University of Florence

International Doctorate in Structural Biology

Cycle XXIII (2008–2010)

Ph.D. thesis

of

Anna Pavelková

NMR Characterization of Protein–Protein Interactions

Tutor: Prof. Isabella Caterina Felli Coordinator: Prof. Ivano Bertini

S.S.D. CHIM/03

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

There are many people who deserve aknowledgement for support that made my task easier and also more enjoyable, they are too numerous to be mentioned here. I would like to devote thanks to Prof. Ivano Bertini for giving me the opportunity to work at CERM and to be involved in interesting projects introducing me to the world of modern nuclear magnetic resonance in structural biology. Special thanks belong to Prof. Isabella C. Felli for dedicated care through informal discussions, lending advice, and other assistance.

List of Abbreviations

δ	chemical shift
τ_c	rotational correlation time
1D, 2D, 3D	one-, two-, three-dimensional
AIR	ambiguous interaction restraints
ALR	augmenter of liver regeneration, FAD-linked sulfhydryl oxidase
ATP7A	P-type ATPase
BL21 (DE3)	competent cells
CLEANEX	Chemical EXchange spectroscopy
CSI	chemical shift index
DG	distance geometry
DNA	deoxyribonucleic acid
DTT	dithiothreitol
DYANA	DYnamic Algorithm for Nmr Applications
EDTA	ethylene diamine tetraacetic acid
FAD	flavin adenine dinucleotide
FPLC	fast protein liquid chromatography
FT	Fourier transform
HADDOCK	High Ambiguity Driven biomolecular DOCKing

HSQC	heteronuclear single-quantum correlation spectroscopy
IDPs	intrinsically disordered proteins
IDRs	intrinsically disordered regions
IMS	mitochondrial intermembrane space
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K_d	dissociation constant
kDa	kilo-Dalton
LB	Luria Bertani
M	mol/l
M9	growth medium (minimal medium)
MBDs	metal-binding domain
MBP	maltose-binding protein
MNK	Menkes protein
MW	molecular weight
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
OD	optical density (absorbance)
PPIs	protein–protein interactions
ppm	parts per million
R_1	longitudinal or spin-lattice relaxation rate
R_2	transversal or spin-spin relaxation rate
RDCs	residual dipolar couplings
REM	restrained energy minimization
RMD	restrained molecular dynamics

rmsd	root mean square deviation
RNA	ribonucleic acid
SA	dynamical simulated annealing
T_1	longitudinal or spin-lattice relaxation time
T_2	transversal or spin-spin relaxation time
TEV	Tobacco etch virus
TGN	trans-Golgi network
TOCSY	total correlation spectroscopy
Tris	tris(hydroxymethyl)-amino-methan
WT	wild type
E. coli	*Escherichia coli*

Contents

Chapter 1

Introduction

The new challenges in the field of structural biology are evolving towards very ambitious goals such as the complete characterization at atomic resolution of complex biological pathways in environments that closely resemble living organisms in order to understand the molecular mechanisms at the basis of life. In this context NMR plays a strategic role for the structural and dynamic characterization at atomic resolution of complex biological macromolecules [1, 2, 3], in environments as complex as whole cells [4, 5, 6, 7, 8]. Moreover it can provide unique information on highly flexible disordered proteins [2, 9, 10] and/or on weak intermolecular interactions [11, 12] which are difficult to characterize at atomic resolution with other techniques. The presence of flexibility and disorder [13, 14, 15, 9, 10] as well as the weak intermolecular interactions [16, 17, 18, 19] play a central role in a variety of different biological processes, and in particular in intermediate steps of biochemical reaction pathways.

Among the many challenges of contemporary structural biology we decided to focus on the general subject of metal ions in biological systems [20, 3]. Indeed it is well known that metal ions, although present in living systems in small amounts, are very important for life as they are involved in carrying out many different functions. Many metals are utilized as cofactors in a variety of proteins [21], however their presence in living systems should be tightly controlled as otherwise they may become toxic [22]. Living organisms have developed processes which can selectively handle metal transport and incorporation in the recipient proteins under a very tight control [22, 23, 24, 25]. This allows that the right metal is incorporated in the right protein. This is accomplished by specific trafficking pathways [20] through which a metal ion, as it enters the cell, is coordinated by a specific protein which takes it to a selected unique partner protein, either as final metal recipient or as one of the steps in the metal transfer. Shared metal co-

ordination together with molecular recognition [26] between the two partner proteins drive the formation of protein–protein complexes which eventually evolve with the transfer of the metal ion from the protein with lower affinity to that with the higher [27]. The protein–protein interaction through which the metal transfer occurs is supposed to be associated with a small, negative change in free energy to ensure a weak transient character of the protein complex. The formation of the metal-mediated complex in this case represents only an intermediate step along the metal transfer pathway (Figure 1.1). A nice recent example in this context is provided by the characterization at atomic resolution of many of the proteins involved in the metal transfer pathway of copper [12, 28, 11, 29, 30, 31]. In particular, the interactions between the soluble copper chaperones and their target proteins (ATPases) have extensively been characterized showing that copper 'trafficking' is mediated by weak protein–protein interactions that only occur in the presence of the metal ion (metal-mediated interactions).

A comprehensive understanding of these protein–protein interactions requires a detailed structural analysis. Although a number of experimental techniques can report on the existence of a protein–protein interaction, very few can provide detailed structural information. There are three main methods used today to determine the structure of macromolecular protein complexes: X-ray crystallography, NMR spectroscopy and to a lesser extent cryoelectron microscopy. Inherent limitations of crystallography are the requirement for the complex to crystallize in the first place, which may be difficult if one/both of the components of the complex has/have significant amounts of disorder, and the possibility that crystals do not contain the biologically relevant conformation of the proteins. In cases where the protein–protein interaction is weak ($K_d > \mu M$), NMR is essentially the only approach that allows the determination of high-resolution structures [1].

In order to better understand the processes of metal binding and transfer, we have focused on copper trafficking from the cytoplasm to the *trans*-Golgi network (TGN), where copper is inserted in a number of enzymes [32]. In the cytoplasm, copper(I) is bound to a metallochaperone which delivers the metal to a variable number of soluble metal-binding domains (MBDs) of specific ATPases located at the TGN membrane [22]. The first system characterized was yeast Atx1/Ccc2 and systems with functional homology have been noted in humans, mice, sheep, *Arabidopsis thaliana*, etc. [11, 33, 34]. In the case of Atx1/Ccc2 the complete three-dimensional structure of the metal-mediated complex was determined [11]. The determination of the 3D structure required a combination of advanced techniques to detect inter-protein constraints and the use of mutagenesis of cysteines to characterize the metal transfer between the partners [1, 35, 36]. This was then used to analyse in detail the

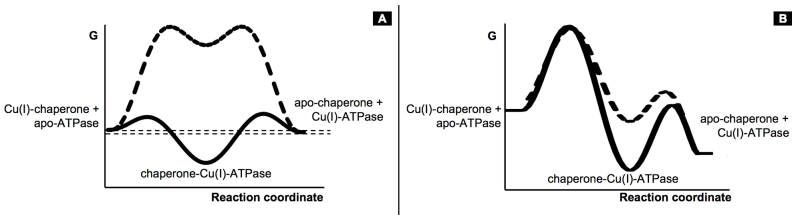


Figure 1.1: The energy profile for the copper(I) transfer reaction for different types of metallochaperone:ATPase interactions. In A, the equilibrium constant for copper(I) exchange between the chaperone and the ATPase is close to 1; while in B the copper affinity of the ATPase is higher then that of the chaperone. In both panels, the dashed line describes the situation where the intermediate adduct does not accumulate to a significant extent in the test tube, and therefore might not be detecteble, whereas the thick line describes the situation where the adduct is more stable than the separate proteins.

inter-protein surfaces and compare these to the possible other complexes to understand the driving forces of interaction. In humans, the metallochaperone is called HAH1 (Atox1) and two target ATPases exist: ATP7A and ATP7B [37]. Both ATP7A and ATP7B have six soluble MBDs, which have different affinities for copper(I) [38, 31]. The present research aims at shedding light on the nature of the interaction between the metallochaperone and MBDs of the ATPases in the general frame of understanding the energetics of copper(I)-mediated interactions. The interaction of the cytosolic tail of MNK with copper(I) and with copper(I)-HAH1 has been extensively studied [30, 29, 39]. We focused on the determination of the 3D structure of the complex between the human copper(I)-chaperone HAH1 and the first metal-binding domain of the P-type copper-transporting ATPase, ATP7A.

The determination of the 3D solution structure of the metal-mediated complex between the soluble human copper chaperone (HAH1) and one of the soluble metal binding domains of one of the target ATPases (MNK1) has provided the initial subject for this doctoral thesis (Project 1). The 3D solution structure determined through a strategy combining advanced NMR experiments and site specific mutagenesis (intermolecular NOEs and mutation of the metal binding cysteines) revealed important details of the protein–protein interfaces that were then used to draw general conclusions on the structural features and energetics of metal mediated protein–protein interactions. The protein–protein interface is finely tuned to prevent direct molecular recognition in the absence of the metal ion but to favor their interaction through the metal ion and the stabilization of the complex through inter-protein interactions.

While this gives us an important message on an important functional unit,

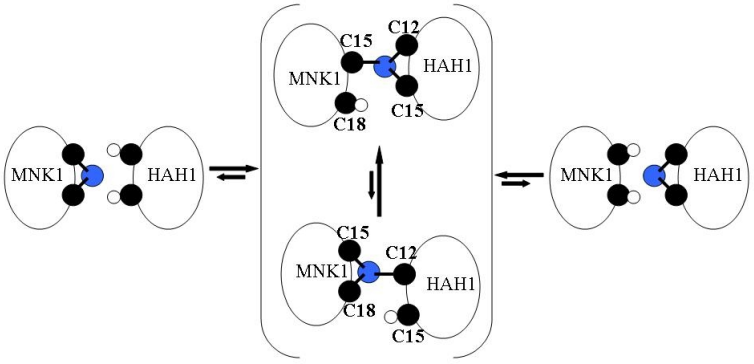


Figure 1.2: The mechanism for Cu(I) transfer between HAH1 and MNK1. Cu(I)-binding sites in metallochaperones are usually quite exposed to ensure easy transfer to the partner, receiving protein, whose coordinating residues should find copper easily accessible. A fine balance between a filled coordination sphere that stabilizes copper coordination and the easy detachment of one ligand when the ligands of the incoming partner invade the copper coordination sphere is primarily achieved by completing the coordination sphere with more labile ligand with respect to the others.

the metal-binding domain (MBD), used in general by nature to bind metal ions, it does not clarify how the selectivity for the different metal ions as well as the direction of the metal transfer pathway is controlled. As evident by the previous study, chaperones and metal-binding domains of ATPases share the same overall 3D fold as well as the metal binding motif, CXXC (Figure 1.2). Moreover, this module is not only used to control the concentration and transfer of copper, but also for other metal ions.

Several proteins involved in different metal transport pathways feature similar 3D folds and metal-coordination sites. As an example, the overall 3D folds and metal-binding sites of the proteins devoted to controlling Cu(I) trafficking are very similar to those involved in the homeostasis of other metal ions, such as Zn(II). This in principle might lead to the coordination of the wrong metal ion. And indeed *in vitro* some proteins exhibit a higher affinity for a non physiological metal ion. However the directionality of metal transfer pathways must be very selective as it is well known that the incorporation of a non natural metal ion in a protein may cause severe problems leading to disease. *In vivo* metal specificity is ensured by metal availability in a compartment-specific level. The concentration of the free metal ion (or small ligand-bound) in the cells is indeed, in most cases very low. Therefore, in addition to the affinity of different proteins for the metal ions, also protein–protein interactions determining molecular recognition processes play an important role in controlling the transfer of metal ions

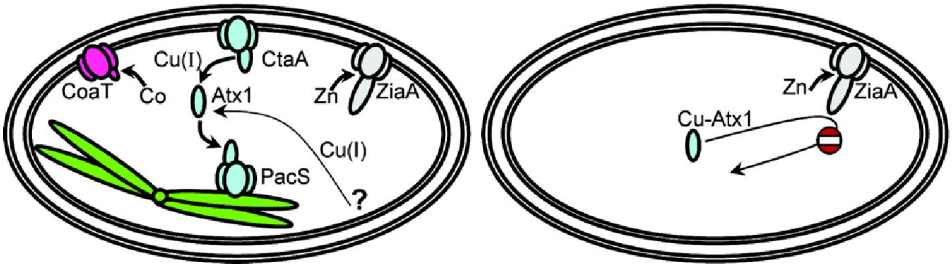


Figure 1.3: Copper trafficking to thylakoids of *Synechocystis* PCC 6803. Proposed mechanism of copper(I) import into thylakoid requires two copper(I) transporting P_1-type ATPases CtaA and PacS in conjunction with small soluble copper(I) metallochaperone ScAtx1. ScAtx1 directly interacts with the soluble N-terminal domains of P_1-type ATPases (left). Failure of ZiaA to interact with ScAtx1 through N-terminal metal-binding domain (right).

from one partner to another, playing a key role in regulating metal transfer processes. In order to better understand the subtle features through which the metal transfer pathway is controlled, we decided to focus on the factors modulating the interaction between a metal chaperone and different ATPases within the cyanobacterial system *Synechocystis* PCC 6803.

Cyanobacteria are a bacterial group that have known enzymatic demand for cytoplasmic copper import. They contain internal membrane-bound compartments called thylakoids, which constitute the site of photosynthetic reactions and respiratory electron transport involving copper proteins such as plastocyanin and caa_3-type cytochrome oxidase. Copper supply into thylakoids is needed to enable the formation of fully functional holoforms of copper proteins for photosyntesis. In the proposed mechanism of copper import into thylakoids *Synechocystis* PCC 6803 uses two copper transporting P-type ATPases, CtaA and PacS, together with the small soluble copper metallochaperone ScAtx1. ScAtx1, interacting with the N-terminal cytosolic domains of ATPases, is presumed to acquire copper from $CtaA_N$ and donate it to $PacS_N$ [28]. In addition to a chaperone ATPase system devoted to transfer copper, ScAtx1 and PacS *Synechocystis* has two further P-type ATPases ZiaA and CoaT, which are involved in metal export into the periplasm (zinc and cobalt, respectively). Surprisingly, the N-terminal cytosolic domain of ZiaA ($ZiaA_N$) shows a lower affinity for zinc, the physiological metal ion, than for copper, rising questions on which are the molecular features controlling metal uptake and delivery *in vivo*. This means that the nature of the amino acids at the surface of the partner proteins plays a key role in the formation of favourable/unfavourable contacts that determine whether a complex will be formed or not (molecular recognition), independently of the

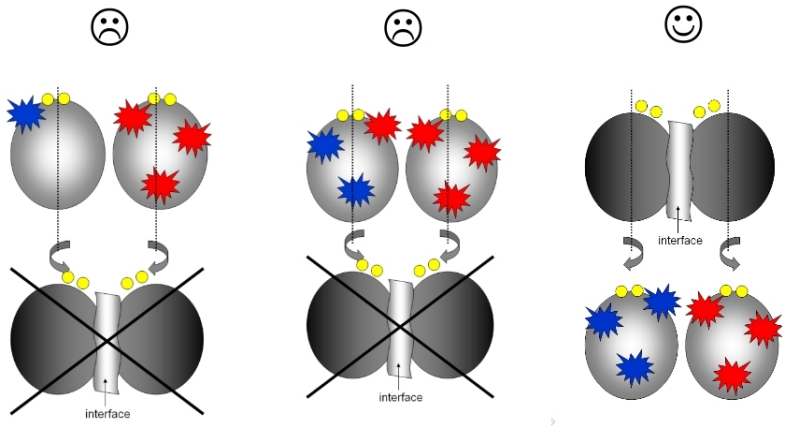


Figure 1.4: For the formation of the metal-mediated complex the two proteins should be in the proper relative orientation with the two CXXC motives facing each other.

nature of the metal-binding site (the same kind is this case). This fine tuning of molecular surfaces is thus a structural feature to regulate metal transfer pathways. The case of the copper chaperone ScAtx1 which does not transfer Cu(I) to a metal-binding domain of a zinc ATPase, $ZiaA_N$, even if the latter has a higher affinity for Cu(I), represents a striking example of how nature controls metal transfer. To understand the subtle features determining and tuning protein recognition, we have substituted a variable number of residues in $ZiaA_N$ with the goal of mimicking the surface of $PacS_N$ and monitor how these surface mutations modulate the capability of the copper chaperone ScAtx1 to select and therefore to interact with the appropriate metal transfer partner (Project 2). These examples show how a conserved 3D structural module is used in living systems for a general property such as metal binding and metal transfer and how fine tuning and selectivity is achieved by minor modifications of a common 3D fold. This shows the importance of a well defined 3D structural model for function.

Recent evidence shows that a well defined 3D structure is not always necessary for function and that in some cases more flexible disordered states characterized by rapid interconversion between several partially structured conformers may provide functional advantages [40, 41, 42, 43, 14]. Many examples of functionally relevant disordered proteins or protein fragments, which are generally referred to as intrinsically disordered proteins (IDPs) or protein regions (IDRs) have recently appeared in the literature challenging well established concepts in structural biology and demanding for a new view of the structure function relationship that also includes the concept of flexibility and disorder. Therefore, while the functional importance of structural

scaffolds is still recognized, more flexible modules may offer complementary functional advantages especially in intermediate steps of reaction pathways, allowing a protein to be more versatile.

The intrinsically disordered proteins are highly abundant in nature. They have been shown to play key roles in remarkable range of cellular processes, including signaling, cell cycle control, molecular recognition, transcription, and replication. Many of them are associated with various human diseases such as neurodegenerative diseases, cancer and cardiovascular diseases. In many of these cases, the uncontrolled transition to disordered states or the aggregation of disordered protein modules leads to disease. Well-known examples are neurodegenerative diseases such as Alzheimer's, Parkinson's and prion diseases.

Present funtional classification attempts have shown that IDPs/IDRs may function by two basically different mechanisms, either directly due to their structural disorder (entropic chains), or via molecular recognition of partner molecules (another protein, RNA, DNA, membrane or small ligands). In the process of molecular recognition, intrinsically disordered proteins are thought to undergo disorder-to-order transition (also termed induced folding) and acquire a folded, ordered structure. However, the relationship between intrinsic disoder and protein function is still largely unclear and more investigation is required to better understand the biological role of structural disorder.

The presence of disordered fragments, which until recently were discarded from structural studies as they were not considered to be very relevant for function, is actually much more common than one expects. Indeed many complex systems are actually constituted by a combination of structured and unstructured modules, with the unstructured modules often playing key roles in molecular recognition, signaling, etc. One example is provided by the many flexible linkers connecting the metal binding domains in the above mentioned ATPases. However, while focusing on the subject the metal-mediated interactions, the focus gradually shifted to other kinds of cysteine-rich proteins. Indeed, besides their metal binding properties, cysteines also have very interesting redox properties as they can form disulfide bonds. This feature is used in nature to perform and control a variety of redox reactions that occur in living systems. As an example, a variety of different proteins with cysteine-rich motifs, were identified that play an important role in energy metabolism in mitochondria. These proteins perform their function by formation and cleavage of disulfide bonds which, in turn, dramatically affect their structural properties which are generally characterized by a very heterogeneous and dynamic behavior. Indeed the formation of disulfide bonds generally stabilizes a specific 3D fold while their cleavage generally renders the proteins more flexible and disordered. Therefore NMR plays a key role

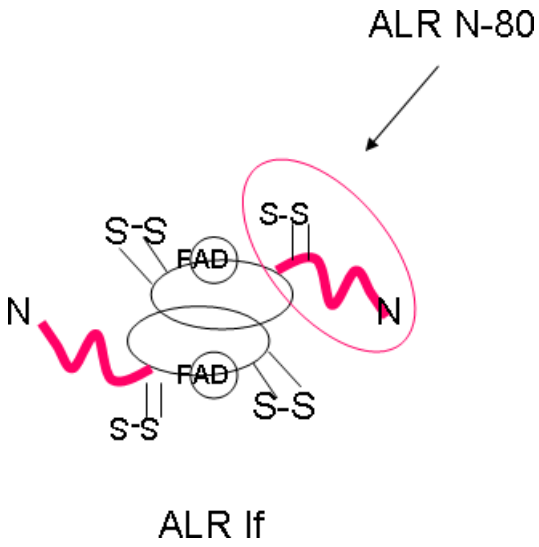


Figure 1.5: Schematic representation of the ALR protein.

for their complete structural and dynamic characterization.

In case of complex systems several cystein-rich motifs can be found within the same proteins and some of them are likely to be in intrinsically disordered protein fragments. One example is provided by ALR [44, 45], a FAD-dependent sulfhydryl oxidase of the mitochondrial intermembrane space (IMS). It is an essential component of the redox regulated MIA40/ALR import and assembly pathway used by many of the cysteine-containing intermembrane space proteins. Upon import of a Cys-reduced substrate, MIA40 interacts with the substrate via intermolecular disulphide bond and shuttles a disulphide to its substrate [46]. ALR functions in catalyzing reoxidation of the reduced MIA40 and release of the substrate.

The ALR enzyme is characterized by a highly conserved central catalytic core of ∼100 amino acids, which includes an active site CXXC motif, $CX_{16}C$ disulphide bond, and residues involved in FAD binding. The $CX_{16}C$ disulphide bond is proposed to play a structural role in stabilizing the FAD binding and/or protein folding. In addition to a well folded protein domain, the protein features an N-terminal segment containing two cysteines that is supposed to have an important role in molecular recognition. In order to contribute to the understanding of the function of this complex system, containing several cysteine-rich sites, the N-terminal portion was expressed separately and characterized to lay the basis for the complete characterization of the system (Project 3).

Chapter 2

Materials and Methods

In general, the characterization of a protein through NMR consists of several steps:

- Biochemical work such as protein cloning, expression, purification, isotopic enrichment
- Initial characterization through ^{1}H-^{15}N HSQC and ^{15}N relaxation experiments to optimize sample conditions
- Acquisition and processing of a set of multidimensional NMR experiments
- Sequence-specific assignment of the backbone and of side chains
- Extraction of experimental restraints that can be used for the structure calculation process, e.g. NOE derived distances, dihedral angle restraints derived from J-couplings, etc.
- Structure calculations (simulated annealing, restrained molecular dynamics) and structure validation
- If the protein lacks a stable 3D structure, determination of secondary structure propensities and characterization of local mobility along polypeptide chain

2.1 Protein expression and purification

The technique of recombinant DNA is widely employed to overexpress proteins and obtain protein samples for NMR. A recombinant protein production process consists of gene cloning, expression and protein purification.

The choice of the cloning method to insert a gene into an expression vector is important, principally because proteins are so heterogenous in their behavior that it is impossible to predict whether a protein of interest will express well, be soluble, easy to purify or possess activity.

The choice of the host strains depends more on the nature of the heterologous protein. The *Escherichia coli* protein expression system is one of the most useful methods employed for NMR sample preparation. *Escherichia coli*-based expression systems are favoured for their well-characterized genetics, easy manipulation, and existance of many strains and expressing vectors which allows rapid cloning and expression of proteins at low cost. However, most eukaryotic proteins are unlikely to be properly folded when expressed in *E. coli*. The use of alternative expression systems often allows the expression of proteins that are insoluble when expressed in *E. coli*. At present, the other choices for expressing these proteins, such as yeast, insect or human cells, have disadvantages. Insect and human cell culture, in its current form, is expensive and time-consuming, and current yeast systems lack the capacity to add many post-translational modifications that are important for proteins from higher organisms.

The variables which affect the expression of recombinant protein are: host strain, growth medium and induction parameters (temperature, IPTG concentration and duration of induction step).

A typical expression experiment consists of the following steps:

- Growing of a starter culture
- Inoculation of the main culture and incubation until OD_{600} reaches 0.4–1
- Induction of protein expression[1]
- Harvesting of the cell pellet by centrifugation (20 min at 6000 g). Cell pellets are stored at -20 °C

[1]Protein expression is induced by the addition of the proper inducer or by changing the growth conditions. From this point the cells will use most of their resources for the production of the target protein and will not grow much further.

NMR structure determination is greatly facilitated by the use of isotopically labelled proteins [2], in which ^{14}N and ^{12}C are replaced with ^{15}N and ^{13}C. To prevent incorporation of unlabelled isotopes into the protein (uniformly labelled proteins), bacteria are grown on a minimal medium completely lacking in carbon or nitrogen, to which a single source of these elements is then added, typically in the form of ammonium-sulfate or ammonium-chloride for nitrogen and glucose for carbon. In special cases selective labelling can be employed.

The more that is known about characteristics of a protein, the more easily it can be isolated and purified. Purification strategies are routine for proteins that are highly expressed for use in NMR or X-ray crystallography. An affinity-chromatography step using a removable tag, such as hexahistadine, usually provides a major step of the purification. The tag can then be removed by inserting a linker region between the tag and the native protein sequence, which contains a specific site for endoprotease cleavage. One or two additional simple chromatographic steps yield protein samples that are sufficiently pure for acquisition of an initial HSQC spectrum.

The proteins that were studied together with relevant details regarding their expression and purification procedures are reported below.

2.1.1 MNK1/HAH1 protein–protein interaction

Recombinant protein MNK1 was expressed corresponding to amino acids 5–77 of the full-length ATP7A protein, using essentially the same protocol adopted for other individual MNK domains described in [47]. HAH1 protein samples were produced as previously reported [37].

2.1.2 ScAtx1 and ZiaA$_N$ protein mutants

Recombinant protein ZiaA$_N$ was expressed similarly to the previously described procedure [28], as a soluble protein in Rosetta strain (Cat. No. 71402) exposed to Zn^{2+}. Lysates were loaded onto a Hi-Trap nickel column. Collected fractions were subjected to FPLC size exclusion chromatography using Hi-Load 26/60 or 16/60 Superdex 75 column. Positive eluted fractions

[2]The NMR spectroscopy of biomolecules is based on excitation of nuclei with nuclear spin quantum number equal to 1/2 by strong oscillating magnetic fields. Only ^{1}H and ^{31}P nuclei are naturally abundant in biomacromolecules. ^{12}C, ^{14}N, or ^{16}O have either zero magnetic moment or relaxation properties unfavorable for biomolecular NMR studies. Small proteins can be analyzed by proton NMR but to study larger proteins and to obtain sufficient data to determine side chain conformations, it is necessary to introduce ^{13}C and ^{15}N in the protein.

were pooled altogether and purified in the last run of Hi-trap nickel column using the elution of the gradient of imidazole.

The expression plasmids for the selected mutants of $ZiaA_N$ were obtained through the Quickchange Site-directed mutagenesis kit (Stratagene). These include: mutant QC2: Asp18Arg, Lys23Ala, Leu24Ser, Lys25Ser; mutant QC3: Asp18Arg, Lys23Ala, Leu24Ser, Lys25Ser, Thr20Ala, Ser21Ala; mutant QC4: Asp18Arg, Lys23Ala, Leu24Ser, Lys25Ser, Thr20Ala, Ser21Ala, Gly28Arg, Ser29Ala). The expression and purification protocol for the mutants was essentially the same as for the N-terminal domain of the ZiaA WT.

Recombinant protein ScAtx1 was expressed similarly to the previously described procedure [48]. Lysates were applied to Sephadex G-75, fractions were eluted in 25 mM Tris-HCl, pH 7.0. Pooled fractions were applied to Q-Sepharose and subsequently eluted.

2.1.3 The N-terminal ALR protein fragment

The recombinant construct pET16HisMBPTEVN80 ALR was expressed as a soluble protein in *E. coli* Origami B strain. Lysates were loaded onto a Ni-beads slurry (used buffer: 50 mM NaCl, 50 mM Tris pH 8.00, 5 mM imidazole). Collected fractions were dialyzed and subsequently subjected to the cleavage with the TEV protease (buffer used for TEV cleavage: 50 mM NaCl, 50 mM Tris pH 8.00, 1 M DTT). The cleaved protein was dialyzed again (into 50 mM NaCl, 50 mM Tris pH 8.00, 5 mM imidazole) to remove the DTT for the further purification step by the negative affinity chromatography. If needed, an additional purification step on the Q Sepharose could be implemented.

The final concentration of the ^{15}N labeled and ^{13}C, ^{15}N labeled human N-80 ALR samples was in the 0.3–0.8 mM range, in 50 mM phosphate buffer at pH 7.00 with 0.25 mM EDTA. Reduction was achieved by gradual addition of DTT (monitored by NMR) up to a final concentration of 20 mM and incubated overnight. D_2O was added to the buffer for the lock signal.

2.2 NMR experiments and analysis

2.2.1 Sample requirements

Suitable protein samples for biomolecular NMR studies should satisfy the following requirements. The typical volume of the purified protein sample should be 180–600 μl with the protein in the 'millimolar' concentration range

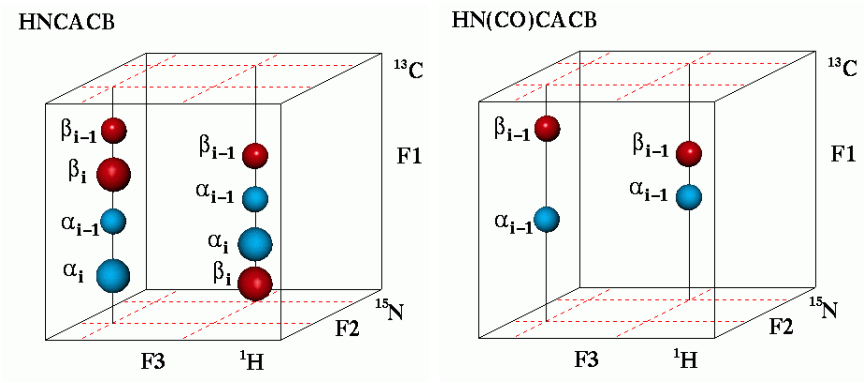


Figure 2.1: Schematic representation of the 3D heteronuclear NMR spectra HNCACB and HN(CO)CACB.

(above 0.1 mM) dissolved in an acqueous buffer solution. Ideally, a well behaving protein should be highly soluble and stable for at least several days at room temperature.

2.2.2 Resonance assignment

The sequence-specific resonance assignment of a protein provides the basis for the analysis of structure and dynamic properties by NMR spectroscopy. The use of multidimensional triple resonance NMR experiments for the protein resonance assignment has become standard [49, 50]. For globular proteins of moderate size ($<$ 200 aa), triple resonance NMR data analysis is, in many cases, a routine task (Figure 2.1). The resonance assignment can be divided into two steps, i.e., backbone sequential assignment and side chain assignment.

Based on the primary sequence, the sequential assignment is usually achieved by comparing multidimensional NMR spectra providing the necessary correlations to identify spin systems within each amino acid (intra-residue) and then to link them in a sequence-specific manner (inter-residue). The information about intra- and inter-residual correlation is obtained using the triple resonance experiments based on well-defined heteronuclear scalar couplings along the polypeptide chain [51, 52, 53, 54, 55]. The most popular sequential assignment strategy relies on a set of experiments CBCANH (or HNCACB) and CBCA(CO)NH (HN(CO)CACB), Figure 2.2.2. The names of the experiments are essentially self-explanatory, as they indicate all nuclear spins that are involved in the coherence transfer pathway exploiting scalar couplings and frequency labelling in the various dimensions of the

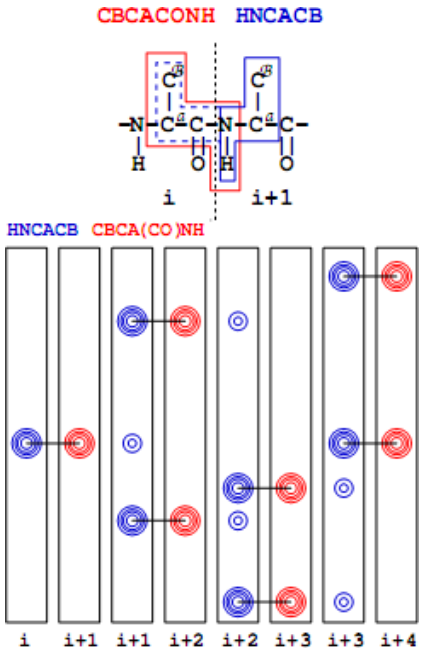


Figure 2.2: Principle of the assignment based on CBCANH and CBCA(CO)NH spectra.

experiments.

The simultaneous acquisition of C^{α} with C^{β}, in one of the indirect dimensions, decreases sequential assignment ambiguities. The information about amino acid type is obtained from the combination of the C^{α} and C^{β} chemical shifts, which provide a first indicator of the amino acid type and facilitates the sequential assignment [56]. The HNCO 3D experiment can then be used to complete backbone assignment by the identification of the C' chemical shift.

To obtain the detailed structure of a protein, the ^{1}H-^{1}H NOE cross peaks should be unambiguosly assigned. The unambiguos analysis of NOE cross peaks requires the side-chain resonance assignment. Several 3D NMR experiments are used for the resonance assignment of protein side chains [57]. In the HCCH-TOCSY experiment [58], proton resonances are correlated throughout the complete spin system of an amino acid by ^{13}C–^{13}C scalar couplings. The type of an amino acid can be recognized from the peak pattern.

A characteristic pattern of signals results for each amino acid from which the amino acid type can be identified. To identify side chains of aromatic amino acids additional experiments are necessary, e.g. HB(CBCGCD)HD and HB(CBCGCDCE)HE, which correlate the aliphatic and the aromatic parts of the side chain [59]. It is possible to assign aromatic side chain

protons using experiments, which correlate these protons with intra-residue H^{β} and C^{β} frequencies, in addition to an HCCH experiment with the ^{13}C radiofrequency carrier placed in the aromatic region.

The most important double- and triple-resonance experiments used for sequential and side-chain resonance assignment and the key correlations that can be detected in each experiment are summarized in Table 2.1.

Table 2.1: Correlations observed in the double and triple resonance experiments used for sequential and side chain assignment (^1H approach)

Experiment	Correlation	J-coupling
CBCANH	$^{13}C_i^{\beta}/^{13}C_i^{\alpha}$–$^{15}N_i$–$^1H_i^N$, $^{13}C_{i-1}^{\beta}/^{13}C_{i-1}^{\alpha}$–$^{15}N_i$–$^1H_i^N$	$^1J_{NC_\alpha}$, $^1J_{NH}$, $^1J_{CC}$
CBCA(CO)NH	$^{13}C_{i-1}^{\beta}/^{13}C_{i-1}^{\alpha}$–$^{15}N_i$–$^1H_i^N$	$^1J_{NC'}$, $^1J_{C_\alpha C'}$, $^1J_{NH}$, $^1J_{CH}$, $^1J_{CC}$
HNCA	$^{13}C_i^{\alpha}$–$^{15}N_i$–$^1H_i^N$, $^{13}C_{i-1}^{\alpha}$–$^{15}N_i$–$^1H_i^N$	$^1J_{NC_\alpha}$, $^1J_{NH}$
HNCO	$^{13}C_{i-1}^{'}$–$^{15}N_i$–$^1H_i^N$	$^1J_{NC,}$, $^1J_{NH}$
HN(CO)CA	$^{13}C_{i-1}^{\alpha}$–$^{15}N_i$–$^1H_i^N$	$^1J_{NC'}$, $^1J_{C_\alpha C'}$, $^1J_{NH}$
HN(CA)CO	$^{13}C_i^{'}$–$^{15}N_i$–$^1H_i^N$, $^{13}C_{i-1}^{'}$–$^{15}N_i$–$^1H_i^N$	$^1J_{NC_\alpha}$, $^2J_{NC_\alpha}$, $^1J_{NH}$, $^1J_{C_\alpha C'}$
HN(CA)NN(H)	$^{15}N_{i+1}$–$^{15}N_i$–$^1H_i^N$, $^{15}N_{i-1}$–$^{15}N_i$–$^1H_i^N$	$^1J_{NC_\alpha}$, $^2J_{NC_\alpha}$, $^1J_{NC_\alpha}$
HCCH-TOCSY	$^1H^j$–$^{13}C^j$…$^{13}C^{j\pm n}$–$^1H^{j\pm n}$	$^1J_{CC}$, $^1J_{CH}$

2.2.3 Exclusively heteronuclear NMR experiments

Direct-detection of heteronuclei, and of ^{13}C in particular, offers a valuable alternative to the standard triple resonance assignment strategy based on the ^1H detection approach. Unfolded systems, where the chemical shift dispersion of ^1H resonances may be unfavourable, can efficiently be characterized using ^{13}C NMR. NMR spectroscopy is a uniquely powerful tool for studying the dynamic structural ensembles of intrinsically disordered proteins.

The starting point of the assignment procedure [60, 61, 62] consists in identifying backbone C^{α} and $C^{'}$ correlations through CACO experiment. The $C^{'}$ is then connected to the C^{β} through a CBCACO experiment. Once all the spin systems have been identified, correlation with backbone nitrogen nuclei is performed using the CON experiment, which correlates the backbone $C^{'}$ carbon atom with the peptidic nitrogen atom of the following residue. The combination of these 2D experiments yields the 3D experiment (CBCACON) enabling to enhance the resolution which is very important in complex systems. The sequence-specific assignment is then accomplished with a CANCO/CBCANCO experiment that correlates the $C^{'}$ shift of one residue with the ^{15}N shift of the following residue and the C^{α}/C^{β} shifts of the one residue and the following residue. In the case of unfolded systems, since transverse proton relaxation does not generally constitute a limiting factor,

the sensitivty of these experiments can be enhanced by using proton polarization as a starting pool, still keeping only heteronuclear chemical shifts in all the dimensions of multidimensional NMR experiments [63]. The COCON and (H)NCANCO experiments provide additional possibilities which result particularly useful to resolve ambiguities in case of unfolded systems. Finally, the CCCON allows to extend the sequence-specific assignment also to the side chains.

This strategy also provides the assignment of Asp, Asn, Glu, and Gln carbonyl and carboxylate carbon atoms, as well as quaternary carbon atoms of aliphatic side chains.

Table 2.2: Correlations observed in the double and triple resonance experiments used for sequential and side chain assignment (^{13}C approach)

Experiment	Correlation	J-coupling
CON	$^{13}C'_{i-1}$–$^{15}N_i$	$^{1}J_{NC'}$
(H)CACO	$^{13}C^{\alpha}_{i}$–$^{13}C'_{i}$	$^{1}J_{C_\alpha C'}$
(H)CBCACO	$^{13}C^{\beta}_{i}$/$^{13}C^{\alpha}_{i}$–$^{13}C'_{i}$	$^{1}J_{C_\alpha C'}$, $^{1}J_{CC}$
(H)CBCACON	$^{13}C^{\beta}_{i-1}$/$^{13}C^{\alpha}_{i-1}$–$^{13}C'_{i-1}$–$^{15}N_i$	$^{1}J_{C_\alpha C'}$, $^{1}J_{NC'}$, $^{1}J_{CC}$
(H)CANCO	$^{13}C^{\alpha}_{i}$–$^{13}C'_{i-1}$–$^{15}N_i$, $^{13}C^{\alpha}_{i-1}$–$^{13}C'_{i}$–$^{15}N_i$	$^{1}J_{C_\alpha C'}$, $^{1}J_{NC'}$, $^{2}J_{NC^\alpha}$
(H)CBCANCO	$^{13}C^{\beta}_{i}$/$^{13}C^{\alpha}_{i}$–$^{13}C'_{i-1}$–$^{15}N_i$, $^{13}C^{\beta}_{i-1}$/$^{13}C^{\alpha}_{i-1}$–$^{13}C'_{i}$–$^{15}N_i$	$^{1}J_{C_\alpha C'}$, $^{1}J_{NC'}$, $^{1}J_{CC}$, $^{1}J_{NC^\alpha}$, $^{2}J_{NC^\alpha}$
(H)CCCON	$^{13}C^{\alpha}_{i-1}$–$^{13}C'_{i-1}$–$^{15}N_i$, $^{13}C^{\beta,\gamma,\delta,\varepsilon}_{i-1}$–$^{13}C'_{i-1}$–$^{15}N_i$	$^{1}J_{NC'}$, $^{3}J_{CC}$

2.2.4 Chemical shift analysis

Chemical shift perturbation is the most widely used NMR method to map protein interfaces [36]. The signals of one protein, generally monitored through 2D experiments such as ^{1}H-^{15}N HSQC or alternatively the CON, are followed upon titration of the partner. The interaction causes environmental changes on the protein interfaces and, hence, affects the chemical shifts of the nuclei in this area. Although shift perturbation measurements reveal which residues are located at an intermolecular interface, they do not in themselves provide information about the relative orientation of the two partners. In addition to the mapping of the interface, NMR titrations provide the information on the affinity, stoichiometry, and specificity of binding as well as the kinetics of binding. How the chemical shifts of the protein change during the titration is determined by the kinetics of the interaction.

If the complex dissociation is very fast compared to the chemical shift difference of a nucleus in the free and bound forms, there is, even during the titration, only a single set of resonances whose chemical shifts are the

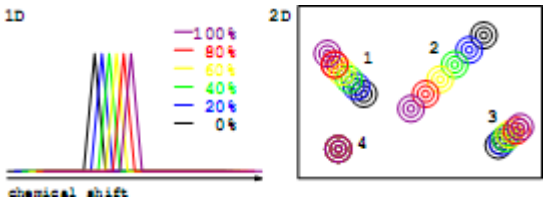


Figure 2.3: Theoretical example of changes in chemical shifts caused by titration of a protein in the fast exchange regime monitored by 1D and 2D (HSQC) spectra.

weighted average of the free and bound chemical shifts. Here the resonances of nuclei at the interface change in a continuous fashion during the titration (Figure 2.3). This regime is referred to as fast chemical exchange and is often observed for weaker interactions.

If the complex dissociation is very slow, one observes one set of resonances for the free protein and one set for the bound protein. During the titration, the 'free set' will decrease the intensity while 'bound set' will increase. Most of the resonances of the two sets will overlap with each other, but the resonances that change chemical shift will indicate the interaction interface. This regime is referred to as slow chemical exchange. In slow exchange one does not automatically know the assignment of the new peaks so it is necessary to perform sequence-specific assignment of the protein in the bound state.

Of course intermediate situations do exist. However in this case extensive line broadening may render signals difficult to observe and the accurate interpretation of the data becomes more difficult.

2.2.5 Determination of structural and dynamic NMR observables

As a result of the sequence-specific assignment, chemical shifts are available which, as it is increasingly becoming evident, do contain structural information. This is indeed used to predict secondary structural elements in well folded proteins and in case of intrinsically disordered modules [64, 65, 66, 67, 68]. The chemical shifts, in particular heteronuclear ones, can provide information on the secondary structure propensities of protein segments. However, the major source of structural information derives from the ^{1}H-^{1}H NOE effect.

NOEs

Traditionally, NMR-based protein structure determination relies on the nuclear Overhauser effect[3] (NOE), an interaction between two nuclei, usually two hydrogen atoms, that arises if the nuclei are sufficiently close in space (usually less than 6–7 Å). Hundreds or thousands of such interactions are typically measured for a well-ordered protein during the structure determination process.

In principle, a protein complex can be treated in exactly the same manner as a single polypeptide chain, for the purposes of structure determination. NOEs between individual subunits in the protein complex can likewise be used to define the structure of a protein complex. Indeed, the 3D structure of each partner in the complex can be determined by exploiting isotopic labelling of one of the two partners to simplify the spectra of each partner if necessary. In this way, with this kind of isotopic labeling strategies, intermolecular (and intramolecular) NOEs can be unambiguously identified.

Recent pulse sequences developed for measuring NOEs in molecular complexes of labelled and unlabelled components utilize building blocks consisting of two critical periods 1) filtering and 2) editing of magnetizations [35, 69]. Period of filtering serves for rejection, for example, a ^{13}C-filter removes the signals of protons attached to ^{13}C from the spectrum, but passes all others. However, filtering the proton magnetization from ^{1}H–^{13}C pair is less efficient because of a wide range of one-bond ^{1}H–^{13}C scalar coupling (120–220 Hz), which can generate spectral artifacts. In contrast, the term 'edit' is used to denote selection of isotope-attached proton magnetization.

In principle, the determination of the structure of a complex requires sufficient experimental data to define both the structures of the individual subunits within the complex and their arrangement relative to each other through intermolecular NOEs. However, in case of weak intermolecular interactions, for a variety of reasons, such a full dataset might not be easily obtainable as a result, for example, of either signal broadening due to chemical exchange or the lack of detectable intermolecular NOE effects. In such situations, data-driven docking procedures can be used to obtain structural models of the complex using structures of the two partners as input [16].

Long-range NOEs, traditionally used for obtaining distance constraints

[3]In most situations, the dominant mechanism through which a nucleus relaxes from the excited state in an NMR experiment involves the interaction of that nucleus with the magnetic dipoles of other nuclei. This so-called dipole-dipole relaxation effect (or NOE) causes magnetization to be transferred from one nucleus to another and is most commonly observed as a cross-peak linking a pair of hydrogen atoms (protons) in a NOESY experiment (Figure 2.4).

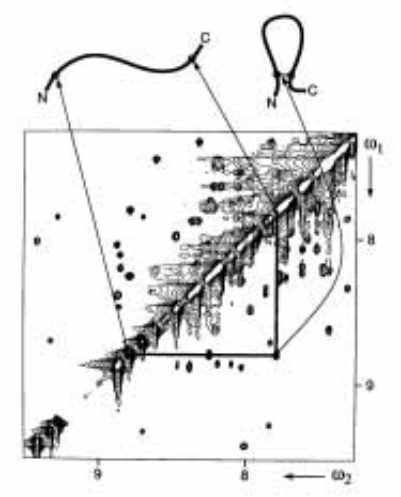


Figure 2.4: 2D NOESY spectrum identifying pairs of protons that are in close proximity.

in well-structured proteins, have proven difficult to observe in disordered proteins. The contacts in disordered proteins may be too short lived to allow for efficient buildup of NOE signals, and the very high degeneracy of side chain chemical shifts makes it very difficult to reliably assign long-range NOEs involving side chain nuclei. Transient structure formation in disordered proteins can be explored at atomic resolution by measuring solvent exposure (Figure 2.5). A useful method to investigate accessibility to solvent is the monitoring of rapid hydrogen exchange using magnetization transfer between amide and solvent protons. CLEANEX method [70] is one example revealing an unambiguous water-NH exchange. In the HSQC map, only NHs exchanging with water will appear providing information on the region exposed to the solvent. Additionally, when CLEANEX spectra are acquired at various mixing times and the reference spectrum is measured as well, the exchange rates from the initial slope analysis can be obtained. Moreover, local variations in backbone dynamics are correlated with propensities for local compaction of the polypeptide chain that results in constriction of backbone motions.

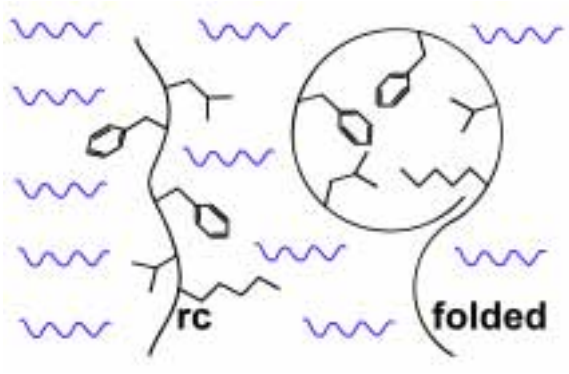


Figure 2.5: Scheme illustrating that the solvent (blue) has free access to all parts of a random-coil polypeptide chain (rc), whereas it is excluded from the core of a folded globular protein.

NMR relaxation

The measurement of ^{15}N relaxation[4] rates (R_1, R_2) and of ^{15}N $\{^1$H$\}$ NOE effects provides a very clean tool to have information on effective local correlation times. These measurements give a very quick and reliable information on:

1. Is the protein structured or unstructured?
2. If structured, which is the effective correlation time? This can be used to see if the protein is monomeric or if it tends to aggregate as well as to follow interactions.
3. For unstructured proteins the ^{15}N $\{^1$H$\}$ NOE effect is a particularly useful indicator of protein fragments characterized by different flexibility.

[4]Relaxation is the process by which the spins return to the thermal equilibrium from a non-equilibrium state. Equilibrium is the state, in which the populations of the energy levels are those predicted by Boltzmann distribution and there is no coherence present in the spin system. Relaxation is the process by which equilibrium is regained by the interaction of the spin system with the molecular environment.

2.3 Structure determination

2.3.1 Structure calculations

A variety of methods have been developed to calculate protein structures using restraints derived from NMR data. These protocols aim to determine coordinates for the protein atoms that will satisfy the input distance and dihedral angle restraints in an unbiased fashion while exploring all regions of conformational space compatible with observed NMR parameters. The structure calculation is repeated many times to determine an ensemble of low-energy structures consistent with the input data.

I have used the program DYANA (CYANA) [71, 72] for structure calculation of proteins. DYANA (DYnamic Algorithm for Nmr Applications) was developed for efficient calculation from distance constraints and torsion angle constraints collected by NMR experiments. The program performs simulated annealing by molecular dynamics in torsion angle space and uses a fast recursive algorithm to integrate the equations of motion. The structure calculation itself is an iterative procedure and the approach may slightly vary according to the system under study. In general, the 'protocol' for structure calculation can be divided into following steps:

- rough estimation of distances from NOE intensities
- the resultant constraints are used for structure calculation
- structure refinement and quantitative analysis of NOEs
- analysis and eventually correction of violating NOEs

As soon as the structure is close to the final one the convergence of the structure calculations becomes faster and target function decreases. NMR structure is represented by a family of conformers which are in good agreement with the structural constraints imposed. The precision of the structure is measured by the root-mean-square-deviation (rmsd) of the coordinates of the protein atoms for each conformer of the family from the mean structure and the accuracy of the structure is measured by an average target function of the family.

The quality of the structure is estimated by programs developed for this purpose [73]. The programs check the stereochemical quality of the protein structures by comparing the geometry of the residues in a given protein structure with stereochemical paramets derived from well-refined, high resolution structures. The commonly used software is PROCHECK [74] (with incorporated module for NMR structures, PROCHECK-NMR), that is intended

as a tool for highlighting regions of a protein structure where the geometry is unusual and hence which may need closer examination. The controlled parameters include correct values for covalent bond lengths and bond angles, the percentage of residues with ϕ/ψ values in the most favored regions of the Ramachandran plot etc.

2.3.2 Data-driven docking

The full structural analysis of a protein–protein complex is generally very time consuming and requires very high quality NOE data. In contrast, data such as chemical shift mapping, RDCs and paramagnetic data can be acquired rapidly and do not even require assignments of the sidechain protons to be made. If the structures of the individual proteins are already known, then data from these approaches can be used to calculate a model structure that is consistent with the data. Several computational approaches have been developed to carry out data-driven docking [75, 76, 77]. Docking algorithms comprise a rigid-body approach in which the two proteins are brought together using a target energy function that includes a term describing the available NMR data.

HADDOCK algorithm [77] incorporates ambiguous interaction restraints (AIRs), which allow chemical shift mapping or mutagenesis data to be used as restraints in the calculation protocol. AIRs are defined between any residue thought to be involved in the interface and all of the putative interfacial residues from the partner protein. HADDOCK begins with a rigid-body docking step, but differs from other approaches in that this step is followed by a stage of simulated annealing. Here, restrained molecular dynamics is carried out, allowing selected sidechains and even regions of the backbone to move freely in order to improve complementarity and electrostatic interactions at the interface.

2.3.3 Residual secondary structure

Secondary structure in disordered proteins is typically transient and confined to short individual helical or extended segments with ensemble-averaged structured populations. Chemical shifts, with their exquisite sensitivity to environment and structure, remain the most powerful tool for detection of such segments. Deviations of observed shifts from those estimated for random coil shifts can be used to assess both the location and population of transient (or fully formed) secondary structure [78].

Correlations between ^{13}C chemical shifts (C^{α}, C^{β}, and C' carbons) and the secondary structure were identified. A method of secondary structure

prediction based on these secondary shifts has been proposed by Wishart [64, 79].

Chapter 3

Results and Discussions

3.1 Project 1: Copper(I)-mediated protein–protein interactions result from suboptimal interaction surfaces

- Paper 1
  Banci, L., I. Bertini, V. Calderone, N. Della-Malva, I. Felli, S. Neri, A. Pavelkova, and A. Rosato: 2009, 'Copper(I)-mediated protein–protein interactions result from suboptimal interaction surfaces'. *Biochemical Journal* **422**, 37–42.

BJ www.biochemj.org

BJ Structure

Biochem. J. (2009) **422**, 37–42 (Printed in Great Britain) doi:10.1042/BJ20090422

online data

Copper(I)-mediated protein–protein interactions result from suboptimal interaction surfaces

Lucia BANCI*†‡, Ivano BERTINI*†[1], Vito CALDERONE*, Nunzia DELLA-MALVA*‡, Isabella C. FELLI*†, Sara NERI*, Anna PAVELKOVA* and Antonio ROSATO*†

*Magnetic Resonance Center (CERM), University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy, †Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy, and ‡FIORGEN Foundation, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

The homoeostasis of metal ions in cells is the result of the contribution of several cellular pathways that involve transient, often weak, protein–protein interactions. Metal transfer typically implies the formation of adducts where the metal itself acts as a bridge between proteins, by co-ordinating residues of both interacting partners. In the present study we address the interaction between the human copper(I)-chaperone HAH1 (human ATX1 homologue) and a metal-binding domain in one of its partners, namely the P-type copper-transporting ATPase, ATP7A (ATPase, Cu^+ transporting, α polypeptide). The adduct was structurally characterized in solution, in the presence of copper(I), and through X-ray crystallography, upon replacing copper(I) with cadmium(II). Further insight was obtained through molecular modelling techniques and site-directed mutagenesis. It was found that the interaction involves a relatively small interface (less than 1000 Å^2, 1 Å = 0.1 nm) with a low fraction of non-polar atoms. These observations provide a possible explanation for the low affinity of the two apoproteins. It appears that electrostatics is important in selecting which domain of the ATPase is able to form detectable amounts of the metal-mediated adduct with HAH1.

Key words: human ATX1 homologue (HAH1), ATPase, Cu^+ transporting, α polypeptide (ATP7A), ATPase, Cu^+ transporting, β polypeptide (ATP7B), Menkes disease, metal homoeostasis.

INTRODUCTION

Many functional processes in the cell occur through weak, transient interactions among proteins. In these cases, the interaction is associated with a small, negative change in free energy. Indeed, if the free energy change were large and negative, the complex would be stable and non-transient, whereas if it were positive, no interaction would occur. The study of these interactions is crucial for the thorough understanding of the mechanisms at the basis of life and of their alterations caused by mutations. Metal homoeostasis in cells, including uptake, controlled distribution to metalloproteins, storage and excretion of metal ions, is a process involving many transient interactions.

Copper is utilized as a cofactor in a variety of proteins [1], but it can be potentially toxic *in vivo* and thus its intracellular concentration is presumably strictly controlled [2–4]. Its traffic and homoeostasis occur through copper-mediated protein–protein interactions [5], where copper(I) is bound to ligands from both proteins of the complex. The free energy of formation of copper(I)-mediated protein–protein adducts is the outcome of the balance of the metal-donor(s) bond energies, of the hydrophobic and hydrophilic interaction energies at the interface and of the entropic implications related to the establishment of the above interactions.

In order to better understand the processes of metal binding and transfer, we have focused on copper trafficking from the cytoplasm to the TGN (*trans*-Golgi network), where copper is inserted in a number of enzymes [6]. In the cytoplasm, copper(I) is bound to a metallochaperone, which delivers the metal to a variable number of soluble MBDs (metal-binding domains) on specific ATPases located at the TGN membrane [2]. In yeast, the metallochaperone is called Atx1 and the ATPase is Ccc2 [7]. In humans, the metallochaperone is called HAH1 (human ATX1 homologue, also known as Atox1) and two target ATPases exist, ATP7A (Cu^+ transporting, α polypeptide) and ATP7B (Cu^+ transporting, β polypeptide) [8]. Both ATP7A and ATP7B have six soluble MBDs. The research in the present study aims at shedding light on the nature of the interaction between the metallochaperone and one MBD of the ATPases within the general frame of understanding the energetics of copper(I)-mediated interactions. With this in mind, we addressed the copper(I)-mediated complex between HAH1 and the first MBD of ATP7A (MNK1). We also investigated whether introducing cadmium(II) instead of copper(I) had an impact on the interaction. The structures of the copper(I)- and cadmium(II)-mediated adducts are reported, the former in solution and the latter as a crystal (whose existence in solution was validated by ^{113}Cd NMR). We also probed some of the key interactions observed in the structures by site-directed mutagenesis.

EXPERIMENTAL

Sample preparation

We cloned and expressed the MNK1 domain, corresponding to amino acids 5–77 of the full-length ATP7A protein, using essentially the same protocol adopted for other individual MNK domains described in [9]. WT (wild-type) and mutant HAH1 protein samples were produced as previously reported in [10]. Further details are supplied in the Supplementary section (see http://www.BiochemJ.org/bj/422/bj4220037add.htm).

Abbreviations used: 1D, one-dimensional; 3D, three-dimensional; ATP7A, ATPase, Cu^+ transporting, α polypeptide; ATP7B, ATPase, Cu^+ transporting, β polypeptide; HAH1, human ATX1 homologue; HSQC, heteronuclear single-quantum coherence; MBD, metal-binding domain; MNK1, first metal-binding domain of ATP7A; NOE, nuclear Overhauser effect; RMSD, root mean square deviation; TGN, *trans*-Golgi network; WT, wild-type.

[1] To whom correspondence should be addressed (email ivanobertini@cerm.unifi.it).

The structure of HAH1–Cu(I)–MNK1 and HAH1–Cd(II)–MNK1 were deposited in the PDB (as 2K1R, including chemical shift data, and 3CJK respectively).

Structural analysis

All NMR experiments were collected at 298 K except on Bruker Avance spectrometers equipped with a triple resonance cryoprobe. The experiments for the sequence-specific assignment of resonances were acquired on the samples in which only one of the partners was ^{13}C,^{15}N labelled, while the other was unlabelled. The 1D (one-dimensional) ^{113}Cd experiment was acquired with a BBO (broadband observe) probe at 600 MHz. Supplementary Tables S1 and S2 (see http://www.BiochemJ.org/bj/422/bj4220037add.htm) report details on the NMR experiments performed. The dynamics of the protein backbone was investigated through ^{15}N relaxation [11] and allowed us to estimate the overall rotational correlation time of the molecules (see Supplementary Table S3 at http://www.BiochemJ.org/bj/422/bj4220037add.htm). Combined NMR chemical shift variations were calculated from the experimental ^{1}H and ^{15}N chemical shift variations [$\Delta\delta(^{1}H)$ and $\Delta\delta(^{15}N)$ respectively] measured between corresponding peaks, through the following equation [12]:

$$\Delta\delta^{\text{combined}} = \sqrt{\frac{(\Delta\delta(^{1}H))^2 + \frac{1}{25}(\Delta\delta(^{15}N))^2}{2}}$$

The intramolecular cross-peaks used for structure calculations were extracted from 3D (three-dimensional) NOESY-HSQC (heteronuclear single-quantum coherence) spectra acquired on samples where the apo-form of one of the partners, enriched in ^{15}N and ^{13}C, was mixed with a 1.5 molar excess of the unlabelled copper(I)-form of the other partner. The intermolecular cross-peaks were integrated in filtered NOESY experiments acquired on samples containing 1:1:1 mixture of copper(I)/enriched partner/unlabelled partner.

Crystals of HAH1–Cd(II)–MNK1 belonging to the spacegroup $P2_12_12_1$ were obtained using the sitting drop vapour diffusion method at 16 °C from a solution containing 0.1 M sodium citrate, 20 % PEG [poly(ethylene glycol)]-6000, 0.8 mM complex concentration, at pH 4.7.

Structural models of all MBDs of the MNK protein were built using the program Modeller v9.5 [13], and then used in protein–protein docking calculations with the program HADDOCK [14].

The experimental structures of HAH1–Cu(I)–MNK1 and HAH1–Cd(II)–MNK1 are available from the PDB (2K1R, including chemical shift data, and 3CJK respectively). The coordinates of the modelled MBD-HAH1 structures are available from http://www.cerm.unifi.it/modstruc/MNKdomains_HAH1modeladducts.zip. Further details on structure determination and analysis are given in the Supplementary section (see http://www.BiochemJ.org/bj/422/bj4220037add.htm).

RESULTS AND DISCUSSION

Formation of the HAH1–Cu(I)–MNK1 and HAH1–Cd(II)–MNK1 adducts

An equimolar mixture of HAH1 and MNK1 of ATP7A formed a metal-mediated adduct with 1:1:1 stoichiometry when presented with Cu^{+}. The complex was in fast (on the NMR timescale) exchange with the free proteins. A similar adduct is presumably formed by HAH1 and MNK4 [15,16], whereas this does not happen with domains 2, 5 and 6 of ATP7A [17,18]. By measuring the reorientation time for molecular tumbling in a solution of MNK1 (5.5 ns) and of HAH1 (4.6 ns), and comparing it with the corresponding value for the complex (8.9 ns), it can be deduced that ~88 % of the protein is involved in the intermolecular complex. In addition, it is known that the relative affinity for copper(I) of the various MBDs of ATP7A is similar, and approximately 10-fold that of HAH1 [17–19]. From these results, it can be estimated that the constant for the equilibrium:

$$\text{Cu(I)-MNK1} + \text{HAH1} \leftrightarrow \text{MNK1-Cu(I)-HAH1}$$

is close to 10^5 M^{-1}.

We additionally investigated the interaction of the HAH1/MNK1 mixture, prepared in the absence of metal ions, with Zn^{2+} and Cd^{2+}. Addition of Zn^{2+} led to extensive sample precipitation, whereas Cd^{2+} induced formation of an adduct at essentially 100 % completion, as shown by the HSQC spectra of both partners in the mixture. 1D ^{113}Cd NMR spectra were recorded to obtain hints on the chemical identity and number of the donor atoms. A shift of 680 p.p.m. was measured (results not shown), which is consistent with the ^{113}Cd environment seen in the crystal structure [20]. The adduct was in slow, rather than fast equilibrium, with the free proteins. HAH1 and MNK6 are known not to form an adduct when mixed at 1:1 ratio in the presence of equimolar copper(I) [17]. Also with Cd^{2+}, an adduct between HAH1 and MNK6 could not be detected. In addition, we did not observe the formation of Cu^{+}- or Cd^{2+}-mediated HAH1 dimers, which have been reported under different experimental conditions [21,22].

The structures of HAH1 and MNK1 within the HAH1–Cu(I)–MNK1 adduct

The structure of the HAH1–Cu(I)–MNK1 complex was solved using a combination of: (i) intramolecular NOEs (nuclear Overhauser effects) and dihedral angle restraints measured on protein samples containing the apo-form of one of the two partners enriched in ^{15}N,^{13}C in the presence of a 1.5 molar excess of the copper(I)-form of the unlabelled partner; (ii) intermolecular NOEs measured on protein samples containing the same proteins as for (i) but at a 1:1 ratio; and (iii) distance restraints between the Sγ atoms of the cysteine residues and the copper(I) ion based on the results of mutagenesis experiments (see section on metal co-ordination). The data from (i) allowed us to obtain the solution structure of each protein when 100 % in the complex, whereas the distance restraints from (ii) originate solely from the complex. This procedure allowed us to confidently assess the structural rearrangements due to complex formation, which has not been possible for the Atx1–Cu(I)–Ccc2a adduct [5]. The structure of the HAH1–Cd(II)–MNK1 complex was solved at 1.8 Å (1 Å = 0.1 nm). Statistics for the two structures are given in Supplementary Tables S4 and S5 respectively (see http://www.BiochemJ.org/bj/422/bj4220037add.htm).

The solution structure of copper(I)–HAH1 (PDB entry 1TL4) [10] could be superimposed very well on to the present structure of HAH1 in the HAH1–Cu(I)–MNK1 adduct (Figure 1), with a backbone RMSD (root mean square deviation) between the two mean structures of 0.95 Å (see Supplementary Table S6 at http://www.BiochemJ.org/bj/422/bj4220037add.htm). Significant differences could be noted at the end of the first β strand for residues 8–10, located before the metal-binding loop and for the region of loop V, encompassing residues 57–60. Other changes regarded the orientation of side chains at the protein–protein interface, e.g. the side chain of Arg^{21} experienced a significant rearrangement to accommodate the partner. In HAH1 alone, the side chain of Lys^{60} rearranges upon metallation, pointing towards the loaded metal site [10]; in the present structures, this side chain points towards the partner molecule, with a conformation similar to that observed in apo-HAH1. The NMR structure of copper(I)–MNK1 (PDB entry 1KVJ) [23] had a backbone RMSD similar to

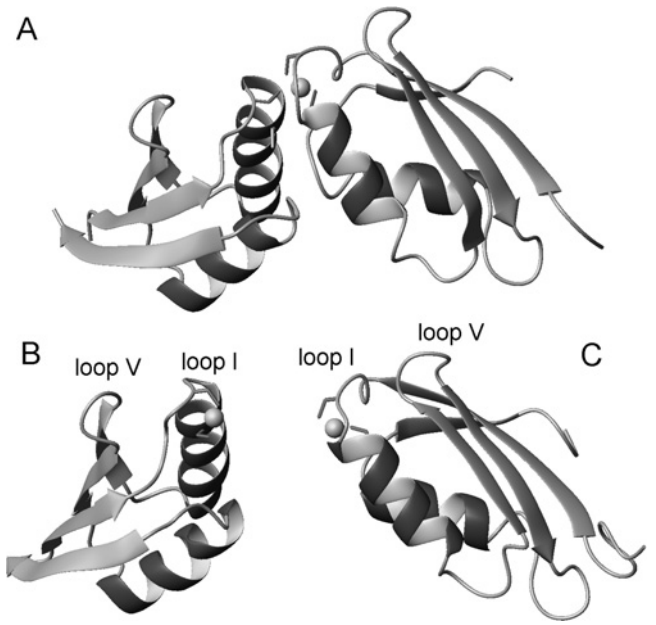


Figure 1 Solution structure of HAH1–Cu(I)–MNK1 complex compared with the structures of the isolated proteins

The solution structures of (**A**) HAH1–Cu(I)–MNK1, (**B**) Cu(I)–HAH1 (PDB code: 1TL4) and (**C**) Cu(I)–MNK1 (PDB code: 1KVJ) are compared. The side chains of the metal-binding cysteines are shown in dark grey. The fourth cysteine, which does not bind the metal in the present structure, is shown in light grey.

the one in the present study of 1.4 Å (see Supplementary Tables S7 and S8 at http://www.BiochemJ.org/bj/422/bj4220037add.htm). Note that 1KVJ had five amino acids at the N-terminus more than the one used in the present study and was solved under different experimental conditions. Significant differences were also observed for the backbone conformation of loops I and V. In addition, a reorientation of helix α_1 was apparent. A rearrangement of the same helix has been reported for yeast Ccc2a upon interaction with its partner [5].

The structures of HAH1–Cu(I)–MNK1 and HAH1–Cd(II)–MNK1

Figures 1 and 2 show respectively the structures of HAH1–Cu(I)–MNK1 and HAH1–Cd(II)–MNK1 determined in the present study. The two structures were similar, with the HAH1–Cu(I)–MNK1 interface being smaller and slightly rotated with respect to HAH1–Cd(II)–MNK1. A list of contacts across the interface for the two complexes is given in Table 1. Most contacts, except those in the metal-binding loop I, are conserved in both structures.

Table 2 reports an analysis of the protein–protein interface for HAH1–Cd(II)–MNK1, HAH1–Cu(I)–MNK1 (shown in Figure 3), yeast Atx1–Cu(I)–Ccc2a, and Cd-HAH1$_2$. Some consistent features are observed: (i) the total interface area is between 800 and 1000 Å^2, i.e. 400–500 Å^2 per partner, covering slightly more than 10 % of the total surface; (ii) the fraction of non-polar atoms at the interface is, on average, between 54 % and 64 %; and (iii) the propensity score of the interfaces is negative. The propensity score is calculated on the basis of the residue composition of the interfaces, using the pre-computed propensity of each amino acid [24]. The latter represents a measure of how much more (positive propensity) or less (negative propensity) frequently a given amino acid is at a protein–protein interface than just on the protein surface. The interface areas were smaller than 'standard' areas for protein–protein complexes, which are

Table 1 Interprotein contacts in the structures HAH1–Cu(I)–MNK1 and HAH1–Cd(II)–MNK1

Short distances (< 4.0 Å) occurring in > 50 % of the family members in the NMR structures of the complexes are reported. (+), positively charged residue; (−), negatively charged residue.

HAH1–Cu(I)–MNK1		HAH1–Cd(II)–MNK1	
Residue in HAH1	Residue in MNK1	Residue in HAH1	Residue in MNK1
Cys^{12}	Cys^{15}		
Cys^{12}	Ser^{17}		
Cys^{12}	Cys^{18}	Cys^{12}	Cys^{18}
		Gly^{14}	Gly^{12}
		Gly^{14}	Thr^{14}
Gly^{14}	Cys^{15}		
		Gly^{14}	Cys^{18}
Cys^{15}	Cys^{18}		
		Ala^{18}	Met^{65}
Ala^{18}	Gly^{66}	Ala^{18}	Gly^{66}
Ala^{18}	Phe^{67}	Ala^{18}	Phe^{67}
		Arg^{21} (+)	Asp^{63} (−)
Arg^{21} (+)	Gly^{66}	Arg^{21} (+)	Gly^{66}
Arg^{21} (+)	Phe^{67}	Arg^{21} (+)	Phe^{67}
Arg^{21} (+)	Asp^{68} (−)	Arg^{21} (+)	Asp^{68} (−)
Val^{22}	Asp^{64} (−)	Val^{22}	Asp^{64} (−)
Val^{22}	Met^{65}	Val^{22}	Met^{65}
		Val^{22}	Gly^{66}
		Lys^{57} (+)	Gln^{25}
Lys^{57} (+)	Asp^{64} (−)	Lys^{57} (+)	Asp^{64} (−)
		Lys^{57} (+)	Met^{65}
Thr^{58}	Thr^{21}	Thr^{58}	Thr^{21}
Thr^{58}	Met^{65}	Thr^{58}	Met^{65}
		Gly^{59}	Gln^{25}
Lys^{60} (+)	Trp^{20}	Lys^{60} (+)	Trp^{20}
Lys^{60} (+)	Thr^{21}	Lys^{60} (+)	Thr^{21}

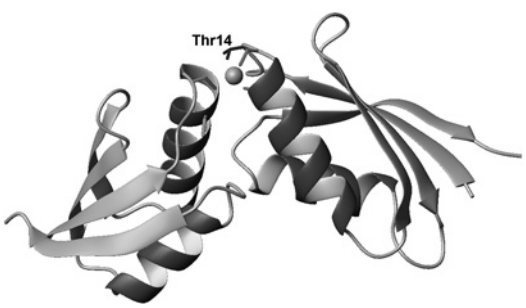


Figure 2 Crystal structure of HAH1–Cd(II)–MNK1

The crystal structure of HAH1–Cd(II)–MNK1 is shown in the same orientation of HAH1–Cu(I)–MNK1 in Figure 1(A). The side chains of the metal-binding cysteines are shown in dark grey, the side chain of metal-binding Thr^{14} is labelled and is shown in black. The fourth cysteine is shown in light grey.

typically at least 1000 Å^2 [24,25]. However, because these are relatively small proteins, the interface/total area ratio is within the lower end of the distribution of values from complexes in the PDB. The fraction of non-polar atoms at the interface was also relatively low. The negative propensity score indicates that the interfaces analysed in the present study were composed mainly of residues that, when at the protein surface, are probably not at the interaction interface. This is due to an over-representation of glycine, alanine and charged residues at the interface, in particular in HAH1 (Table 2). For example, in HAH1–Cd(II)–MNK1 the core of the interface comprised Cys^{12}, Gly^{13}, Lys^{57}, Thr^{58} and Gly^{59} of HAH1, all of which have negative propensities, and Met^{65}, Gly^{66}, Phe^{67} and Asp^{68} of MNK1, of which only methionine

Table 2 Properties of the protein–protein interface of various metal-mediated complexes

Results for the experimental structures were calculated with the program Proface [34].

	HAH1–Cu(I)–MNK1			HAH1–Cd(II)–MNK1			Yeast Atx–Cu(I)–Ccc2a			HAH1 dimer		
Property	HAH1	MNK1	Total	HAH1	MNK1	Total	Atx1	Ccc2a	Total	HAH1 (Chain A)	HAH1 (Chain B)	Total
Interface area (Å^2)	527	490	1017	449	434	883	416	426	842	494	478	972
Interface area / surface area	0.12	0.11	0.12	0.11	0.10	0.10	0.1	0.11	0.11	0.12	0.12	0.12
Number of atoms	54	56	110	50	44	94	46	51	97	51	52	103
Number of residues	16	14	30	15	14	29	15	15	30	15	15	30
Fraction of non-polar atoms	0.61	0.54	0.57	0.70	0.64	0.67	0.59	0.49	0.54	0.59	0.58	0.58
Non-polar interface area (Å^2)	249	210	459	253	219	472	185	161	345	223	218	441
Fraction of fully buried atoms	0.28	0.25	0.26	0.16	0.18	0.17	0.28	0.29	0.29	0.22	0.17	0.19
Residue propensity score	− 2.18	− 0.21	− 2.39	− 1.87	0.46	− 1.41	− 2.32	− 0.17	− 2.49	− 1.61	− 1.67	− 3.28
Local density	35.04	33.86		32.76	28.95		29.91	33.53		31.84	30.92	

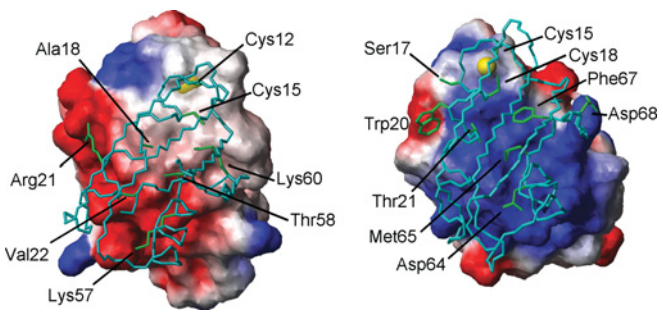


Figure 3 Electrostatic potentials at the protein-protein interface

The electrostatic potential surfaces of MNK1 (left) and HAH1 (right) in the HAH1–Cu(I)–MNK1 structure are shown. The backbone and side chains (as in Table 1) of the partner are also shown.

and phenylalanine are favoured at interfaces. Cumulatively, these results provide a consistent, general description of how the metallochaperone/ATPase pair(s) is designed to prevent the formation of stable adducts in the absence of the metal ion.

We then performed molecular modelling followed by protein–protein flexible docking to create models of the structures of the adducts of all six MBDs of ATP7A with HAH1. We did not include the metal explicitly, as we wanted to evaluate the different energetic properties of the protein–protein interface under the assumption of a constant contribution from copper(I) co-ordination. Table 3 shows a comparison of the calculated energies of interaction at the end of the docking simulation. Note that to enable a direct comparison, the structure of the MNK1–Cu(I)–HAH1 was also recalculated with the same strategy. The cumulative HADDOCK scores identified the first, fourth and fifth MBDs as those for which the interaction should be most favourable. Of these three, only for the first and fourth MBDs was the convergence rate [14] greater than 50 %. When looking at the calculated energies, it appeared that the largest significant difference among all the complexes formed by the six MBDs was in the predicted electrostatic energy of interaction, which was more favourable for the first and fourth MBDs by 150 to 300 kcal · mol^{-1} (Table 3). The differences among other energetic terms were within 10–20 kcal · mol^{-1}. This constitutes a demonstration of the key role of electrostatics in domain selection.

We further characterized the MNK1–Cu(I)–HAH1 complex by mutating some amino acids of HAH1 at the protein–protein interface. Arg^{21}, Val^{22} and Thr^{58} of HAH1 were targeted. Arg^{21} is conserved, or conservatively substituted by lysine, in eukaryotic metallochaperones [26], and is involved in an intermolecular salt-bridge with Asp^{68} of MNK1 (Table 1). A negatively charged residue is commonly, but not strictly, found at position 68 of eukaryotic MBDs. We mutated Arg^{21} to glutamine in order to remove its positive charge. Val^{22} is present, or is conservatively replaced by isoleucine or leucine, in the large majority of all metallochaperones [26]. It is replaced by alanine in the *Caenorhabditis elegans* homologue protein. The side chain of Val^{22} points towards the partner and makes contacts with the backbone of amino acids 64–66 of MNK1. We mutated Val^{22} to alanine. Finally, Thr^{58} is fairly conserved in eukaryotic metallochaperones [26]. Its side chain makes hydrogen bonds with the backbone of Leu^{55}, whereas the backbone atoms are involved in intermolecular contacts. We mutated Thr^{58} to alanine. All three mutants could bind copper(I) and were capable of interacting with MNK1. The mutations of Arg^{21} and Val^{22} had little impact on the structure, as shown by the high similarity of the HSQC spectra of the WT and mutant proteins (results not shown). The strength of the interaction for the WT and mutant HAH1 systems could be straightforwardly evaluated by analysing the combined chemical shift variations obtained at the same HAH1 and MNK1 concentrations (Figure 4). Both the

Table 3 Statistics for the structures of the adducts between each of the six MBDs of ATP7A and HAH1

The results were calculated with the docking program HADDOCK and only the best cluster of solutions was taken into account.

	MNK1	MNK2	MNK3	MNK4	MNK5	MNK6
HADDOCK score	− 103.7 ± 5.7	− 65.4 ± 3.0	− 57.2 ± 4.7	− 103.8 ± 3.2	− 91.2 ± 7.3	− 48.9 ± 3.7
RMSD from the overall lowest-energy structure (Å)	1.2 ± 0.9	0.7 ± 0.5	1.2 ± 0.7	0.8 ± 0.5	0.6 ± 0.4	1.0 ± 0.7
van der Waals energy (kcal · mol^{-1})	− 28.1 ± 6.8	− 36.1 ± 2.8	− 36.4 ± 3.5	− 27.7 ± 0.8	− 46.9 ± 5.5	− 26.7 ± 2.0
Electrostatic energy (kcal · mol^{-1})	− 391 ± 43	− 242 ± 17	− 89 ± 13	− 382 ± 12	− 206 ± 20	− 246 ± 20
Desolvation energy (kcal · mol^{-1})	1.3 ± 4.1	18.4 ± 3.5	− 3.6 ± 3.3	− 0.7 ± 0.9	− 6.6 ± 2.3	26.3 ± 6.3
Restraints violation energy (kcal · mol^{-1})	13 ± 11	7.4 ± 3.6	5.4 ± 3.5	9.6 ± 2.5	35 ± 27	7.1 ± 2.6

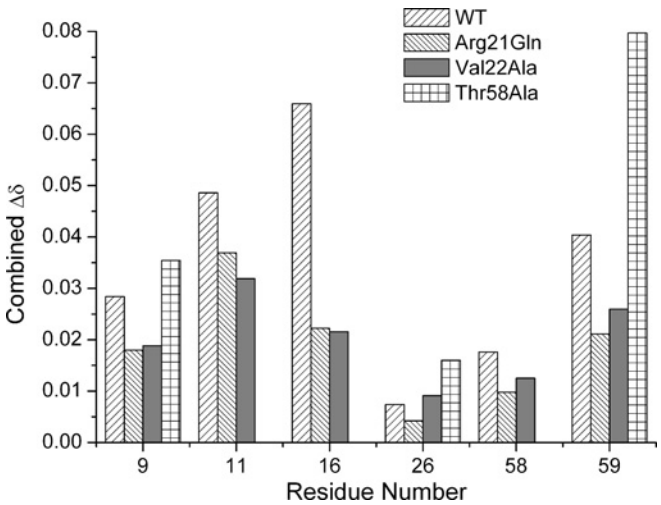


Figure 4 Effect of HAH1 mutations on the formation of the HAH1–Cu(I)–MNK1 complex

Combined chemical shift variations of selected residues for WT-HAH1 and the three interface mutants. Not all peaks could be detected in the Thr58Ala mutant.

$Arg^{21}Gln$ and $Val^{22}Ala$ mutants featured a weaker interaction with the partner; the decrease of the combined chemical shift variations was similar for the two mutants. For $Arg^{21}Gln$, this effect can be ascribed to the loss of the electrostatic interaction. In the case of $Val^{22}Ala$, the reduced interaction can be ascribed to the loss of the favourable contribution due to the desolvation of the hydrophobic side chain of Val^{22} upon complex formation. Instead, the HSQC spectrum of the Thr58Ala mutant was significantly perturbed, and showed the occurrence of multiple protein conformations. In addition, several peaks were broadened beyond detection at the first addition of MNK1 (results not shown). Despite these complications, it appeared that the combined chemical shift variations upon interaction with Cu(I)–HAH1 were as large as that of the WT protein, but the maxima did not correspond to the same amino acids (Figure 4). The HAH1–Cu(I)–MNK1 adduct was thus formed by the $Thr^{58}Ala$ mutant to an extent similar to that of the WT protein, but possibly with a distorted conformation.

Metal co-ordination

We used site-directed mutagenesis to investigate the role of individual cysteine residues in the formation of the HAH1–Cu(I)–MNK1 adducts. To this end, we observed through HSQC spectra whether the metal-mediated adduct was formed upon addition of the metal ion (at equimolar ratio) to a 1:1 mixture containing the WT form of one protein and a mutant of the other protein lacking one of the two copper(I)-binding cysteine residues. Only three cysteine residues were needed for the formation of the HAH1–Cu(I)–MNK1 adduct. The two cysteines of each partner that are more N-terminal in sequence (Cys^{12} of HAH1 and Cys^{15} of MNK1) were strictly required for the interaction, whereas the third cysteine residue could be either Cys^{15} of HAH1 or Cys^{18} of MNK1 (results not shown).

The crystal structure of HAH1–Cd(II)–MNK1 (determined at pH 4.7) provides independent information on Cd^{2+} coordination. The metal ion was tetra-co-ordinated by the side chains of the two cysteines of MNK1 and of Cys^{12} of HAH1. The fourth ligand was provided by the hydroxy group of Thr^{14} of MNK1, leading to an elongated tetrahedral geometry (results not shown). Thus MNK1 provided three ligands and HAH1 only one. The Cd^{2+}-Oγ-Cβ angle was about 150°, suggesting that the hydroxyl group may be deprotonated. Site-directed mutagenesis showed that both cysteines of MNK1 were necessary for the formation of the complex. In the Cd–$HAH1_2$ structure [22], the Cd^{2+} ion is instead at the centre of an essentially regular tetrahedron formed by the four Sγ atoms of the metal-binding cysteines. In both adducts, the net charge at the metal site is -2. Instead, the charge at the more solvent-exposed site in isolated Cu(I)–HAH1 is -1. In addition, upon complex formation the distance from the metal site of the Nε of Lys^{60} of HAH1 increased. These effects could be compensated by the favourable orientation of the dipole moment of the first helix of the incoming MNK1 molecule.

Concluding remarks

The whole body of results on the interaction between metallochaperones and their partner ATPases, as well as studies with copper(I) chelators [27], demonstrate that metal homoeostasis in cells involves a subtle balance between the control of metal reactivity and the capability to transfer the metal between proteins with an appropriate kinetics. The use of metallochaperones to deliver copper to specific intracellular targets has been likened to the action of an enzyme lowering the activation barrier for copper transfer to their specific partners [28]. Additionally, the fine tuning of the identity of the amino acids at the interface may allow the metallochaperone-based system to select among different metal ions. The selection can result from the complete abrogation of the formation of the metal-mediated adduct or from a kinetic 'trap', in which the adduct gets locked [29]. The mutagenesis analysis from the present study indicates that no single amino acidic substitution plays a decisive role. Instead, the network of van der Waals and electrostatic interactions taking place among the interface residues as well as, presumably, the variation of the conformational space available to the same residues determine the configuration of the adduct and its energy of formation. Structural modelling suggested that for the various MBDs of ATP7A, the electrostatic contribution is crucial to determine the energetics of the interaction with the metallochaperone.

Protein recognition takes place only in the presence of the copper(I) ion, as a result of the small interface area and low propensity of the surface of HAH1 for forming a stable intermolecular complex. This, together with the fact that the affinity for copper(I) of the MBDs of ATP7A is approx. 10-fold higher than that of HAH1 [17–19], suggests that upon interaction the metal ion is effectively transferred to the ATPase. However, for some MBDs within the soluble tail, such as MNK1, an intermolecular adduct is formed at detectable levels, involving one cysteine residue from HAH1. Whereas the latter can be regarded as a weakly trapped intermediate along the metal-trafficking pathway, its formation might also be instrumental to determine rearrangements of the inter-domain interactions of the ATPase. The latter inter-domain interactions and their variations during the catalytic cycle are important for the enzymatic function of the protein [30,31], as well as to determine the intracellular localization of the ATPase [32]. The formation of adducts between HAH1 and the first or fourth MBD of ATP7A can thus modulate the interaction of the N-terminal cytosolic tail with the other protein domains, ultimately constituting a means for the cell to modulate enzyme turnover [33].

AUTHOR CONTRIBUTION

Lucia Banci, Ivano Bertini and Antonio Rosato defined the experimental strategy and wrote the article. Nunzia Della-Malva produced the samples used for structure

determinations. Sara Neri produced and analysed HAH1 mutants. Vito Calderone solved the crystallographic structure of HAH1–Cd(II)–MNK1. Isabella Felli and Anna Pavelkova performed the NMR measurements. Anna Pavelkova solved the NMR structure of HAH1–Cu(I)–MNK1.

FUNDING

This work was supported by the Ministero dell'Istruzione, dell'Università e della Ricerca [project number PRIN 2005], the European Commission [contract number 031220, SPINE2-COMPLEXES]; and by the Ente Cassa di Risparmio di Firenze [projects Basi Molecolari di patologie umane correlate a disfunzioni della catena respiratoria and Relazione varianti proteiche strutturali-malattie genetiche]. The HADDOCK web server is supported by the European Commission through the eNMR project [contract number 213010].

REFERENCES

1 Linder, M. C. (1991) Biochemistry of Copper, Plenum Press, New York
2 O'Halloran, T. V. and Culotta, V. C. (2000) Metallochaperones: an intracellular shuttle service for metal ions. J. Biol. Chem. **275**, 25057–25060
3 Harrison, M. D., Jones, C. E., Solioz, M. and Dameron, C. T. (2000) Intracellular copper routing: the role of copper chaperones. Trends Biochem. Sci. **25**, 29–32
4 Puig, S. and Thiele, D. J. (2002) Molecular mechanisms of copper uptake and distribution. Curr. Opin. Chem. Biol. **6**, 171–180
5 Banci, L., Bertini, I., Cantini, F., Felli, I. C., Gonnelli, L., Hadjiliadis, N., Pierattelli, R., Rosato, A. and Voulgaris, P. (2006) The Atx1-Ccc2 complex is a metal-mediated protein-protein interaction. Nat. Chem. Biol. **2**, 367–368
6 Yuan, D. S., Stearman, R., Dancis, A., Dunn, T., Beeler, T. and Klausner, R. D. (1995) The Menkes/Wilson disease gene homologue in yeast provides copper to a ceruloplasmin-like oxidase required for iron uptake. Proc. Natl. Acad. Sci. U.S.A. **92**, 2632–2636
7 Lin, S. J., Pufahl, R., Dancis, A., O'Halloran, T. V. and Culotta, V. C. (1997) A role for the *Saccharomyces cerevisiae* ATX1 gene in copper trafficking and iron transport. J. Biol. Chem. **272**, 9215–9220
8 Hamza, I., Schafer, M., Klomp, L. W. and Gitlin, J. D. (1999) Interaction of the copper chaperone HAH1 with the Wilson disease protein is essential for copper homeostasis. Proc. Natl. Acad. Sci. U.S.A. **96**, 13363–13368
9 Banci, L., Bertini, I., Del Conte, R., D'Onofrio, M. and Rosato, A. (2004) Solution structure and backbone dynamics of the Cu(I) and apo-forms of the second metal-binding domain of the Menkes protein ATP7A. Biochemistry **43**, 3396–3403
10 Anastassopoulou, J., Banci, L., Bertini, I., Cantini, F., Katsari, E. and Rosato, A. (2004) Solution structure of the apo-and copper(I) loaded human metallo-chaperone HAH1. Biochemistry **43**, 13046–13053
11 Ishima, R. and Torchia, D. A. (2000) Protein dynamics from NMR. Nature Struct. Biol. **7**, 740–743
12 Garrett, D. S., Seok, Y. J., Peterkofsky, A., Clore, G. M. and Gronenborn, A. M. (1997) Identification by NMR of the binding surface for the histidine-containing phosphocarrier protein HPr on the N-terminal domain of enzyme I of the escherichia coli phosphotransferase system. Biochemistry **36**, 4393–4398
13 Sali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. **234**, 779–815
14 Dominguez, C., Boelens, R. and Bonvin, A. M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. **125**, 1731–1737
15 Banci, L., Bertini, I., Cantini, F., Chasapis, C., Hadjiliadis, N. and Rosato, A. (2005) A NMR study of the interaction of a three-domain construct of ATP7A with copper(I) and copper(I)-HAH1: the interplay of domains. J. Biol. Chem. **280**, 38259–38263
16 Banci, L., Bertini, I., Cantini, F., Della Malva, N., Migliardi, M. and Rosato, A. (2007) The different intermolecular interactions of the soluble copper-binding domains of the Menkes protein, ATP7A. J. Biol. Chem. **282**, 23140–23146
17 Banci, L., Bertini, I., Cantini, F., Migliardi, M., Rosato, A. and Wang, S. (2005) An atomic level investigation of the disease-causing A629P mutant of the Menkes protein ATP7A. J. Mol. Biol. **352**, 409–417
18 Banci, L., Bertini, I., Chasapis, C., Ciofi-Baffoni, S., Hadjiliadis, N. and Rosato, A. (2005) An NMR study of the interaction between the human copper(I) chaperone and the second and fifth metal-binding domains of the Menkes protein. FEBS J. **272**, 865–871
19 Yatsunyk, L. A. and Rosenzweig, A. C. (2007) Copper(I) binding and transfer by the N-terminus of the Wilson disease protein. J. Biol. Chem. **282**, 8622–8631
20 Oz, G. L., Pountney, D. L. and Armitage, I. M. (1998) NMR spectroscopy studies of I = 1/2 metal ions in biological systems. Biochem. Cell. Biol. **76**, 223–234
21 Tanchou, V., Gas, F., Urvoas, A., Cougouluegne, F., Ruat, S., Averseng, O. and Quemeneur, E. (2004) Copper-mediated homo-dimerisation for the HAH1 metallochaperone. Biochem. Biophys. Res. Commun. **325**, 388–394
22 Wernimont, A. K., Huffman, D. L., Lamb, A. L., O'Halloran, T. V. and Rosenzweig, A. C. (2000) Structural basis for copper transfer by the metallochaperone for the Menkes/Wilson disease proteins. Nat. Struct. Biol. **7**, 766–771
23 DeSilva, T. M., Veglia, G. and Opella, S. J. (2005) Solution structures of the reduced and Cu(I) bound forms of the first metal binding sequence of ATP7A associated with Menkes disease. Proteins **61**, 1038–1049
24 Bahadur, R. P., Chakrabarti, P., Rodier, F. and Janin, J. (2004) A dissection of specific and non-specific protein-protein interfaces. J. Mol. Biol. **336**, 943–955
25 Lo Conte, L., Chothia, C. and Janin, J. (1999) The atomic structure of protein-protein recognition sites. J. Mol. Biol. **285**, 2177–2198
26 Arnesano, F., Banci, L., Bertini, I., Ciofi-Baffoni, S., Molteni, E., Huffman, D. L. and O'Halloran, T. V. (2002) Metallochaperones and metal transporting ATPases: a comparative analysis of sequences and structures. Genome Res. **12**, 255–271
27 Hussain, F., Olson, J. S. and Wittung-Stafshede, P. (2008) Conserved residues modulate copper release in human copper chaperone Atox1. Proc. Natl. Acad. Sci. U.S.A. **105**, 11158–11163
28 Huffman, D. L. and O'Halloran, T. V. (2000) Energetics of copper trafficking between the Atx1 metallochaperone and the intracellular copper-transporter, Ccc2. J. Biol. Chem. **275**, 18611–18614
29 Liu, T., Reyes-Caballero, H., Li, C., Scott, R. A. and Giedroc, D. P. (2007) Multiple metal binding domains enhance the Zn(II) selectivity of the divalent metal ion transporter AztA. Biochemistry **46**, 11057–11068
30 Lutsenko, S., Barnes, N. L., Bartee, M. Y. and Dmitriev, O. Y. (2007) Function and regulation of human copper-transporting ATPases. Physiol. Rev. **87**, 1011–1046
31 Argüello, J. M., Eren, E. and Gonzalez-Guerrero, M. (2007) The structure and function of heavy metal transport P1B-ATPases. Biometals **20**, 233–248
32 Hamza, I., Prohaska, J. and Gitlin, J. D. (2003) Essential role for Atox1 in the copper-mediated intracellular trafficking of the Menkes ATPase. Proc. Natl. Acad. Sci. U.S.A. **100**, 1215–1220
33 Mandal, A. K. and Argüello, J. M. (2003) Functional roles of metal binding domains of the *Archaeoglobus fulgidus* Cu^{+}-ATPase CopA. Biochemistry **42**, 11040–11047
34 Saha, R. P., Bahadur, R. P., Pal, A., Mandal, S. and Chakrabarti, P. (2006) ProFace: a server for the analysis of the physicochemical features of protein-protein interfaces. BMC Struct. Biol. **6**, 11

Received 13 March 2009/12 May 2009; accepted 19 May 2009
Published as BJ Immediate Publication 19 May 2009, doi:10.1042/BJ20090422

Biochem. J. (2009) **422**, 37–42 (Printed in Great Britain) doi:10.1042/BJ20090422

online data

SUPPLEMENTARY ONLINE DATA

Copper(I)-mediated protein–protein interactions result from suboptimal interaction surfaces

Lucia BANCI*†‡, Ivano BERTINI*†[1], Vito CALDERONE*, Nunzia DELLA-MALVA*‡, Isabella C. FELLI*†, Sara NERI*, Anna PAVELKOVA* and Antonio ROSATO*†

*Magnetic Resonance Center (CERM), University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy, †Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy, and ‡FIORGEN Foundation, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

EXPERIMENTAL

Sample preparation

We cloned and expressed the MNK1 domain of ATP7A, corresponding to amino acids 5–77 of the full-length protein using essentially the same protocol adopted for other individual MNK domains [1]. WT and mutant HAH1 protein samples were produced as previously reported in [2]. The expression plasmids for all mutants were obtained from the Quikchange site-directed mutagenesis kit (Stratagene).

Apoprotein samples were reduced with excess DTT (dithiothreitol) and washed with 50 mM sodium phosphate buffer, pH 7.0. The final molar ratio of DTT to protein was typically 2–5:1. Copper(I)-containing protein samples were prepared by incubating the apoprotein, after removal of DTT, with a slight excess of the acetonitrile complex of copper(I). Sample preparations were carried out in a nitrogen atmosphere chamber (Coy Lab). The final protein concentration for all the NMR samples used for experiments aimed at structure determination was around 0.6 mM, whereas the concentration of samples for HSQC titrations was around 0.2 mM.

The formation of the complex was investigated by adding, under a nitrogen atmosphere, increasing amounts of one of the partners in its apoform to the copper(I)-form of the other partner. In the various experiments, one of the two proteins was enriched in ^{15}N or in ^{15}N and ^{13}C, while the other was not enriched. For interface mutants, MNK1 was added to mutant, ^{15}N-enriched copper(I)-HAH1. The reaction was monitored through ^{15}N-^{1}H-HSQC spectra.

NMR experiments

All NMR experiments were collected at 298 K, except when otherwise indicated, on Bruker Avance spectrometers equipped with a triple resonance cryoprobe. The experiments for the sequence-specific assignment of resonances were acquired on samples in which only one of the partners were ^{13}C,^{15}N labelled, while the other was unlabelled. The 1D ^{113}Cd experiment was acquired with a BBO probe at 600 MHz. Tables S1 and S2 report details on the NMR experiments performed. The dynamics of the protein backbone was investigated through ^{15}N relaxation [3] and allowed estimation of the overall rotational correlation time of the molecules (Table S3).

Combined NMR chemical shift variations were calculated from the experimental ^{1}H and ^{15}N chemical shift variations ($\Delta\delta(^{1}\mathrm{H})$ and $\Delta\delta(^{15}\mathrm{N})$ respectively) measured between corresponding peaks, through the following equation [4]:

$$\Delta\delta^{\mathrm{combined}} = \sqrt{\frac{(\Delta\delta(^{1}\mathrm{H}))^2 + \frac{1}{25}(\Delta\delta(^{15}\mathrm{N}))^2}{2}}$$

Solution structure calculations

The intramolecular cross-peaks used for structure calculations were extracted from 3D NOESY-HSQC spectra acquired on samples where the apoform of one of the partners, enriched in ^{15}N and ^{13}C, was mixed with a 1.5 molar excess of the unlabelled copper(I)-form of the other partner. The intermolecular cross-peaks were integrated in filtered NOESY experiments acquired on samples containing 1:1:1 mixture of copper(I)/enriched partner/unlabelled partner.

The computational approach was the same used previously for the structural characterization of the yeast Atx1–Cu(I)–Ccc2a adduct in [5]. Structure calculations were performed with the program CYANA-2.1 [6]. The intensities of cross-peaks were converted into upper distance limits through the standard procedure available in CYANA. The data from different spectra were treated independently. Φ and ψ dihedral angle restraints were derived from the chemical shift index [7]. The MNK1 and HAH1 were connected through a linker of 228 dummy residues with null van der Waals radius. An additional linker of 111 residues was used to include the copper(I) ion. The copper(I) ion was 'bound' to the Sγ atoms of the cysteine residues, identified from mutagenesis experiments as being necessary for complex formation by imposing upper distance limits of 2.30 Å. The 20 conformers with the lowest CYANA target function out of 100 calculated were subjected to molecular dynamics refinement in explicit water using the program AMBER, version 8.0 [8].

Crystallization, data collection and X-ray structure determination

Crystals of HAH1–Cd(II)–MNK1 belonging to the spacegroup $P2_12_12_1$ were obtained using the sitting drop vapour diffusion method at 16 °C from a solution containing 0.1 M sodium citrate, 20 % PEG-6000, 0.8mM complex concentration, at pH 4.7.

We collected data at the ESRF (European Synchrotron Radiation Facility, Grenoble, France), beamline ID23–1. The dataset was collected at 100 K and the crystals used for data collection were cryo-cooled using a solution containing 15 % ethylene glycol in the mother liquor. The crystal diffracted to 1.8 Å resolution, and had one molecule of complex in the asymmetric unit, a solvent content of about 50 % and a mosaicity of about 0.8°.

[1] To whom correspondence should be addressed (email ivanobertini@cerm.unifi.it).

Table S1 Acquisition parameters for NMR experiments performed on HAH1–Cu(I)–MNK1 containing ^{13}C,^{15}N-MNK1

The magnetic field strength of the instrument used for the measurements is indicated by using the relative ^1H frequency, B_0 reports the operating ^1H frequency of the spectrometer. d_1 is the recycle delay. Experiments for the measurement of ^{15}N relaxation rates (^{15}N R_1 and ^{15}N R_2) were acquired on ^{15}N labelled samples.

	Dimension of acquired data			Spectral width (p.p.m.)					
Experiments	F3	F2	F1	F3	F2	F1	Number of scans	d_1 (s)	B_0 (MHz)
^1H-^{15}N-HSQC		1024 (^1H)	128 (^{15}N)		11.984	50.019	2	1.5	900
^1H-^{13}C-HSQC		1024 (^1H)	512 (^{13}C)	13.950	80.000		4	1.3	900
CBCANH	1024 (^1H)	48 (^{15}N)	96 (^{13}C)	13.019	41.000	80.011	32	1.3	800
CBCA(CO)NH	1024 (^1H)	48 (^{15}N)	96 (^{13}C)	13.019	41.000	80.011	16	1.3	800
HNCO	1024 (^1H)	48 (^{15}N)	64 (^{13}C)	13.019	41.000	16.031	4	1.3	800
^1H-^{13}C-HSQC$_{arom}$		2048 (^1H)	256 (^{13}C)		17.983	100.000	4	1.5	900
^{15}N-edited [^1H-^1H]-NOESY	2048 (^1H)	1 (^{15}N)	700 (^1H)	17.983	41.227	18.041	64	1.5	900
^{13}C-edited [^1H-^1H]-NOESY	2048 (^1H)	1 (^{13}C)	700 (^1H)	17.982	160.211	18.041	64	1.5	900
^{15}N R_1		1024 (^1H)	96 (^{15}N)		15.019	35.000	128	1.5	400
^{15}N R_2		1024 (^1H)	96 (^{15}N)		15.019	35.000	128	1.5	400
HCCH TOCSY	2048 (^1H)	64 (^{13}C)	224 (^{13}C)	13.949	75.078	75.078	8	1.4	800
(HB)CB(CGCD)HD		1024 (^1H)	110 (^{13}C)		10.014	35.002	256	1.4	800
[^1H-^1H]-NOESY$_{all}$ 2D		2048 (^1H)	800 (^1H)		19.988	19.988	64	1.5	900
^{13}C-edited-NOESY$_{inter}$ 3D	2048 (^1H)	256 (^1H)	48 (^{13}C)	18.036	18.000	80.000	64	1.5	700
^{13}C-edited [^1H-^1H]-NOESY 3D	2048 (^1H)	64 (^{13}C)	256 (^1H)	18.036	80.000	18.036	8	1.5	800
^{15}N-edited [^1H-^1H]-NOESY 3D	2048 (^1H)	48 (^{15}N)	180 (^1H)	15.022	38.064	15.022	16	1.3	800

Table S2 Acquisition parameters for NMR experiments performed on HAH1–Cu(I)–MNK1 containing ^{13}C,^{15}N-HAH1

The magnetic field strength of the instrument used for the measurements is indicated by using the relative ^1H frequency, B_0 reports the operating ^1H frequency of the spectrometer. d_1 is the recycle delay. Experiments for the measurement of ^{15}N relaxation rates (^{15}N R_1 and ^{15}N R_2) were acquired on ^{15}N labelled samples.

	Dimension of acquired data			Spectral width (p.p.m.)					
Experiments	F3	F2	F1	F3	F2	F1	Number of scans	d_1 (s)	B_0 (MHz)
^1H-^{15}N-HSQC		1024 (^1H)	128 (^{13}C)		17.954	60.003	2	1.5	800
^1H-^{13}C-HSQC		1024 (^1H)	512 (^{13}C)		13.950	80.000	4	1.3	700
CBCANH	1024 (^1H)	48 (^{15}N)	96 (^{13}C)	13.019	41.000	80.011	32	1.3	800
CBCA(CO)NH	1024 (^1H)	48 (^{15}N)	96 (^{13}C)	13.019	41.000	80.011	16	1.3	800
HNCO	1024 (^1H)	48 (^{15}N)	64 (^{13}C)	13.019	41.000	16.031	4	1.3	800
^1H-^{13}C-HSQC$_{arom}$	2048 (^1H)	256 (^{13}C)			17.983	100.000	4	1.5	800
^{15}N-edited [^1H-^1H]-NOESY	2048 (^1H)	1 (^{15}N)	800 (^1H)	15.022	38.064	15.022	56	1.5	800
^{13}C-edited [^1H-^1H]-NOESY	2048 (^1H)	600 (^1H)	1 (^{13}C)	16.935	16.935	80.1639	96	1.5	800
^{15}N R_1		1024 (^1H)	96 (^{15}N)		15.019	35.000	128	1.5	400
^{15}N R_2		1024 (^1H)	96 (^{15}N)		15.019	35.000	128	1.5	400
HCCH TOCSY	2048 (^1H)	86 (^{13}C)	256 (^{13}C)	18.036	75.141	75.141	8	1.3	700
[^1H-^1H]-NOESY$_{all}$ 2D		1024 (^1H)	1024 (^1H)		16.022	16.022	32	1.5	500
^{13}C-edited-NOESY$_{inter}$ 3D	2048 (^1H)	256 (^1H)	48 (^{13}C)	18.036	18.000	80.004	16	1.5	700
^{13}C-edited [^1H-^1H]-NOESY 3D	1024 (^1H)	64 (^{13}C)	200 (^1H)	15.022	86.289	15.022	16	1.3	800
^{15}N-edited [^1H-^1H]-NOESY 3D	2048 (^1H)	48 (^{15}N)	192 (^1H)	15.022	38.064	15.022	16	1.5	800
[^1H-^1H]-NOESY$_{inter}$ 2D	2048 (^1H)	600 (^1H)	1 (^{13}C)	18.036	18.000	80.000	64	1.5	700

Table S3 Acquisition parameters for NMR experiments performed on the equimolar mixture of HAH1/MNK1

The 1D ^1H NMR experiment was acquired with standard parameters, and the 2D ^1H-^{15}N-HSQC experiments were repeated at different Cd(II) concentrations to monitor complex formation. For the 1D ^{113}Cd spectrum, $Cd(ClO_4)_2$ was used as a standard for NMR referencing (0 p.p.m. position in the ^{113}Cd spectrum).

	Dimension of acquired data		Spectral width (p.p.m.)			
Experiments	F2	F1	F2	F1	Number of scans	d_1 (s)
^1H-^{15}N-HSQC (^{15}N NMK1, unlabelled HAH1) at 500 MHz, TCI probe	1024 (1H)	128 (15N)	15.011	37.943	8	1.00
^1H-^{15}N-HSQC (^{15}N HAH1, unlabelled MNK1) at 500 MHz, TCI probe	1024 (1H)	128 (15N)	16.935	40.015	16	1.20
1D 113Cd (both partners unlabelled) at 600 MHz, BBO probe		16384 (113Cd)		600	212992	1.0

Table S4 Measured values of the correlation time (τ_c) for protein reorientation for MNK1 in solution, isolated HAH1 and HAH1 in the HAH1–Cu(I)–MNK1 complex

	Apo-MNK1	MNK1 in HAH1-Cu(I)-MNK1	Apo-HAH1	HAH1 in HAH1–Cu(I)–MNK1
τ_c (per ns)	5.5	8.2	4.5	8.6

Table S5 Statistical analysis of the solution structure of HAH1–Cu(I)–MNK1

The number of meaningful constraints for each class is reported in parentheses. The structural analysis is based on results from the Ramachandran plot analysis.

Property	REM (energy-minimized average of 20 structures)
RMS violations per meaningful distance constraint (Å)	
Intraresidue (452)	0.0105 ± 0.0029
Sequential (557)	0.0072 ± 0.0025
Medium range (413)*	0.0084 ± 0.0016
Long range (784)	0.0083 ± 0.0018
Total (2206)	0.0088 ± 0.0010
RMS violations per meaningful dihedral angle constraints (deg)	
Phi (91)	0.65 ± 0.42
Psi (91)	0.79 ± 0.47
Average number of constraints per residue	16
Average number of violations per structure	
Intraresidue	3.35 ± 1.49
Sequential	3.60 ± 1.43
Medium range	3.90 ± 1.34
Long range	7.95 ± 2.56
Phi	2.70 ± 2.17
Psi	1.50 ± 0.92
Average no. of NOE violations larger than 0.3 Å	0.00 ± 0.00
Average no. of NOE violations between 0.1 Å and 0.3 Å	4.65 ± 1.80
Structural analysis	
% of residues in most favourable regions	88.2
% of residues in allowed regions	10.7
% of residues in generously allowed regions	0.9
% of residues in disallowed regions	0.2

*Medium range distance constraints are those between residues (i,i + 2), (i,i + 3), (i,i + 4) and (i,i + 5).

Table S6 X-ray data collection and refinement statistics

Numbers in parentheses refer to the high resolution shell.

	HAH1–Cd(II)–MNK1
Space group	$P2_12_12_1$
Cell dimensions (Å, °)	a = 47.71, b = 55.27, c = 63.24, $\alpha = \beta = \gamma = 90$
Resolution (Å)	26.3 – 1.8
Unique reflections	16071 (2307)
Overall completeness (%)	99.9 (99.9)
Rsym (%)	7.2 (40.8)
Multiplicity	13.7 (14.1)
I/(σI)	5.4 (2.0)
B-factor from Wilson plot (Å^2)	26.10
R_{cryst} / R_{free} (%)	22.5 / 28.2
Protein atoms	1093
Ions	1
Water molecules	75
RMSD bonds (Å)	0.02
RMSD angles (°)	2.18
Average B-factor (Å^2)	28.74

Table S7 Backbone RMSD values (Å) for HAH1 in the various structures available

For solution structures, the mean structure was used. For the HAH1 dimer structure, the first model was used.

	apo-HAH1	Cu(I)–HAH1	MNK1–Cu(I)–HAH1 (NMR)	MNK1–Cd(II)–HAH1 (X-ray)	HAH1–Cd(II)–HAH1 (X-ray)
apo-HAH1		0.64	0.93	0.95	0.98
Cu(I)–HAH1	0.64		0.95	0.82	0.88
MNK1–HAH1 (NMR)	0.93	0.95		0.68	0.66
MNK1–HAH1 (X-ray)	0.85	0.82	0.68		0.33
MNK1–Cd(II)–HAH1 (X-ray)	0.98	0.88	0.66	0.33	

Table S8 Backbone RMSD values (Å) for MNK1 in the various structures available

For solution structures, the mean structure was used.

	apo-MNK1	Cu(I)–MNK1	MNK1–Cu(I)–HAH1 (NMR)	MNK1–Cd(II)–HAH1 (X-ray)
apo-MNK1		1.58	1.59	1.51
Cu(I)–MNK1			1.38	0.99
MNK1–HAH1 (NMR)	1.59	1.38		1.21
MNK1–HAH1 (X-ray)	1.51	0.99	1.21	

The data was processed using the program MOSFLM [9] and scaled using the program SCALA [10] with the TAILS and SECONDARY corrections on (the latter restrained with a TIE SURFACE command) to achieve an empirical absorption correction. The correct orientation and position of the two protein molecules within the asymmetric unit was determined with the program MOLREP [11]. The isotropic refinement was carried out using REFMAC5 [12], except for the cadmium atom for which anisotropy was taken into account. In between the refinement cycles the model was subjected to manual rebuilding. Water molecules were added using the ARP/WARP software suite (http://www.embl-hamburg.de/ARP/).

The quality of the solution and crystal structures was evaluated using the programs PROCHECK [13,14] and WHATCHECK [15].

Modelling of the adduct of Cu(I)-HAH1 with the various MNK domains

Structural models of all MBDs of the MNK protein were built using the program Modeller v9.5 [16]. Owing to the high sequence similarity among the domains, the default modelling procedure could be applied successfully. The second to the sixth MBDs were modelled using the structure of MNK1 as the template. To enable a fair comparison of the subsequent docking calculations, we also generated a model for the structure of MNK1, to be used as input to HADDOCK (see below), using the modelled structure of the sixth MBD as the template.

For docking calculations, was used the HADDOCK web server [17]. We ran six docking simulations, in which the same structure of HAH1 was docked to the six structural models of the MBDs of ATP7A generated with Modeller. The active and passive residues [17] of the domains were defined by transferring the information experimentally obtained for MNK1 in the HAH1–Cu(I)–MNK1 adduct to the corresponding amino acids in the alignments of

MNK1 to the other MBDs. The copper(I) ion was not explicitly defined in calculations, as we were interested in comparing the energetic contribution due to the protein–protein interaction only. To properly orient the side chains of the cysteine residues, the corresponding $S\gamma$–$S\gamma$ distances were restrained to be within 4.2 Å.

REFERENCES

1 Banci, L., Bertini, I., Del Conte, R., D'Onofrio, M. and Rosato, A. (2004) Biochemistry **43**, 3396–3403
2 Anastassopoulou, J., Banci, L., Bertini, I., Cantini, F., Katsari, E. and Rosato, A. (2004) Solution structure of the apo and copper(I)-loaded human metallochaperone HAH1. Biochemistry **43**, 13046–13053
3 Ishima, R. and Torchia, D. A. (2000) Protein dynamics from NMR. Nature Struct. Biol. **7**, 740–743
4 Garrett, D. S., Seok, Y. J., Peterkofsky, A., Clore, G. M. and Gronenborn, A. M. (1997) Identification by NMR of the binding surface for the histidine-containing phosphocarrier protein HPr on the N-terminal domain of enzyme I of the *Escherichia coli* phosphotransferase system. Biochemistry **36**, 4393–4398
5 Banci, L., Bertini, I., Cantini, F., Felli, I. C., Gonnelli, L., Hadjiliadis, N., Pierattelli, R., Rosato, A. and Voulgaris, P. (2006) The Atx1-Ccc2 complex is a metal-mediated protein-protein interaction. Nat. Chem. Biol. **2**, 367–368
6 Herrmann, T., Güntert, P. and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J. Mol. Biol. **319**, 209–227

Received 13 March 2009/12 May 2009; accepted 19 May 2009
Published as BJ Immediate Publication 19 May 2009, doi:10.1042/BJ20090422

7 Wishart, D. S., Sykes, B. D. and Richards, F. M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry **31**, 1647–1651
8 Case, D. A., Cheatham, III, T. E., Darden, T., Gohlke, H., Luo, R., Merz, Jr, K. M., Onufriev, A., Simmerling, C., Wang, B. and Woods, R. (2005) The AMBER biomolecular simulation programs. J. Computat. Chem. **26**, 1668–1688
9 Leslie, A. G. W. (1991) Molecular data processing. In Molecular data processing (Moras, D., Podjarny, A. D. and Thierry, J.-C., eds), pp. 50–61, Oxford University Press, Oxford
10 Evans, P. R. (1997) Joint CCP4 and ESF-EACBM Newsletter. Protein Crystallogr. **33**, 22–24
11 Vagin, A. and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement. Acta Crystallogr. D. Biol. Crystallogr. **56**, 1622–1624
12 Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D. Biol. Crystallogr. **53**, 240–255
13 Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemistry of protein structures. J. Appl. Crystallogr. **26**, 283–291
14 Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR **8**, 477–486
15 Vriend, G. (1990) What if: a molecular modeling and drug design program. J. Mol. Graphics **8**, 52–56
16 Sali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. **234**, 779–815
17 Dominguez, C., Boelens, R. and and Bonvin, A. M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J. Am. Chem. Soc. **125**, 1731–1737

3.2 Project 2: Protein interfaces determining metal-transfer pathways

- Paper 2 (advanced draft to be submitted)
  Banci, L., I. Bertini, I. Felli, A. Pavelkova, N. Robinson and K. Waldron

Introduction

Living organisms have developed processes which can selectively handle metal transport and incorporation into their target proteins. This may be accomplished by specific trafficking pathways [1, 2, 3, 4, 5]. For example, copper is coordinated by a specific protein, as it enters the cell, which takes it to a selected unique partner protein, either as final metal recipient or as one of the steps in the metal transfer. This occurs through protein–protein interactions mediated by the metal ion. Two partner proteins involved in metal transfer processes interact only in the presence of and through the metal ion [6, 7], while, when metal-free, they do not interact thus preventing non productive complexes [6, 8]. Shared metal coordination together with molecular recognition between the two partner proteins drive the formation of protein–protein complexes which eventually evolve with the transfer of the metal ion from the protein with lower affinity to that with higher [9] or similar affinity.

Several proteins involved in different metal transport pathways feature similar metal-coordination sites [8]. This in principle might lead to the coordination of the wrong metal ion. And indeed *in vitro* some proteins exhibit a higher affinity for a non physiological metal ion [10].

Proteins involved in the homeostasis of Cu(I) and their mutual interactions have been recently extensively characterized [6, 7, 11, 12, 13, 14]. In particular, metal-mediated interactions through which Cu(I) is transferred from the copper chaperone to one of the soluble domains of membrane-bound ATPases, ensure reliable copper delivery to the target proteins.

It may happen that the overall 3D folds and/or metal-binding sites of the proteins devoted to controlling Cu(I) trafficking are all very similar one to the other and also similar to those involved in the homeostasis of other metal ions, such as Zn(II). However, the directionality for each metal must be very selective as the incorporation of a non natural metal ion in a protein may cause altered cell physiology. In order to better understand the subtle features through which the metal transfer pathway is controlled, we decided to focus on the factors modulating the interaction between a metal chaperone and two different ATPases within the cyanobacterial system *Synechocystis* PCC 6803.

Cyanobacteria are a bacterial group that have known enzymatic demand for cytoplasmic copper import. They contain internal membrane-bound compartments called thylakoids, which constitute the site of photosynthetic reactions and respiratory electron transport involving copper proteins such as plastocyanin and caa_3-type cytochrome oxidase. Copper supply into thylakoids is needed to enable the formation of fully functional holoforms of copper proteins for photosyntesis.

The proposed mechanism of copper import into thylakoids *Synechocystis* PCC 6803 consists of two copper transporting P-type ATPases, CtaA and PacS [15], together with the small soluble copper metallochaperone ScAtx1. ScAtx1, interacting with the N-terminal cytosolic domains of ATPases, is presumed to acquire copper from $CtaA_N$ and donate it to $PacS_N$ [16]. *Synechocystis* has two further P-type ATPases ZiaA and CoaT, which are involved in metal (zinc and cobalt, respectively) export into the periplasm. Surprisingly, the N-terminal cytosolic domain of ZiaA ($ZiaA_N$) shows a lower affinity for zinc, the physiological metal ion, than for copper [16], rising the question on which are the molecular features controlling metal uptake and delivery *in vivo*. Failure of $ZiaA_N$ to interact with ScAtx1 [17], the copper chaperone, is thought to prevent ZiaA from acquiring copper. It is true, however, that Cu(I)-ScAtx1 and $PacS_N$ form a metal-mediated complex which allows copper to be transferred from a protein to the other. Cu(I)-ScAtx1 does not form any detectable complex with $ZiaA_N$ nor copper transfer is actually detected.

In vivo ScAtx1 does not transfer Cu(I) to $ZiaA_N$, even if the latter has a higher affinity for Cu(I) than both ScAtx1 and $PacS_N$. This represents a striking example of how nature controls metal transfer through kinetics mechanisms. To understand the subtle features determining and tuning protein recognition, we have stepwise changed amino acids on the surface of $ZiaA_N$ in order to make it tailor to $PacS_N$ and we have monitored the formation of the metal-mediated complex. Finally, a few mutations have been performed to bring the extent of the complex to 90%. We believe that the metal-mediated complex formation is a key step to control the entry, though not necessarily its extent.

Materials and Methods

Sample preparation

The expression plasmids for the selected mutants of $ZiaA_N$ (Figure 1; mutant QC2: Asp18Arg, Lys23Ala, Leu24Ser, Lys25Ser; mutant QC3: Asp18Arg, Lys23Ala, Leu24Ser, Lys25Ser, Thr20Ala, Ser21Ala; mutant QC4: Asp18Arg, Lys23Ala, Leu24Ser, Lys25Ser, Thr20Ala, Ser21Ala, Gly28Arg, Ser29Ala), were obtained through the Quikchange Site-directed Mutagenesis kit (Stratagene). The expression and purification protocols of $ZiaA_N$ and its mutants were essentially the same as those already reported [16].

Also recombinant protein ScAtx1 was expressed similarly to the previously described procedure [17].

For both proteins ($ZiaA_N$ and ScAtx1) isotopic enrichment with ^{15}N and with ^{15}N and ^{13}C was obtained by growing *E. coli* cells (BL21 (DE3)) in M9 minimal medium supplemented with [^{13}C]glucose and $(^{15}NH_4)Cl/(^{15}NH_4)_2SO_4$. The proteins are produced as apo, oxidized forms which were reduced with excess of dithiothreitol (DTT) and washed with 50 mM sodium phosphate buffer, pH 7.0.

To 1:1 mixture of the two apo-proteins ScAtx1 and $ZiaA_N$ (either WT or one of the mutants) increasing amounts of copper(I) were added up to final ratio 1:1:1, to monitor whether a metal-mediated complex forms. As a control, copper(I) was also added to the isolated proteins. In the various experiments, either only one of the proteins was enriched in ^{15}N or in ^{15}N and ^{13}C or, in some cases, one protein was enriched in ^{15}N and the other in ^{15}N and ^{13}C. The formation of the complex was monitored through ^{15}N-^{1}H HSQC spectra and confirmed through the determination of ^{15}N relaxation rates.

NMR experiments

All NMR experiments were collected at 298 K on Bruker Avance spectrometers equipped with a triple resonance cryoprobe. NMR experiments were performed on various [^{13}C,^{15}N]-labelled and [^{15}N]-labelled $ZiaA_N$/ScAtx1 samples at 0.5-1.0 mM concentration. 10% D2O was added for the lock signal. ^{15}N-^{1}H HSQC spectra were used to follow chemical shift changes during the formation of the complex. The formation of the complex was also monitored through the determination of ^{15}N relaxation rates of backbone amides. Relaxation rates R_1 and R_2 were determined by fitting the cross-peak intensities measured as a function of the delay within the pulse to a single-exponential decay. The overall rotational correlation time τ_c values were estimated from the R_2/R_1 ratio. To identify the coordination mode of copper(I)-binding histidine of ScAtx1, a ^{1}H-^{15}N HSQC experiment for measuring $^{2}J_{N^{\varepsilon}H^{\delta}}$, $^{2}J_{N^{\varepsilon}H^{\varepsilon}}$, $^{2}J_{N^{\delta}H^{\varepsilon}}$, and $^{2}J_{N^{\delta}H^{\delta}}$ coupling constants was performed. Double and triple resonance 3D experiments were collected to perform sequence-specific assignments or to verify and extend the available ones [7, 18, 19], for ScAtx1, $ZiaA_N$ WT, and $ZiaA_N$ QC4. ^{15}N and ^{13}C separated 3D NOESY-HSQC experiments were acquired to collect structural constraints. ^{13}C-direct detection experiments were carried out on a 16.4 T Bruker AVANCE 700 spectrometer, operating at 700.06 MHz for ^{1}H and 176.03 MHz for ^{13}C and equipped with a TXO ^{13}C, ^{15}N, ^{1}H cryogenically cooled probe. Complete tables describing in detail the experiments acquired on the various samples investigated and the parameters used are reported in Supplementary Material (Tables S2–5).

Structure determination

Structure calculations. The three-dimensional structure of the $ZiaA_N$ QC4 mutant in the apo form was determined, using distance constraints obtained from the 3D ^{15}N- and ^{13}C-NOESY-HSQC spectra. Structure calculations were performed with the program CYANA-2.1 [20, 21]. Each peak list was calibrated independently. ϕ and ψ dihedral angle constraints were derived from the chemical shift index. Restrained energy minimization was performed for each member of the final family of conformers using the AMBER 8 package [22]. The quality of the structure was evaluated using program PROCHECK [23] (Table S6). Structure calculations and analysis were performed on a cluster of Linux PCs. All spectra were processed using standard Bruker software XWINNMR/TopSpin and analysed using the XEASY program.

HADDOCK docking calculations. HADDOCK [24] calculations, performed with standard protocols to obtain a model of the complex between $ZiaA_N$QC4-Cu(I)-ScAtx1, used, as input, the averaged minimized structures of the apo-$ZiaA_N$ QC4 and the monomeric form of Cu(I)-ScAtx1 [17]. The 'active' and 'passive' residues were selected on the basis of the mapping the backbone chemical shift changes. An initial threshold was used to select the most affected residues, which were considered as active. This threshold was then changed to include other residues experiencing chemical shift changes, in agreement with the surface of the interaction. Each run of docking was performed for the four possible sets of coordinating cysteines to copper in $ZiaA_N$ QC4 and in ScAtx1, by including additional distance-unambiguous restraints of 2.3 $\pm$ 0.2 Å. His 61 was not involved in copper binding, because ^{2}J ^{1}H-^{15}N HSQC experiments show that it is detached from copper(I) coordination when Cu(I)-ScAtx1 interacts with apo-$ZiaA_N$ QC4.

Results and Discussion

Design of the mutants

In order to understand why $ZiaA_N$ and $PacS_N$ behave different with respect to ScAtx1, the sequence alignment was analyzed. $ZiaA_N$ and $PacS_N$ shows pronounced differences, as expected, in the amino acids immediately preceding and following the CXXC metal binding motif (Figure 1). The region around the metal-binding site is that displaying the major differences in the primary sequence and thus was that on which we focused (Figure 1). Indeed, the residue preceding the CXXC motif, which in $PacS_N$ is an arginine, becomes an aspartate in $ZiaA_N$, i.e. with opposite charge properties. Also

the properties of two residues between the two metal binding cysteines differ, being two hydrophilic residues in $ZiaA_N$ (Thr, Ser) while they are hydrophobic residues, with a slightly reduced steric hindrance in $PacS_N$ (Ala, Ala). Pronounced differences between the two proteins are also observed in the 7 amino-acid stretch just after the metal binding cysteine motif. While in $ZiaA_N$ positively charged bulky residues (two lysines) are in the first part (KLK) and small hydrophilic ones in the second (GS), the opposite occurs in $PacS_N$ (ASS; RA). The location of these residues affects both the possibility of small rearrangements of the metal-binding site as well as the characteristics of the surface of the protein.

The variation of electrostatic and steric properties featured by specific amino-acids involved in the protein (molecular) recognition are essential for determining the specificity of the metal-mediated protein–protein interactions. The electrostatic surfaces of $PacS_N$ and ScAtx1 show consistent complementarity on interaction interfaces. In particular $PacS_N$ features a positively charged residue (R13) before the first metal binding cysteine, which is an optimal partner to E13 of partner ScAtx1. This interaction could contribute to direct the protein–protein recognition. In $ZiaA_N$ WT the residue preceding the first metal-binding cysteine is negatively charged (D18) which is causing a repulsion with E13 of ScAtx1 thus contributing to prevent the interaction between the $ZiaA_N$ and ScAtx1. Moreover a positively charged motif (KLK) of $ZiaA_N$ WT is found right after the metal-binding loop: this could promote an electrostatic interaction with Glu 13 of ScAtx1 that prevents the proper orientation of the two partners for successful metal binding. The introduction of the ASS pattern in place of the KLK motif of $ZiaA_N$ removes the strong positive charged and reduces the steric hindrance of this motif rendering the surface more sterically available. Therefore the first mutated residues (QC2 mutant) were the residues immediately preceding (D $\rightarrow$ R) and following (KLK $\rightarrow$ ASS) the metal-binding motif (CXXC). In the second round, also the residues between the two metal-binding cysteines, T and S were mutated to A, thus removing the hydrophilicity of this part. Finally, the next residues in the primary sequence that show pronounced differences (GS) were changed to RA, as in $PacS_N$, possibly reintroducing important electrostatic interactions with ScAtx1 that may be important to form productive complexes.

Whereas in the QC4 mutant, the presence of Arg 28 at the place of glycine modulates considerably the protein surface promoting possible electrostatic contacts with Asp 27 and Glu 26 of the ScAtx1.

On these grounds, the mutants of $ZiaA_N$ WT (Figure 1) were designed to progressively mimic the residues of $PacS_N$ (both in the CXXC binding loop and in the 7 amino-acid stretch) in order to determine their role in the

interaction with the ScAtx1. The three mutants are referred to as $ZiaA_N$ QC2, $ZiaA_N$ QC3, $ZiaA_N$ QC4 (Figure 1).

The three $ZiaA_N$ mutants are stable and soluble. They maintain the native fold as it appears from the good dispersion of the cross peaks in the ^{1}H-^{15}N HSQCs and by the fact that, besides spectral variations for the mutated residues and those in the neighbouring residues, the overall NMR spectra remain the same indicating that the surface properties can be modulated without perturbing the overall fold.

Monitoring metal-mediated complex formation

In order to assess the role of mutations in $ZiaA_N$ that progressively mimic $PacS_N$ in the formation of a metal mediated complex with ScAtx1, 1:1 protein–protein mixtures of ScAtx1 (unlabelled) and $ZiaA_N$ variants (WT, QC2, QC3, QC4, ^{15}N labelled) were titrated with Cu(I). When the two apo proteins are mixed no significant chemical shift changes are observed with respect to the spectra of the two isolated ones indicating that no detectable interaction occurs in the absence of a metal ion. Addition of copper ion causes changes in the HSQC spectra, depending on the $ZiaA_N$ mutant. Indeed, addition of copper to a mixture containing $ZiaA_N$ WT and ScAtx1 produces changes that are very similar to those observed upon addition of Cu(I) to the two isolated proteins. When the mutants are present instead new peaks are observed whose intensity increases with increasing Cu(I). This shows that a new species (in slow exchange with the others on the chemical shift time scale) is formed respect to the simple formation at the isolated Cu(I)-bound forms of the proteins. The intensity of the new peaks at a 1:1:1 relative concentration is maximal in the case of the QC4 mutant, indicating that the amount of this new species increases progressively as the mutations mimic more closely the $PacS_N$ interface and the binding loop.

The rate of interconversion is slow on the NMR timescale. Also the equilibrium between Cu(I)-ScAtx1 and $PacS_N$ was found to be slow. The relative ratio of the forms present at equilibrium, determined by the relative intensity of the new peaks formed respect to those of the apo proteins, progressively increases passing from $ZiaA_N$ WT (no new form) to the QC4 $ZiaA_N$ mutant (almost all new form). The NMR signals of the new species differ in chemical shift from those in the isolated species for residues located in and around the metal-binding region and at the protein-protein interface; some signals in this region disappear due to line broadening. In the case of the QC4 $ZiaA_N$ mutant, 12 signals disappear and 14 change chemical shift. This new species is the complex between the two proteins formed in the presence of copper, ScAtx1:Cu(I):$ZiaA_N$ QC4.

Therefore, while there is no formation of a ScAtx1-Cu(I)-ZiaA$_N$ WT metal-mediated complex, the amount of complex present at equilibrium significantly increases with the increase in the mutations, being maximal for ZiaA$_N$ QC4. Experiments repeated in the same conditions with ^{15}N labelled ScAtx1 and unlabelled ZiaA$_N$ mutant confirm these conclusions. In the absence of copper no protein–protein interaction is detected.

The progressive mutations of an increasing number of amino-acids of ZiaA$_N$ confirm that substitution of a single amino acid is not sufficient to change the affinity of the complex and that a number of other residues need to be mutated for increasing the amount of the formed complex. This indicates that the complex between the two proteins is mediated by fine tuning of several inter-protein interactions. This is also reassuring as, otherwise, just a point mutation would be sufficient to mess up regulatory mechanisms controlling the delivery of metal ions to the right protein targets.

Determination of a structural model of the metal-mediated complex

The structural characterization of the ZiaA$_N$ QC4 mutant/Cu/ScAtx1 complex was performed to obtain a model of the metal-mediated protein–protein complex involved in molecular recognition.

The structure of the apo form of ZiaA$_N$ QC4 mutant maintains, as predicted from the HSQC experiment, a classical ferredoxin-like fold, typical of soluble N-terminal domains of heavy metal ATPases [6, 9] and of their metallochaperones. The carboxy-terminal tail, which contains a His-rich region, appears, from both dynamics and structural investigations, highly disordered as previously described for ZiaA$_N$ WT [16]. Compared to the native form, the backbone structure is largely maintained with the most pronounced difference involving helix α_1, which shows a slight shift along the main helical axis. This is probably due to a large change in the steric hindrance of the mutated residues (four out of 8 mutated) in helix α_1 that can be compensated by small relative shifts of the main secondary structural elements one respect to the other (helix α_1 respect to β_2 strand). The majority of the side chains of the mutated residues are exposed to the surface of the molecule, thus modulating its properties without perturbing the overall fold. In comparison with Asp 18 of native ZiaA$_N$, the longer side chain of Arg 18 is more exposed out of the molecule. The metal-binding loop maintains the overall conformation. Ala 20, Ala 21, introducing between the two metal-binding cysteines neutral side chains of approximately the same length as the original polar amino acids (Thr and Ser), maintain the orientation of their side

chains. The substitution of the KLK motif located after the metal-binding cysteines with the ASS drastically reduces the steric hindrance of this protein segment, removing the exposed flexible positive charge (the other lysine is less exposed and points towards the interior of the protein) from the protein surfaces. Finally, substitution of GS with RA, introduces a positive bulky residue on the surface, in place of a glycine in a position that differs by a helix turn respect to the Lys 25 in $ZiaA_N$ WT. This is important to define the surface charge distribution, which is likely to provide the first driving force to orient the two molecules in the proper direction to favour metal binding.

The chemical shift changes observed upon formation of the ScAtx1-Cu(I)-$ZiaA_N$ QC4 metal-mediated complex, together with the available solution structures of the ScAtx1 (Cu(I) form) and of $ZiaA_N$ QC4 mutant (apo form), were used as input data to calculate a model of the complex through the program HADDOCK. Calculations repeated with various possible patterns of cysteine coordination, indicated two possible modes for copper coordination (either the copper ion is coordinated by two cysteine residues of the $ZiaA_N$ QC4 together with Cys 15 of ScAtx1, or the copper ion is bound to Cys 12 and Cys 15 of the ScAtx1 with Cys 19 of the $ZiaA_N$ QC4). Analysis of ^{1}H-^{15}N HSQC spectra of the various forms of ScAtx1 together with ^{1}H-^{15}N HSQC experiments for measuring ^{2}J and ^{3}J coupling constants shows that His 61 is not involved in metal coordination and moves far from the copper as a consequence of the presence of the interaction with the QC4 mutant of $ZiaA_N$ (Figure 4). In the complex, the average buried contact surface has an area ranging from 1030 to 1350 $Å^2$. The interacting region involves mainly loops 1 and 5, and helices α_1 and α_2 on both proteins (Figure 3).

The analysis of the residues at the interface of the ScAtx1-Cu(I)-$ZiaA_N$ QC4 complex confirms the importance of the designed mutations to increase the stability of the complex. If $ZiaA_N$ WT is superimposed to the QC4 mutant in the complex , Asp 18 provides a repulsion to Glu 13 of ScAtx1. The substitution of the negatively charged Asp 18 in $ZiaA_N$ WT by an arginine with an opposite charge and a longer side chain changes the local electrostatic properties of the protein surface and providing the freedom of movement necessary to promote the interaction with Glu 13 of the ScAtx1. Three additional mutations were introduced in the QC2 mutant (KLK to ASS). Ala 23 is the only mutated residue which is not exposed outside the protein, (in ZiaAN WT the bulky positive Lys 23 is pointing inside of the molecule). The change of lysine to alanine however does not perturb the local conformation as small variations in the neighbouring amino acids compensate for difference in bulkiness. Ser 24 and Ser 25 are in contact with Ala 18 of ScAtx1 and with the exposed residues of loop 5 (neutral Gly 60 and Ser 58). Lys 25

in place of Ser 25 in the $ZiaA_N$ QC4 introduces a large bulky residue with a positive charge, which changes the shape and charge distribution on the surface. Arg 18 contributes to the recognition of the two partners, however the increase in the stability of the complex requires the presence of other residues at the interface. The mutation of the Thr/Ser to Ala 20 and Ala 21, the two residues between the two metal-binding cysteines (CXXC), which, as noted before, do not drastically influence the stability of the complex, still allow the cysteines to be in the proper orientation to transfer the metal ion. The substitution of the neutral glycine with Arg 28 has a major contribution to the complex stabilization. The presence of the long bulky arginine at the interface gives rise to many potential electrostatic contacts (with Gln 24, Gln 51, Glu 26, Asp 27) which probably play a key role in favouring the initial relative orientation of the two proteins that leads to metal coordination. The substitution of Ser 29 to alanine constitutes a less pronounced change by just removing an -OH group and does not significantly perturb the structure of the surface. Some contacts involve hydrophobic residues, namely Gly 70 and Tyr 71 of $ZiaA_N$ QC4 with Ala 18 of ScAtx1 at the centre of the interaction surface. Some residues providing suitable electrostatic contacts were already present in $ZiaA_N$ WT namely Arg 32 close to Asn 25 and Ala 28 of ScAtx1.

By superimposing the solution structure of apo-$PacS_N$ to $ZiaA_N$ QC4 in the complex with ScAtx1 it is indeed evident that the main interactions stabilizing the metal-mediated complex have been reintroduced through the mutations. The structures are similar with most contacts at the interface present in both structures. The majority of the observed intermolecular contacts involves polar charged residues (already discussed above) indicating some electrostatic interactions between residues with opposite signs. The orientation of the Arg 13 (apo-$PacS_N$) side chain varies within the family of the conformers implying possible freedom of movement, the side chain of Arg 18 keeps fixed position in the model of ScAtx1-Cu(I)-$ZiaA_N$ QC4 heterodimer. The metal-binding loop of $ZiaA_N$ QC4 in the model of the adduct is slightly moved away from the approaching partner ScAtx1 comparing to apo-$PacS_N$. Slight shift of the first α-helix along the main axis ($ZiaA_N$ QC4 in ScAtx1-Cu(I)-$ZiaA_N$ QC4) with respect to apo-$PacS_N$ produces slight shift in position of individual amino-acid side chains creating the interface.

Summarizing, the progressive mutations of an increasing number of residues of $ZiaA_N$ in order to mimic the interaction surface of $PacS_N$ with the chaperone ScAtx1 indicate the importance of the charge distribution at the interface to favour a reciprocal relative orientation of the two proteins that will lead to the formation of the metal-mediated complex. Indeed, for the formation of a metal-mediated complex the two proteins should be in the proper relative orientaion with the two CXXC motives facing each other. The example of

the series of mutants of $ZiaA_N$ that progressively stabilize the metal-mediated complex with ScAtx1 shows that 1) more than one point mutation is necessary to promote complex formation; 2) that the electrostatic interactions play a key role in favouring a suitable relative orientation of the two molecules one respect to the other; 3) that one of these electrostatic contacts (i.e. Arg 18/Glu 13) is not enough and that a couple of them at opposite edges of the interface contribute to properly orient the two proteins.

This could be a general way through which mutation of surface residues are used to properly orient two proteins to favour the right relative orientation to form the metal-mediated complex. So just changing the position of positive charges on the surface and eliminating or disturbing key contacts at the extreme sides of the interaction surface, may provide a general tool used by nature to control metal transfer pathways.

The variability of key surface amino acids appears to be natural tool to modulate the surface properties that drive molecular recognition. This provides an elegant way to design metal transfer pathways by still using the same 3D folds and metal-binding motifs.

Acknowledgements

This work has been supported by the European Commission [contract 031220, SPINE 2-COMPLEXES] and by the FIRB proteomica.

Bibliography

[1] TV O'halloran and VC Culotta. Metallochaperones, an intracellular shuttle service for metal ions. *Journal of Biological Chemistry*, 275:25057–25060, 2000.

[2] BE Kim, T Nevitt, and DJ Thiele. Mechanisms for copper acquisition, distribution and regulation. *Nature Chemical Biology*, 4:176–185, 2008.

[3] L Banci, I Bertini, S Ciofi-Baffoni, T Hadjiloi, M Martinelli, and P Palumaa. Mitochondrial copper(i) transfer from cox17 to sco1 is coupled to electron transfer. *Proceedings of the National Academy of Sciences of the United States of America*, 105:6803–6808, 2008.

[4] S Tottey, DR Harvie, and NJ Robinson. Understanding how cells allocate metals using metal sensors and metallochaperones. *Accounts of Chemical Research*, 38:775–783, 2005.

[5] L Banci, I Bertini, F Cantini, and S Ciofi-Baffoni. Cellular copper distribution: a mechanistic systems biology approach. *Cellular and Molecular Life Sciences*, 67(15):2563–2589, 2010.

[6] L Banci, I Bertini, F Cantini, IC Felli, L Gonnelli, N Hadjiliadis, R Pierattelli, A Rosato, and P Voulgaris. The atx1-ccc2 complex is a metal-mediated protein–protein interaction. *Nature Chemical Biology*, 2:367–368, 2006.

[7] L Banci, I Bertini, V Calderone, N Della-Malva, IC Felli, S Neri, A Pavelkova, and A Rosato. Copper(i)-mediated protein–protein interactions result from suboptimal interaction surfaces. *Biochemical Journal*, 422:37–42, 2009.

[8] L Banci, I Bertini, KS Mcgreevy, and A Rosato. Molecular recognition in copper trafficking. *Natural Product Reports*, 27:695–710, 2010.

[9] L Banci, I Bertini, S Ciofi-Baffoni, T Kozyreva, K Zovo, and P Palumaa. Affinity gradients drive copper to cellular destinations. *Nature*, 465:645–U145, 2010.

[10] S Tottey, KJ Waldron, SJ Firbank, B Reale, C Bessant, K Sato, TR Cheek, J Gray, MJ Banfield, C Dennison, and NJ Robinson. Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature*, 455:1138–U17, 2008.

[11] L Banci, I Bertini, F Cantini, C Massagni, M Migliardi, and A Rosato. An nmr study of the interaction of the n-terminal cytoplasmic tail of the wilson disease protein with copper(i)-hah1. *Journal of Biological Chemistry*, 284:9354–9360, 2009.

[12] F Arnesano, L Banci, I Bertini, and AR Thompsett. Solution structure of copc: A cupredoxin-like protein involved in copper homeostasis. *Structure*, 10:1337–1347, 2002.

[13] L Banci and A Rosato. Structural genomics of proteins involved in copper homeostasis. *Accounts of Chemical Research*, 36:215–221, 2003.

[14] C Andreini, I Bertini, G Cavallaro, GL Holliday, and JM Thornton. Metal ions in biological catalysis: from enzyme databases to general principles. *Journal of Biological Inorganic Chemistry*, 13:1205–1218, 2008.

[15] L Banci, I Bertini, S Ciofi-Baffoni, NG Kandias, NJ Robinson, GA Spyroulias, XC Su, S Tottey, and M Vanarotti. The delivery of copper for thylakoid import observed by nmr. *Proceedings of the National Academy of Sciences of the United States of America*, 103:8320–8325, 2006.

[16] L Banci, I Bertini, S Ciofi-Baffoni, L Poggi, M Vanarotti, S Tottey, KJ Waldron, and NJ Robinson. Nmr structural analysis of the soluble domain of ziaa-atpase and the basis of selective interactions with copper metallochaperone atx1. *Journal of Biological Inorganic Chemistry*, 15:87–98, 2010.

[17] L Banci, I Bertini, S Ciofi-Baffoni, XC Su, GPM Borrelly, and NJ Robinson. Solution structures of a cyanobacterial metallochaperone - insight into an atypical copper-binding motif. *Journal of Biological Chemistry*, 279:27502–27510, 2004.

[18] TD Rae, PJ Schmidt, RA Pufahl, VC Culotta, and TV O'halloran. Undetectable intracellular free copper: The requirement of a copper chaperone for superoxide dismutase. *Science*, 284:805–808, 1999.

[19] F Arnesano, L Banci, I Bertini, F Cantini, S Ciofi-Baffoni, DL Huffman, and TV O'halloran. Characterization of the binding interface between the copper chaperone atx1 and the first cytosolic domain of ccc2 atpase. *Journal of Biological Chemistry*, 276:41365–41376, 2001.

[20] Peter Guntert. Calculating protein structures from nmr data. *Methods in Molecular Biology; Protein NMR techniques*, pages 157–194, 1997.

[21] P Guntert, C Mumenthaler, and K Wuthrich. Torsion angle dynamics for nmr structure calculation with the new program dyana. *Journal of Molecular Biology*, 273:283–298, 1997.

[22] E Enggist, L Thony-Meyer, P Guntert, and K Pervushin. Nmr structure of the heme chaperone ccme reveals a novel functional motif. *Structure*, 10:1551–1557, 2002.

[23] RA Laskowski, JAC Rullmann, MW Macarthur, R Kaptein, and JM Thornton. Aqua and procheck-nmr: Programs for checking the quality of protein structures solved by nmr. *Journal of Biomolecular NMR*, 8:477–486, 1996.

[24] C Dominguez, R Boelens, and AMJJ Bonvin. Haddock: A protein–protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, 125:1731–1737, 2003.

Supplementary Material

Table 1: R_1 and R_2 relaxation rates (s^{-1}) for amide ^{15}N nuclei measured at 500 MHz for apo-ZiaA$_N$ QC4, apo-ScAtx1 and ScAtx1-Cu(I)-ZiaA$_N$ QC4, correlation time for molecular tumbling, τ_c (ns) as estimated from R_2/R_1 ratio is also reported.

	$\mathbf{R_1}$ **(s^{-1})**	$\mathbf{R_2}$ **(s^{-1})**	τ_c **(ns)**
apo-**ScAtx1**[1]	2.81 ± 0.06	7.00 ± 0.13	∼4.01
ScAtx1-Cu(I)-ZiaA$_N$ QC4	1.81 ± 0.12	14.07 ± 0.07	∼10.20
apo-**ZiaA$_N$ WT**[2]	1.65 ± 0.11	9.54 ± 1.46	∼6.96
ZiaA$_N$ WT-Cu(I)-ScAtx1	2.21 ± 0.10	11.59 ± 0.14	∼6.27
ZiaA$_N$ QC4-Cu(I)-ScAtx1	1.74 ± 0.13	15.47 ± 0.06	∼10.5

[1]ScAtx1 is one of the chaperones that has a tendency to form homodimer, the increase in τ_c can be attributed dimer $(ScAtx1)_2$ present at equilibrium.

[2]The value of τ_c of apo-ZiaA$_N$ QC4 was very similar (τ_c ∼7.4 ns).

Table 2: Acquisition parametrs for NMR experiments performed on ScAtx1-Cu(I)-ZiaA$_N$ QC4 containing ^{13}C, ^{15}N-apo-ZiaA$_N$

Experiments	**Dimension**			**SW (ppm)**			**NS**	**d_1 (s)**	**B_0[3] (MHz)**
	F3	**F2**	**F1**	**F3**	**F2**	**F1**			
^{1}H-^{15}N HSQC		1024 (^{1}H)	128 (^{15}N)		13.9486	36.0602	2	1.00	800
CBCANH	2048 (^{1}H)	80 (^{15}N)	128 (^{13}C)	13.9486	36.0180	75.0780	8	1.20	800
CBCA(CO)NH	2048 (^{1}H)	64 (^{15}N)	128 (^{13}C)	13.9486	36.0180	75.0780	8	1.20	800
HNCO	1024 (^{1}H)	64 (^{15}N)	128 (^{13}C)	16.0230	36.0602	24.1237	16	1.00	800
^{15}N R_1[4]		1024 (^{1}H)	128 (^{15}N)		16.0216	38.0015	16	2.50	500
^{15}N R_2[4]		1024 (^{1}H)	128 (^{15}N)		16.0216	38.0015	16	2.50	500

Table 3: Acquisition parametrs for NMR experiments performed on ScAtx1-Cu(I)-ZiaA$_N$ QC4 containing ^{13}C, ^{15}N-apo-ScAtx1

Experiments	**Dimension**			**SW (ppm)**			**NS**	**d_1 (s)**	**B_0 (MHz)**
	F3	**F2**	**F1**	**F3**	**F2**	**F1**			
^{1}H-^{15}N HSQC		1024 (^{1}H)	128 (^{15}N)		15.0031	39.9999	2	1.00	700
CBCANH	1024 (^{1}H)	48 (^{15}N)	96 (^{13}C)	14.0396	34.2883	80.1075	32	1.20	900
CBCA(CO)NH	1024 (^{1}H)	48 (^{15}N)	96 (^{13}C)	14.0396	34.2883	80.1075	32	1.20	900
HNCO	1024 (^{1}H)	64 (^{15}N)	96 (^{13}C)	11.9824	36.0277	12.6324	8	1.00	700
^{15}N R_1		1024 (^{1}H)	128 (^{15}N)		16.0216	38.0015	16	2.50	500
^{15}N R_2		1024 (^{1}H)	128 (^{15}N)		16.0216	38.0015	16	2.50	500
$^2J_{NH}$ [^{1}H-^{15}N]-HSQC		2048 (^{1}H)	256 (^{15}N)		16.0845	170.2083	128	1.00	700

Table 4: Acquisition parametrs for NMR experiments performed on ^{13}C, ^{15}N-apo-ZiaA$_N$ QC4

Experiments	**Dimension**			**SW (ppm)**			**NS**	**d_1 (s)**	**B_0 (MHz)**
	F3	**F2**	**F1**	**F3**	**F2**	**F1**			
^{1}H-^{15}N HSQC		2048 (^{1}H)	128 (^{15}N)		13.9486	36.0602	2	1.00	800
^{1}H-^{13}C HSQC		2048 (^{1}H)	1024 (^{13}C)		16.0230	165.6661	2	2.00	500
CBCANH	1024 (^{1}H)	64 (^{15}N)	96 (^{13}C)	14.0020	35.0077	64.0222	16	1.50	500
CBCA(CO)NH	1024 (^{1}H)	64 (^{15}N)	96 (^{13}C)	14.3642	35.0077	64.0000	8	1.50	500
HNCO	1024 (^{1}H)	64 (^{15}N)	96 (^{13}C)	14.3642	40.0049	22.0847	8	1.00	500
^{15}N R_1		1024 (^{1}H)	128 (^{15}N)		16.0216	38.0015	16	2.50	500
^{15}N R_2		1024 (^{1}H)	128 (^{15}N)		16.0216	38.0015	16	2.50	500
HCCH TOCSY	2048 (^{1}H)	128 (^{13}C)	196 (^{13}C)	13.9486	75.0000	75.0780	8	1.50	800
^{13}C-edited [^{1}H-^{1}H]-NOESY 3D	1024 (^{1}H)	76 (^{13}C)	256 (^{1}H)	13.9486	75.0780	13.9486	8	1.20	800
^{15}N-edited [^{1}H-^{1}H]-NOESY 3D	1024 (^{1}H)	64 (^{15}N)	256 (^{1}H)	13.9486	25.0052	13.9486	8	1.20	800

[3]The magnetic field strength of the instrument used for the measurements is indicated by using the relative ^{1}H frequency.

[4]Experiments for the measurement of ^{15}N relaxation rates R_1, R_2 were acquired on ^{15}N labelled samples.

Table 5: Acquisition parametrs for NMR experiments performed on ^{13}C, ^{15}N-apo-ScAtx1

Experiments	**Dimension**			**SW (ppm)**			**NS**	**d_1 (s)**	**B_0 (MHz)**
	F3	**F2**	**F1**	**F3**	**F2**	**F1**			
^{1}H-^{15}N HSQC		1024 (^{1}H)	128 (^{15}N)		13.9486	36.0602	2	1.00	800
^{1}H-^{13}C HSQC		2048 (^{1}H)	1024 (^{13}C)		16.0230	180.0000	4	2.00	800
CBCANH	2048 (^{1}H)	64 (^{15}N)	128 (^{13}C)	13.9486	36.0180	64.0000	8	1.20	800
CBCA(CO)NH	1024 (^{1}H)	64 (^{15}N)	128 (^{13}C)	13.9486	36.0180	64.0000	8	1.20	800
HNCO	1024 (^{1}H)	64 (^{15}N)	128 (^{13}C)	13.9486	36.0180	16.0000	4	1.20	800
^{15}N R_1		1024 (^{1}H)	128 (^{15}N)		16.0216	38.0015	8	2.50	500
^{15}N R_2		1024 (^{1}H)	128 (^{15}N)		16.0216	38.0015	8	2.50	500
HCCH TOCSY	2048 (^{1}H)	96 (^{13}C)	192 (^{13}C)	13.9486	75.0000	75.0780	4	1.20	800
CON 2D		1024 (^{13}C)	512 (^{15}N)		49.9987	35.0108	16	2.50	700
(H)CACO 2D		1024 (^{13}C)	256 (^{13}C)		49.9987	30.2120	16	2.50	700
(H)CBCACON 3D	1024 (^{13}C)	64 (^{15}N)	192 (^{13}C)	49.9987	35.0631	74.1495	4	1.50	700

Table 6: Statistical analysis of the solution structure of apo-ZiaA_N QC4.

	REM
RMS violations per meaningful distance constraint (Å)[5]:	**(20 structures)**
Intraresidue (128)	0.0000 ± 0.0000
Sequential (196)	0.0160 ± 0.0056
Medium range (168)[6]	0.0295 ± 0.0032
Long range (269)	0.0169 ± 0.0036
Total (761)	0.0192 ± 0.0019
Average number of constraints per residue:	11.04
Average number of violations per structure:	
Intraresidue	0.00 ± 0.00
Sequential	4.10 ± 2.02
Medium range	9.05 ± 2.42
Long range	5.75 ± 2.00
Average No. of NOE violations larger than 0.3 Å	0.00 ± 0.00
Average No. of NOE violations between 0.1–0.3 Å	10.95 ± 2.60
Structural analysis[7]	
% of residues in most favorable regions	82.3
% of residues in allowed regions	16.6
% of residues in generously allowed regions	0.5
% of residues in disallowed regions	0.6

[5]The number of meaningful constraints for each class is reported in parenthesis.

[6]Medium range distance constraints are those between residues (i,i+2), (i,i+3), (i,i+4) and (i,i+5).

[7]As it results from the Ramachandran plot analysis.

Table 7: ^{1}H, ^{15}N and ^{13}C resonance assignments of apo-ScAtx1 from *Synechocystis* PCC 6803

residue	**N (H^N)**	**C^α (H^α)**	**C^β (H^β)**	**C'**	**Others**
M1		59.230	29.893	169.810	C^γ 29.470; C^ε 17.568
		4.017	1.927; 1.962		Q^ε 0.809
T2	121.488	59.236	67.439	170.560	$C^{\gamma 2}$ 18.692
	8.813	5.445	3.724		$Q^{\gamma 2}$ 0.934
I3	126.663	56.305	39.221	169.810	$C^{\gamma 2}$ 14.027; $C^{\gamma 1}$ 24.598; $C^{\delta 1}$ 11.101
	9.158	4.459	1.588		$H^{\gamma 12}$ 0.961; $H^{\gamma 13}$ 1.335; $Q^{\gamma 2}$ 0.713; $Q^{\delta 1}$ 0.757
Q4	123.780	51.008	28.716	171.892	C^γ 31.653; $N^{\varepsilon 2}$ 111.949
	8.371	5.368	1.735		$H^{\gamma 2}$ 2.006; $H^{\gamma 3}$ 2.099; $H^{\varepsilon 21}$ 6.787; $H^{\varepsilon 22}$ 7.306
L5	125.136	49.828	42.625	172.810	C^γ 23.477; $C^{\delta 1}$ 19.931; $C^{\delta 2}$ 23.185
	8.943	1.249; 1.325			H^γ 1.338; $Q^{\delta 1}$ 0.754; $Q^{\delta 2}$ 0.543
T6	120.492	59.812	66.327	170.932	$C^{\gamma 2}$ 18.521
	9.300	3.949			$Q^{\gamma 2}$ 1.003
V7	127.595	55.167	28.727		$C^{\gamma 1}$ 19.925; $C^{\gamma 2}$ 18.283
	8.548	5.270	1.977		$Q^{\gamma 1}$ 0.812; $Q^{\gamma 2}$ 0.699
P8		61.596	29.843	173.933	C^γ 24.040; C^δ 48.048
		4.087	1.927; 2.280		Q^γ 1.865; $H^{\delta 2}$ 3.471; $H^{\delta 3}$ 4.043
T9	104.159	59.213	64.494	173.187	$C^{\gamma 2}$ 19.909
	7.173	4.077			$Q^{\gamma 2}$ 1.169
I10	121.894	60.352	34.030	171.685	$C^{\gamma 2}$ 14.654; $C^{\gamma 1}$ 24.619; $C^{\delta 1}$ 11.660
	7.506	3.687	1.630		$H^{\gamma 12}$ 0.480; $H^{\gamma 13}$ 1.418; $Q^{\gamma 2}$ 0.533; $Q^{\delta 1}$ 0.414
A11	133.129	49.845	18.189	174.436	
	9.230	4.445	1.433		
C12	113.432	53.489	28.138	171.805	
	8.119	4.890	3.067; 3.294		
E13	121.011	57.420	26.221	175.433	C^γ 33.426
	9.425	3.905	1.984		
A14	124.026	52.735	15.232	178.564	
	8.658	4.045	1.468		
C15	122.023	59.252	26.957	174.810	
	7.854	4.211	2.716; 3.299		
A16	120.561	52.283	15.228	177.439	
	7.470	3.609	1.354		
E17	121.010	56.291	26.357	175.190	C^γ 33.158
	8.226	3.924	2.001; 2.075		$H^{\gamma 2}$ 2.136; $H^{\gamma 3}$ 2.294
A18	122.640	52.176	15.234	178.741	
	7.854	3.958	1.578		
V19	118.269	63.923	28.721	173.559	$C^{\gamma 1}$ 18.168; $C^{\gamma 2}$ 19.918
	8.019	3.257	1.796		$Q^{\gamma 1}$ 0.584; $Q^{\gamma 2}$ 0.463
T20	116.742	63.903	65.109	172.930	$C^{\gamma 2}$ 19.955
	7.801		4.197		$Q^{\gamma 2}$ 1.170
K21	120.224	56.280	29.263	175.442	C^γ 22.218; C^δ 25.780; C^ε 39.253
	8.317	3.863	1.722		γ2,3 1.332; 1.456; Q^δ 1.557; Q^ε 2.860
A22	120.480	51.634	15.246	177.813	
	7.618	3.957	1.311		
V23	116.567	63.913	28.706	175.309	$C^{\gamma 1}$ 18.204; $C^{\gamma 2}$ 19.353
	7.550	3.221	1.929		$Q^{\gamma 1}$ 0.572; $Q^{\gamma 2}$ 0.764
Q24	117.768	53.948	24.312	175.060	C^γ 31.671; $N^{\varepsilon 2}$ 108.572
	8.351	4.281	1.842; 1.972		$H^{\gamma 2}$ 2.243; $H^{\gamma 3}$ 2.357; $H^{\varepsilon 21}$ 6.569; $H^{\varepsilon 22}$ 7.196
N25	116.709	52.165	35.580	173.685	$N^{\delta 2}$ 111.949
	7.989	4.308	2.687; 2.756		$H^{\delta 21}$ 6.678; $H^{\delta 22}$ 7.523
E26	116.140	53.334	27.540	173.607	C^γ 32.256
	7.128	4.261	1.734; 2.019		$H^{\gamma 2}$ 2.163; $H^{\gamma 3}$ 2.163

Table 7: ^{1}H, ^{15}N and ^{13}C resonance assignments of apo-ScAtx1 from *Synechocystis* PCC 6803 (cont'd)

residue	**N (H^N)**	**C^α (H^α)**	**C^β (H^β)**	**C'**	**Other**
D27	117.618	49.853	38.992	172.934	
	7.697	4.553	2.398		
A28	127.512	51.573	15.578	176.311	
	8.700	4.228	1.356		
Q29	113.311	52.134	25.794	173.058	C^γ 31.642; $N^{\varepsilon 2}$ 111.949
	7.879	4.242	1.992; 2.211		$H^{\gamma 2}$ 2.242; $H^{\gamma 3}$ 2.358; $H^{\varepsilon 21}$ 6.733; $H^{\varepsilon 22}$ 7.517
A30	123.518	49.996	15.233	175.312	
	7.617	4.449	1.268		
T31	115.322	57.956	67.421	170.812	$C^{\gamma 2}$ 18.512
	8.574	4.511	3.975		$Q^{\gamma 2}$ 1.064
V32	124.787	58.051	31.008	171.562	$C^{\gamma 2}$ 18.154
	8.735	4.541	1.857		$Q^{\gamma 2}$ 0.512
Q33	127.523	51.335	28.143	171.560	C^γ 30.520; $N^{\varepsilon 2}$ 111.105
	8.827	4.615	1.798; 1.960		Q^γ 2.154; $H^{\varepsilon 21}$ 6.651; $H^{\varepsilon 22}$ 7.388
V34	124.622	58.623	30.501	171.431	$C^{\gamma 1}$ 18.013; $C^{\gamma 2}$ 18.690
	8.924	4.309	1.851		$Q^{\gamma 1}$ 0.659; $Q^{\gamma 2}$ 0.529
D35	127.770	49.267	38.706	173.562	
	8.313	4.824	2.329; 2.989		
L36	125.260	54.511	38.705	175.195	C^γ 24.024; $C^{\delta 1}$ 22.262; $C^{\delta 2}$ 20.514
	8.814	3.621	1.381; 1.597		H^γ 1.480; $Q^{\delta 1}$ 0.674; $Q^{\delta 2}$ 0.447
T37	113.781	62.199	65.224	172.811	$C^{\gamma 2}$ 18.742
	8.361	3.986	4.168		$Q^{\gamma 2}$ 1.103
S38	115.156	55.154	61.567	172.295	
	7.704	4.248	3.642; 3.784		
K39	116.467	54.528	25.781	170.938	C^γ 22.259; C^δ 24.681; C^ε 40.188
	7.875	3.467	1.479		γ2,3 1.114; 1.200; δ2,3 1.994; 2.046; Q^ε 2.917
K40	116.404	52.532	31.074	173.808	C^γ 21.710; C^δ 25.771; C^ε 39.113
	7.389		1.297; 1.568		γ2,3 1.141; 1.192; δ2,3 1.313; 1.383; Q^ε 2.732
V41	126.748	57.476	30.458	171.809	$C^{\gamma 1}$ 18.612; $C^{\gamma 2}$ 19.239
	9.483		1.777		$Q^{\gamma 1}$ 0.578; $Q^{\gamma 2}$ 0.683
T42	124.571	59.525	66.282	171.809	$C^{\gamma 2}$ 18.034
	9.217	4.843	3.888		$Q^{\gamma 2}$ 0.946
I43	129.135	56.894	38.802	172.187	$C^{\gamma 2}$ 15.804; $C^{\gamma 1}$ 25.772; $C^{\delta 1}$ 12.343
	9.330	4.966	1.543		$H^{\gamma 12}$ 0.804; $H^{\gamma 13}$ 1.467; $Q^{\gamma 2}$ 0.776; $Q^{\delta 1}$ 0.661
T44	125.065	58.644	66.392	170.685	$C^{\gamma 2}$ 17.563
	8.728	5.148	3.961		$Q^{\gamma 2}$ 1.006
S45	119.473	53.987	63.304	169.812	
	8.464	3.299; 3.867			
A46	125.689	49.255	15.830	175.436	
	8.479	4.314	1.266		
L47	120.154	51.573	39.633	173.439	C^γ 24.015; $C^{\delta 1}$ 23.162; $C^{\delta 2}$ 19.326
	8.561	4.107	1.249; 1.589		H^γ 1.523; $Q^{\delta 1}$ 0.667; $Q^{\delta 2}$ 0.674
G48	104.304	40.960		171.683	
	8.035	3.762; 4.176			
E49	119.962	57.480	26.383	175.187	C^γ 33.423
	8.624	3.568	1.849; 2.046		Q^γ 2.061
E50	117.968	57.203	25.720	176.245	C^γ 33.302
	8.890	3.856	1.870; 1.945		$H^{\gamma 2}$ 2.179; $H^{\gamma 3}$ 2.287
Q51	119.018	56.283	25.808	176.687	C^γ 31.656; $N^{\varepsilon 2}$ 112.794
	7.618	3.972	1.835; 2.211		Q^γ 2.358; $H^{\varepsilon 21}$ 6.869; $H^{\varepsilon 22}$ 7.877
L52	120.629	55.037	39.247	175.559	C^γ 22.910; $C^{\delta 1}$ 21.115; $C^{\delta 2}$ 23.406
	7.958	3.834	0.962; 1.869		H^γ 1.607; $Q^{\delta 1}$ 0.729; $Q^{\delta 2}$ 0.638

Table 7: ^{1}H, ^{15}N and ^{13}C resonance assignments of apo-ScAtx1 from *Synechocystis* PCC 6803 (cont'd)

residue	**N ($\mathbf{H^N}$)**	**$\mathbf{C^\alpha}$ ($\mathbf{H^\alpha}$)**	**$\mathbf{C^\beta}$ ($\mathbf{H^\beta}$)**	**$\mathbf{C'}$**	**Other**
R53	118.364	58.034	26.942	176.563	C^δ 39.871
	8.585	3.606	1.721; 1.802		$H^\delta 2$ 3.022; $H^\delta 3$ 3.109
T54	116.320	63.365	65.657	173.559	$C^{\gamma 2}$ 18.769
	7.990	3.840	4.113		$Q^{\gamma 2}$ 1.133
A55	125.287	52.172	14.114	177.533	
	7.901	4.078	1.272		
I56	118.668	63.317	35.186	175.187	$C^{\gamma 2}$ 15.787; $C^{\gamma 1}$ 26.961; $C^{\delta 1}$ 10.519
A57	122.261	51.569	14.676	180.186	
	8.011		1.443		
S58	116.186	58.056	59.798	171.931	
	8.246	4.192	3.958		
A59	122.632	48.615	15.633	173.809	
	7.329	4.450	1.431		
G60	105.702	42.162		170.932	
	7.957	3.935			
H61	118.665	53.313	32.230	171.058	
	6.943	4.396	2.283; 2.469		
E62	120.709	52.192	29.296	171.187	C^γ 32.835
	8.024	4.386	1.693; 1.855		$H^{\gamma 2}$ 2.033; $H^{\gamma 3}$ 2.080
	8.024	4.386	1.693; 1.855		$H^{\gamma 2}$ 2.033; $H^{\gamma 3}$ 2.080
V63	119.979	57.532	30.497	173.435	$C^{\gamma 1}$ 19.494; $C^{\gamma 2}$ 20.474
	8.022	4.548	1.892		$Q^{\gamma 1}$ 0.832; $Q^{\gamma 2}$ 1.012
E64	132.856	56.290	28.153		C^γ 34.512
	8.477	3.944	1.716; 1.929	179.072	$H^{\gamma 2}$ 1.714; $H^{\gamma 3}$ 2.065

Table 8: ^{1}H, ^{15}N and ^{13}C resonance assignments of apo-ZiaA$_N$ QC4 mutant from *Synechocystis* PCC 6803

residue	**N (H^N)**	**C^α (H^α)**	**C^β (H^β)**	**C'**	**Others**
M1					
T2					
Q3					
S4					
S5					
P6		60.261	29.242	173.927	C^γ 24.170; C^δ 47.941
		4.257	1.804		Q^γ 1.859; Q^δ 3.629
L7	118.998	51.010	40.226	173.410	C^γ 23.977; $C^{\delta1}$ 22.432; C^δ 19.786
	7.911	4.338	0.960; 1.396		H^γ 1.450; $Q^{\delta1}$ 0.605; $Q^{\delta2}$ 0.551
K8	122.504	51.741	32.723	171.512	C^γ 22.133; C^δ 26.530; C^ε 39.212
	8.663	4.365	1.123; 1.014		$\gamma^{2,3}$ 1.096; 0.938; Q^δ 1.450; Q^ε 2.813
T9	119.736	58.756	67.757	171.170	$C^{\gamma2}$ 18.837
	8.480	5.372			$Q^{\gamma2}$ 0.932
Q10	126.291	51.236	31.484	170.651	C^γ 31.519
	9.292		1.886		Q^γ 2.267
Q11	120.873	50.760	28.237	173.069	C^γ 31.224
	8.549	5.646	1.750; 1.858		Q^γ 2.241
M12	119.994	51.598	34.258	170.824	C^γ 30.368
	9.287		1.560; 1.996		Q^γ 2.486
Q13	121.883	51.753	27.243	171.863	C^γ 33.273
	8.803		1.776; 1.913		
V14	125.751	58.255	30.467	173.586	$C^{\gamma1}$ 18.913; $C^{\gamma2}$ 19.196
	8.952				$Q^{\gamma1}$ 0.660; $Q^{\gamma2}$ 0.714
G15	117.125	41.245		170.997	
	9.348	3.521			
G16	106.451	42.256		173.067	
	8.279				
M17	118.615	54.740	31.243	174.102	C^γ 29.380
R18	126.252	54.239	29.744	172.032	C^γ 24.494; C^δ 40.322
	10.802		1.804; 1.831		$H^{\gamma2}$ 1.696; $H^{\gamma3}$ 1.532; $H^{\delta2}$ 3.193; $H^{\delta3}$ 3.085
C19	114.240	53.261	27.739		
	8.417		2.949; 3.003		
A20		48.257	16.247		
A21		51.756	15.233	177.553	
		4.147	1.438		
C22	121.235	61.748	26.245	174.966	
	7.646	3.821	2.513; 3.112		
A23	118.748	52.258	15.741	176.173	
	6.625	3.603			
S24	110.513	58.259	59.754	173.764	
	7.749	4.147	3.847		
S25	116.507		59.751	174.448	
	8.001	3.521			
I26	122.015	63.255	35.237	173.930	$C^{\gamma2}$ 14.802; $C^{\gamma1}$ 27.409; $C^{\delta1}$ 11.566
	7.587	3.330			$H^{\gamma12}$ 1.586; $H^{\gamma13}$ 0.415; $Q^{\gamma2}$ 0.497; $Q^{\delta1}$ 0.061

Table 8: ^{1}H, ^{15}N and ^{13}C resonance assignments apo-ZiaA$_{N}$ QC4 mutant from *Synechocystis* PCC 6803 (cont'd)

residue	**N ($\mathbf{H}^{N}$)**	**$\mathbf{C}^{\alpha}$ ($\mathbf{H}^{\alpha}$)**	**$\mathbf{C}^{\beta}$ ($\mathbf{H}^{\beta}$)**	**$\mathbf{C}'$**	**Other**
E27	116.597	57.753	26.707	175.916	C^{γ} 33.567
	8.010	3.466			
R28	114.276	55.752	27.746	176.517	C^{γ} 24.184; C^{δ} 40.318
	8.269	3.956	1.721; 1.831		$H^{\gamma 2}$ 1.723; $H^{\gamma 3}$ 1.669; Q^{δ} 3.112
A29	118.298	51.749	15.729	178.073	
	7.560	3.957	1.341		
L30	115.589	54.221	38.248	175.658	C^{γ} 24.136; $C^{\delta 1}$ 20.661; $C^{\delta 2}$ 23.067
	7.971	3.984	1.015; 1.668		H^{γ} 1.805; $Q^{\delta 1}$ 0.632; $Q^{\delta 2}$ 0.551
E31	115.478	55.727	26.241	174.793	C^{γ} 34.442
	8.114	3.738	1.940		Q^{γ} 2.484
R32	113.803	53.250	27.744	173.757	C^{γ} 24.484; C^{δ} 40.613
	6.875	4.147	1.913		Q^{γ} 1.750; Q^{δ} 3.085
L33	122.421	51.249	38.739	173.412	C^{γ} 23.297; $C^{\delta 1}$ 22.364; $C^{\delta 2}$ 21.546
	7.499	4.202	1.722		H^{γ} 1.669; $Q^{\delta 1}$ 0.661; $Q^{\delta 2}$ 0.606
K34	126.893	55.242	28.746	173.412	C^{γ} 21.839; C^{δ} 26.240; C^{ε} 39.144
	8.344	3.874	1.694		$\gamma^{2,3}$ 1.369; 1.286; Q^{δ} 1.614; Q^{ε} 2.921
G35	110.778	42.488		172.715	
	8.521	3.438; 4.310			
V36	122.133	60.757	28.250	172.032	$C^{\gamma 1}$ 20.377; $C^{\gamma 2}$ 18.907
	7.902	3.874			$Q^{\gamma 1}$ 0.687; $Q^{\gamma 2}$ 0.551
A37	131.332	50.246	17.722	174.619	
	9.043	4.365	1.190		
E38	113.896	52.577	30.666	170.651	C^{γ} 33.018
	7.552	4.501	1.817		$H^{\gamma 2}$ 2.021; $H^{\gamma 3}$ 1.968
A39	125.126	48.280	20.237	172.033	
	9.031	5.101	1.150		
S40	113.457	53.254	62.255	170.824	
	8.494	4.855	3.522		
V41	128.910	57.265	30.743	172.204	$C^{\gamma 1}$ 20.472; $C^{\gamma 2}$ 18.636
	10.164				$Q^{\gamma 1}$ 0.715; $Q^{\gamma 2}$ 0.497
T42	122.769	57.760	66.255	172.030	$C^{\gamma 2}$ 19.651
	8.470	4.392			$Q^{\gamma 2}$ 1.178
V43	127.807	60.268	29.227	173.067	$C^{\gamma 1}$ 20.081; $C^{\gamma 2}$ 17.732
	8.799	3.302			$Q^{\gamma 1}$ 0.823; $Q^{\gamma 2}$ 0.688
A44	118.596	52.225	15.995	176.167	
	8.552	3.984	1.287		
T45	99.906	57.254	67.261	173.237	$C^{\gamma 2}$ 18.849
	6.711	4.338			$Q^{\gamma 2}$ 1.014
G46	110.668	43.225		169.096	
	7.833	3.304; 4.021			
R47	117.375	53.250	28.739	172.032	C^{γ} 24.165; C^{δ} 40.320
	7.313		1.559; 1.641		Q^{γ} 1.368; Q^{δ} 3.003
L48	128.288	50.250	42.081	172.033	C^{γ} 24.756; $C^{\delta 1}$ 24.485; $C^{\delta 2}$ 21.491
T49	122.364	58.752	66.756	170.475	$C^{\gamma 2}$ 18.558
	9.151	4.964			$Q^{\gamma 2}$ 0.960
V50	124.874	55.762	32.452	169.614	$C^{\gamma 1}$ 19.193; $C^{\gamma 2}$ 16.559
	9.083	5.182			$Q^{\gamma 1}$ 0.933; $Q^{\gamma 2}$ 0.905
T51	124.376	58.248	67.265	171.684	$C^{\gamma 2}$ 20.052
	8.929	5.782			$Q^{\gamma 2}$ 0.987
Y52	123.892	52.245	38.754	168.408	
	9.608	5.291	2.539; 2.621		

Table 8: ^{1}H, ^{15}N and ^{13}C resonance assignments apo-ZiaA$_{N}$ QC4 mutant from *Synechocystis* PCC 6803 (cont'd)

residue	**N (H^N)**	**C^α (H^α)**	**C^β (H^β)**	**C'**	**Other**
D53	118.197	46.763	38.750		
	8.514	4.965	2.431; 2.921		
P54		60.263	29.243	175.311	C^γ 24.184; C^δ 48.520
		5.373	2.212		$H^{\delta 2}$ 4.065; $H^{\delta 3}$ 3.901
K55	116.414	54.726	28.865	174.802	C^γ 22.428; C^δ 25.935; C^ε 39.142
	8.282	4.120	1.886		$\gamma^{2,3}$ 1.423; 1.314; Q^δ 1.614; Q^ε 2.895
Q56	115.083	53.756	30.752	172.031	C^γ 31.522
	7.876	4.311	1.859; 2.022		Q^γ 2.186
V57	119.458	56.772	31.228	167.718	$C^{\gamma 1}$ 19.287; $C^{\gamma 2}$ 17.461
	8.182	4.393			$Q^{\gamma 1}$ 0.878; $Q^{\gamma 2}$ 0.633
S58	116.355	52.251	63.750	172.377	
	7.610		3.794; 4.147		
E59	120.084	56.743	26.240	175.308	C^γ 34.139
	8.928	3.358	1.858; 2.131		$H^{\gamma 2}$ 2.458; $H^{\gamma 3}$ 1.832
I60	117.418	60.749	34.515	174.448	$C^{\gamma 2}$ 14.508; $C^{\gamma 1}$ 25.943; $C^{\delta 1}$ 9.509
	7.587	3.738			$H^{\gamma 12}$ 1.408; $H^{\gamma 13}$ 1.150; $Q^{\gamma 2}$ 0.742; $Q^{\delta 1}$ 0.742
T61	119.137	63.264	65.245	174.104	$C^{\gamma 2}$ 19.859
	7.468	3.766			$Q^{\gamma 2}$ 1.122
I62	120.707	62.760	35.251	173.756	$C^{\gamma 2}$ 13.631; $C^{\gamma 1}$ 26.537; $C^{\delta 1}$ 9.520
	7.263	3.031			$H^{\gamma 12}$ 1.259; $H^{\gamma 13}$ 0.278; $Q^{\gamma 2}$ 0.524; $Q^{\delta 1}$ -0.212
Q63	117.253	56.755	24.734	176.172	C^γ 31.225
	7.941	3.602	1.942; 2.159		Q^γ 2.485
E64	118.829	56.259	26.700	176.866	C^γ 33.868
	8.641	3.874	1.886; 2.077		$H^{\gamma 2}$ 2.458; $H^{\gamma 3}$ 2.077
R65	119.310	54.256	25.748	175.310	C^γ 23.889; C^δ 39.415
	7.886	4.093	1.805		$H^{\gamma 2}$ 1.614; $H^{\gamma 3}$ 1.341; $H^{\delta 2}$ 3.194; $H^{\delta 3}$ 3.058
I66	117.321	61.745	34.681	175.311	$C^{\gamma 2}$ 15.675; $C^{\gamma 1}$ 26.239; $C^{\delta 1}$ 11.283
	7.450	3.547			$H^{\gamma 12}$ 0.905; $H^{\gamma 13}$ 1.641; $Q^{\gamma 2}$ 0.741; $Q^{\delta 1}$ 0.605
A67	122.392	51.250	14.735	179.626	
	8.135	4.529	1.368		
A68	120.647	51.259	15.221	176.690	
	7.941	4.012	1.422		
L69	115.869	52.245	40.241	173.931	C^γ 24.167; $C^{\delta 1}$ 22.128; $C^{\delta 2}$ 20.660
	7.260	4.147	1.466		H^γ 1.450; $Q^{\delta 1}$ 0.769; $Q^{\delta 2}$ 0.715
G70	103.463	41.472		170.651	
	7.683	3.412; 3.849			
Y71	118.647	54.020	36.502	171.513	
	6.811		2.213; 3.385		
T72	109.597	57.249	68.254	170.998	$C^{\gamma 2}$ 18.673
	8.146				$Q^{\gamma 2}$ 0.986
L73	122.710	50.252	40.743	171.168	C^γ 24.495; $C^{\delta 1}$ 22.719; $C^{\delta 2}$ 21.839
	8.411	5.046	1.504		H^γ 1.559; $Q^{\delta 1}$ 0.797; $Q^{\delta 2}$ 0.796
A74	124.219	52.254	15.225	173.931	
	8.244	4.338	1.205		
E75	120.724	51.459	26.756		C^γ 33.273
	8.399	4.474	1.805; 1.914		$H^{\gamma 2}$ 2.187; $H^{\gamma 3}$ 1.831
P76		60.260	28.744		C^γ 24.483; C^δ 47.939
		4.311	2.104		$H^{\gamma 2}$ 1.940; $H^{\gamma 3}$ 1.804; $H^{\delta 2}$ 3.766; $H^{\delta 3}$ 3.603
K77	120.770	53.246	27.616	173.930	C^γ 21.837; C^δ 26.331; C^ε 39.144
	8.390	4.158	1.614; 1.695		Q^γ 1.314; Q^δ 1.532; Q^ε 2.867
S78	116.063	55.258	60.760	171.857	
	8.274	3.656			

Table 8: ^{1}H, ^{15}N and ^{13}C resonance assignments apo-ZiaA$_{N}$ QC4 mutant from *Synechocystis* PCC 6803 (cont'd)

residue	**N (H^{N})**	**C^{α} (H^{α})**	**C^{β} (H^{β})**	**C′**	**Other**
S79	118.464	55.257	60.760	171.515	
	8.346	3.712			
V80	120.683	59.251	29.753	173.241	$C^{\gamma 1}$ 18.031; $C^{\gamma 2}$ 17.438
	8.050	4.093			$Q^{\gamma 2}$ 0.796
T81	117.919	58.725	66.776	173.589	$C^{\gamma 2}$ 18.903
	8.168	4.229			$Q^{\gamma 2}$ 1.042
L82	124.603	52.247	39.701	173.241	C^{γ} 23.948; $C^{\delta 1}$ 21.823; $C^{\delta 2}$ 20.592
	8.263	4.256	1.463		H^{γ} 1.475; $Q^{\delta 1}$ 0.795; $Q^{\delta 2}$ 0.742
N83	123.694	51.753	38.230	173.583	
	7.932		2.540; 2.649		
G84	113.080	42.625		171.343	
	8.270	3.739			
H85	117.920	53.246	27.241		
	8.222				
K86	121.915	53.242	29.710	173.070	C^{γ} 21.720; C^{ε} 39.145
	8.398				Q^{γ} 1.232
H87	120.381				
	8.210				
P88		63.239	28.719	174.790	C^{γ} 24.169; C^{δ} 47.653
			1.668		Q^{γ} 1.832; Q^{δ} 3.330
H89	118.616	53.236	27.257	172.554	
	8.522				
S90	115.965				
	8.137		3.957		
H91		53.242	27.235	172.204	
R92	121.677	53.257	27.241	174.447	
	8.125		1.913		
E93	121.431	53.224	27.723		
	8.080	4.202			
E94	121.504	53.702	27.252	174.274	C^{γ} 33.274
	8.397	4.147	1.777; 1.912		$H^{\gamma 2}$ 1.967; $H^{\gamma 3}$ 1.914
G95	108.753	42.276		171.213	
	8.348				
H96	118.168	53.255	27.259	175.658	
	8.062				
S97	115.951	55.258	60.754		
	8.153				
H98	118.564				
	8.279				
S99		55.256	60.755	176.517	
H100	120.490	53.254	27.254	172.896	
	8.314				
G101	109.355	42.266			
	8.321				
A102	123.087	49.736	16.237	175.484	
	8.202	4.214	1.259		
G103	107.279	42.241		171.342	
	8.288				
E104	119.508	53.759	27.236	173.413	
	8.060				

Table 8: ^{1}H, ^{15}N and ^{13}C resonance assignments apo-ZiaA$_{N}$ QC4 mutant from *Synechocystis* PCC 6803 (cont'd)

residue	**N ($\mathbf{H^N}$)**	**$\mathbf{C^\alpha}$ ($\mathbf{H^\alpha}$)**	**$\mathbf{C^\beta}$ ($\mathbf{H^\beta}$)**	**$\mathbf{C'}$**	**Other**
F105	119.556	54.743	36.499	172.547	
	8.127	4.474			
N106	119.567	50.225	35.753	172.547	
	8.182	4.528			
L107	121.719	52.249	39.261	174.446	C^γ 23.901; $C^{\delta1}$ 21.834; $C^{\delta2}$ 20.512
	8.013				$Q^{\delta1}$ 0.747; $Q^{\delta2}$ 0.688
K108	121.346	53.240	29.733	173.413	C^γ 21.836; C^δ 25.888; C^ε 39.144
	8.158	4.229			$\gamma^{2,3}$ 1.181; 1.123; Q^δ 1.482
Q109	121.266	52.756	26.732	172.716	C^γ 33.499
	8.246	4.201	1.995		Q^γ 2.077
E110	122.336	53.246	27.241	172.545	C^γ 33.279
	8.321	4.147			Q^γ 2.130
L111	128.259		40.257		C^γ 24.188; $C^{\delta1}$ 22.136; $C^{\delta2}$ 20.668
	7.832	4.065	1.450		H^γ 1.450; $Q^{\delta1}$ 0.769; $Q^{\delta2}$ 0.715

Table 9: Statistical analysis of the structural model of the ScAtx1-Cu(I)-ZiaA$_N$ QC4 complex obtained through HADDOCK

RMSD (Å)	**No. of structures**	**E_{vdw} (kcal.mol^{-1})**	**E_{elec} (kcal.mol^{-1})**	**E_{AIR} (kcal.mol^{-1})**	**Buried surface area (Å)**
1.4 ± 0.9	150	-42.5 ± 1.9	-336.7 ± 37.1	349.3 ± 76.65	1367 ± 133.9

3.3 Project 3: Structural and dynamic characterization of intrinsically disordered proteins: The example of the N-terminal ALR fragment

- Paper 3 (draft in preparation)

Introduction

Characterization of disordered states of proteins by NMR is particularly challenging because the polypeptide chain is inherently flexible and rapidly interconverts between multiple conformations [12, 9, 10, 13]. As a result the chemical shift dispersion of most resonances is poor, and sequence-specific assignment of resonances may become difficult. Recent improvements in probehead technology together with development of new NMR experiments help overcome this problem making feasible direct detection of heteronuclei and thus offering an alternative solution for demanding situations.

A suite of ^{13}C-detected NMR experiments [14, 15, 16, 17] recently developed at CERM, now provide an extra tool for the sequence-specific assignment and structure and dynamics characterization of unfolded protein fragments [18, 19]. The heteronuclear shifts which are intrinsically characterized by a large chemical shift dispersion and the reduced sensitivity of heteronuclei to the exchange broadening are two features that make exclusively heteronuclear experiments particularly well suited for the characterization of unfolded/disordered segments. For this reason, we have used them, in combination with the ^{1}H-detected ones, for the characterization of the N-terminal protein fragment preceding the catalytic core of the ALR protein.

ALR is a FAD-dependent sulfhydryl oxidase of the mitochondrial intermembrane space (IMS). It is an essential component of the redox regulated MIA40/ALR import and assembly pathway used by many of the cysteine-containing intermembrane space proteins. Upon import of a Cys-reduced substrate, Mia40 interacts with the substrate via intermolecular disulphide bond and shuttles a disulphide to its substrate [20]. ALR functions in catalyzing reoxidation of the reduced Mia40 and release of the substrate.

The ALR enzymes is characterized by a highly conserved central catalytic core of $\sim$100 amino acids, which includes an active site CXXC motif, $CX_{16}C$ disulphide bond, and residues involved in FAD binding. The $CX_{16}C$ disulphide bond is proposed to play a structural role in stabilizing the FAD binding and/or protein folding. Apart from the catalytic core, the other parts of the protein seem flexible and unfolded.

The disordered fragments of complex macromolecular machines are often involved in molecular recognition, signalling, etc., this may play an important role for the overall protein function. Since it may not be trivial to characterize at once the whole macromolecular machine, as an intermediate step, the N-terminal fragment was expressed, purified and characterized through NMR in the two biologically relevant redox states (reduced and oxidized), in order to lay the basis for further experimental characterization of this complex system.

Materials and Methods

Protein Expression and Purification

The recombinant construct pET16HisMBPTEVN80 ALR was expressed as a soluble protein in *E. coli* Origami B strain. Lysates were loaded onto a Ni-beads slurry (used buffer: 50 mM NaCl, 50 mM Tris pH 8.00, 5 mM imidazole). Collected fractions were dialyzed and subsequently subjected to the cleavage with the TEV protease (buffer used for TEV cleavage: 50 mM NaCl, 50 mM Tris pH 8.00, 1 M DTT). The cleaved protein was dialyzed again (into 50 mM NaCl, 50 mM Tris pH 8.00, 5 mM imidazole) to remove the DTT for the further purification step by the negative affinity chromatography. If needed, an additional purification step on the Q Sepharose could be implemented.

The final concentration of the ^{15}N labeled and ^{13}C, ^{15}N labeled human N-80 ALR samples was in the 0.3–0.8 mM range, in 50 mM phosphate buffer at pH 7.00 with 0.25 mM EDTA. Reduction was achieved by gradual addition of DTT (monitored by NMR) up to a final concentration of 20 mM and incubated overnight. D_2O was added to the buffer for the lock signal.

NMR experiments

Multidimensional NMR experiments were recorded either on a Bruker AVANCE 700 spectrometer equipped with a triple-resonance cryogenically cooled probehead (TCI), or on a Bruker AVANCE 800 spectrometer equipped with a triple-resonance cryogenically cooled probehead (TXI), or on a Bruker AVANCE 700 spectrometer equipped with a triple-resonance cryogenically cooled probehead optimized for ^{13}C sensitivity (TXO). All the experiments were recorded at 295.1 K, unless otherwise specified.

The 3D experiments used for the sequence-specific assignment as well as the experiments to determine ^{15}N relaxation rates (R_1, R_2, ^{1}H-^{15}N NOE) and exchange rates with the solvent (CLEANEX) [21] are listed in Table 1. Standard pulses and parameters generally used for protein biomolecular applications were used. All ^{13}C detected NMR experiments employ the IPAP approach [14] to suppress the C^{α}–C' coupling in the direct acquisition dimension. To this end the in-phase (IP) and antiphase (AP) components are acquired and stored separately, doubling the number of FIDs recorded in one of the indirect dimensions.

Chemical shifts were referenced using the ^{13}C signal of dioxane (69.3 ppm) as well as the ^{1}H and ^{13}C shifts of DSS. Nitrogen chemical shifts were referenced indirectly to the ^{1}H standard using the conversion factor derived from

the ratio of NMR frequencies. Data were acquired and processed by using Topspin 1.3 software. Spectral analysis was carried out with the aid of the program CARA/NEASY.

Results and Discussions

The ^{1}H-^{15}N HSQC and CON spectra of the reduced N-80 ALR protein (N-80 ALR_{SH}) are shown in Figure 1. The chemical shift dispersion is typical of a disordered system. Nevertheless the first count of the cross peaks of the two spectra already allows us to identify the correlations for a large fraction of the amino acids (63/73 cross peaks in ^{1}H-^{15}N HSQC spectra, 79/79 in CON spectra) indicating promising conditions for detailed structural and dynamic characterization of the protein. An experimental strategy combining ^{1}H and ^{13}C detected multidimensional NMR experiments was used to achieve the sequence-specific assignment. As a result the complete sequence-specific assignment of H^{N}, N, C', C^{α}, $C^{\beta}$was achieved (100 % N, 100 % H^{N}, 100 % C^{α}, 100 % C^{β}, 99 % C') and is reported in Table 2. The chemical shifts, in particular heteronuclear ones, provide the initial information for the structural and dynamic characterization of the protein giving, if any, evidence of regions characterized by secondary structural propensities. The difference of experimental chemical shifts respect to those predicted on the basis of the primary sequence using random-coil chemical shifts properly corrected to include the contributions of previous an following amino acids, are reported in Figure 2. The peak values are observed for residues preceding prolines, which could be due to poor prediction of reference random-coil chemical shifts in these cases [22, 23]. Other than these, a region comprising residues 23–37 shows typical values indicating α-helical secondary structural propensity. In addition a short amino acid fragment preceding the CXXC motif (63–68) also shows some indication ofα-helical secondary structure while the remaining data are quite scattered and thus are difficult to be interpreted in a straightforward way in terms of secondary structural propensities. Therefore the analysis of other NMR observables is important to contribute to the structural and dynamic characterization of this disordered protein. The ^{1}H-^{15}N NOEs provide a clear indication of parts of the polypeptide chain characterized by different degrees of local flexibility (fast motions). Positive ^{1}H-^{15}N NOE values are identified for residues 4–45 and 62–70. The initial part of the polypeptide chain (4–45) includes the fragment 23–37 for which chemical shifts indicate α-helical secondary structural propensity as well as a stretch of bulky hydrophobic residues (13–18), which could also contribute to reduced flexibility. The other fragment in which positive ^{1}H-^{15}N NOE values are detected is located in the last part of the polypeptide chain where a short fragment with α-helical secondary structural propensity was identified, right before the CXXC motif. Interestingly the CXXC motif is preceeded by an Arg rich fragment (bulky and potentially charged residues). These two regions are separated by an amino acid fragment (45–62) characterized by

negative ^{1}H-^{15}N NOE values, which indicate higher local flexibility. Negative ^{1}H-^{15}N NOEs are also observed for the initial and terminal residues of the polypeptide chain. Exchange rates of amide protons with the solvent are in agreement with a more protected region between residues 13–19 (characterized by bulky amino acids) and 23–35 (withα-helical secondary structural propensity). High exchange effects are instead observed for residues 10–12 and 20–22, which could indicate the presence of loops in fragments that are overall characterized by reduced flexibility and higher structural propensity (13–19 and 23–37). The region 35–59 (with the exception of residues 50–54) is also influenced by efficient exchange processes in which also negative NOEs were identified (45–62).

This N-80 ALR fragment could thus be described as constituted by and N-terminal module (4–45), which includes a fragment rich in bulky amino acids (13–19) as well as a fragment withα-helical secondary structural propensity (23–37) and by a C-terminal module (62–74) containing a more structured, Arg rich part and the CXXC motif. These two less flexible, more structured modules appear separated by a central fragment characterized by fast motions (45–62) as evidenced by ^{1}H-^{15}N NOEs.

It is thus interesting to investigate what happens upon oxidation of the protein which can be achieved by removal of the reducing agent DTT.

The most simple 2D ^{13}C detected experiments (CON, CACO, CBCACO), as well as ^{1}H-^{15}N HSQC experiment were initially collected to evaluate the extent of variations observed upon oxidation. The protein still behaves as a largely unstructured protein and the spectral variations upon oxidation involve a subset of residues (Figure 5). This already indicates that no marked folding occurs upon oxidation, as identified in other cases. The complete sequence-specific assignment of the protein in the oxidized form can thus be achieved (Table 3) by collecting a subset of the 3D experiments reported in Table 1, taking advantage of the sequence-specific assignment of the reduced form (Table 2). Among them the CBCACON and CBCANCO were selected as they provide complete information on the polypeptide chain.

The chemical shift differences between the reduced and oxidized forms of the protein are reported in Figure 6. These clearly indicate that the major variations in chemical shifts are observed for the residues in the close vicinity of the CXXC binding motif (71–74), including the characteristic chemical shifts changes of C$^{\beta}$s of cysteines occurring upon disulfide bond formation [24]. The ^{15}N relaxation rates (Figure 7), in particular R_2 rates indicate an overall increase throughout the primary sequence, which could either be due an overall more structured conformation or to partial population of a dimeric form at equilibrium. As the variations in chemical shifts between the two forms are not so pronounced a much more structured conformation

can probably be ruled out leaving open the hypothesis of partial dimerization could contribute to this observation. ^{1}H-^{15}N NOEs indicate that the overall modular behavior of the protein is maintained, with a reduced degree of fast motions in the C-terminal module of the protein around the CXXC motif. Comparison of exchange data shows as the most notable difference that exchange processes are reduced in the C-terminal region also indicating a more protected structural ensemble around the CXXC motif.

Finally, in order to evaluate the importance of studying this N-terminal module of the human ALR protein, the 2D ^{1}H-^{15}N HSQC experiment of N-80 ALR_{SH} is overlaid with that of the full length ALR reduced protein.

Concluding the present characterization of the N-80 ALR fragment in the oxidized and reduced forms shows that:

1. This fragment is largely unstructured in both oxidation states.

2. The protein can be described as constituted by an N-terminal module (4–45) comprising a region characterized by a bulky amino acid cluster (13–19) and an α-helical secondary structural propensity (23–37) and by a C-terminal module (62–74) containing the CXXC domain. The two modules are separated by a more flexible fragment characterized by smaller NOE values (45–62).

3. The redox reaction leaves the N-terminal module essentially unaffected and produces the major changes in the C-terminal module. The latter is characterized by higher NOE values and reduced exchange effects, both features indicating a less flexible, more protected conformation. A general increase in ^{15}N R_2 values may indicate partial dimer formation.

4. The overlay of the HSQC spectra of this fragment with those of the whole ALR protein indicates that the chemical shift assignment obtained here is going to provide a suitable starting point for the study of the role of this N-terminal fragment in the complete macromolecular machinery.

Bibliography

[1] T Mittag, LE Kay, and JD Forman-Kay. Protein dynamics and conformational disorder in molecular recognition. *Journal of Molecular Rocognition*, 23:105–116, 2010.

[2] Peter Tompa and Lajos Kalmar. Power law distribution defines structural disorder as a structural element directly linked with function. *J Mol Biol*, 403:346–50, 2010.

[3] VN Uversky, CJ Oldfield, U Midic, HB Xie, B Xue, S Vucetic, LM Iakoucheva, Z Obradovic, and AK Dunker. Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics*, 10, 2009.

[4] P Tompa, C Szasz, and L Buday. Structural disorder throws new light on moonlighting. *Trends in Biochemical Sciences*, 30:484–489, 2005.

[5] P Tompa. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters*, 579:3346–3354, 2005.

[6] M Vendruscolo, J Zurdo, CE Macphee, and CM Dobson. Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems. *Philosophical Transactions of the Royal Society a Mathematical Physical and Engineering Sciences*, 361:1205–1222, 2003.

[7] M Vendruscolo, E Paci, M Karplus, and CM Dobson. Structures and relative free energies of partially folded states of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 100:14817–14821, 2003.

[8] T Mittag and JD Forman-Kay. Atomic-level characterization of disordered protein ensembles. *Current Opinion in Structural Biology*, 17:3–14, 2007.

[9] HJ Dyson and PE Wright. Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Unfolded Proteins*, 62:311–340, 2002.

[10] HJ Dyson and PE Wright. Unfolded proteins and protein folding studied by nmr. *Chemical Reviews*, 104:3607–3622, 2004.

[11] VN Uversky. Seven lessons from one idp structural analysis. *Structure*, 18:1069–1071, 2010.

[12] T Mittag, J Marsh, S Orlicky, M Borg, X Tang, F Sicheri, HS Chan, LE Kay, M Tyers, and JD Forman-Kay. "fuzzy" complexes: How much disorder can a biologically relevant complex tolerate, and can it even be beneficial? *Biochemistry and Cell Biology-Biochemie et Biologie Cellulaire*, 88:403–403, 2010.

[13] STD Hsu, LD Cabrita, P Fucini, CM Dobson, and J Christodoulou. Probing protein folding on the ribosome by solution state nmr spectroscopy. *Journal of Biomolecular Structure & Dynamics*, 26:94, 2009.

[14] W Bermel, I Bertini, IC Felli, R Pierattelli, and PR Vasos. A selective experiment for the sequential protein backbone assignment from 3d heteronuclear spectra. *Journal of Magnetic Resonance*, 172:324–328, 2005.

[15] W Bermel, I Bertini, IC Felli, M Piccioli, and R Pierattelli. ^{13}c-detected protonless nmr spectroscopy of proteins in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 48:25–45, 2006.

[16] W Bermel, I Bertini, IC Felli, R Kummerle, and R Pierattelli. Novel ^{13}c direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins (vol 178, pg 56, 2006). *Journal of Magnetic Resonance*, 180:321–321, 2006.

[17] W Bermel, I Bertini, IC Felli, YM Lee, C Luchinat, and R Pierattelli. Protonless nmr experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *Journal of the American Chemical Society*, 128:3918–3919, 2006.

[18] W Bermel, I Bertini, V Csizmok, IC Felli, R Pierattelli, and P Tompa. H-start for exclusively heteronuclear nmr spectroscopy: The case of intrinsically disordered proteins. *Journal of Magnetic Resonance*, 198:275–281, 2009.

[19] V Csizmok, IC Felli, P Tompa, L Banci, and I Bertini. Structural and dynamic characterization of intrinsically disordered human securin by nmr spectroscopy. *Journal of the American Chemical Society*, 130:16873–16879, 2008.

[20] L Banci, I Bertini, C Cefaro, S Ciofi-Baffoni, A Gallo, M Martinelli, DP Sideris, N Katrakili, and K Tokatlidis. Mia40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria. *Nature Structural & Molecular Biology*, 16:198–206, 2009.

[21] TL Hwang, PCM van Zijl, and S Mori. Accurate quantitation of water-amide proton exchange rates using the phase-modulated clean chemical exchange (cleanex-pm) approach with a fast-hsqc (fhsqc) detection scheme. *Journal of Biomolecular NMR*, 11:221–226, 1998.

[22] S Schwarzinger, GJA Kroon, TR Foss, J Chung, PE Wright, and HJ Dyson. Sequence-dependent correction of random coil nmr chemical shifts. *Journal of the American Chemical Society*, 123:2970–2978, 2001.

[23] HY Zhang, S Neal, and DS Wishart. Refdb: A database of uniformly referenced protein chemical shifts. *Journal of Biomolecular NMR*, 25:173–195, 2003.

[24] DS Wishart and BD Sykes. The ^{13}c chemical-shift index - a simple method for the identification of protein secondary structure using ^{13}c chemical-shift data. *Journal of Biomolecular NMR*, 4:171–180, 1994.

Figure 1. A schematic representation of the ALR protein reported on top is used to show which is the fragment that was studied in this work (N-80 ALR). The ^{1}H-^{15}N HSQC and CON spectra acquired on the reduced form of the N-80 ALR protein fragment (N-80 ALR_{SH}) are shown in the bottom. It is worth noting that the two spectra are displayed with the same ^{15}N scale in the indirect dimension and with comparable scales for ^{1}H and ^{13}C in the direct dimension to have the same spectral widths (in Hz). This way of displaying the two spectra shows the improved chemical shift dispersion of the CON spectrum respect to the HN HSQC one.

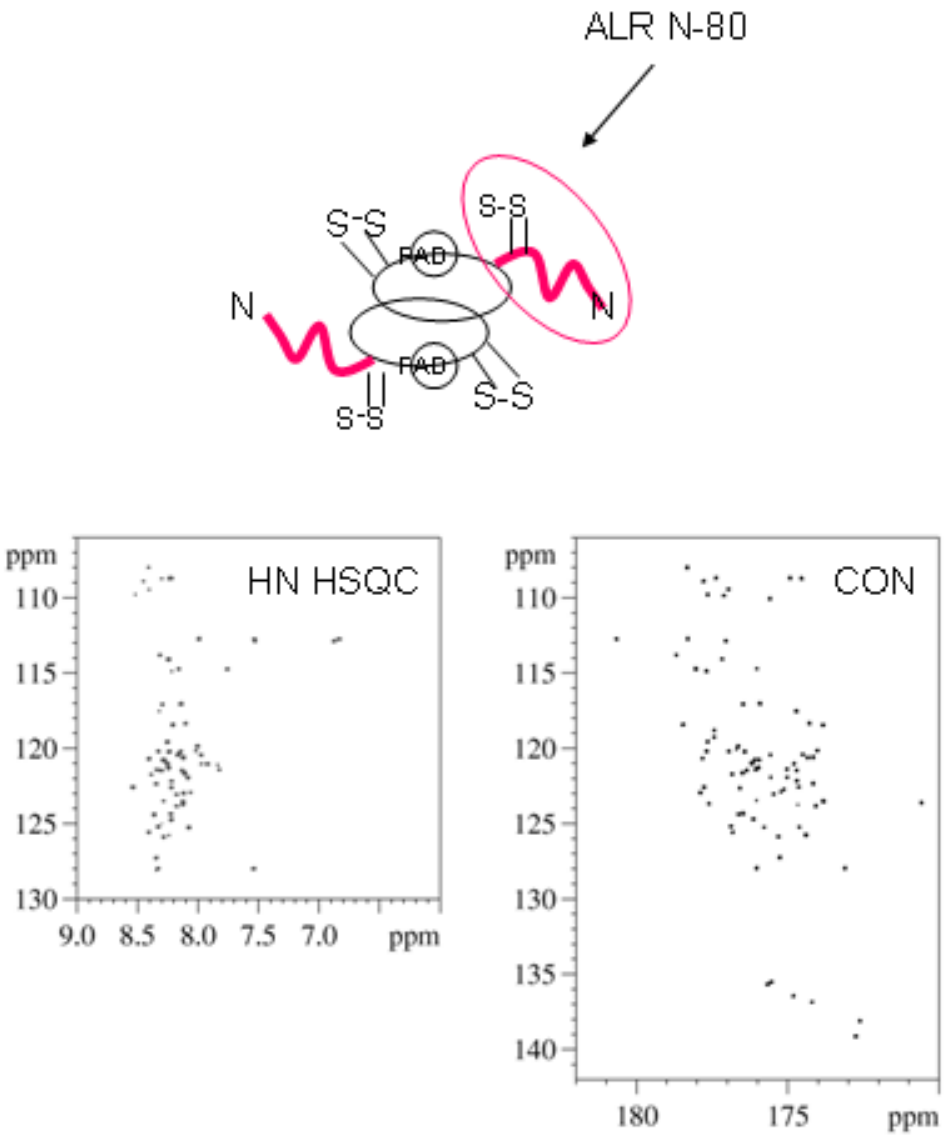


Figure 2. Differences (in ppm) of the observed chemical shifts with respect to reference values characteristic of the random-coil conformation of small peptides, properly corrected to account for the contributions by sequential neighbors, are shown: C^{a} carbons (top) and $C^{'}$ carbons (bottom) for N-80 ALR_{SH} protein fragment.

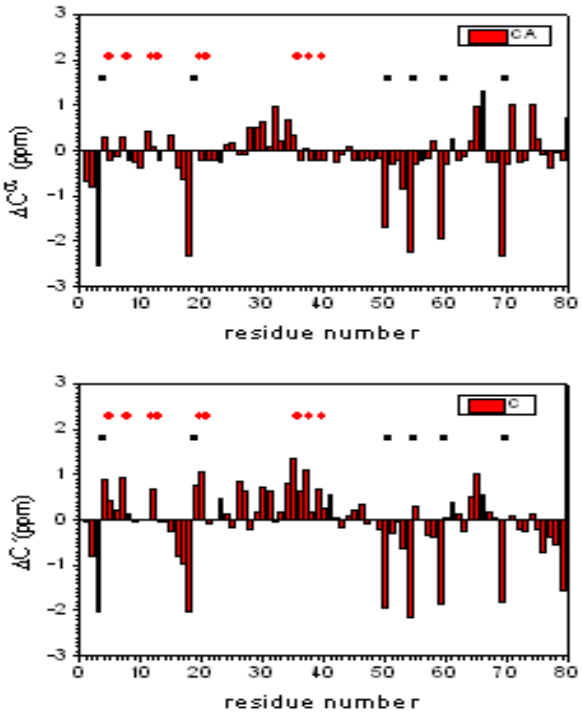


Figure 3. ^{15}N relaxation rates (^{15}N R_1, R_2, ^{1}H-^{15}N NOEs) for N-80 ALR_{SH} protein fragment.

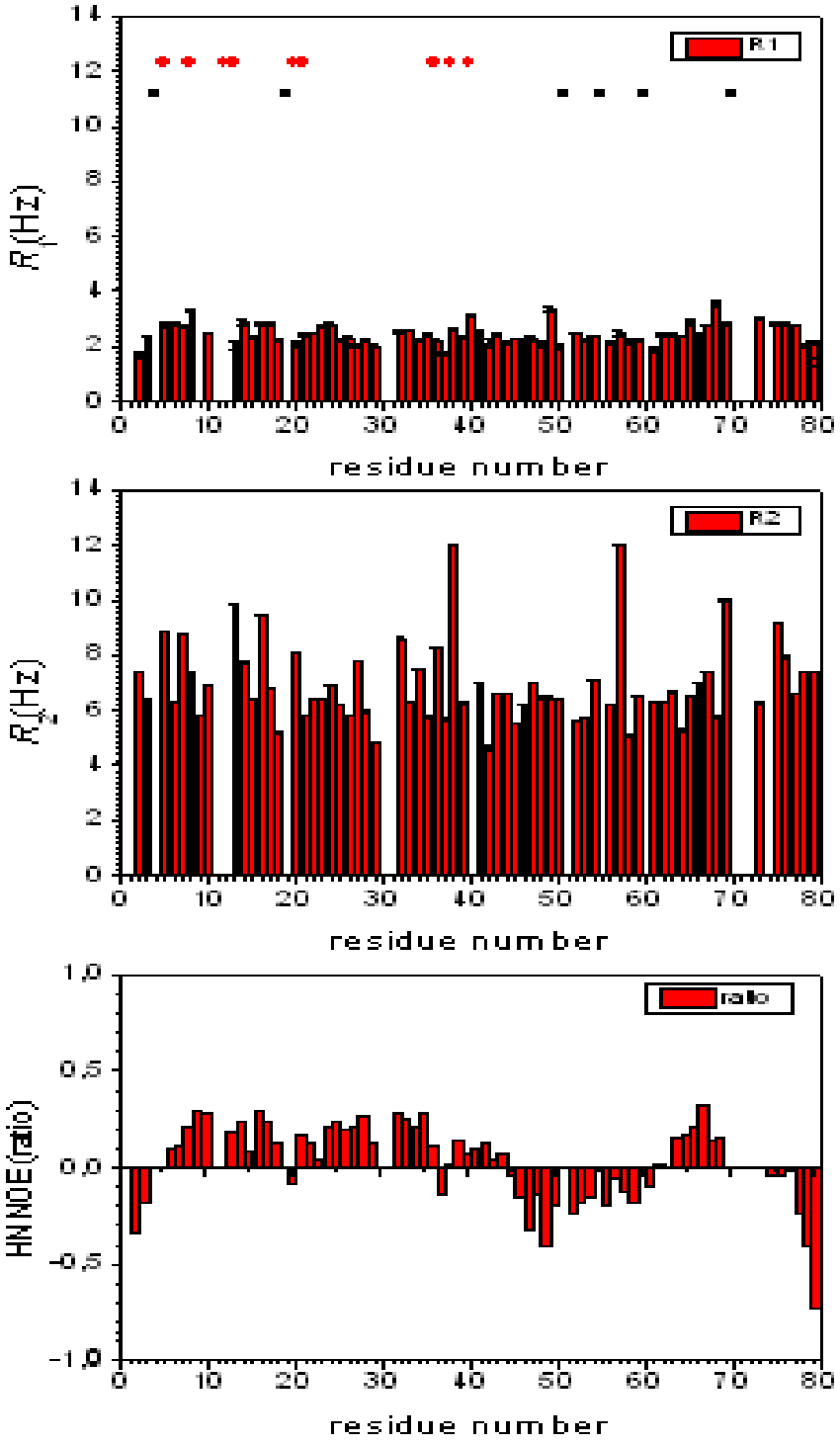


Figure 4. Ratios between cross peak intensities measured in CLEANEX experiments and in reference experiments are shown (top) for a few residues versus the CLEANEX mixing time (2.5 ms, 5.0 ms, 7.5 ms, 10 ms, 12.5 ms) and for all residues versus the residue number (bottom) for N-80 ALR_{SH} protein fragment.

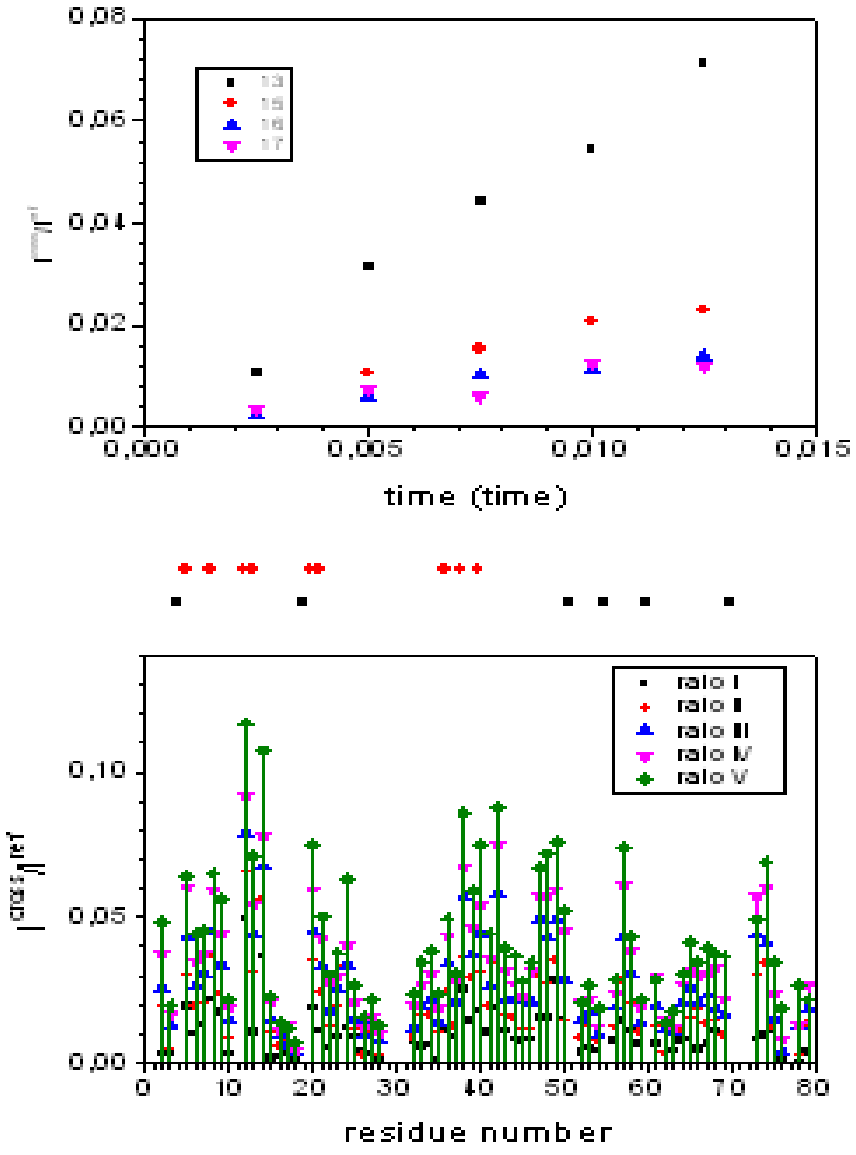


Figure 5. Superposition of CON and CBCACO spectra of N-80 ALR_{SH}/ ALR_{SS} protein fragment (green/blue contours respectively).

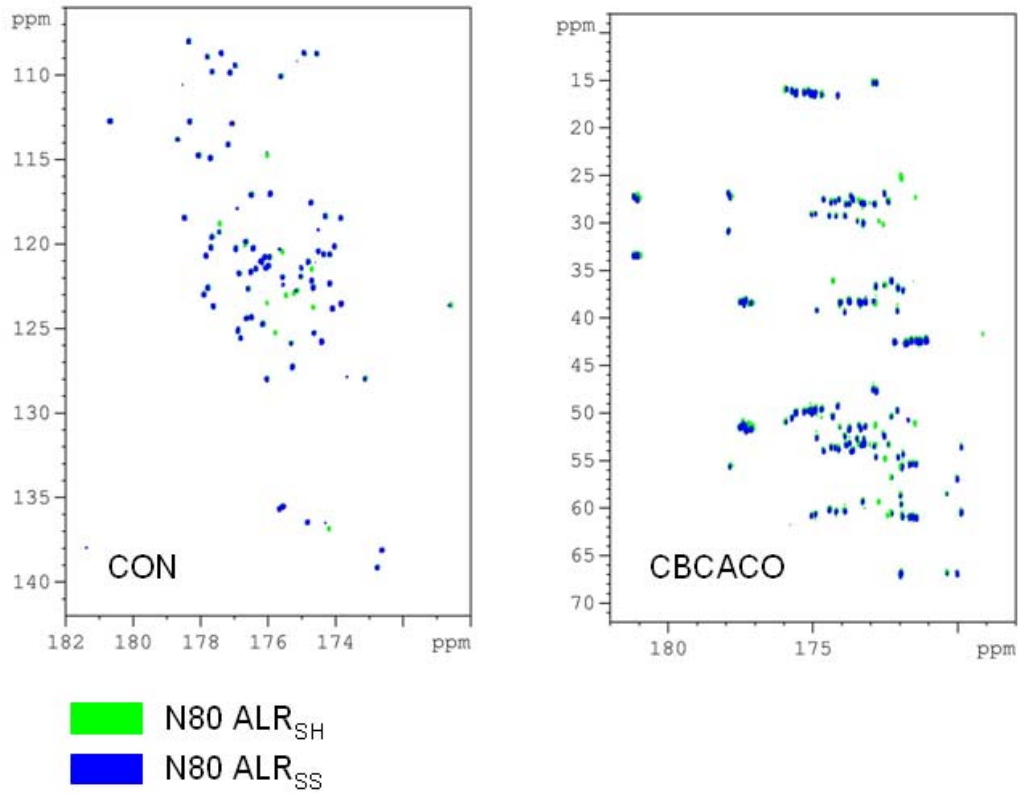


Figure 6. Chemical shift differences between the oxidized and reduced forms of N-80 ALR.

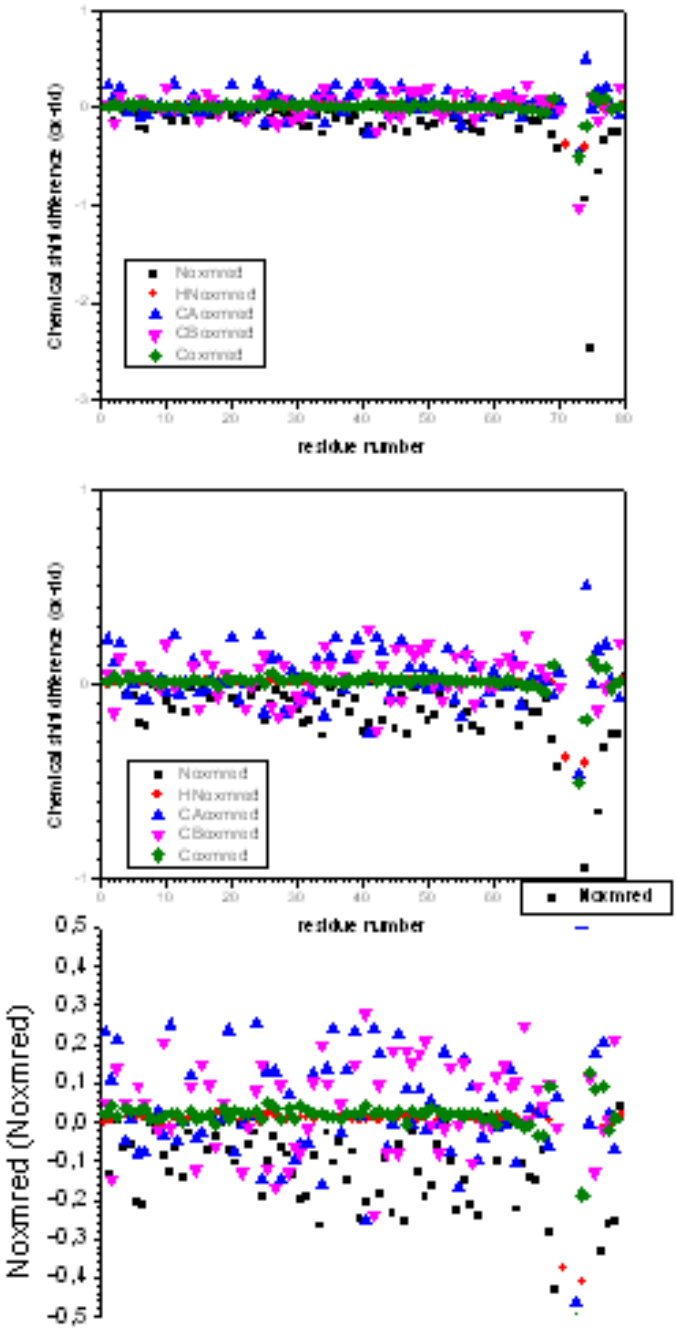


Figure 7. ^{15}N relaxation rates (^{15}N R_1, R_2, ^{1}H-^{15}N NOEs) for N-80 ALR_{SS} protein fragment.

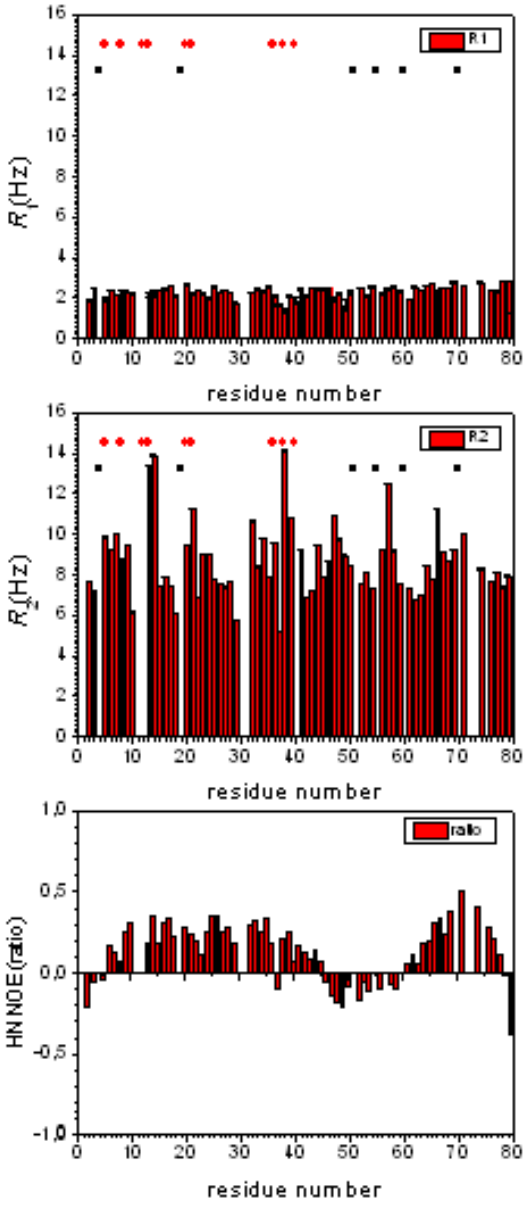


Figure 8. The ratios indicating amide proton exchange with the solvent for N-80ALRss.

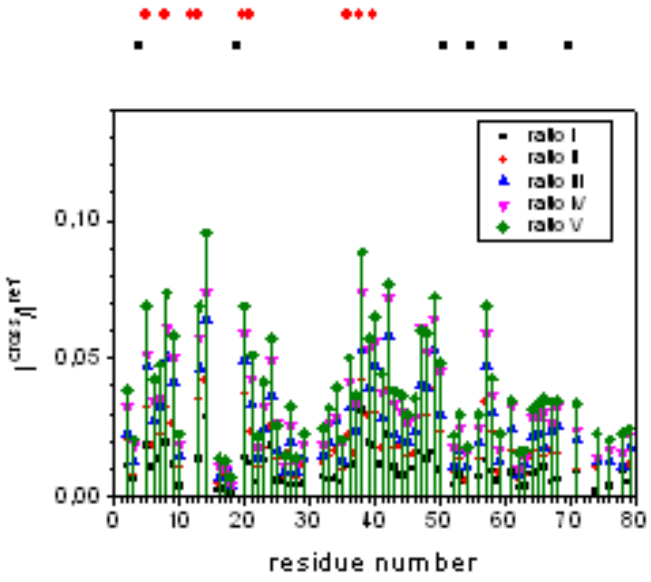
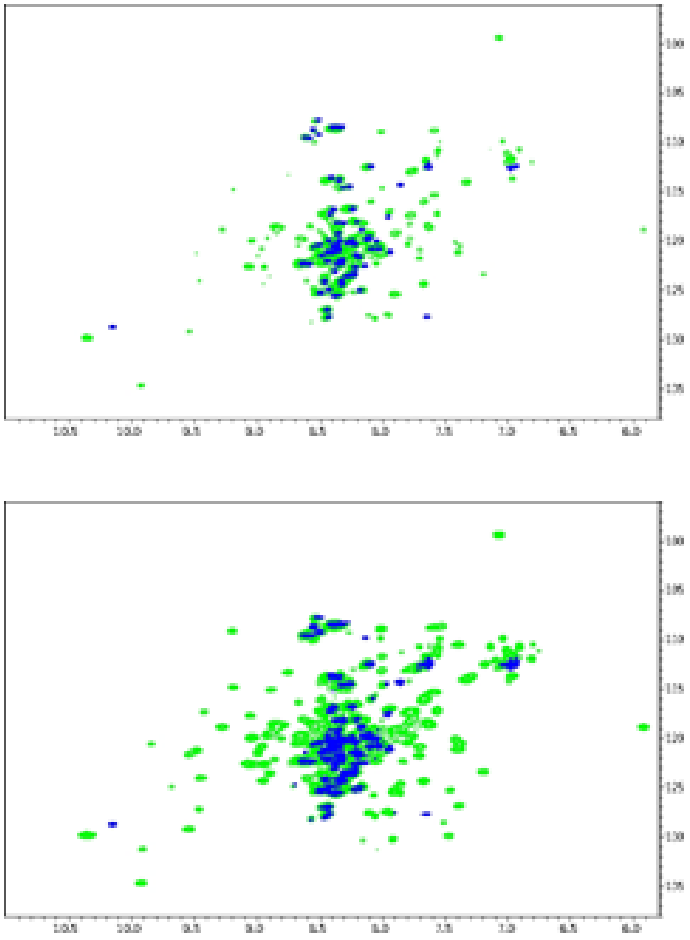


Figure 9. Overlay of ^{1}H-^{15}N HSQC spectra of the oxidized full-length ALR protein (green, 308K) and the N-80 ALR fragment (blue, 295K). The different panels display the same data with different contour levels for the full-length protein.



Supplementary Material

Table 1: Experimental details of the ^{1}H detected and ^{13}C detected NMR experiments acquired on purified human N-80 ALR_{SH} at 295.1 K.

Experiments		**Dimension**			**SW (ppm)**			**NS**	**d_1 (s)**
		F3	**F2**	**F1**	**F3**	**F2**	**F1**		
^{13}C	(H)CACO	660 (^{13}C)	2048 (^{13}C)		50.0	60.2		8	1.5
	(H)CBCACO	360 (^{13}C)	860 (^{13}C)		76.1	24.0		8	0.9
	CON	400 (^{15}N)	1024 (^{13}C)		30.1	40.0		16	1.5
	(H)CBCACON	200 (^{13}C)	76 (^{15}N)	1024 (^{13}C)	60.4	36.0	50.0	8	1.4
	(H)CBCANCO	188 (^{13}C)	364 (^{15}N)	1024 (^{13}C)	60.4	36.0	30.1	16	1.3
	(H)COCON	128 (^{13}C)	64 (^{15}N)	512 (^{13}C)	12.0	35.0	30.1	16	1.3
^{1}H	^{1}H-^{15}N HSQC	256 (^{15}N)	2048 (^{1}H)		40.0	18.6		2	1.0
	HNCO	192 (^{13}C)	160 (^{15}N)	2048 (^{1}H)	22.1	36.0	13.9	2	1.0
	HNCANNH	70 (^{13}C)	70 (^{15}N)	2048 (^{1}H)	25.0	25.0	13.9	32	1.0
	HN(CA)CO	96 (^{13}C)	100 (^{15}N)	2048 (^{1}H)	11.0	26.0	13.9	4	1.0
	CBCA(CO)NH	192 (^{13}C)	160 (^{15}N)	2048 (^{1}H)	75.1	36.0	13.9	4	1.0
	CBCANH	192 (^{13}C)	160 (^{15}N)	2048 (^{1}H)	75.1	36.0	13.9	4	1.0
	^{15}N R_1	512 (^{15}N)	2048 (^{1}H)		22.0	16.1		8	3.0
	^{15}N R_2	512 (^{15}N)	2048 (^{1}H)		24.0	16.1		8	3.0
	^{1}H-^{15}N NOE	512 (^{15}N)	2048 (^{1}H)		22.0	16.1		32	6.0
	^{1}H-^{15}N CLEANEX	128 (^{15}N)	1024 (^{1}H)		22.0	12.0		16	3.0

Table 2: Backbone chemical shifts (ppm), ALR N-80 oxidized

Residue	N	H^N	C^α	C^β	C'
M1	-	8.26	55.23	33.05	175.34
A2	125.74	8.28	52.17	19.36	176.91
A3	125.17	8.33	50.51	18.17	175.68
P4	135.59	-	63.74	31.87	177.81
G5	108.85	8.46	45.31	-	174.36
E6	120.39	8.13	56.66	30.45	176.86
R7	121.52	8.39	56.65	30.69	176.99
G8	109.42	8.41	45.32	-	-
R9	120.11	8.02	56.18	30.92	175.97
F10	120.67	8.26	57.59	39.65	175.59
H11	121.80	8.23	56.41	30.92	175.64
G12	109.98	8.03	45.55	-	-
G13	108.57	8.30	45.31	-	-
N14	118.42	8.27	53.34	38.95	175.05
L15	121.80	8.09	55.24	42.25	176.66
F16	119.83	8.01	57.59	39.88	174.82
F17	120.95	7.93	57.35	40.12	174.65
L18	125.17	8.08	52.64	42.02	174.84
P19	136.43	-	63.50	31.87	177.68
G20	109.70	8.52	45.55	-	174.94
G21	108.57	8.23	45.22	-	173.85
A22	123.49	8.12	52.63	19.36	177.71
R23	120.11	8.33	56.17	30.92	176.50
S24	117.01	8.29	58.76	63.87	174.67
E25	122.36	8.54	56.88	30.21	176.45
M26	120.11	8.25	55.70	32.58	176.24
M27	120.95	8.26	55.71	32.81	176.00
D28	121.23	8.25	54.54	41.31	176.11
D29	120.67	8.29	54.76	41.07	176.52
L30	121.52	8.11	55.47	42.02	177.63
A31	123.49	8.13	52.88	19.12	178.33
T32	112.51	7.99	62.55	69.63	174.71
D33	122.08	8.22	54.52	41.31	176.50
A34	124.05	8.23	53.35	19.11	178.49
R35	118.42	8.21	56.89	30.45	177.40
G36	108.57	8.24	45.55	-	174.53
R37	120.39	8.17	56.43	30.69	177.13
G38	109.70	8.48	45.45	-	174.1
A39	123.77	8.17	52.87	19.36	178.35
G40	107.73	8.41	45.33	-	174.19
R41	120.39	8.12	56.19	30.92	176.61
R42	122.64	8.48	56.40	30.70	176.08
D43	121.23	8.34	54.29	41.30	176.17

Table 2: Backbone chemical shifts (cont'd), ALR N-80 oxidized

Residue	N	H^N	C^α	C^β	C'
A44	124.61	8.23	52.87	19.12	177.80
A45	122.36	8.22	52.64	19.12	177.91
A46	122.92	8.12	52.87	19.13	178.06
S47	114.48	8.16	58.31	63.97	174.42
A48	125.74	8.25	52.64	19.36	177.72
S49	114.76	8.22	58.30	63.96	174.31
T50	118.14	8.10	59.95	69.86	172.77
P51	138.96	-	63.25	32.10	176.65
A52	124.33	8.36	52.63	19.12	177.68
Q53	119.55	8.26	55.46	29.75	175.30
A54	127.14	8.35	50.51	18.17	175.56
P55	135.30	-	63.03	32.11	177.19
T56	113.92	8.25	61.61	70.10	174.74
S57	117.30	8.32	58.30	63.97	174.18
D58	122.08	8.35	54.29	41.29	175.94
S59	117.01	8.14	56.42	63.49	172.62
P60	138.12	-	63.26	32.10	176.97
V61	120.11	8.15	62.31	32.81	176.05
A62	127.99	8.33	52.63	19.35	177.84
E63	120.67	8.41	56.88	30.45	176.37
D64	121.23	8.30	54.29	41.31	176.82
A65	125.46	8.40	53.81	18.87	178.66
S66	113.64	8.31	59.72	63.50	175.02
R67	121.23	7.84	56.18	30.45	176.15
R68	120.95	7.99	56.18	30.68	176.02
R69	123.20	8.28	54.05	30.45	174.30
P70	136.43	-	63.26	32.10	176.25
C71	119.27	8.07	-	-	-
R72	-	-	-	-	-
A73	-	-	52.17	18.18	176.93
C74	117.86	7.86	59.01	38.70	174.53
V75	118.98	-	62.32	32.81	175.58
D76	122.36	8.25	54.28	41.07	175.67
F77	120.11	7.95	57.83	39.42	175.33
K78	122.64	8.04	56.18	33.28	176.03
T79	114.48	7.76	61.37	69.86	173.14
W80	127.99	7.55	-	-	-

Table 3: Backbone chemical shifts (ppm), ALR N-80 reduced

Residue	N	H^N	C^α	C^β	C'
M1	-	8.27	55.00	33.00	175.32
A2	125.88	8.28	52.06	19.50	176.87
A3	125.17	8.32	50.30	18.04	175.67
P4	135.65	-	63.79	31.82	177.77
G5	108.91	8.44	45.31	-	174.33
E6	120.60	8.12	56.74	30.36	176.83
R7	121.73	8.38	56.73	30.64	176.96
G8	109.42	8.40	45.31	-	174.03
R9	120.13	8.02	56.16	30.94	175.97
F10	120.76	8.27	57.63	39.45	175.58
H11	121.93	8.24	56.16	30.93	175.62
G12	110.05	8.02	45.60	-	174.55
G13	108.72	8.29	45.31	-	173.86
N14	118.45	8.25	53.22	38.86	175.02
L15	121.93	8.08	55.28	42.38	176.66
F16	119.86	7.99	57.62	39.74	174.81
F17	121.03	7.91	57.33	40.03	174.63
L18	125.21	8.06	52.64	42.09	174.85
P19	136.43	-	63.49	31.82	177.65
G20	109.77	8.51	45.32	-	174.94
G21	108.68	8.21	45.30	-	173.82
A22	123.49	8.12	52.64	19.49	177.69
R23	120.17	8.32	56.16	30.94	176.48
S24	117.04	8.28	58.51	63.79	174.64
E25	122.56	8.54	57.03	30.06	176.44
M26	120.21	8.23	55.57	32.70	176.19
M27	120.99	8.24	55.58	32.98	176.10
D28	121.30	8.28	54.69	41.20	176.48
D29	120.76	8.28	54.69	41.20	176.48
L30	121.65	8.10	55.57	42.08	177.62
A31	123.68	8.12	52.93	19.20	178.29
T32	112.70	7.98	62.61	69.65	174.68
D33	122.16	8.21	54.40	41.20	176.48
A34	124.31	8.21	53.52	18.91	178.46
R35	118.45	8.20	56.75	30.35	177.39
G36	108.68	8.22	45.31	-	174.51
R37	120.44	8.16	56.46	30.64	177.13
G38	109.85	8.47	45.31	-	174.08
A39	123.84	8.17	52.64	19.21	178.33
G40	107.97	8.40	45.31	-	174.16
R41	120.60	8.12	56.44	30.64	176.57
R42	122.63	8.47	56.16	30.93	176.06
D43	121.42	8.33	54.11	41.20	176.14

Table 3: Backbone chemical shifts (cont'd), ALR N-80 reduced

Residue	N	H^N	C^α	C^β	C'
A44	124.70	8.21	52.93	19.20	177.77
A45	122.59	8.21	52.64	18.94	177.90
A46	122.98	8.11	52.65	19.21	178.03
S47	114.74	8.15	58.22	63.79	174.42
A48	125.76	8.24	52.64	19.21	177.69
S49	114.89	8.21	58.21	63.79	174.29
T50	118.33	8.09	59.97	69.65	172.75
P51	139.13	-	63.20	32.11	176.61
A52	124.39	8.36	52.64	19.20	177.65
Q53	119.55	8.24	55.28	29.76	175.28
A54	127.25	8.33	50.59	18.03	175.54
P55	135.53	-	63.20	32.11	177.17
T56	114.07	8.24	61.45	69.94	174.72
S57	117.51	8.31	58.21	64.07	174.16
D58	122.32	8.34	54.39	41.20	175.92
S59	117.00	8.13	56.46	63.49	172.62
P60	138.11	-	63.19	32.11	176.96
V61	120.21	8.13	62.32	32.70	176.05
A62	127.98	8.32	52.64	19.21	177.82
E63	120.67	8.40	56.75	30.36	176.36
D64	121.46	8.29	54.39	41.20	176.83
A65	125.56	8.40	53.82	18.63	178.68
S66	113.78	8.31	59.68	63.49	175.02
R67	121.38	7.82	56.16	30.37	176.18
R68	121.03	7.97	56.16	30.64	176.05
R69	123.49	8.28	54.11	30.35	174.21
P70	136.86	-	63.19	32.11	-
C71	-	8.44	58.51	28.30	174.68
R72	123.73	8.42	56.16	30.94	175.80
A73	125.25	8.36	52.63	19.21	177.43
C74	118.80	8.27	58.51	28.00	174.72
V75	121.46	8.13	62.32	32.70	175.45
D76	123.02	8.17	54.11	41.20	175.58
F77	120.44	7.97	57.63	39.44	175.24
K78	122.91	8.05	56.16	33.27	176.05
T79	114.74	7.75	61.44	69.65	173.13
W80	127.95	7.53	58.50	30.06	180.61

Table 4: R_1 and R_2 relaxation rates (s^{-1}) for amide ^{15}N nuclei measured at 700 MHz for N-80 ALR_{SH} and N-80 ALR_{SS} protein fragment.

	R_1 (s^{-1})	R_2 (s^{-1})
N-80 ALR_{SH}	2.41 ± 0.34	6.91 ± 0.94
N-80 ALR_{SS}	2.29 ± 0.23	8.62 ± 0.76

Chapter 4

Conclusions

Among the methods that can provide information at atomic resolution, NMR plays a central role. By studying weak magnetic interactions between nuclear spins, NMR gives access to information on different aggregation states as well as in conditions that closely mimic the environment *in vivo*. NMR provides the information at atomic resolution on the structural and dynamic features, as well as on the kinetic and thermodynamic properties of biological macromolecules and their interactions. A complete description of the molecules and molecular processes, ranging from the folding process trough internal motions of the protein, to protein–protein interactions, can be obtained with NMR spectroscopy. All these aspects are essential for a deep understanding of the molecular basis of the function of biological systems contributing significantly to advances in life sciences.

The thesis reviews three projects in which I have been involved during my Ph.D. training. Results were obtained in our laboratory and I was involved in most of the aspects of the research including expression, purification and isotopic enrichment of the studied proteins followed by the acquisition, processing and analysis of NMR data and the structure determination process.

The first part of the work has focused on the general theme of the role of metal-mediated protein–protein interactions in controlling metal ion transfer and homeostasis. The role of surface residues was shown to be of key importance in determining protein–protein recognition. Once the proper relative orientation of the two partners is achieved, with the two CXXC motives facing each other, the shared metal coordination then determines metal ion transfer. The determination of the 3D structure of the complex between the soluble human copper chaperone HAH1 and the first cytosolic domain of the MNK ATPase has revealed that the interaction surface is designed for properly orienting the two partners, still not to favor a highly stable complex. Indeed, the protein–protein complex forms (is populated at equilibrium) only

in the presence of metal ion (metal-mediated interaction). The fine tuning between the energetics of surface interprotein interactions and metal-ligand interactions is therefore the key aspect determining the recognition and then transfer of the metal ion. In this way, through weak, but tightly controlled interactions, the copper chaperone HAH1 is able to ensure copper delivery to MNK1. These findings nicely show how the copper transfer pathway is controlled through minor modifications of the surfaces of the proteins in the metal transfer pathway.

The same structural motif (same 3D fold and metal binding motif) however is also used by nature to bind and control the homeostasis of metal ions, such as zinc. Moreover in many cases the proteins do show *in vitro* a higher affinity for the 'non-natural' metal ion. Therefore, after studying the subtle features controlling metal ion transfer, the following step consists in understanding how the metal transfer pathways are designed to prevent transfer to the wrong proteins.

The copper chaperone and ATPase system of *Synechocystis* PCC 6803 in which also a zinc ATPase is present (ZiaA) with very similar structural features to the copper ATPase (PacS), provided an ideal case to address this aspect. Indeed, the soluble domain of the zinc ATPase *in vitro* has even higher affinity for copper than for zinc posing the question of how the incorporation of the wrong metal ion (Cu) is prevented.

To address this question, the surface amino acids of $ZiaA_N$ were progressively mutated to mimic the interface of the soluble cytosolic domain of the copper transporting ATPase (PacS) in the same organism, in order to evaluate whether our hypothesis of the role of surface amino acids in triggering the metal-mediated interaction was valid. The progressive mutations of surface amino acids of $ZiaA_N$ to mimic $PacS_N$ were important in promoting complex formation with the copper chaperone (no complex formed for the Atx1-Cu(I)-$ZiaA_N$ WT and high complex stability for the Atx1-Cu(I)-$ZiaA_N$ QC4 mutant).

The fact that not only one mutation, but eight of them, distributed all around the interaction surface, was necessary to reintroduce a complex formation between Atx1 and $ZiaA_N$ shows the importance of the complementary surface to properly orient the proteins with the metal binding sites facing each other in order to form the metal-mediated complex and to promote metal transfer.

Therefore our results clearly show how the selectivity of metal transfer pathways is achieved. Simple complementarity of interprotein surfaces of proteins all sharing the same 3D fold and metal binding motifs, are promoting metal ion transfer to the final recipients and preventing metal transfer to 'wrong' proteins, (ie. to proteins devoted to controlling metal homeostasis of

other metal ions).

As the biotechnological methods and NMR instrumentation progress, the perspective of this work would be to investigate biochemical metal transfer pathways directly in environments mimicking closely *in vivo* situations (in living cells).

While working on this subject, which nicely shows the importance of well defined 3D structural modules for function, the idea that a stable 3D structure might not be a requirement for function was emerging in the literature shifting our attention to the characterization of highly flexible disordered proteins or protein fragments.

Indeed, the so-called intrinsically disordered proteins (IDPs), are recently attracting increasing attention for their functional role in many important biological processes and diseases. The presence of structural disorder raises novel questions with respect to the way in which these proteins carry out their functions and interact with their cellular partners. Intrinsically disordered proteins are predicted to represent a significant fraction of the human genome, while a detailed atomic resolution structural and dynamic characterization is available for only a subset of these proteins. The reason for the limited experimental characterization of this important class of proteins is intimately related to their characteristic dynamic behaviour. Only a subset of experimental techniques can be successfully used to study systems characterized by high local mobility. It is indeed currently extremely difficult to achieve the simultaneous characterization of structural and dynamical aspects and the development of meaningful molecular descriptions of these proteins remains a key challenge of structural biology. In this context, NMR spectroscopy has proven to be a unique tool to study intrinsically disordered proteins at atomic resolution.

In the present work, recently developed NMR methods based on ^{1}H and ^{13}C direct detection, were used to contribute to the characterization of an intrinsically disordered region of a complex protein ALR involved in the oxidative protein folding process in the intermembrane space (IMS) of mitochondria. The chemical shifts provided an initial insight into secondary structural propensities sampled in several regions of the polypeptide chain. An additional analysis of other NMR observables, in particular ^{1}H-^{15}N NOEs provided an indication of parts of the polypeptide chain characterized by different degrees of local flexibility suggesting the tendency of the chain to become more structured or to be less flexible with respect to the remaining part of the polypeptide chain.

The atomic resolution characterization of the N-terminal fragment of the ALR protein obtained through NMR revealed structural and dynamic features that allowed us to make hypothesis on the role of this fragment in

the redox reaction in which the ALR protein is involved. Interestingly, the considerations coming from NMR data provide strong experimental support for previous hypothesis based on independent biochemical evidence. Therefore this characterization provides important information to understand the funtion of the complex modular protein.

Looking to the future, the progressive instrumental and methodological developments in the NMR field and promising technologies in protein science are steadily bringing us closer to the broader more modern ambitious goal of understanding about how life works all the way from whole organisms, to their molecular components, to the structure, function, stability and interactions of the molecules at atomic resolution.

Bibliography

[1] J Qin, O Vinogradova, and AM Gronenborn. Protein–protein interactions probed by nuclear magnetic resonance spectroscopy. *Nuclear Magnetic Resonance of Biological Macromolecules*, 339:377–389, 2001.

[2] T Mittag and JD Forman-Kay. Atomic-level characterization of disordered protein ensembles. *Current Opinion in Structural Biology*, 17:3–14, 2007.

[3] L Banci, I Bertini, and S Ciofi-Baffoni. Copper trafficking in biology: an nmr approach. *HFSP Journal*, 3:165–175, 2009.

[4] V Dotsch. Investigation of proteins in living bacteria with in-cell nmr experiments. *Bioactive Conformation II*, 273:203–214, 2008.

[5] S Reckel, R Hansel, F Lohr, and V Dotsch. In-cell nmr spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 51:91–101, 2007.

[6] Z Serber, L Corsini, F Durst, and V Dotsch. In-cell nmr spectroscopy. *Nuclear Magnetic Resonance of Biological Macromolecules, Part C*, 394:17–41, 2005.

[7] Z Serber, P Selenko, R Hansel, S Reckel, F Lohr, JE Ferrell, G Wagner, and V Dotsch. Investigating macromolecules inside cultured and injected cells by in-cell nmr spectroscopy. *Nature Protocols*, 1(6):2701–2709, 2006.

[8] P Selenko and G Wagner. Looking into live cells with in-cell nmr spectroscopy. *Journal of Structural Biology*, 158:244–253, 2007.

[9] HJ Dyson and PE Wright. Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Unfolded Proteins*, 62:311–340, 2002.

[10] HJ Dyson and PE Wright. Unfolded proteins and protein folding studied by nmr. *Chemical Reviews*, 104:3607–3622, 2004.

[11] L Banci, I Bertini, F Cantini, IC Felli, L Gonnelli, N Hadjiliadis, R Pierattelli, A Rosato, and P Voulgaris. The atx1-ccc2 complex is a metal-mediated protein–protein interaction. *Nature Chemical Biology*, 2:367–368, 2006.

[12] L Banci, I Bertini, V Calderone, N Della-Malva, IC Felli, S Neri, A Pavelkova, and A Rosato. Copper(i)-mediated protein–protein interactions result from suboptimal interaction surfaces. *Biochemical Journal*, 422:37–42, 2009.

[13] T Mittag, J Marsh, S Orlicky, M Borg, X Tang, F Sicheri, HS Chan, LE Kay, M Tyers, and JD Forman-Kay. "fuzzy" complexes: How much disorder can a biologically relevant complex tolerate, and can it even be beneficial? *Biochemistry and Cell Biology-Biochemie et Biologie Cellulaire*, 88:403–403, 2010.

[14] VN Uversky, CJ Oldfield, U Midic, HB Xie, B Xue, S Vucetic, LM Iakoucheva, Z Obradovic, and AK Dunker. Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics*, 10, 2009.

[15] HRA Jonker, RW Wechselberger, R Boelens, R Kaptein, and GE Folkers. The intrinsically unstructured domain of pc4 modulates the activity of the structured core through inter- and intramolecular interactions. *Biochemistry*, 45:5067–5081, 2006.

[16] AMJJ Bonvin, R Boelens, and R Kaptein. Nmr analysis of protein interactions. *Current Opinion in Chemical Biology*, 9:501–508, 2005.

[17] Joachim Von eichborn, Stefan Gunther, and Robert Preissner. Structural features and evolution of protein–protein interactions. *Genome Inform*, 22:1–10, 2010.

[18] D Reichmann, O Rahat, M Cohen, H Neuvirth, and G Schreiber. The molecular architecture of protein–protein binding sites. *Current Opinion in Structural Biology*, 17:67–76, 2007.

[19] MJJ Blommers, A Strauss, M Geiser, P Ramage, H Sparrer, and W Jahnke. Nmr-based strategies to elucidate bioactive conformations of weakly binding ligands. *Bioactive Conformation*, 273:1–14, 2008.

[20] L Banci, I Bertini, F Cantini, and S Ciofi-Baffoni. Cellular copper distribution: a mechanistic systems biology approach. *Cellular and Molecular Life Sciences*, 67:2563–2589, 2010.

[21] C Andreini, I Bertini, G Cavallaro, GL Holliday, and JM Thornton. Metal ions in biological catalysis: from enzyme databases to general principles. *Journal of Biological Inorganic Chemistry*, 13:1205–1218, 2008.

[22] TV O'halloran and VC Culotta. Metallochaperones, an intracellular shuttle service for metal ions. *Journal of Biological Chemistry*, 275:25057–25060, 2000.

[23] BE Kim, T Nevitt, and DJ Thiele. Mechanisms for copper acquisition, distribution and regulation. *Nature Chemical Biology*, 4:176–185, 2008.

[24] L Banci, I Bertini, S Ciofi-Baffoni, T Hadjiloi, M Martinelli, and P Palumaa. Mitochondrial copper(i) transfer from cox17 to sco1 is coupled to electron transfer. *Proceedings of the National Academy of Sciences of the United States of America*, 105:6803–6808, 2008.

[25] S Tottey, DR Harvie, and NJ Robinson. Understanding how cells allocate metals using metal sensors and metallochaperones. *Accounts of Chemical Research*, 38:775–783, 2005.

[26] L Banci, I Bertini, KS McGreevy, and A Rosato. Molecular recognition in copper trafficking. *Natural Product Reports*, 27:695–710, 2010.

[27] L Banci, I Bertini, S Ciofi-Baffoni, T Kozyreva, K Zovo, and P Palumaa. Affinity gradients drive copper to cellular destinations. *Nature*, 465:645–U145, 2010.

[28] L Banci, I Bertini, S Ciofi-Baffoni, L Poggi, M Vanarotti, S Tottey, KJ Waldron, and NJ Robinson. Nmr structural analysis of the soluble domain of ziaa-atpase and the basis of selective interactions with copper metallochaperone atx1. *Journal of Biological Inorganic Chemistry*, 15:87–98, 2010.

[29] L Banci, I Bertini, F Cantini, CT Chasapis, N Hadjiliadis, and A Rosato. A nmr study of the interaction of a three-domain construct of atp7a with copper(i) and copper(i)-hah1 - the interplay of domains. *Journal of Biological Chemistry*, 280:38259–38263, 2005.

[30] L Banci, I Bertini, F Cantini, N Della-Malva, T Herrmann, A Rosato, and K Wurthrich. Solution structure and intermolecular interactions of the third metal-binding domain of atp7a, the menkes disease protein. *Journal of Biological Chemistry*, 281:29141–29147, 2006.

[31] L Banci, I Bertini, F Cantini, N Della-Malva, M Migliardi, and A Rosato. The different intermolecular interactions of the soluble copper-binding domains of the menkes protein, atp7a. *Journal of Biological Chemistry*, 282:23140–23146, 2007.

[32] DS Yuan, R Stearman, A Dancis, T Dunn, T Beeler, and RD Klausner. The menkes-wilson-disease gene homolog in yeast provides copper to a ceruloplasmin-like oxidase required for iron uptake. *Proceedings of the National Academy of Sciences of the United States of America*, 92:2632–2636, 1995.

[33] L Banci, I Bertini, CT Chasapis, A Rosato, and L Tenori. Interaction of the two soluble metal-binding domains of yeast ccc2 with copper(i)-atx1. *Biochemical and Biophysical Research Communications*, 364:645–649, 2007.

[34] F Arnesano, L Banci, I Bertini, F Cantini, S Ciofi-Baffoni, DL Huffman, and TV O'halloran. Characterization of the binding interface between the copper chaperone atx1 and the first cytosolic domain of ccc2 atpase. *Journal of Biological Chemistry*, 276:41365–41376, 2001.

[35] C Zwahlen, P Legault, SJF Vincent, J Greenblatt, R Konrat, and LE Kay. Methods for measurement of intermolecular noes by multinuclear nmr spectroscopy: Application to a bacteriophage lambda n-peptide/boxb rna complex. *Journal of the American Chemical Society*, 119:6711–6721, 1997.

[36] ERP Zuiderweg. Mapping protein–protein interactions in solution by nmr spectroscopy. *Biochemistry*, 41:1–7, 2002.

[37] I Anastassopoulou, L Banci, I Bertini, F Cantini, E Katsari, and A Rosato. Solution structure of the apo and copper(i)-loaded human metallochaperone hah1. *Biochemistry*, 43:13046–13053, 2004.

[38] L Banci, I Bertini, F Cantini, AC Rosenzweig, and LA Yatsunyk. Metal binding domains 3 and 4 of the wilson disease protein: solution structure and interaction with the copper(i) chaperone hah1. *Biochemistry*, 47:7423–7429, 2008.

[39] L Banci, I Bertini, S Ciofi-Baffoni, CT Chasapis, N Hadjiliadis, and A Rosato. An nmr study of the interaction between the human copper(i) chaperone and the second and fifth metal-binding domains of the menkes protein. *FEBS Journal*, 272:865–871, 2005.

[40] P Tompa. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters*, 579:3346–3354, 2005.

[41] P Tompa, C Szasz, and L Buday. Structural disorder throws new light on moonlighting. *Trends in Biochemical Sciences*, 30:484–489, 2005.

[42] Peter Tompa and Lajos Kalmar. Power law distribution defines structural disorder as a structural element directly linked with function. *J Mol Biol*, 403:346–50, 2010.

[43] Z Dosztanyi, B Meszaros, P Tompa, and I Simon. Instrinsically unstructured proteins in complexes. *Biophysical Journal*, pages 398A–398A, 2007.

[44] VN Daithankar, SA Schaefer, M Dong, BJ Bahnson, and C Thorpe. Structure of the human sulfhydryl oxidase augmenter of liver regeneration and characterization of a human mutation causing an autosomal recessive myopathy. *Biochemistry*, 49(31):6737–6745, 2010.

[45] VN Daithankar, SR Farrell, and C Thorpe. Augmenter of liver regeneration: Substrate specificity of a flavin-dependent oxidoreductase from the mitochondrial intermembrane space. *Biochemistry*, 48(22):4828–4837, 2009.

[46] L Banci, I Bertini, C Cefaro, S Ciofi-Baffoni, A Gallo, M Martinelli, DP Sideris, N Katrakili, and K Tokatlidis. Mia40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria. *Nature Structural & Molecular Biology*, 16:198–206, 2009.

[47] L Banci, I Bertini, R Del Conte, M D'Onofrio, and A Rosato. Solution structure and backbone dynamics of the cu(i) and apo forms of the second metal-binding domain of the menkes protein atp7a. *Biochemistry*, 43:3396–3403, 2004.

[48] L Banci, I Bertini, S Ciofi-Baffoni, XC Su, GPM Borrelly, and NJ Robinson. Solution structures of a cyanobacterial metallochaperone - insight into an atypical copper-binding motif. *Journal of Biological Chemistry*, 279:27502–27510, 2004.

[49] LE Kay, M Ikura, R Tschudin, and A Bax. 3-dimensional triple-resonance nmr spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance*, 89:496–514, 1990.

[50] M Sattler, J Schleucher, and C Griesinger. Heteronuclear multidimensional nmr experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 34:93–158, 1999.

[51] Bax and Ikura. An efficient 3d nmr technique for correlating the proton and nitrogen-15 backbone amide resonances with the alpha-carbon of the preceding residue in uniformly nitrogen-15 to carbon-13 enriched proteins. *Journal of Biomolecular NMR*, 1:99–104, 1991.

[52] A Bax and S Grzesiek. Methodological advances in protein nmr. *Accounts of Chemical Research*, 26:131–138, 1993.

[53] M Ikura, LE Kay, and A Bax. A novel-approach for sequential assignment of h^1, c^{13}, and n^{15} spectra of larger proteins - heteronuclear triple-resonance 3-dimensional nmr spectroscopy - application to calmodulin. *Biochemistry*, 29:4659–4667, 1990.

[54] S Grzesiek and A Bax. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance*, 99:201–207, 1992.

[55] S Grzesiek and A Bax. Improved 3d triple-resonance nmr techniques applied to 31-kda protein. *Journal of Magnetic Resonance*, 96:432–440, 1992.

[56] S Grzesiek and A Bax. Correlating backbone amide and side-chain resonances in larger proteins by multiple relayed triple-resonance nmr. *Journal of the American Chemical Society*, 114:6291–6293, 1992.

[57] M Tessari, M Mariani, R Boelens, and R Kaptein. (h)xyh-cosy and (h)xyh-e.cosy experiments for backbone and side-chain assignment and determination of $^3j_{HH}$ coupling-constants in ^{13}, ^{15}n-labeled proteins. *Journal of Magnetic Resonance Series B*, 108:89–93, 1995.

[58] LE Kay, GY Xu, AU Singer, DR Muhandiram, and JD Formankay. A gradient-enhanced hcch tocsy experiment for recording side-chain ^{1}h and ^{13}c correlations in h_2o samples of proteins. *Journal of Magnetic Resonance Series B*, 101:333–337, 1993.

[59] T Yamazaki, JD Formankay, and LE Kay. 2-dimensional nmr experiments for correlating $^{13}c^{\beta}$ and $^1h^{\delta\varepsilon}$ chemical-shifts of aromatic residues in ^{13}c-labeled proteins via scalar couplings. *Journal of the American Chemical Society*, 115:11054–11055, 1993.

[60] W Bermel, I Bertini, IC Felli, R Pierattelli, and PR Vasos. A selective experiment for the sequential protein backbone assignment from 3d heteronuclear spectra. *Journal of Magnetic Resonance*, 172:324–328, 2005.

[61] W Bermel, I Bertini, IC Felli, R Kummerle, and R Pierattelli. Novel ^{13}c direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins (vol 178, pg 56, 2006). *Journal of Magnetic Resonance*, 180:321–321, 2006.

[62] W Bermel, I Bertini, IC Felli, YM Lee, C Luchinat, and R Pierattelli. Protonless nmr experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *Journal of the American Chemical Society*, 128:3918–3919, 2006.

[63] W Bermel, I Bertini, V Csizmok, IC Felli, R Pierattelli, and P Tompa. H-start for exclusively heteronuclear nmr spectroscopy: The case of intrinsically disordered proteins. *Journal of Magnetic Resonance*, 198:275–281, 2009.

[64] DS Wishart and BD Sykes. The ^{13}c chemical-shift index - a simple method for the identification of protein secondary structure using ^{13}c chemical-shift data. *Journal of Biomolecular NMR*, 4:171–180, 1994.

[65] HY Zhang, S Neal, and DS Wishart. Refdb: A database of uniformly referenced protein chemical shifts. *Journal of Biomolecular NMR*, 25:173–195, 2003.

[66] Y Shen, PN Bryan, YN He, J Orban, D Baker, and A Bax. De novo structure generation using chemical shifts for proteins with high-sequence identity but different folds. *Protein Science*, 19:349–356, 2010.

[67] Y Shen, R Vernon, D Baker, and A Bax. De novo protein structure generation from incomplete chemical shift assignments. *Journal of Biomolecular NMR*, 43:63–78, 2009.

[68] Y Shen, O Lange, F Delaglio, P Rossi, JM Aramini, GH Liu, A Eletsky, YB Wu, KK Singarapu, A Lemak, A Ignatchenko, CH Arrowsmith, T Szyperski, GT Montelione, D Baker, and A Bax. Consistent blind protein structure generation from nmr chemical shift data. *Proceedings of the National Academy of Sciences of the United States of America*, 105:4685–4690, 2008.

[69] AL Breeze. Isotope-filtered nmr methods for the study of biomolecular structure and interactions. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 36:323–372, 2000.

[70] TL Hwang, PCM van Zijl, and S Mori. Accurate quantitation of water-amide proton exchange rates using the phase-modulated clean chemical exchange (cleanex-pm) approach with a fast-hsqc (fhsqc) detection scheme. *Journal of Biomolecular NMR*, 11:221–226, 1998.

[71] Peter Guntert. Calculating protein structures from nmr data. *Methods in Molecular Biology; Protein NMR techniques*, pages 157–194, 1997.

[72] P Guntert, C Mumenthaler, and K Wuthrich. Torsion angle dynamics for nmr structure calculation with the new program dyana. *Journal of Molecular Biology*, 273:283–298, 1997.

[73] CAEM Spronk, SB Nabuurs, E Krieger, G Vriend, and GW Vuister. Validation of protein structures derived by nmr spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 45:315–337, 2004.

[74] RA Laskowski, JAC Rullmann, MW Macarthur, R Kaptein, and JM Thornton. Aqua and procheck-nmr: Programs for checking the quality of protein structures solved by nmr. *Journal of Biomolecular NMR*, 8:477–486, 1996.

[75] L Krippahl, JJ Moura, and PN Palma. Modeling protein complexes with bigger. *Proteins-Structure Function and Genetics*, 52:19–23, 2003.

[76] PN Palma, L Krippahl, JE Wampler, and JJG Moura. Bigger: A new (soft) docking algorithm for predicting protein interactions. *Proteins-Structure Function and Genetics*, 39:372–384, 2000.

[77] C Dominguez, R Boelens, and AMJJ Bonvin. Haddock: A protein–protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, 125:1731–1737, 2003.

[78] S Schwarzinger, GJA Kroon, TR Foss, J Chung, PE Wright, and HJ Dyson. Sequence-dependent correction of random coil nmr chemical shifts. *Journal of the American Chemical Society*, 123:2970–2978, 2001.

[79] DS Wishart and BD Sykes. Chemical-shifts as a tool for structure determination. *Nuclear Magnetic Resonance*, 239:363–392, 1994.