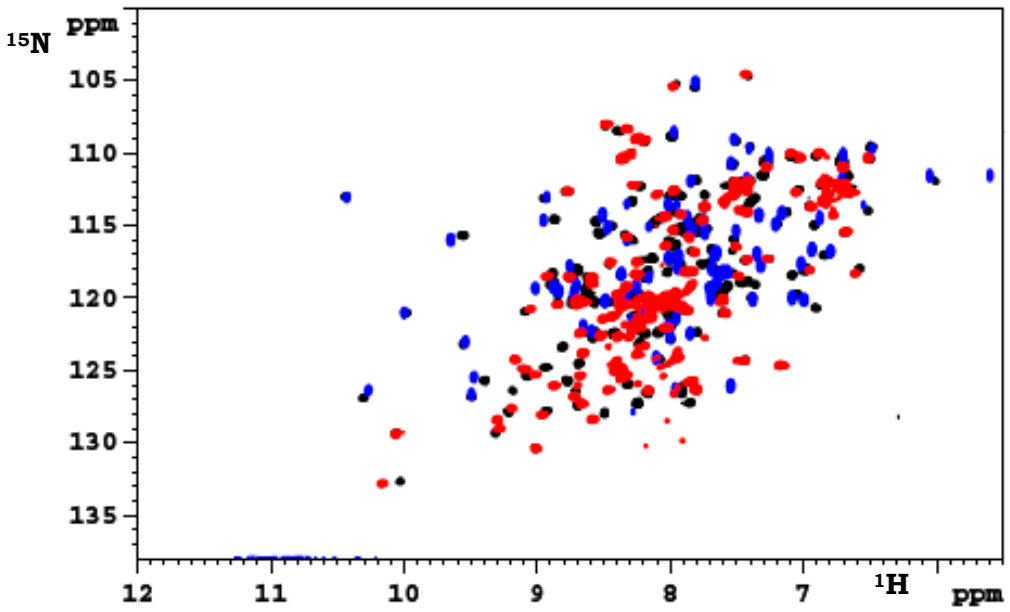

Figure 4 The superimposition of $^{1}H,^{15}N$-HSQC Spectra at 500 MHz, 298 K, H_2O pH6 of **0,3 mM CABP-MNK1**
0,3 mM MNK1
0,3 mM CABP

On the basis of the previous experience with CalbindinD$_{9k}$, Yb^{3*} was initially chosen as paramagnetic metal ion in order to verify the feasibility of the metal substitution in this chimeric protein. Protein precipitation, sometimes associated with the presence in solution of the Ln^{3+} ions, was prevented by working with diluted protein solution (80μM) at pH 6 in deionized water without any reductant.

It is known from literature that a stoichiometric Ln^{3+}/Ca^{2+}substitution occurs in CABP protein.

Also for the diamagnetic protein, the different experimental conditions, adopted for metal substitution with respect to that used for the protein assignment (see BRMB database) made the re-assignment of MNK1 domain by visual inspection difficult. Conversely, the re-assignment of CABP was easily performed thanks to the availability of spectra recorded in the same conditions.

The metal replacement was verified by monitoring the chemical shift alterations of the CalbindinD$_{9k}$ resonances. Shifts up to 1.5 ppm were detected in the 2D HSQC spectra.

The good correspondence found between the spectra of Ca^{2+}-Yb^{3+}CABP and Ca^{2+}-Yb^{3+}CABP-MNK1(Fig.5) is a clear indication that the new construct does not alter the metal binding capability, and more in general the structure of the CalbindinD_{9k}.

The comparison between the Ca^{2+}-Yb^{3+}CABP-MNK1 and the wild-type MNK1 showed that also several peaks of the latter protein were shifted as a consequence of the presence of the paramagnetic metal ion bound to the CABP domain. Some paramagnetic peaks of the CABP domain were re-assigned on the basis of the spectra recorded on the isolated domain. At variance with the MNK1 portion, as mentioned above, the assignment was largely prevented by the different conditions adopted to perform the metal replacement with respect to the ones used in the assignment of the isolated domain. The full assignment of the resonances of the MNK1 spectra is currently in progress.

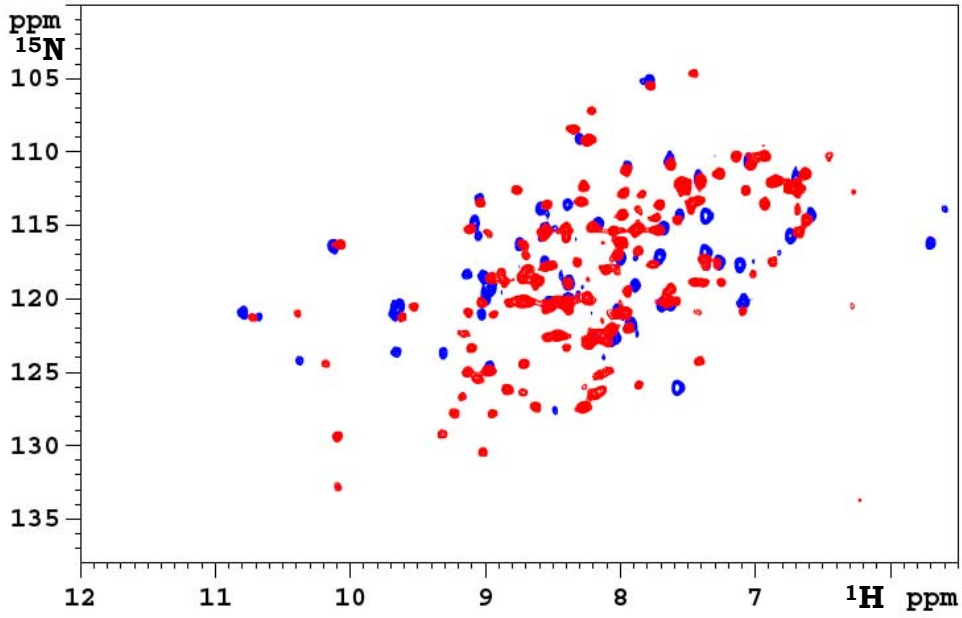

Figure 5 The superimposition of 1H,15N-HSQC Spectra at 500 MHz, 298 K, H_2O pH6 of

80 µM Yb^{3+}-Ca^{2+}CABP-MNK1

80 µM Yb^{3+}-Ca^{2+}CABP

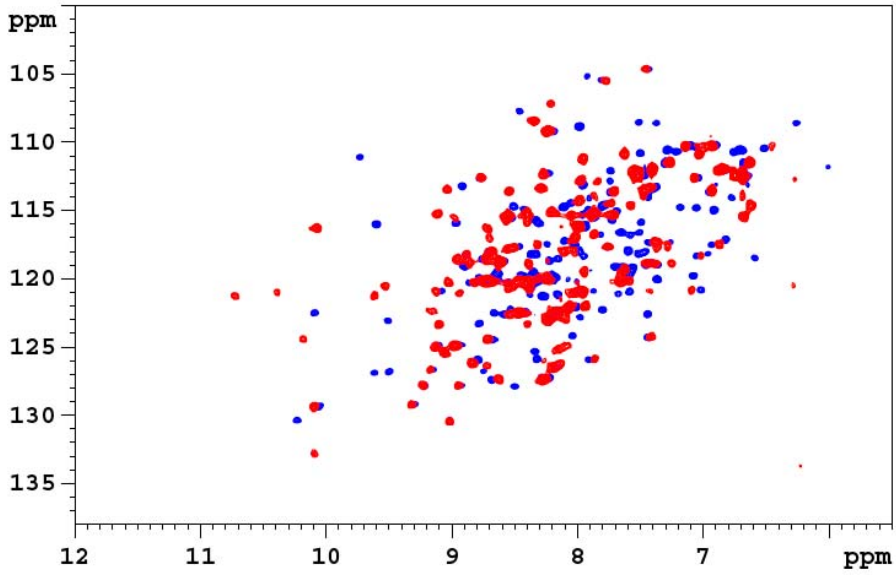

Figure 6 The superimposition of 1H,15N-HSQC Spectra at 500 MHz,298 K, H_2O pH6 of

80 μM Yb^{3+}-Ca^{2+}CABP-MNK1

80 μM La^{3+}-Ca^{2+}CABP-MNK1

In order to determine the magnitude of pcs the diamagnetic analogue of CABP-MNK1 was prepared by titrating the protein with La^{3+} and the metal replacement was monitored by $^{1}H,^{15}N$-HSQC spectra (Fig.6).

Also Er^{3+} and Tb^{3+} were incorporated in CABP-MNK1 with use of this procedure and the pcs for the assigned resonances of CABP-MNK1 were collected (Table 1). Pcs were calculated from the difference of the chemical shifts in the paramagnetic sample and those in the diamagnetic samples.

Residue	DOMAIN	CHEMICAL SHIFT HN		
		Yb^{3+}	Er^{3+}	Tb^{3+}
G 8	CABP	-0.038	-	-
E 17	CABP	0.480	-	-
G 18	CABP	0.113	-	-
K 25	CABP	1.515	-	-
K 29	CABP	1.288	-	-
L 31	CABP	0.934	-	-
Q 33	CABP	0.590	-	-
T 34	CABP	0.455	-	-
S 38	CABP	0.248	-	-
G 42	CABP	0.246	-	-
S 44	CABP	0.879	-	-
T 45	CABP	0.335	-	-
L 46	CABP	0.513	-	-
L 49	CABP	1.295	-	-
F 50	CABP	1.512	-	-
L 69	CABP	0.050	-	-
S 80	MNK1	0.037	0.201	-0.600
V 81	MNK1	0.029	0.200	-0.478
V 85	MNK1	0.03	0.154	-0.390
G 87	MNK1	-0.044	0.052	-0.179
C 93	MNK1	-0.03	0.066	-0.354
Q 102	MNK1	0.025	0.149	-0.390
K 103	MNK1	-0.016	0.078	-0.025
G 106	MNK1	-0.025	0.094	-1.001
H 108	MNK1	0.026	0.190	-1.147
V 112	MNK1	0.000	0.179	-0.325
L 114	MNK1	0.000	0.050	-0.227
I 121	MNK1	0.017	0.180	-0.483

Table 1 The pcs for the assigned resonaces of CABP domain in the CABP-MNK1

The pcs values observed for CABP residues have been used to obtain the magnetic susceptibility tensor for each paramagnetic metal; the pcs values observed for Menkes1 residues, on the other side, provide the position and orientation of the magnetic susceptibility tensors referred to such protein, so that the relative position of the two proteins can be obtained from the superposition of the tensors themselves.

The pcs observed for CABP residues in the Yb^{3+} derivative has been fit using the $CalbindinD_{9k}$ structure (1KQV PDB structure) to obtain the magnetic anisotropy values $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, which have been calculated to be -9.2×10^{-32} and 3.1×10^{-32} m^3, respectively. Such values are in agreement with values already obtained for Yb^{3+}-Ca^{2+}-CABP. The quality of the fit (see Fig. 7) is satisfactory.

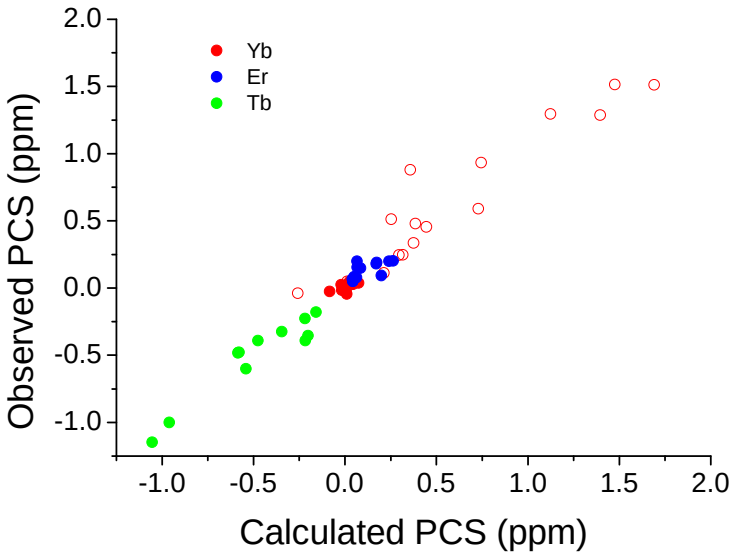

Figure 7. Observed PCS values for Yb (red), Tb (green) or Er (blue) versus PCS values calculated using the protein conformation shown in Fig. 8.A. PCS relative to CABP residues are shown as empty symbols.

The fit of the pcs values measured for the Menkes1 residues has been performed by fixing the anisotropy values to those obtained from the CABP tag residues and leaving as fitting parameters the three Euler angles defining the orientation of the magnetic susceptibility anisotropy tensor and the position of the metal ion. The fit in this case does not clearly

provide a single, well defined, solution since several positions for the metal ion are possible with about the same error in the fit. This unlike result is due both to the fact that a degeneracy in the solution is intrinsic in the problem if data referred to one metal ion only are used, and to the fact that the observed pcs are quite small and all similar among them. All the calculated solutions, however, indicate that the CABP tag is bent with respect to the Menkes1 structure, i.e. it is not attached in an extended position. The distances calculated for the metal ion are in fact similar to those which can be modeled from the conjunction of the CalbindinD$_{9k}$ and Menkes1 structure, pointing out that there has not been extensive averaging due to relative motion between the CABP tag and the Menkes1 protein.

The same calculations have been performed using pcs observed for Er^{3+} and Tb^{3+}. Also in these cases similar results are obtained, with metal ion positions quite close with respect to the Menkes1 structure (much closest with respect to the maximum distance allowed by an extended conformation) indicating that a solution common to all data relative to the three metals can be achieved. Analogously, if the fits on the Menkes1 residues are performed by fixing the position of the metal ion to an averaged distance among those allowed by the fusion protein, magnetic susceptibility anisotropy values larger than those expected are calculated, pointing out that motional averaging is not effective in reducing the observed pcs data.

The conformation of the CABP tagged Menkes1 protein was thus calculated through the program PARAMAGNETIC DYANA, by fixing the structures of both the CalbindinD$_{9k}$ and the Menkes1 domains and using pcs to determine their relative position. The structure of CABP was fixed to the NMR 1KQV PDB structure, that of Menkes1 was fixed to the NMR 1KVJ structure. All dihedral angles were kept fixed during the simulated annealing protocol except those at the connection between the two domains, i.e. the psi angle of GLN 76 of CABP and the phi and psi angles of the following MET residue of Menkes-1, corresponding to MET 5. Three pseudoresidues representing the magnetic anisotropy tensors and the

metal ions corresponding to Yb^{3+}, Er^{3+} and Tb^{3+} were added. Restraints were introduced in the calculation to force these metal ions to be placed in the same position and to be coordinated to the CABP ligands. The magnetic susceptibility anisotropies were fixed according to the values obtained from pcs of CABP residues (for Yb^{3+}) or to the values available in literature[14] as calculated from pcs of Ln-substituted CABP (41.7×10^{-32} and -11.5×10^{-32} m^3 for Tb^{3+}, and 12.2×10^{-32} and -7.5×10^{-32} m^3 for Er^{3+}).

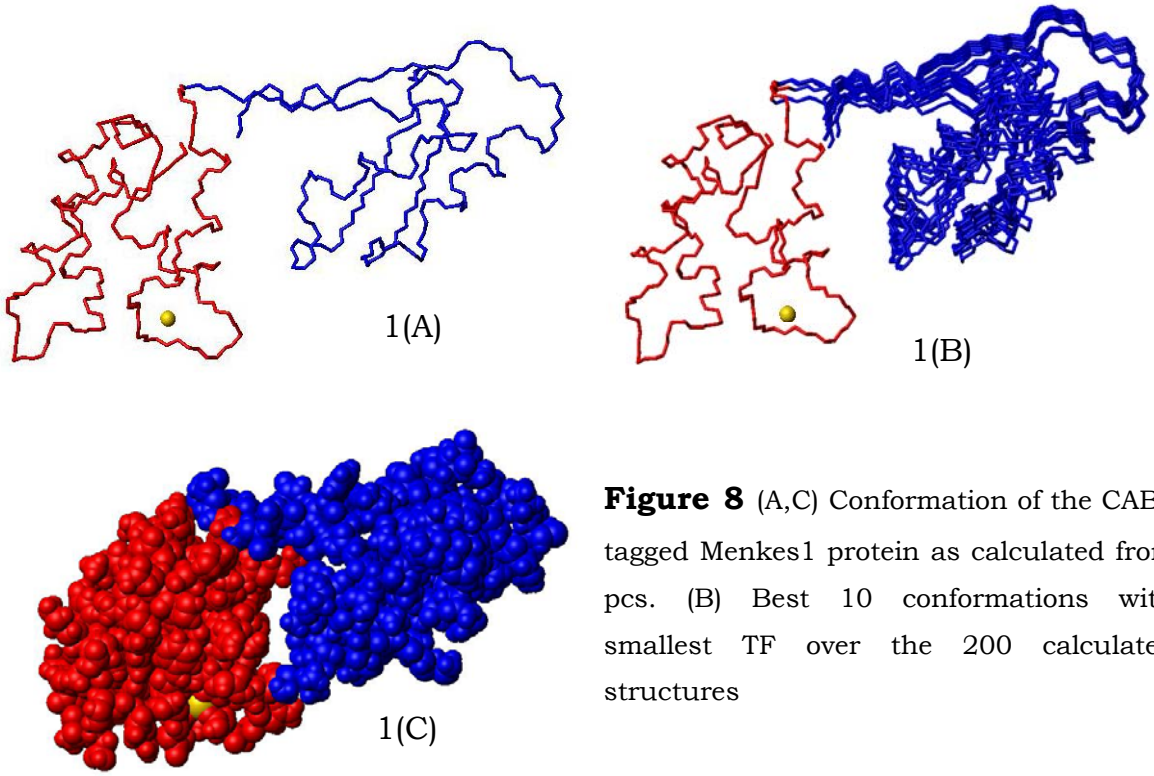

Figure 8 (A,C) Conformation of the CABP tagged Menkes1 protein as calculated from pcs. (B) Best 10 conformations with smallest TF over the 200 calculated structures

The lowest TF structure (3.42) is reported in Fig. 8A, C. Structures in a slight different position are obtained with TF in the range 3.45-3.61 (Fig. 8B). The indetermination in the position of the Menkes1 structure with respect to the $CalbindinD_{9k}$ structure thus results to be actually modest, indicating the low flexibility of the tag. The agreement between calculated and observed pcs values is shown in Fig. 8.

Relative rigidity of the fused protein supports the use of $CalbindinD_{9k}$ as a new tag to insert paramagnetic lanthanides in diamagnetic proteins so permitting the use of additional experimental restraints in structural characterization and in protein-protein interaction studies.

Chapter V

SUMMARY and PERSPECTIVES

Genome sequencing projects have provided researchers with an unprecedented number of molecular information that promises to revolutionize our understanding of life and lead to new treatments of its disorders. However, genome sequences alone offer only limited insights into the biochemical pathways that determine cell and tissue function. These complex metabolic and signaling networks are largely mediated by proteins. The vast number of uncharacterized proteins suggests that our knowledge of cellular biochemistry is far from complete. In the postgenomic era, structural biology becomes a central discipline for the explanation, linking and exploitation of biological data in the life sciences. Indeed a large number of different scientific fields like bioinformatics, biochemistry and molecular biology are necessary in structural biological studies.

In these three years three different projects were been developed with the common goal devoted to the investigation of metalloproteins using a multidisciplinary approach. In particular our attention was focused on two different area: the genomic analysis and the structural proteomic studies through the paramagnetic approach.

In the first project we focused our attention on the degree of Single Nucleotide Polymorphism (SNP) in metalloproteins, since several studies have demonstrated the relationship between DNA sequence variations, as mutations or SNPs, and diseases[69]. In particular our research was devoted to analyse the human somatic Cytochrome c gene. Indeed the SNP density of the coding region of this gene (CYCS), reported by dbSNP database, is 1 order of magnitude higher than the average SNP density. From the evolutional point of view, such variability could be the indication of an exceptional propensity of the human Cyt *c* protein to individual plasticity, although one detailed analysis of these data has been evidenced the possible existence of Cyt *c* variants (cSNP) affecting biological functions without being incompatible with life.

Since all these cSNPs were obtained by computational methods only[20], the aim of our research was the experimental validation of these putative

cSNP. Surprisingly, after our project of resequencing none of the all putative cSNPs, reported by dbSNP, was found in our individuals panel (190 chromosomes) and no other new SNPs were found.
Moreover, our re-examination of all SNPs deposited in dbSNP by the same computational methods (EST-based), used for the prediction of snps in CYCS gene, casts serious doubts on the validity of this bioiformatic approach. It is likely that it is not the computational approach itself that is faulty, but rather the reliability of the raw data, i.e. of the deposited EST sequences.
In conclusion, quality of the EST sequences is not good enough to discriminate a real SNP from an error and the data, provided from the computational methods EST-based, are not reliable.
Moreover, from our resequencing of the human CYCS gene in a 95-individuals panel we derived a likely monomorphism for the ORF of this gene in the human population, resulting in the absence of allelic variants of the Cyt *c* protein.

In the second part of our research the attention was focused on the employment of several paramagnetic tags to obtain structural informations of protein complexes.
In this frame, we initially exploited a peptide (LBT). The LBT sequence was attached to the HAH1 protein and several constructs were expressed. In this research, we used the adduct formed by the copper(I)-chaperone HAH1 and the first metal-binding domain of the human ATPase ATP7A (MNK1) as a model system to investigate the efficiency of the LBT tag in the analysis of protein-protein interactions. The insertion of the LBT on HAH1 did not interfere with the formation of the metal mediated adduct. Differen lanthanide metal ions (Er^{3+},Tb^{3+}, Tm^{3+}, $Dy^{3+}$and La^{3+}) were incorporated in the LBT-HAH1-(CuI)-MNK1 complex. The pcs data reveled a significant conformational freedom of the LBT with respect to the HAH1 protein. In fact the values of the magnetic susceptibility anisotropy tensor calculated from PCS were significantly smaller than those expected. Despite these disappointing results, the PCS measured on copper(I)-MNK1

forming a complex with HAH1-LBT-Ln^{3+} have been exploited in calculation aimed to determine the structure of the adduct.
The variability in the position of the LBT with respect to HAH1 leads to different configurations of the HAH1:MNK1 adduct. We therefore tried blocking the configuration of the LBT based on pcs data of HAH1 alone, and then keeping this configuration throughout structural calculations for HAH1:MNK1. Calculations are in progress.

In the latter project, the entire CABP was successfully fused with MNK1 in order to demonstrate that the CABP can be used as paramagnetic tag. The results obtained so far indicate that CABP has a good affinity for the lanthanide metal ions and this affinity doesn't significantly decrease, despite the fusion with MNK1. Indeed the comparison of the 2D HSQC spectra corresponding to the isolated CABP and to the CABP fused with MNK1 show that the fusion does not alter the general structure of the single proteins. Moreover the analysis of the pcs for the assigned resonances, collected with different lanthanide metal ions (Yb^{3+}, Er^{3+} and Tb^{3+}), suggests that the CABP protein experience a reduced mobility with respect to the tagged protein (MNK1). This is probably related to the steric volume occupied by tag and protein.

The present results suggest that the fusion of CABP with target proteins is a powerful tool to obtain paramagnetic restrains useful to determine the structure of protein complexes.
Moreover, this tag can be fused to a multi-domain protein, where the domains reorient themself in different spatial position, in order to investigate the relative position of the domains and the entity of the inter-domain mobility.
The next step will be the fusion of CABP to the entire cytosolic N –terminal tail of ATP7A. In fact this protein construct contains six metal binding domais (MNK1-6) where the domain are connected each other by flexible loops. The induced PCS will provide us additional information to investigate mobility, metal binding properties, and protein-protein

inetractions.

Efforts will be also dedicated to search other protein targets where the use of CABP as paramagnetic tag could be useful in solving biological and structural questions.

LIST OF PUBLICATIONS

REFERENCE LIST

Reference List

1. Allegrozzi M, Bertini I, Janik MBL, Lee YM, Lin GH, and Luchinat C. Lanthanide-induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 angstrom from the metal ion. Journal of the American Chemical Society 2000;122:4154-61.

2. Anastassopoulou I, Banci L, Bertini I, Cantini F, Katsari E, and Rosato A. Solution structure of the Apo and copper(I)-loaded human metallochaperone HAH1. Biochemistry 2004;43:13046-53.

3. Arnesano F, Banci L, and Piccioli M. NMR structures of paramagnetic metalloproteins. Quarterly Reviews of Biophysics 2005;38:167-219.

4. Arnesano F, Banci L, and Piccioli M. NMR structures of paramagnetic metalloproteins. Quarterly Reviews of Biophysics 2005;38:167-219.

5. Balayssac S, Jimenez B, and Piccioli M. Assignment strategy for fast relaxing signals: complete aminoacid identification in thulium substituted Calbindin D-9K. Journal of Biomolecular Nmr 2006;34:63-73.

6. Banci L, Bertini I, Cantini F, Della-Malva N, Migliardi M, and Rosato A. The different intermolecular interactions of the soluble copper-binding domains of the Menkes protein, ATP7A. Journal of Biological Chemistry 2007;282:23140-6.

7. Banci L, Bertini I, Cantini F et al. The Atx1-Ccc2 complex is a metal-mediated protein-protein interaction. Nature Chemical Biology 2006;2:367-8.

8. Banci L, Bertini I, Cremonini MA et al. PSEUDYANA for NMR structure calculation of paramagnetic metalloproteins using torsion angle molecular

dynamics. Journal of Biomolecular Nmr 1998;12:553-7.

9. Banci L, Bertini I, Rosato A, and Varani G. Mitochondrial cytochromes c: a comparative analysis. Journal of Biological Inorganic Chemistry 1999;4:824-37.

10. Bell DA, Taylor JA, Butler MA et al. Genotype-Phenotype Discordance for Human Arylamine N-Acetyltransferase (Nat2) Reveals A New Slow-Acetylator Allele Common in African-Americans. Carcinogenesis 1993;14:1689-92.

11. Bertini, I. Handbook on Metalloproteins. Singel, A and Singel, H. 2001.
Ref Type: Generic

12. Bertini I, Cavallaro G, and Rosato A. Cytochrome c: Occurrence and functions. Chemical Reviews 2006;106:90-115.

13. Bertini I, Donaire A, Jimenez B et al. Paramagnetism-based versus classical constraints: An analysis of the solution structure of Ca Ln calbindin D-9k. Journal of Biomolecular Nmr 2001;21:85-98.

14. Bertini I, Janik MBL, Lee YM, Luchinat C, and Rosato A. Magnetic susceptibility tenser anisotropies for a lanthanide ion series in a fixed protein matrix. Journal of the American Chemical Society 2001;123:4181-8.

15. Bertini I, Lee YM, Luchinat C, Piccioli M, and Poggi L. Locating the metal ion in calcium-binding proteins by using cerium(III) as a probe. Chembiochem 2001;2:550-8.

16. Bertini I, Luchinat C, Parigi G, and Pierattelli R. NMR spectroscopy of paramagnetic metalloproteins. Chembiochem 2005;6:1536-49.

17. Bertini I, Luchinat C, and Piccioli M. Paramagnetic probes in metalloproteins. Nuclear Magnetic Resonance of Biological Macromolecules, Pt B 2001;339:314-40.

18. Boccia LM, Lillicrap D, Newcombe E, and Mueller CR. Binding of the Ets factor GA-binding protein to an upstream site in the factor IX promoter is a critical event

in transactivation. Molecular and Cellular Biology 1996;16:1929-35.

19. Brodersen DE, Etzerodt M, Madsen P et al. EF-hands at atomic resolution: the structure of human psoriasin (S100A7) solved by MAD phasing. Structure 1998;6:477-89.

20. Buetow KH, Edmonson MN, and Cassidy AB. Reliable identification of large numbers of candidate SNPs from public EST data. Nature Genetics 1999;21:323-5.

21. Changela A, Chen K, Xue Y et al. Molecular basis of metal-ion selectivity and zeptomolar sensitivity by CueR. Science 2003;301:1383-7.

22. Clifford R, Edmonson M, Hu Y, Nguyen C, Scherpbier T, and Buetow KH. Expression-based genetic/physical maps of single-nucleotide polymorphisms identified by the cancer genome anatomy project. Genome Research 2000;10:1259-65.

23. Collins FS, Brooks LD, and Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation (vol 8, pg 1229, 1998). Genome Research 1999;9:210.

24. Collins FS, Lander ES, Rogers J, and Waterston RH. Finishing the euchromatic sequence of the human genome. Nature 2004;431:931-45.

25. Contreras MA, Ubach J, Millet O, Rizo J, and Pons M. Measurement of one bond dipolar couplings through lanthanide-induced orientation of a calcium-binding protein. Journal of the American Chemical Society 1999;121:8947-8.

26. DeSilva TM, Veglia G, and Opella SJ. Solution structures of the reduced and Cu(I) bound forms of the first metal binding sequence of ATP7A associated with Menkes disease. Proteins-Structure Function and Bioinformatics 2005;61:1038-49.

27. Diaz-Moreno I, Diaz-Quintana A, De la Rosa MA, and Ubbink M. Structure of the

complex between plastocyanin and cytochrome f from the cyanobacterium Nostoc sp. PCC 7119 as determined by paramagnetic NMR. The balance between electrostatic and hydrophobic interactions within the transient complex determines the relative orientation of the two proteins (vol 280, pg 18908, 2005). Journal of Biological Chemistry 2005;280:35784.

28. Dominguez C, Boelens R, and Bonvin AMJJ. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. Journal of the American Chemical Society 2003;125:1731-7.

29. Feeney J, Birdsall B, Bradbury AF, Biekofsky RR, and Bayley PM. Calmodulin tagging provides a general method of using lanthanide induced magnetic field orientation to observe residual dipolar couplings in proteins in solution. Journal of Biomolecular Nmr 2001;21:41-8.

30. Finney LA and O'Halloran TV. Transition metal speciation in the cell: Insights from the chemistry of metal ion receptors. Science 2003;300:931-6.

31. Franz KJ, Nitz M, and Imperiali B. Lanthanide-binding tags as versatile protein coexpression probes. Chembiochem 2003;4:265-71.

32. Franz KJ, Nitz M, Lukovic E, and Imperiali B. Development of lanthanide-binding peptides as natively expressed protein probes. Journal of Inorganic Biochemistry 2003;96:131.

33. Geiser M, Cebe R, Drewello D, and Schmitz R. Integration of PCR fragments at any specific site within cloning vectors without the use of restriction enzymes and DNA ligase. Biotechniques 2001;31:88-+.

34. Hao ZY, Duncan GS, Chang CC et al. Specific ablation of the apoptotic functions of cytochrome c reveals a differential requirement for cytochrome c and apaf-1 in apoptosis. Cell 2005;121:579-91.

35. Hohjoh H and Tokunaga K. Allele-specific binding of the ubiquitous transcription

factor OCT-1 to the functional single nucleotide polymorphism (SNP) sites in the tumor necrosis factor-alpha gene (TNFA) promoter. Genes and Immunity 2001;2:105-9.

36. Honarpour N, Gilbert SL, Lahn BT, Wang XD, and Herz J. Apaf-1 deficiency and neural tube closure defects are found in fog mice. Proceedings of the National Academy of Sciences of the United States of America 2001;98:9683-7.

37. Hudson TJ. Wanted: regulatory SNPs. Nature Genetics 2003;33:439-40.

38. Ikegami T, Verdier L, Sakhaii P et al. Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions. Journal of Biomolecular Nmr 2004;29:339-49.

39. Irizarry K, Kustanovich V, Li C et al. Genome-wide analysis of single-nucleotide polymorphisms in human expressed sequences. Nature Genetics 2000;26:233-6.

40. Kluck RM, BossyWetzel E, Green DR, and Newmeyer DD. The release of cytochrome c from mitochondria: A primary site for Bcl-2 regulation of apoptosis. Science 1997;275:1132-6.

41. Kluck RM, Ellerby LM, Ellerby HM et al. Determinants of cytochrome c pro-apoptotic activity - The role of lysine 72 trimethylation. Journal of Biological Chemistry 2000;275:16127-33.

42. Knight JC, Udalova I, Hill AVS et al. A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. Nature Genetics 1999;22:145-50.

43. Kruglyak L and Nickerson DA. Variation is the spice of life. Nature Genetics 2001;27:234-6.

44. Kurland, R. J. Isotropic NMR shifts in transition metal complexes: Calculation of the Fermi contact and pseudocontact terms. McGarvey, B. R. J Magn Reson 2,

286-301. 1970.

Ref Type: Generic

45. Li K, Li YC, Shelton JM et al. Cytochrome c deficiency causes embryonic lethality and attenuates stress-induced apoptosis. Cell 2000;101:389-99.

46. Lippard SJ. Principles of Bioinorganic Chemistry. 1994.

47. Liu XS, Kim CN, Yang J, Jemmerson R, and Wang XD. Induction of apoptotic program in cell-free extracts: Requirement for dATP and cytochrome c. Cell 1996;86:147-57.

48. Lutsenko S, Barnes NL, Bartee MY, and Dmitriev OY. Function and regulation of human copper-transporting ATPases. Physiological Reviews 2007;87:1011-46.

49. Lutsenko S, LeShane ES, and Shinde U. Biochemical basis of regulation of human copper-transporting ATPases. Archives of Biochemistry and Biophysics 2007;463:134-48.

50. Ma C and Opella SJ. Lanthanide ions bind specifically to an added "EF-hand" and orient a membrane protein in micelles for solution NMR spectroscopy. Journal of Magnetic Resonance 2000;146:381-4.

51. Macmanus JP, Hogue CW, Marsden BJ, Sikorska M, and Szabo AG. Terbium Luminescence in Synthetic Peptide Loops from Calcium-Binding Proteins with Different Energy Donors. Journal of Biological Chemistry 1990;265:10358-66.

52. Marcuello E, Altes A, Menoyo A, Del Rio E, and Baiget M. Methylenetetrahydrofolate reductase gene polymorphisms: genomic predictors of clinical response to fluoropyrimidine-based chemotherapy? Cancer Chemotherapy and Pharmacology 2006;57:835-40.

53. Marcus PM, Vineis P, and Rothman N. NAT2 slow acetylation and bladder cancer risk: a metaanalysis of 22 case-control studies conducted in the general

population. Pharmacogenetics 2000;10:115-22.

54. Martin LJ, Hahnke MJ, Nitz M et al. Double-lanthanide-binding tags: Design, photophysical properties, and NMR applications. Journal of the American Chemical Society 2007;129:7106-13.

55. Martin LL, Sculimbrene BR, Nitz M, and Imperiali B. Rapid combinatorial screening of peptide libraries for the selection of lanthanide-binding tags (LBTs). Qsar & Combinatorial Science 2005;24:1149-57.

56. McConnell, H. M. Isotropic nuclear resonance shift. Robertson, R. E. J.Chem.Phys 29, 1361-1365. 1958.

Ref Type: Generic

57. Newmeyer DD and Ferguson-Miller S. Mitochondria: Releasing power for life and unleashing the machineries of death (vol 112, pg 481, 2003). Cell 2003;112:873.

58. Nitz M, Franz KJ, Maglathlin RL, and Imperiali B. A powerful combinatorial screen to identify high-affinity terbium(III)-binding peptides. Chembiochem 2003;4:272-6.

59. Packer BR, Yeager M, Staats B et al. SNP500Cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes. Nucleic Acids Research 2004;32:D528-D532.

60. Pecina P, Houstkova H, Hansikova H, Zeman J, and Houstek J. Genetic defects of cytochrome C oxidase assembly. Physiological Research 2004;53:S213-S223.

61. Pidcock E and Moore GR. Structural characteristics of protein binding sites for calcium and lanthanide ions. Journal of Biological Inorganic Chemistry 2001;6:479-89.

62. Piedrafita FJ, Molander RB, Vansant G, Orlova EA, Pfahl M, and Reynolds WF. An Alu element in the myeloperoxidase promoter contains a composite SP1-thyroid

hormone-retinoic acid response element. Journal of Biological Chemistry 1996;271:14412-20.

63. Ponomarenko JV, Merkulova TI, Orlova GV et al. rSNP_Guide, a database system for analysis of transcription factor binding to DNA with variations: application to genome annotation. Nucleic Acids Research 2003;31:118-21.

64. Reich DE, Gabriel SB, and Altshuler D. Quality and completeness of SNP databases. Nature Genetics 2003;33:457-8.

65. Schmidt TR, Wildman DE, Uddin M, Opazo JC, Goodman M, and Grossman LI. Rapid electrostatic evolution at the binding site for cytochrome c on cytochrome c oxidase in anthropoid primates. Proceedings of the National Academy of Sciences of the United States of America 2005;102:6379-84.

66. Scott RA, Mauk AG. Cytochrome c. A multidisciplinary approach.University Science Books, 1996.

67. Sculimbrene BR and Imperiali B. Lanthanide-binding tags as luminescent probes for studying protein interactions. Journal of the American Chemical Society 2006;128:7346-52.

68. Sellick GS, Longman C, Tolmie J et al. Genomewide linkage searches for Mendelian disease loci can be efficiently conducted using high-density SNP genotyping arrays. Nucleic Acids Research 2004;32.

69. Shastry BS. SNP alleles in human disease and evolution. Journal of Human Genetics 2002;47:561-6.

70. Sherman F. The importance of mutation, then and now: studies with yeast cytochrome c. Mutation Research-Reviews in Mutation Research 2005;589:1-16.

71. Sherry ST, Ward MH, and Sirotkin K. dbSNP - Database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Research

1999;9:677-9.

72. Silvaggi NR, Martin LJ, Schwalbe H, Imperiali B, and Allen KN. Double-lanthanide-binding tags for macromolecular crystallographic structure determination. Journal of the American Chemical Society 2007;129:7114-20.

73. Stephens JC. Haplotype variation and linkage disequilibrium in 313 human genes (vol 293, pg 489, 2001). Science 2001;293:1048.

74. Stitziel NO, Tseng YY, Pervouchine D, Goddeau D, Kasif S, and Liang J. Structural location of disease-associated single-nucleotide polymorphisms. Journal of Molecular Biology 2003;327:1021-30.

75. Su XC, Huber T, Dixon NE, and Otting G. Site-specific labelling of proteins with a rigid lanthanide-binding tag. Chembiochem 2006;7:1599-604.

76. Thomson G, Valdes AM, Noble JA et al. Relative predispositional effects of HLA class II DRB1-DQB1 haplotypes and genotypes on type 1 diabetes: a meta-analysis. Tissue Antigens 2007;70:110-27.

77. Turner DL, Brennan L, Chamberlin SG, Louro RO, and Xavier AV. Determination of solution structures of paramagnetic proteins by NMR. European Biophysics Journal with Biophysics Letters 1998;27:367-75.

78. Vasiliev GV, Merkulov VM, Kobzev VF, Merkulova TI, Ponomarenko MP, and Kolchanov NA. Point mutations within 663-666 bp of intron 6 of the human TDO2 gene, associated with a number of psychiatric disorders, damage the YY-1 transcription factor binding site. Febs Letters 1999;462:85-8.

79. Volkov AN, Ferrari D, Worrall JAR, Bonvin AMJJ, and Ubbink M. The orientations of cytochrome c in the highly dynamic complex with cytochrome b(5) visualized by NMR and docking using HADDOCK (vol 14, pg 799, 2005). Protein Science 2006;15:1563.

80. Wang WY and Malcolm BA. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange (TM) site-directed mutagenesis. Biotechniques 1999;26:680-2.

81. Wang X, Srisailam S, Yee AA et al. Domain-domain motions in proteins from time-modulated pseudocontact shifts. Journal of Biomolecular Nmr 2007;39:53-61.

82. Wohnert J, Franz KJ, Nitz M, Imperiali B, and Schwalbe H. Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings. Journal of the American Chemical Society 2003;125:13338-9.

83. Zhao ZM, Fu YX, Hewett-Emmett D, and Boerwinkle E. Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. Gene 2003;312:207-13.