University of Florence

International Doctorate in Structural Biology

Cycle XXII cycle (2007-2009)

Expression and characterization of proteins involved in rare diseases

Ph.D. thesis of

Chiara Massagni

Tutor	Coordinator
Prof. Ivano Bertini	Prof. Claudio Luchinat

S.S.D. CHIM/03

Table of contents

Chapter 1

Introduction

1.1 Rare Diseases

Rare diseases, including those of genetic origin, are disorders that only affect small numbers of individuals in the world. Low prevalence is taken as prevalence of less than 5 per 10.000 persons in the European Union. The United States definition is very similar to the European one. In the US a patience of rare disease is generally considered to have a prevalence of fewer than 200.000 affected individuals. Certain diseases with 200.000 or more affected individuals may qualify if subpopulations of these conditions are equal to the prevalence standard for rare diseases.

However, the low prevalence of this pathologies does not mean that there is a small number of patients affected by a rare disease. Only in Italy there are thousands of individuals affected by rare diseases and in Europe there are tens of thousands. The number of known and diagnosed rare diseases ranges from 7,000 to 8,000. Moreover some diseases are widespread in certain parts of the world but rare in others. The topic of rare diseases implies necessarily talking about orphan drugs. An orphan medicinal product, or "orphan drug", is a product that is potentially useful in treating a rare disease but does not have a market sufficient to cover the costs of its development. It therefore remains without a sponsor and thus an "orphan". These pharmaceutical products, due to the fragmentation of the single pathologies, often fail to arouse the economic interest of the pharmaceutical companies and the loss of a single research lab can be devastating for affected patients and their families.

Many rare diseases are caused by changes in genes and are called genetic diseases. A genetic disorder is a disease caused by a different form of a gene called a variation, or an alteration of a gene called a mutation. Many diseases have a genetic aspect. Some, including many cancers, are caused by a mutation in a gene or group of genes in a person's cells. These mutations can occur randomly or because of an environmental exposure such as cigarette smoke. Other genetic disorders are inherited. A mutated gene is passed down through a family and each generation of children can inherit the gene that causes the disease. Still other genetic disorders are due to problems with the number of packages of genes called chromosomes. In Down syndrome, for example, there is an extra copy of chromosome 21.

There are many different type of rare diseases and different part of the body can be affected: autoimmune disorders, blood disorders, brain and nervous system, cancers, chromosome disorders, congenital heart defects, connective tissue disorders, immune

disorders, kidney and urinary disorders, leukodystrophy, metabolic disorders, mitochondrial diseases, muscles, bones, and joints, skin disorders. In particular, in the metabolic disorders have to be mentioned Menkes disease and Wilson's disease which are inherited disorders of copper metabolism resulting from the absence or dysfunction of homologous copper-transporting ATPases. Also Cytochrome c Oxidase Deficiency is a very rare inherited metabolic disorder which is characterized by deficiency of the enzyme Cytochrome c Oxidase (CcO). Same cases, as Leigh's disease, imply inherited metabolic disorders that damage the central nervous system (brain, spinal cord, and optic nerves). Another important disease involving brain and nervous system is the Amyotrophic Lateral Sclerosis (ALS) which is a disease that attacks motor neurons that control muscles.

During my PhD I focused my attention on proteins involved in the above mentioned rare diseases in order to better understand causes and development of the disorders.

1.2 Wilson's disease

Wilson disease and Menkes syndrome are human diseases of copper metabolism caused by mutations in the genes for two closely related proteins, the Wilson (1) and Menkes (2) copper (I)- transporting ATPases. These proteins couple ATP hydrolysis to copper translocation across membranes. Menkes syndrome is caused by decreased copper uptake across the small intestine as well as impaired copper distribution to a variety of tissues, including the brain. By contrast, Wilson disease is characterized primarily by copper overload in the liver because of reduced biliary excretion (3). (**Fig. 1**)

Wilson's disease (WD) was first described in 1912 as progressive lenticular degeneration by Samuel Alexander Kinnier Wilson. He described WD as a rare, familial, progressive, and invariably fatal disease, characterized mainly by neurological deficits and liver cirrhosis. WD is an autosomal recessive disorder of copper transport, leading to systemic copper accumulation and multi-organ damage, particularly of the brain and the liver. Permanent organ damage can be effectively prevented if early de-coppering treatment is initiated. By the latter criterion, the National Institutes of Health listed WD as a rare disease. Most literature describes the worldwide prevalence of WD as 1 in 30.000, with a carrier rate of 1 in 90 and incidence ranging from 15 to 30 per million (5).

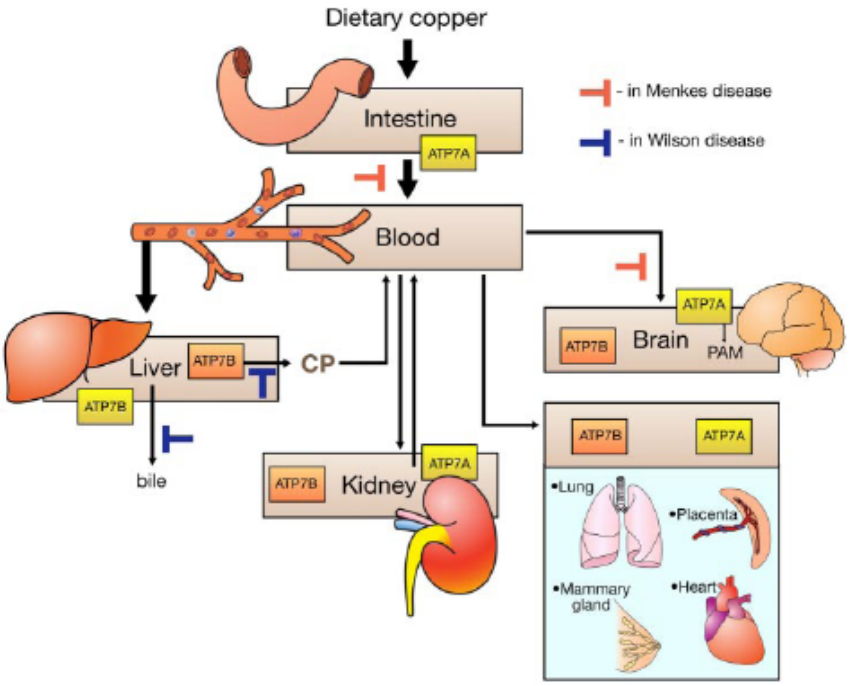


Fig. 1 : *Simplified scheme of copper distribution in the body. CP indicate Ceruloplasmin.(4)*

The conventional diagnosis of WD depends on typical neurological symptoms, the presence of Kayser-Fleischer rings observed by slit-lamp examination and a decreased serum ceruloplasmin concentration (6). The disease onset usually occurs in childhood and young adulthood. However, case reports in the literature include patients from three years of age to 72 years old (7). Clinical manifestations of WD may show considerable variation. Typical presentations involve mainly hepatic, neurological, and psychiatric manifestations.

In 1993 the gene responsible for WD was identified (1, 8, 9), and the gene product was shown to be the copper(I) P-type adenosine triphosphatase ATP7B (10). Indeed also ATP7A (Menkes protein) belong to the group of P-type ATPases and Wilson protein has 67% amino acid identity to the Menkes protein (1) and the same structural characteristics.

The culprit ATP7B gene consists of 21 exons that span a genomic region of about 80 kb and encode a protein of 1465 amino acids. Mutant ATP7B results in defective copper incorporation into ceruloplasmin and a reduction in biliary excretion of copper. Copper(I)-transporting P-Type ATPase ATP7B is expressed primarily in the liver and kidney. The protein plays a dual functional role in the hepatocyte (11). First it mediates the biosynthesis of holoceruloplasmin by delivering copper to apoceruloplasmin within

the trans-Golgi network under low-copper conditions. The other role is to facilitate biliary excretion of excess copper when cellular copper concentrations are elevated. Therefore, an ATP7B mutant fails in the process of holoceruloplasmin synthesis and/or the biliary excretion of copper, which in turn results in a significant reduction of serum ceruloplasmin concentration and an accumulation of copper in the liver, respectively. For example, a WD mutant protein, R778L, has been shown to be extensively mislocalized, presumably to the endoplasmic reticulum (12). Factors other than ATP7B mutations such as COMMD199 may modulate the clinical manifestations, but more convincing data are required. Numerous mutational studies of WD have been conducted, and they have revealed that the mutation spectrum of the ATP7B gene is population specific. The most frequent ATP7B mutation in Caucasian patients is H1069Q, which is present in up to 72% of all alleles (13-17). No such mutation has so far been detected in East Asian WD patients. The most common East Asian-specific mutant is R778L present up to 44% of WD patients (18-22). To date, more than 370 mutations have been reported worldwide, and most are rare and infrequent (www.medicalgenetics.med.ualberta.ca/wilson/index.php)(23).

1.2.1 An overview on Wilson Protein

The Wilson disease protein (WND), as Menkes similarly disease protein, are copper(I)-transporting ATPase involved in copper delivery to the secretory pathway of Golgi. It is usually localized in the trans-Golgi network (TGN) of hepatocytes where it delivers copper to key metalloenzymes, including ceruloplasmin, a multicopper oxidase involved in high affinity iron uptake (24, 25). This process is dependent on interaction with Atox1(or HAH1) a copper chaperone that delivers Cu(I) ions to WND (26, 27). At elevated copper concentrations, WND redistributes to cytoplasmic vesicles and mediates copper efflux from the cell (24, 28). Similarly, the Menkes disease protein (MNK) relocalizes from the trans-Golgi network to the plasma membrane in response to changing copper concentration (29, 30)(**Fig.2**). This is the mechanism by which the copper ingested with the diet is transported by Menkes through the basolateral membrane of enterocytes into the portal blood and hence to other tissues. (31)

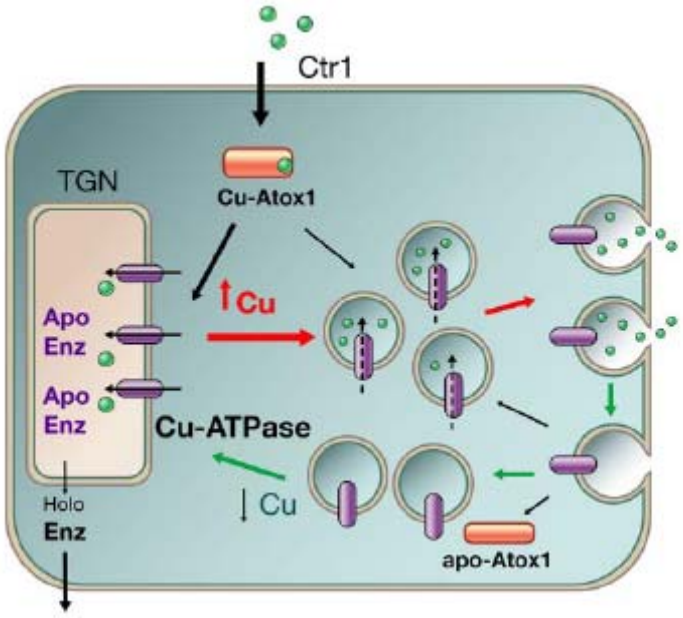


Fig.2: *Localization and function of Cu-ATPases in low (green arrows) and high (red arrows) copper. In basal (low) copper, the Cu- ATPases are located in the trans-Golgi network (TGN) where they deliver copper to secreted copper-dependent enzymes. When copper is elevated, the Cu-ATPases relocate to vesicles. The Cu-ATPase in the vesicles may or may not be active, depending on the intravesicular copper concentration. The copper-containing vesicles are then delivered to the plasma membrane where the accumulated copper is released and the Cu-ATPase is endocytosed. Apo-Atox1 removes regulatory copper from the NH_2-terminal domain and thus may facilitate return of the ATPase to the TGN. (4)*

At the biochemical level, the function of Cu-ATPases is to translocate copper across the membrane from the cytosol into the lumen of appropriate intracellular compartment (either TGN or vesicles). The vectorial copper translocation across the membranes is driven by the hydrolysis of ATP and two copper ions are transported per one hydrolyzed ATP. Both ATP7A and ATP7B belong to the P1B-subfamily of the P-type ATPases which are a superfamily of membrane proteins responsible for the active transport of a variety of cations across cell membranes. This subgroup contains over 100 members (see http://www.patbase.kvl.dk/IB.html) and is characterized by a number of transmembrane helices involved in the formation of an intramembranous channel; by an hydrophilic region protruding into the cytosol, which contains the Actuator domain (A-domain) and the ATP-binding domain, which in turn can be further separated into two smaller domains named Phosphorylation domain (P-domain) and Nucleotide-binding domain (N-domain); and by the presence of six soluble metal-binding domains

(MBDs) at the N- and/or C-termini of the molecule, as well as characteristic conserved sequence motifs GMXCXXC through which it can bind one equivalent of copper (32). The number of the N-terminal metal binding domains (NMBDs) varies from one in bacteria and archea to six in higher eukaryote ATPases (33). The N-terminal copper-binding cytosolic tails of ATP7A and ATP7B are both ca. 630 amino acids long.

Moreover all the protein belonging to the P1B-subfamily show some peculiar features: (i) heavy-metal binding sites in the polar amino-terminal region; (ii) a conserved intramembranous Cys-Pro-Cys, Cys-Pro-His or Cys-Pro-Ser motif (hereafter CPx motif); (iii) a conserved histidine-proline dipeptide (HP locus) 34 to 43 amino acids carboxy-terminal to the CPx motif; and (iv) a unique number and topology of transmembrane helices.

Similarly to all members of the P-type ATPase family, so-called because they catalyze self phosphorylation of a key conserved aspartate residue within the protein, the human Cu-ATPases hydrolyze ATP with the formation of a transient acyl-phosphate intermediate (**Fig. 3**) Phosphorylation takes place at the invariant Asp residue in the signature motif DKTG (**Fig. 3a**). The reaction requires the transfer of copper from the cytosol to the intra-membrane portion of the transporter; while the release of copper to the opposite side of the membrane is accompanied by dephosphorylation (34, 35). Cu-ATPases can also be phosphorylated by inorganic phosphate at the same aspartate residue within the DKTG motif. This reaction is reverse to the dephosphorylation step and is inhibited by copper binding to the intra-membrane site(s) from the luminal milieu (36). Comparing the ability of Cu-ATPase to be phosphorylated from ATP or Pi and the effect of copper on this reaction may serve as a tool for determining the conformational state of the enzyme. During the catalytic cycle the Cu-ATPases are likely to undergo significant conformational changes(35). By analogy with other P-type ATPases, the binding of copper from the cytosolic side is thought to take place when the protein is present in the so-called E1 state, which is characterized by high affinity of the intra-membrane sites for the transported ion. Both transmembrane binding sites bind copper(I) with high affinity (Ka^1=1.12 ± 0.25 fM^{-1} and Ka^2=1.30 ± 0.22 fM^{-1}) (37, 38). Copper binding to the intra-membrane sites is associated with the transfer of c-phosphate of ATP to the Cu-ATPase and transient stabilization of the phosphorylated state of the protein, E1P (**Fig. 3b**).

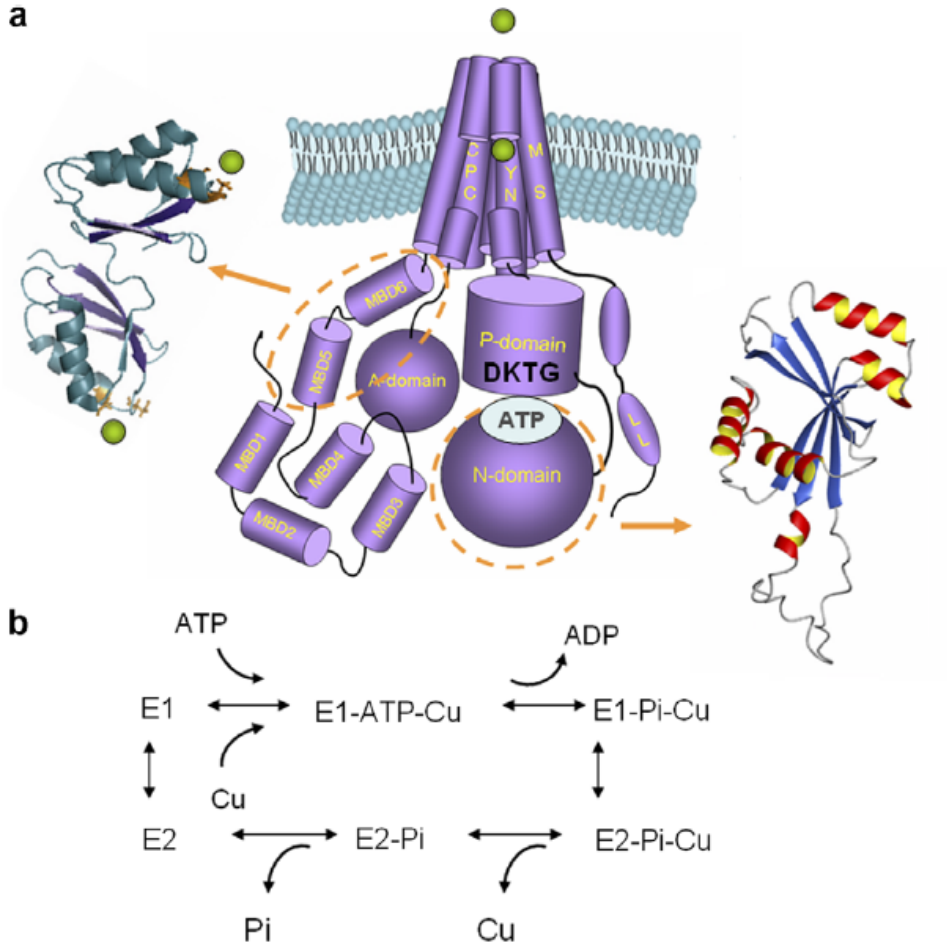


Fig. 3: *Transmembrane organization and catalytic cycle of human Cu-ATPases. (a) Cartoon illustrating the major functional domains of Cu-ATPases. The N-terminal domain contains six copper-binding MBDs (MBD1–6). The transmembrane portion has eight transmembrane segmants (TMS); the position of residues predicted to be involved in copper coordination (CPC, YN, and MxxS) is indicated. The A-domain may link changes in the N-terminal domain with those in the ATPbinding domain and in the transmembrane portion. The ATP-binding domain consists of the P-domain and the N-domain. The domains of ATP7B for which structure has been experimentally determined are indicated by dashed circles and corresponding structures are shown. Two Leu residues in the C-terminal tails required for endocytosis and/or return to TGN are indicated by ‘‘LL’’. (b) The simplified catalytic cycle of human Cu-ATPases. Two major conformational states associated with high affinity for ATP and Cu (E1) and lower affinity for these ligands (E2) as well as phosphorylated intermediates (E1–Pi–Cu and E2–Pi–Cu) are shown (4).*

In this state, access to intra-membrane sites from the cytosol is blocked and copper is sequestered in the ‘‘occluded’’ form. Subsequently, the Cu-ATPase undergoes conformational change to the E2P state, and the affinity for copper is decreased. It is thought that in this state copper is released from the transporter and is taken up by an

acceptor protein. Intermediate proteins are not required for copper transfer from the transporter to acceptors (39), however interaction between ATP7A and SOD3, a putative target protein of ATP7A activity, has been reported (40). This latter observation suggests that the donor–acceptor interaction, although not obligatory, may facilitate metal transfer to the copper-requiring enzymes in the secretory pathway. After copper is released, the Cu-ATPase dephosphorylates (E2 state) and then undergoes conformational transition into a high-affinity state, E1, for initiation of the next transport cycle.

Both WND and MNK ATPases consist of eight transmembrane helices, of which the sixth contains an invariant CPC motif proposed to bind metal ions. Phosphatase- and ATP-binding domains and a phosphorylation site are conserved as well (41). P1B-type transport ATPases are distinguished from other P-type ATPases by the presence of multiple soluble N-terminal domains containing a conserved CXXC metal-binding motif (42).

Both WND and MNK contain six soluble domains, whereas the yeast homolog, Ccc2 (43), contains only two repeats and the Drosophila melanogaster homolog contains four. These domains are present not only in the copper ATPases but also occur in Zn(II), Cd(II), and Pb(II) transporters (41, 42). The Atox1 (HAH1) metallochaperone and its homologs, both eukaryotic and prokaryotic, contain a single CXXC motif (44). Crystal and solution structures of Atox1(45), yeast Atx1(46, 47), bacterial CopZ(48, 49), and single domains of MNK(50) and Ccc2(51) reveal a conserved βαββαβ-fold with the cysteines from the CXXC motif coordinating metal ions on a surface-exposed loop. Notably, the CXXC motifs from two Atox1 molecules coordinate a single Cu(I) ion in the x-ray structure (45). Both the WND and MNK N termini bind 5–6 copper ions (52), and x-ray absorption spectroscopic studies indicate that the copper is present as Cu(I), ligated by the sulfurs from two cysteines (53-55). The roles of the six WND and MNK metal-binding domains in copper metabolism are not well established. Numerous genetic and biochemical studies suggest that the domains are not functionally equivalent. For WND, introduction of wild type or mutated *ATP7B* cDNA into a yeast strain lacking the *CCC2* gene , the yeast homologue of ATP7B, indicate that the sixth domain alone is sufficient for copper loading of Fet3, the yeast homolog of ceruloplasmin (56, 57). Furthermore, the second or third metal-binding domain cannot substitute for the sixth domain, supporting distinct functions for the individual domains (57). Then, a more specific function has been assigned to domains 5 and 6. These

domains are responsible for the cooperative effect of copper on WND catalytic phosphorylation activity and human missense mutations found within MBD1, 5, and 6 of WND are known to give rise to Wilson disease, and these mutants show impaired interaction with HAH1 (58). In contrast to WND, yeast complementation studies on MNK indicate that the first four N-terminal domains are important for copper transport (59). It may be that the domains function is different in the two ATPases. Another potential function for the metalbinding domains is mediating copper-responsive cellular relocalization.

The six soluble domains of both WND and MNK receive copper from the cytoplasmatic metallochaperone HAH1. The KCu binding affinity of the six MBDs of both theMNK (60-62) andWLN (63, 64) proteins range between 1- and 5-fold the binding affinity of the copper(I) metallochaperone HAH1. The exact role and interplay of the six domains is still unclear. In the case of MNK protein, the six-soluble domains interact differently with HAH1. The interactions have been investigated by a variety of techniques ranging from yeast-two hybrid assays to NMR(62, 65-68). These data indicate that the MNK first and fourth domains form a metal-mediated adduct with HAH1, whereas the six domain is simultaneously loaded with copper without formation of adduct. In case of WND is not still clear the behaviour of the six MBD. A pathway was proposed for copper delivery to the N-terminal MBDs of WND from an NMR study of MBD5-6: Cu(I)HAH1 forms a complex with MBD4 and MBD2 that allows transfer of copper between these partners, while any complex formation and no copper transfer occur between Cu(I)HAH1and MBD5–6 suggesting that WLN5–6 must acquire copper from one of the other metal-binding (69). Another study suppose that MBD3-4 have similar affinities for incorporation of Cu(I) from HAH1, even if Cu(I) HAH1 forms detectable amounts of a macromolecular complex with domain 4, whereas domain 3 removes Cu(I) from the metallochaperone without detectable complex formation (70). Domains 1–4 of WND have also been proposed to function in copper-responsive localization (57, 58). Anyway all that study were done studing the interaction of HAH1 with single domains. For this reason, during my PhD I focus my work to obtain the whole 633 residues-long N-terminal tail of WND in order to study simultaneously the behaviour of all six MBD in presence of Cu(I) and Cu(I)-HAH1 complex.

1.3 Cytochrome c Oxidase Deficiency

Mitochondria, ubiquitous organelles of eukaryotic cells, are the main generators of cellular ATP and carry several important metabolic reactions. Mitochondria also play an active role in survival, death signalling and cellular calcium homeostasis. Defects in mitochondrial function are associated with numerous neurodegenerative diseases (Parkinson's, Alzheimer's and Huntington's disease) and, in particular, with mitochondrial diseases. Mitochondria have their own genome which encodes for about 1% of the total mitochondrial proteins, therefore, mitochondrial biogenesis occurs as a result of a highly coordinated action between the nuclear and the mitochondrial genome that requires the import of several proteins into the mitochondria that are encoded by the nuclear genome (71). For this reason, mitochondrial diseases are caused by mutations in mitochondrial genes, in nuclear genes encoding for subunits of the respiratory chain complexes or in genes coding for proteins involved in the nuclear–mitochondrial communication (72).

The mitochondrial respiratory chain is composed of five multisubunit enzymes whose components are encoded in both the nuclear and mitochondrial genomes; of the 82 structural subunits, only 13 are encoded in mtDNA. In addition to the structural components of the respiratory chain, a large number of proteins are involved in the assembly and maintenance of the enzyme complexes. Fully half of the known mitochondrial proteins in the yeast *S.cerevisiae* are involved in some aspect of mitochondrial biogenesis (73), and more than 30 different genetic complementation groups have been identified for the assembly of Cytochrome c Oxidase (CcO) alone (74). More than 50 different pathogenic mutations have been identified in mitochondrial DNA (mtDNA)(75), and much attention has now focused on identification of the nuclear gene defects associated with respiratory chain deficiency.

CcO is the terminal electron acceptor of the mitochondrial respiratory chain. It is a membrane-bound redox-driven proton pump (76) located in the inner mitochondrial membrane, where it catalyzes reduction of molecular oxygen to water and pumps protons across the membrane (77). In this process a membrane electrochemical proton gradient is generated; the energy stored by the proton gradient is subsequently utilized for ATP synthesis. The reduction of O_2 to H_2O in the catalytic centre of CcO generates energy that is necessary for proton translocation from the mitochondrial matrix (a region

of low proton concentration and negative electrical potential) to the intermembrane space (which is in contact with the cytosol, a region of high proton concentration and positive electrical potential). Mammalian CcO consists of 13 polypeptide subunits, 3 of which (COX1-COX3) are encoded by the mitochondrial genome with the remaining 10 subunits encoded by the nuclear genome (78, 79). In addition to the subunits, over 30 distinct proteins are important for the assembly of CcO (74). A number of these accessory proteins are important in the processing and translation of COX1-COX3 mRNA transcripts, in chaperoning the assembly process, and in the synthesis or delivery of cofactors. The cofactors in CcO include two copper sites (CuA and CuB), two *heme-a* moieties, a magnesium and zinc a ion (80)(**Fig.4** and **Fig.5**).

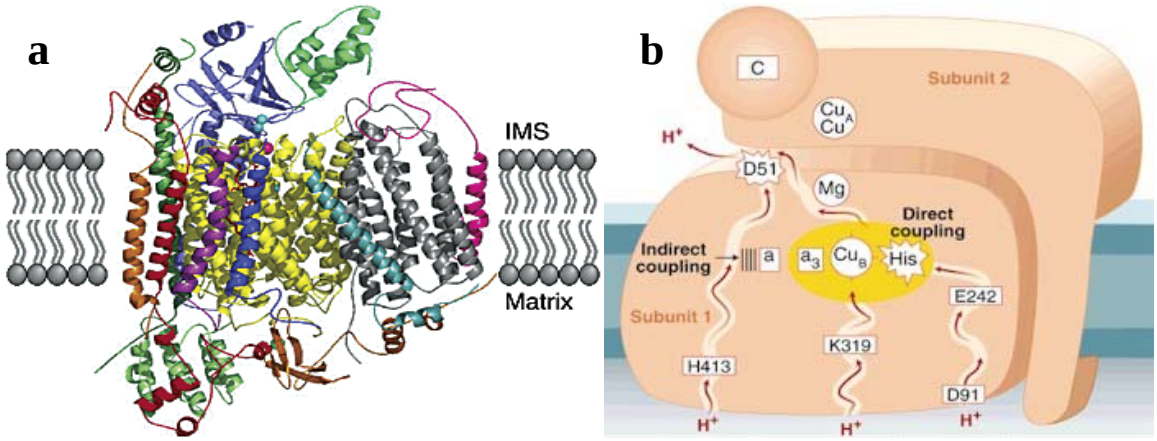


Fig. 4: *The structure of bovine CcO is shown with the 13 subunits colored differently (*a*). The cofactors are largely obscured by the polypeptide chains. Panel* b *shows the cofactors (81).*

Little is known about the delivery of Zn and Mg ions to the mitochondrion or whether specific accessory factors are necessary for the insertion of these ions into CcO. More information are available concerning copper delivery and insertion into CcO. A total of three copper ions have to be inserted into two subunits COX1 and COX2 which contain the CuB and the CuA centers, respectively.

The former has one copper ion buried 13Å below the membrane surface (**Fig.5**), while the second contains a binuclear copper site located into the inner membrane space of the mitochondria in eukaryotes.

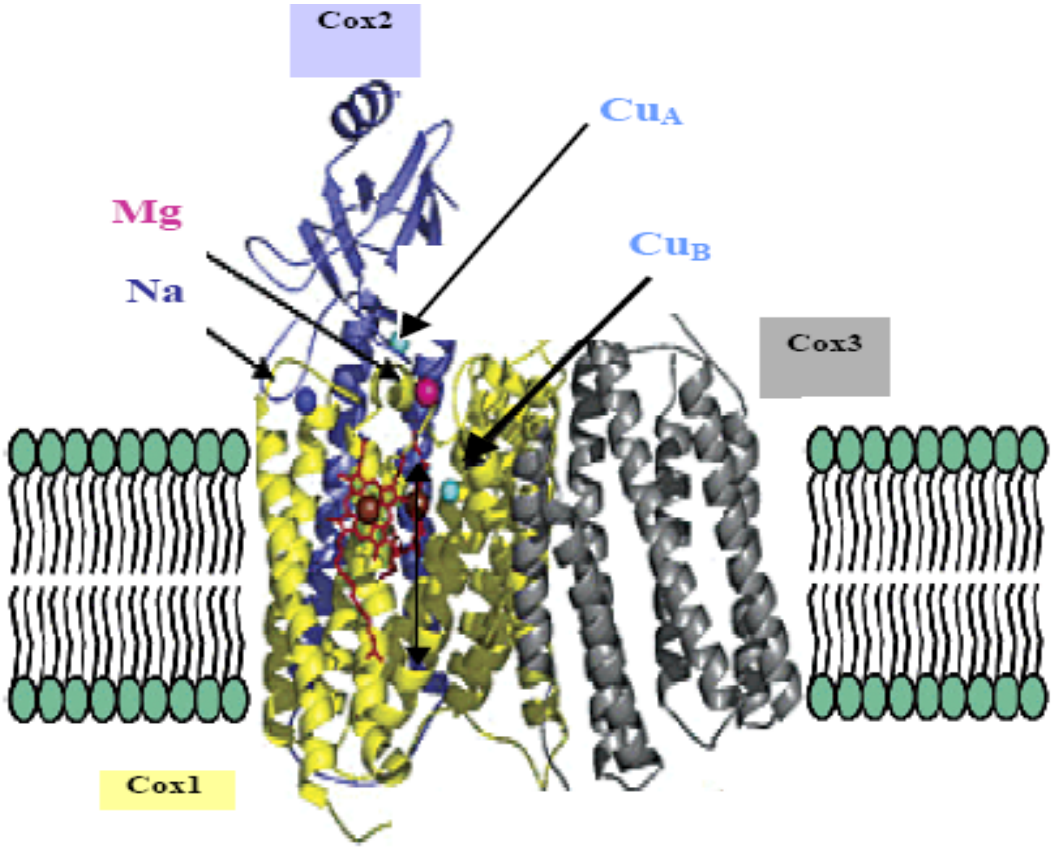


Fig.5: *Structural representation of the mitochondrial-encoded subunits that make up the catalytic core of the enzyme, COX1 (yellow), COX2 (blue) and COX3 (grey)within their cofactors (81).*

Because both COX1 and COX2 subunits of CcO are synthesized inside the mitochondria, the three Cu atoms must be imported from the cytoplasm. COX1 and COX2 are metallated by a pathway that consists of COX17 and the cochaperones, SCO1, SCO2 and COX11. COX17 delivers copper to COX11 for subsequent insertion into the CuB site in COX1, and to SCO1 for insertion into the CuA site of COX2 (**Fig.6**).

CcO deficiencies in humans comprehend a wide variety of disorders and the pathological features and genotype have been extensively reviewed (82).

Genotypically, isolated CcO deficiencies could arise from mutations in any of the 3 subunits encoded in the mtDNA, from any of the 10 structural subunits encoded in the nucleus or from mutations in any of the assembly factors. Phenotypically, all these possibilities create a wide variation on the clinical symptoms that could affect single or multiple organs.

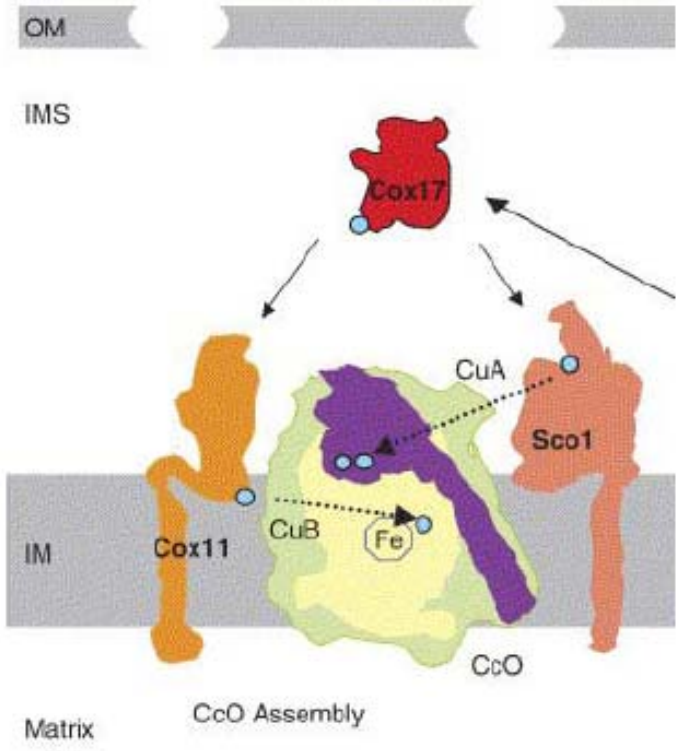


Fig. 6: *Scheme of Cu metalation of COX1 and COX2 mediated by COX11 and SCO1, respectively. COX17 is postulated to transfer Cu(I) to SCO1 prior to the transfer to the CuA site in COX2 and to transfer Cu(I) to COX11 prior to transfer to the buried CuB site in COX1(83).*

Only few CcO deficiencies involve mutations in the mitochondrial encoded subunits (84, 85). The clinical presentations of such mutations encompass myopathy, sideroblastic anemia, amyotrophic lateral sclerosis (ALS)-like syndrome, encephalomyopathy and MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes)(72, 86). Mutations in the mtDNA could affect a single tissue (myopathies) or multiple systems (MELAS and encephalomyopaties) and the severity of the CcO deficiency could vary in different tissues. In terms of defects derived from nuclear mutations, the differences in tissue specificity could be explained by differences in the levels of certain subunits or assembly factors, presence of specific isozymes, specific regulatory mechanisms (phosphorylation and allosteric regulation) or by differential availability of cofactors. Mutations in the nuclear-encoded subunits were not found for many years in patients with CcO deficiency leading to the idea that these mutations were incompatible with live. Recently, Massa and collegues (87) described

two patients (siblings from cousin parents in third degree) suffering from severe infantile encephalopathy with CcO deficiency. (**Fig. 7**)

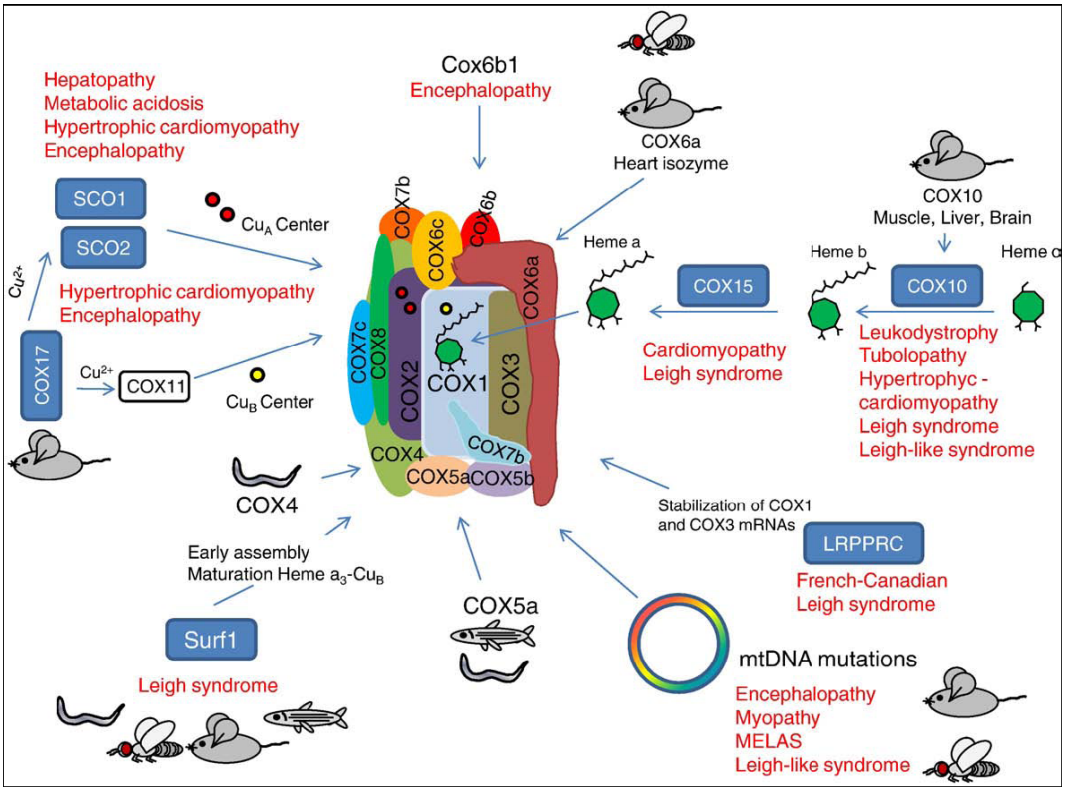


Fig. 7: *Gene defects associated with Cytochrome c Oxidase deficiencies in humans and existing animal models. Clinical pathologies associated with mutations in respective CcO subunits or assembly factors associated with CcO deficiencies are labeled in red. The different animal models created for a corresponding protein or mitochondrial DNA are shown in the figure: mouse, zebrafish, Drosophila, and C. elegans.(88)*

The majority of the genetic defects related to CcO deficiencies have been found in nuclear genes coding for auxiliary proteins. The first identified CcO deficiency with a nuclear origin was Leigh syndrome (LS) is an autosomal recessive inherited neurodegenerative disorder with a characteristic neuropathology consisting of focal, bilateral lesions in one or more areas of the central nervous system, including the brainstem, thalamus, basal ganglia, cerebellum, and spinal cord. The lesions are areas of demyelination, gliosis, necrosis, spongiosis, or capillary proliferation. Clinical symptoms depend on which areas of the central nervous system are involved (89).

LS was originally described as a sub-acute necrotizing encephalopathy which affects infants and young children. The estimated prevalence of LS is 2.05 children per 10 000,

while the estimated prevalence for all mitochondrial respiratory chain disorders is 7.5–8.7 per 10000 (90). Therefore, LS represents one of the most frequent mitochondrial disorders in childhood. LS diagnostic criteria include a progressive neurological disease with motor and intellectual developmental delay, hypotonia, movement disorders, brainstem and/or basal ganglia degeneration and often increased lactate levels in blood and/or cerebrospinal fluid (CSF)(91).

LS patients die between 6 months to 12 years of age. The typical pathology of Leigh syndrome includes the presence of bilateral symmetric necrotic lesions in subcortical areas of the brain that develop into ophthalmoparesis, nystagmus, ataxia, dystonia and optic atrophy (92). The residual levels of CcO activity observed in LS patients range from 10 to 25% of control values with no apparent correlations in terms of severity and tissue specificity. The majority of CcO-deficient LS is associated with mutations in the *SURF1* gene (93), located on chromosome 9q34. The gene encodes the SURF1 protein that is anchored to the inner mitochondrial membrane by two transmembrane domains and is highly conserved in a number of species (94). Although the exact function of SURF1 still remains unclear, it is known that it affects CcO assembly. Patients with SURF1 mutations show accumulation of early CcO assembly intermediates and marked reduction of the fully assembled complex (95).

However, not all the mutations in this gene produced Leigh syndrome, some SURF1 patients did not develop the characteristic neurodegenerative lesions. In some cases, the pathology of Leigh syndrome has been associated with mutations in the *COX10* gene, which encodes *heme-a*:farnesyltransferase, or with mutations in the mtDNA (96). Moreover, LS can result from defects in other genes altering energetic metabolism such as pyruvate dehydrogenase, NADH:ubiquinone oxidoreductase (complex I) or ATP synthase (complex V) (97). A less severe form of LS is the French–Canadian Leigh syndrome (FCLS, exclusive of the Charlevoix and Saguenay-Lac-St. Jean region of Quebec). Clinical presentations of this form include mild regression of psychomotor skills and a fatal lactic acidosis with death occurring between 3 and 10 years of age. In this disorder, the tissues mostly affected are brain (basal ganglia) and liver (98). LRPPRC (leucine-rich pentatricopeptide repeat cassette) was identified as the gene responsible for this disease (99). LRPPRC is a homologue of the yeast P309, another CcO assembly factor, which is required for the expression of COX1 by assisting in the stability and translation of COX1 mRNA.

Human SCO1 and SCO2 code for essential metallochaperones with ill-defined functions in the biogenesis of the CuA site of Cytochrome c Oxidase subunit II (COX2). Mutations in both SCO1 and SCO2 cause severe tissue-specific CcO assembly impairment accompanied by marked copper deficiency (100-102). However, both genes have been shown to be ubiquitously expressed, displaying a similar expression pattern across human tissues. In particular, another CcO deficiency characterized by hypertrophic cardiomyopathy with encephalopathy also with an early onset is due to mutations in the copper related assembly factor SCO2. Although the CcO deficiency was more severe in cardiac and skeletal muscle, autopsy analysis of the brain revealed spongiform neurodegeneration in midbrain, pons and medulla accompanied by gliosis and microglia proliferation in the thalamus, but lacked the typical LS lesions(103).

Mutations in the other copper chaperone, SCO1 have been associated with completely different phenotypes. Valnot and collaborators (102) reported two neonate cases, the first one with severe metabolic acidosis and with hepatic failure consisting in microvesicular steatosis (lipid accumulation) and enlarged liver. The patient died at 2 months of age. A severe CcO deficiency in liver and muscle (less than 1% of control values) was determined in postmortem tissues. The second case was a sibling born in early pregnancy that also presented with metabolic acidosis and neuronal involvement and died at 5 days of age. In both cases, the defect mapped to the SCO1 gene (102). A third case of SCO1 defect was just recently reported (104). This last patient presented with an early onset of hypertrophic cardiomyopathy accompanied by encephalopathy and hepatomegalia and died at 6 months of age. Ultrastructural analysis of the cardiac tissue showed accumulation of abnormal mitochondria of varying sizes, containing hemidense deposits. The liver had microvesicular steatosis and the muscle lacked ragged-red fibers (105). Defects in the *heme-a* biosynthetic pathway lead to different clinical presentation also with fatal outcome early on life (months to 2 years of age). The first COX10 mutation was found in a pediatric patient suffering fromataxia, muscle weakness, hypotonia and pyramidal syndrome, leukodystrophy and proximal tubolopathy accompanied by elevated lactate in both cerebrospinal fluid and blood (102). Other patients with mutations in COX10 presented anemia, sensorineural deafness and hypertrophic cardiomyopathy (106) and Leigh-like syndrome (107). The levels of CcO activity in COX10 patients ranged from 5 to 25% of control, depending on the tissues analyzed. Also few patients with defects of COX15 protein, which is involved in the synthesis of *heme-a*, the *heme* prosthetic group for CcO, have been

documented. The patients present lactic acidosis, hypotonia and cardiomyopathy (106), nystagmus, motor regression, retinopathy and microcephaly (108).

A report of two patients with an early onset of mitochondrial encephalopathy presenting asymmetrical brain atrophy, convulsions, hemiplegia and psychomotor developmental delay associated with low levels of Cytochrome c Oxidase activity in skeletal muscle showed that the defect mapped to the FASTKD2 (fas activated serine–threonine kinase domain 2) gene (109). Although the function of FASTKD2 is not known, this protein appears to be involved in apoptosis.

The assembly factors described above along with mtDNA genes constitute candidate genes screened to diagnose the causes of CcO deficiency in patients. Recently, another yeast assembly gene with a human homolog that has been proposed as a screening candidate for CcO deficiencies is COX18, which is required for the insertion of COX2 into the inner mitochondrial membrane. Unfortunately, mutational screening of COX18 gene in a large cohort of patients with an unknown cause of CcO deficiency and where mutations in other candidate genes were already ruled out, failed to reveal any sequence abnormality for COX18 suggesting that mutations in this gene could be incompatible with life or not as frequent (110). As mentioned earlier, not all the steps of CcO biogenesis and assembly factors discovered in yeast have been identified in humans. This is the case of factors related to mtDNA transcription and translation. A new human CcO assembly factor was identified studying a CcO deficient patient with Leigh syndrome with a slow progression of the disease. The CcO defect was due to a mutation in the CCDC44 gene that impaired COX1 protein synthesis. This gene encodes a protein, conserved in bacteria, that functions as a translational activator of COX1, and therefore was renamed TACO1 after its function (111).

1.3.1 An overview on COX11 protein

The most obvious candidate for delivery of copper to the mitochondrion was the COX17 protein. The latter exists in both the cytoplasm and mitochondrial inner membrane space (IMS)(112). Moreover it is known that yeast lacking COX17 (COX17Δ yeast) is respiratory-deficient due to a complete lack of CcO activity (113). The simple prediction was that COX17 would shuttle Cu(I) into the IMS to be used in the assembly of CcO. This theory was strengthened by the observation that COX17 is a copper-binding protein (114). COX17 does not shuttle Cu(I) ions across the outer

membrane (115). This implies that the primary function of COX17 is carried out within the IMS, where COX17 is metallated by using as a Cu(I) source the mitochondrial matrix copper pool (116). COX17 can also be metallated within the cytosol, where it mediates the Cu(I) metallation of a truncated SCO1 (117). However, this transfer reaction does not have any implications for a function of COX17 within the cytosol. The cytosolic protein might represent inefficient import or retention within the IMS. COX17 is postulated to transfer Cu(I) to SCO1 and COX11 proteins, indeed the latters appear to be co-metallochaperones assisting COX17 in the metalation of CcO (83).

Human COX11 is a copper-binding protein of 276 amino acids ubiquitously expressed. It is a mitochondrial membrane polypeptide originally identified in yeast as essential for CcO activity and *heme-a* stability in subunit I (118). It has been shown that CcO isolated from *Rhodobacter* deficient in COX11 contains no CuB site, suggesting a role of this protein in formation of the binuclear copper–*heme* center (119). Consistent with this, biochemical studies indicate that COX11 is a copper-binding protein (120). While COX11 was found to be a dimmer, it remains possible that copper transfer to and from COX11 occurs through heterodimeric interaction of a single COX11 subunit with COX17 and COX1 respectively. It is unknown if COX11 mediated copper incorporation occurs after CcO assembly or co-translationally, prior to the association of COX1 with other subunits.

COX11 protein presents a single transmembrane helix just downstream of the N-terminal mitochondrial targeting sequence. Spectroscopic and mutagenesis studies indicate that copper ion in COX11 is ligated by two conserved cysteine residues. The C-terminal domain of this protein forms a dimer that coordinates a single Cu ion per monomer. Mutation of these Cys residues reduces Cu(I) binding as well as CcO activity. Thus, the residues important for Cu(I)-binding correlate with in vivo function, suggesting that Cu(I)-binding is important in COX11 function (120).

The structure of the globular domain of a COX11 homolog from *Sinorhizobium meliloti* adopts an immunoglobulin-like β fold (121)(**Fig. 8**).

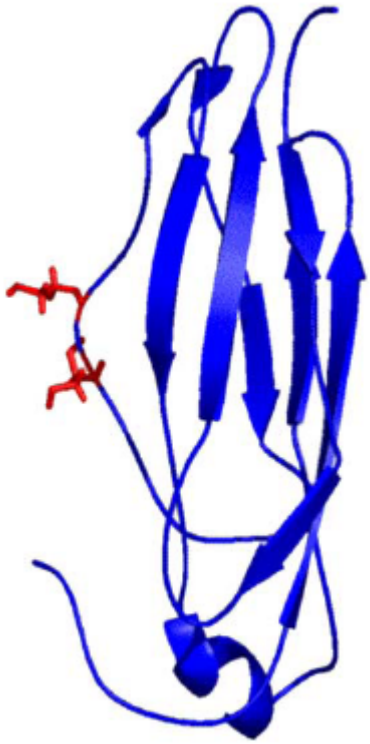


Fig. 8: *Solution structure of COX11. The COX11-homologue from Sinorhizobium meliloti adopts an immunoglobulin-like β fold with the conserved copperbinding cysteine residues on a flexible loop (shown by stick format)* (121)

COX11 dimerizes upon Cu(I) binding to form a binuclear Cu(I) thiolate cluster at the dimer interface (120). COX11 from *Schizosaccharomyces pombe* exists as a fusion protein with the partner protein being homologous to the *S. cerevisiae* protein Rsm22 which is a component of the small subunit of mitochondrial ribosomes (122). Cells lacking Rsm22 are respiratory deficient. The presence of the Rsm22- like domain at the N-terminus of the *S. pombe* COX11 fusion protein suggests that COX11 may function to form the CuB site in nascent COX1 chains emerging from the mitochondrial ribosomes. Transfer of Cu(I) from COX11 to COX1 CuB site buried 13 Å below the membrane surface may occur in nascent COX1 chains extruded across the inner membrane in a cotranslational process. COX11 weakly interacts with the mitochondrial ribosome, so *heme-a*3-CuB site formation likely occurs co-translationally (123).

The above data mainly regards yeast and bacterium COX11 homologous, while detail are missed for human COX11. The function of human COX11 was predicted to be solely attributable to formation of CcO copper centers (81) and it is a possible candidate for human CcO deficiency.

For this reason during my PhD I focussed my work to obtain the soluble C-terminal globular domain of COX11 in order to solve the solution structure by NMR and to better study how it binds Cu(I) and how it is involved in the assembly of CuB site of CcO.

1.3.2 An overview on SURF1 protein

SURF1 is an amphiphilic integral membrane protein of 300 amino acids ubiquitously expressed, but its expression is lower in brain than in other highly aerobic tissues (124). The SURF1 gene is located in a cluster of housekeeping genes known as the Surfeit locus (125) consists of at least six genes (SURF1-6), none of which is related by sequence homology. It presents a characteristic mitochondrial signal peptide (MSP) at its N-terminus. Located in the inner mitochondrial membrane, SURF1 is predicted to form two transmembrane helices connected by a long region facing the intermembrane space (126, 127). Sequence alignments confirm the presence of SURF1 homologs in many eukaryotes and prokaryotes (94). One of the best studied SURF1 proteins is the yeast homolog SHY1p, which has been discovered and characterized in the context of pet mutants (126). Deletion of the gene leads to a strongly decreased CcO level, although the residual enzyme appears fully functional. This points to a role of SHY1p in assembly or stabilization of CcO, most likely during the formation of an early assembly intermediate consisting of the highly conserved core subunit I and II (128). So far, only three bacterial homologs have been inspected in closer detail. In *Paracoccus denitrificans*, two SURF1 homologs were identified and named SURF1c and SURF1q for their specific role in serving a *heme-a*a3-type CcO and a related *heme-b*a3-type quinol oxidase, respectively. With the function of SURF1 in CcO assembly still being speculative, a role in *heme-a* insertion into CcO subunit I seemed conceivable (129, 130). SURF1 may modulate *heme* synthase activity by abstracting the product *heme-a* from the active site of CtaA (*heme-a* synthase in *Paracoccus*; COX15 in eukaryotes) and thus allows continuing *heme* synthesis. It avoids the presence and presumed detrimental action of free *heme-a* in the membrane by providing a safe, yet readily available pool of this cofactor, and specifically chaperones *heme-a* to its target sites in subunit I of CcO and positions the bulky cofactor for presumed co-translational

insertion into both sites within the 12-transmembrane helix scaffold of CcO subunit I (131)(**Fig. 9**).

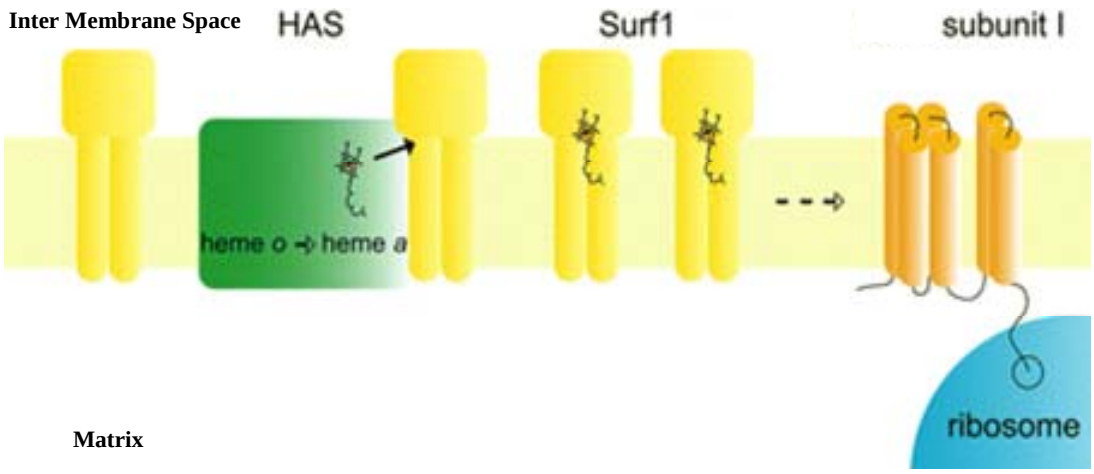


Fig. 9: *Hypothetical role of SURF1 in co-translational incorporation of heme-a into CcO subunit I. Membrane-spanning SURF1 (yellow) in its unloaded apo form interacts with heme-a synthase (HAS, green) to receive the heme-a group. SURF1 is suggested to interact directly with the nascent CcO subunit I polypeptide (orange) emerging from the ribosome (blue) eventually to receive both of its heme groups that are present in its fully folded form. Multiple copies of SURF1 depicted here denote its presumed heme-a pool function.* (131)

The importance of SURF1 in CcO biogenesis is exemplified by the fact that mutations in the human *SURF1* gene, leading to a functional loss of the protein, are responsible for the Leigh syndrome associated with severe CcO deficiency and with loss of CcO activity. The precise role of SURF1 is, however still debated, but as SURF1 is a *heme-a* binding protein, it strongly indicates that SURF1 is involved in direct cofactor incorporation into CcO subunit I. The elucidation of the exact function of SURF1 may spur the development of a treatment for LS.

During my PhD I tried to obtain the full length SURF1 protein in order to study the structure and to understand how it recives the *heme-a* from the *heme-a* synthase and how it release this group to the nascent CcO subunit I. This data should help to clarify how SURF1 protein is involved in CcO assembly and hence in the Leigh's syndrome.

1.4 Familial Amyotrophic Lateral Sclerosis

A common motor neuron disease in human adults is Amyotrophic Lateral Sclerosis (ALS), also called Lou Gehrig's which is a fatal, late-onset, paralytic disorder first described in the late 1800s by the French neurologist Jean-Martin Charcot. The primary hallmark of ALS is the selective killing of motor neurons. Death usually occurs in within five years of symptom onset, typically from respiratory failure. In 90-95% of instances, there is no apparent genetic link and this form of disease is referred as a sporadic ALS, meaning that the afflicted individual has no family history. In the remaining 5-10% of cases, the disease is inherited in a dominant manner (a form referred to as familial ALS, FALS), meaning that a genetic lesion is passed from generation to generation (132, 133).

The characteristic of ALS and FALS diseases is a progressive muscle weakness and athropy each of which reveals the degeneration and death of upper and lower motor neurons in the brain and spinal cord. Among FALS cases, approximately one in fifth have been directly attributed to autosomal dominant mutations in the gene (SOD1), that encodes the copper zinc superoxide dismutase (SOD1) protein (134),(135). Over one hundred mutations in *SOD1* gene has been identified in FALS patients (136, 137) and the list is certain to grow longer as more patients will be analysed (a continually updated list can be found at the online database for SOD1/ALS genetic mutations www.alsa.org). Most of the individual mutations result in substitution of one single amino acid by another and they are scattered throughout the sequence of the protein (**Fig. 10**). None of the mutations eliminate the synthesis of a nearly full-length polipeptide (138). In addition to the these mutations there is also a smaller group of mutations resulting in amino acid deletions. An intriguing aspect of these mutations is that they occur throughout the gene and yet in most instances do not significantly lower the SOD activity of the resulting enzyme (139).

Superoxide dismutases are the major antioxidant enzymes involved in free radical scavenging. These enzymes exist in three distinct forms, where two of them contain copper (SOD1 and SOD3), and the third one contains manganese (SOD2). Wild type (WT) human SOD1 is an exceptionally stable, homodimeric 32kDa protein, located mainly in the cytoplasm, but also present in the peroxisomes, in the mitochondrial intermembrane space and in the nucleus of eukaryotic cells (140, 141).

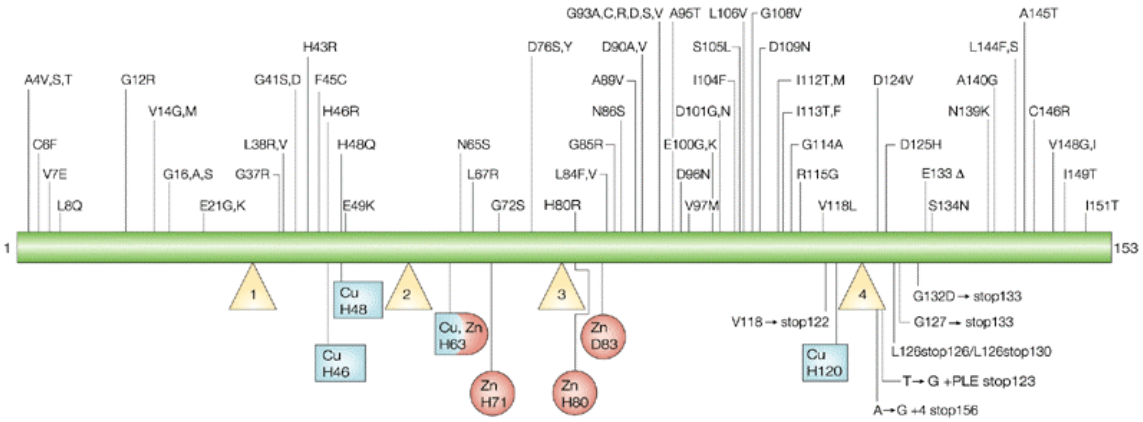


Fig. 10: *ALS-causing mutations lie throughout the SOD1 polypeptide. Mutations in the 153-amino-acid superoxide dismutase polypeptide that are known to cause amyotrophic lateral sclerosis (ALS)*(138).

Each subunit of the dimer binds one copper (binding residues: His46, His48, His63 and His120) and one zinc ion (binding residues: His63, His71, His80 and Asp83) and folds as an eight-stranded Greek-key β-barrel that is stabilized by an intra-subunit disulfide bond (Cys57, Cys146) near the active site **(Figure 11**)(142). In vivo, in the highly reducing cytoplasm environment, the existence of this intrasubunit disulfide bond points to its very low reduction potential. WT human SOD1 actually contains four cysteine residues per monomer. Besides the two cysteines involved in the formation of the intramolecular disulfide bond, two reduced cysteines Cys6 and Cys111 are located on β strand 1 and loop VI respectively. Among the loops connecting the eight β strands, two have structural and functional role. The electrostatic loop (loop VII, residues 121–144) contains charged residues that contribute to guide the negatively charged superoxide substrate towards the catalytic copper site. The long zinc loop (loop IV, residues 49–84) contains all the zinc binding residues including His63, which acts as a ligand to both metals (142). In mammals, SOD1 is ubiquitously expressed in all tissues and within cells it is primarily localized to the cytosol, although lesser amounts are found in the nucleus, peroxisomes, and mitochondria (141). SOD1 detoxifies reactive superoxide anion, a normal by-product of cellular respiration, to molecular oxygen and water (144), so the fundamental role of SOD1 is to act as an antioxidant protein. Loss of enzymatic function could lead to oxidative damage and death of neural cells.

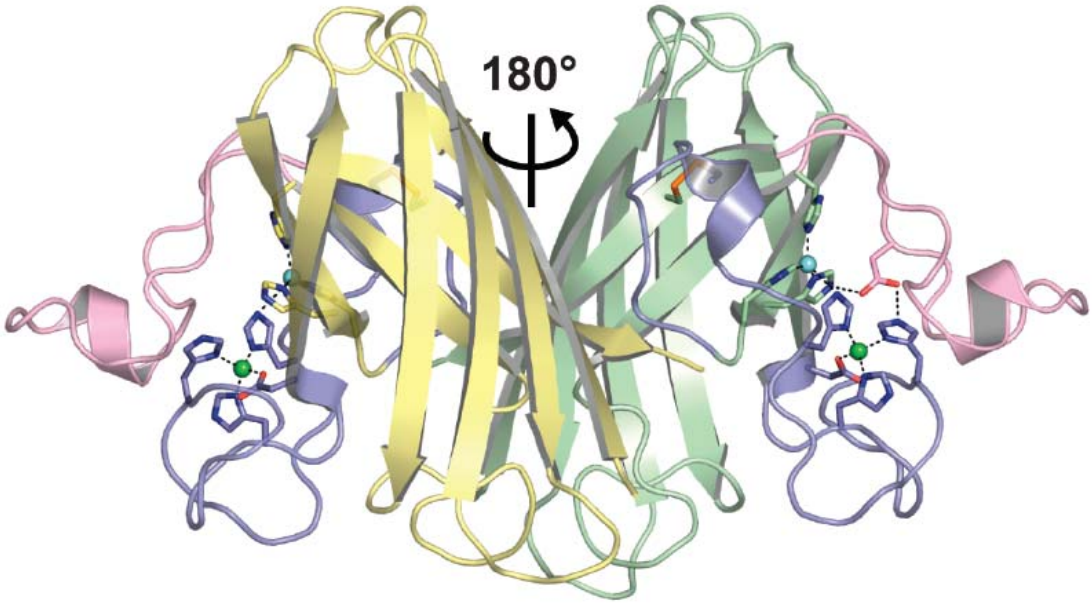


Fig.11: *SOD1 structure. Human Cu-Zn superoxide dismutase [pdb code 2C9V (29)]. The relationship of the two monomers is indicated. Intrasubunit disulfide bonds are shown as orange sticks, the metal-binding loops (loop IV and VII) are shown in blue and pink, respectively. Copper and zinc ions are shown as cyan and green spheres, respectively.(143)*

However, mice lacking SOD1 do not develop motor neuron disease (145) and transgenic mice expressing human FALS SOD1 mutants in addition to their own endogenous SOD1 develop paralytic symptoms strikingly similar to those observed in human patients(139,146,147). Together, these observations imply that pathogenic SOD1 molecules act through the gain of a cytotoxic property and not a loss of function. The insertion of copper into SOD1 requires a copper metallochaperone called CCS (copper chaperone for superoxide dismutase) (149,151). Yeast mutants lacking CCS express a form of SOD1 protein that is essentially apo for copper (151) but contains a single atom of zinc per subunit (149) (152). Mice with disruptions in CCS exhibit marked reductions in SOD1 activity, emphasizing the conserved requirement for CCS in activation of eukaryotic SOD1 (152). In brain SOD1 can not acquire its Cu in the absence of the CCS (153). Due to its biological function CCS is considered as a potential candidate gene for FALS.

Human CCS is a three-domain polypeptide that confers at least two critical stabilizing posttranslational modifications on newly synthesized SOD1: the insertion of the

catalytic copper ion (149), and the oxidation of the disulfide bond found within each SOD1 subunit (154). The presence of a disulfide bond is rare for cytosolic proteins given the strong reducing environment of the cytosol, and recent studies suggest that CCS-mediated oxidation of this disulfide bond occurs concomitant with copper delivery in an oxygen- or superoxide-dependent fashion (154). It is important to note that at the protein level, the ratio of SOD1 to CCS in the cytosol is estimated to be between 15:1 and 30:1 (153), meaning that CCS must cycle through the newly translated SOD1 pool to activate these molecules (155). It is presumed that upon copper delivery to SOD1, CCS becomes recharged with copper via the membrane-bound copper transporter CTR1, although direct protein-protein interactions between CCS and CTR1 have not yet been demonstrated.

Is suspected that pathogenic SOD1 mutations may result in increased levels of immature SOD1 folding intermediates by hindering the action of CCS at various points in the SOD1 maturation cycle. These immature SOD1 folding intermediates may lack some or all posttranslational modifications that are dependent upon CCS activity, including the insertion of copper co-factors and oxidation of the intrasubunit disulfide bond. Two recent in vitro studies indicate that failure to form the intrasubunit disulfide bond may be a key step (156, 157). Similarly, studies of aggregates isolated from cell culture and transgenic mouse models indicate that immature SOD1 molecules eventually end up in the insoluble inclusions (158). One possible explanation for these observations is that mutant forms of SOD1 interact with CCS in a manner that inhibits normal posttranslational modifications (for example, the nascent metal-binding mutants). Alternatively, some mutant forms may be so destabilized that CCS encounters only a fraction of these variants before they exit the soluble fraction or are turned over by the protein quality control machinery (for example, the nascent β-barrel mutants).

1.4.1 An overview on CCS protein

Human CCS is encoded by a single copy gene on chromosome 11, which is ubiquitously expressed in all human tissues. The protein is located in cytoplasm and binds copper ions and one zinc ion. CCS exists as a homodimer that may form a heterodimer with SOD1 during copper loading. Three functional domains have been described in the 274 amino acid long protein product (159, 160).

CCS domain I (residues 1–85) contains a copperbinding motif MXCXXC that is postulated to acquire copper ion from the membrane copper transporter CTR1 (161). Domain II (residues 86–238) is similar to human SOD1 and retains amino acid residues found at the SOD1 dimer interface (162). Because dimer interface residues are conserved, domain II is proposed to be responsible for the specificity of CCS/SOD1 interaction via the formation of a SOD1/CCS heterodimer (163). Domain III (residues 239– 274) contains the copper-binding motif CXC, which is proposed to directly insert copper ion into nascent SOD1 (163). The ‘‘heterodimerization’’ model of CCS activation of SOD1 is shown in **Fig. 12**. It remains unclear what amino acid residues of both proteins participate in copper delivery and disulfide bond oxidation and, most importantly, how might FALS mutations in SOD1 interfere with CCS action, and what are the properties of the resulting immature SOD1 proteins.

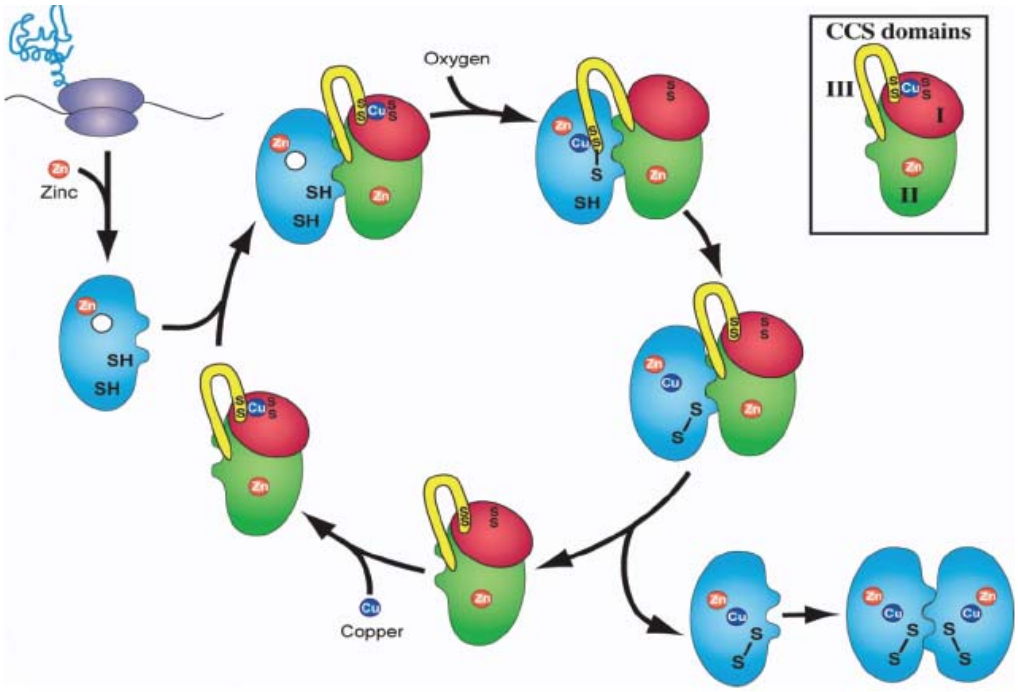


Fig. 12: *Heterodimerization model of CCS action after O'Halloran and colleagues(161). Newly translated SOD1 monomers are shown in blue. CCS domain I is shown in red, CCS domain II is shown in green, and CCS domain III is shown in yellow (inset)(143).*

To probe various mechanistic aspects of CCS action, Blackburn and colleagues titrating purified human CCS with increasing concentrations of Cu(I) (164). Upon addition of a single equivalent of Cu(I) per CCS molecule, they observed that the ‘‘canonical’’ CCS

dimer mediated by the SOD1-like domain II dissociated into monomers and that addition of additional equivalents resulted in the appearance of a ‘‘noncanonical’’ dimer mediated by a Cu_4S_6 copper cluster formed in part by the two CXC motifs of CCS domain III (**Fig. 13**).

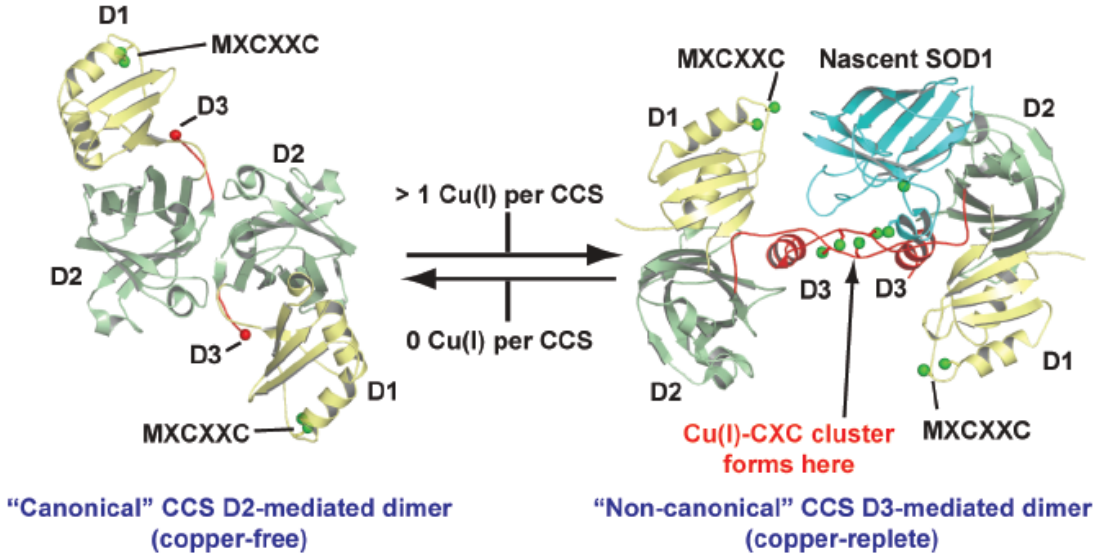


Fig. 13: *Human CCS quaternary structure as a function of copper loading. The CCS canonical dimer [pdb code 1QUP (162)] reorganizes to form the noncanonical dimer [pdb code 1JK9(165)] upon the binding of Cu(I)(164), thereby freeing domain II to interact with nascent SOD1.(143)*

This canonical-to-noncanonical CCS dimer transition, if it occurs in vivo, would seem to act as a functional copper-sensing switch to make CCS domain II available to nascent SOD1 binding only when sufficient copper is available (164). Interestingly, a non-canonical domain III mediated CCS dimer similar to that shown in **Fig. 13** was also observed in the crystal structure of a yeast SOD1/yeast CCS complex, although no copper was present in the crystallization experiment (165). Taken together, these observations suggest an alternate model of CCS action shown in **Fig. 14**. The latter mechanism is particularly intriguing, given that CCS contains its own intrasubunit disulfide bond and bound zinc, which as described above, are both factors that are known to stabilize SOD1 dimers (166) and by extension, would be expected to stabilize CCS dimers. However, this new model has not been unambiguously validated and does

not reveal precise mechanistic details of the posttranslational modification of SOD1 or how the binding of a single copper ion results in allosteric CCS dimer dissociation. For this reason one of my PhD projects was focussed to obtain the full length CCS protein in order to better understand the behaviour of the three domains in presence of Cu(I) by studying the solution structure and the interaction with this metal by NMR. Moreover it will be interesting to study the interaction of CCS protein in presence of SOD.

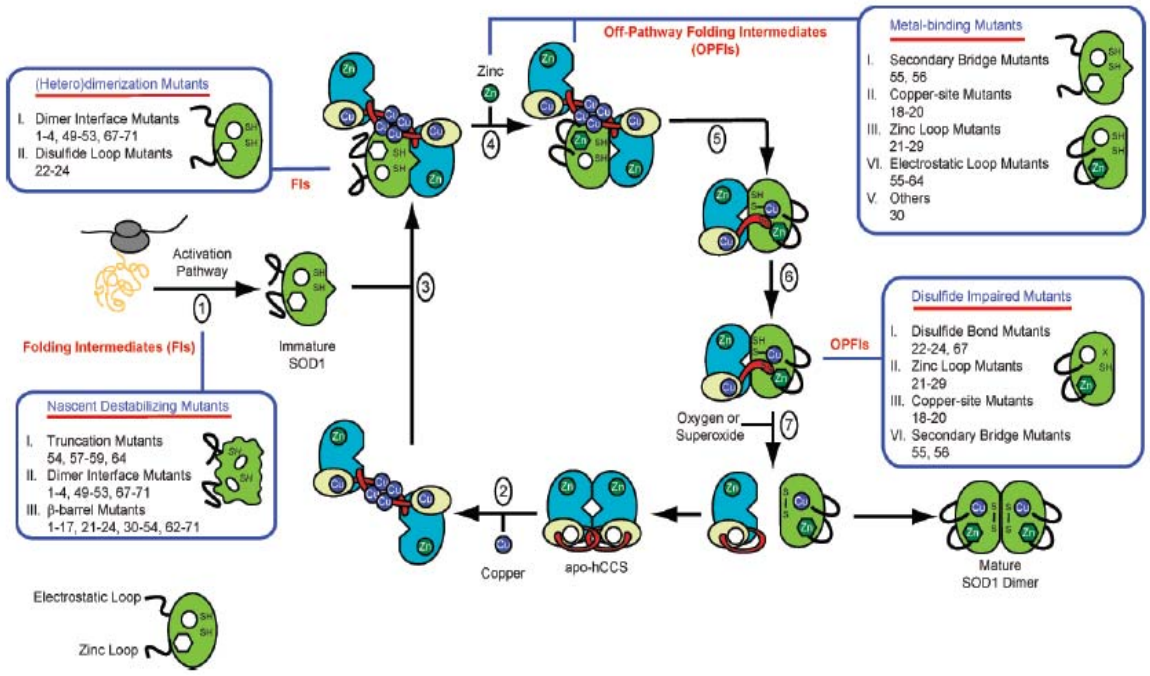


Fig. 14: *An alternate model of CCS action and how the various pathogenic SOD1 mutations may hinder CCS-mediated SOD1 maturation. (1) SOD1 is translated. (2) The canonical CCS dimer is loaded with Cu(I) to generate the noncanonical CCS dimer mediated by a Cu4S6 cluster (164). (3) Nascent SOD1 binds to domain II of CCS in the noncanonical, Cu(I)-loaded CCS dimer. (4) Zinc is loaded into SOD1 (this could also occur as early as step 1). (5–7) Cu(I) from the Cu4S6 cluster is transferred to nascent SOD1 and the intrasubunit disulfide bond in nascent SOD1 is oxidized(161). Upon being depleted with Cu(I), CCS reforms the canonical CCS dimer and the cycle repeats. The pathogenic SOD1 mutations listed in Table 1 may interfere with CCS-mediated SOD1 maturation at the various positions indicated.(143)*

Reference List

1. Bull, P. C., Thomas, G. R., Rommens, J. M., Forbes, J. R. & Cox, D. W. (1993) *Nature Genet.* **5**, 327-337.

2. Vulpe, C. D., Levinson, B., Whitney, S., Packman, S. & Gitschier, J. (1993) *Nature Genet.* **3**, 7-13.

3. Llanos, R. M. & Mercer, J. F. B. (2002) *Dna and Cell Biology* **21**, 259-270.

4. Lutsenko, S., Barnes, N. L., Bartee, M. Y. & Dmitriev, O. Y. (2007) *Physiol Rev.* **87**, 1011-1046.

5. Walshe, J. A. (2006) *Movement Disorders* **21**, 142-147.

6. Sternlieb, I. (1990) *Hepatology* **12**, 1234-1239.

7. Wilson, D. C., Phillips, M. J., Cox, D. W. & Roberts, E. A. (2000) *Journal of Pediatrics* **137**, 719-722.

8. Petrukhin, K., Fischer, S. G., Tanzi, R. E., Chernov, I., Devoto, M. & Brzustowicz LM.et al. (1993) *Nature Genet.* **5**, 338-343.

9. Tanzi, R. E., Petrukhin, K., Chernov, I., Pellequer, J. L., Wasco W., Ross, B., Romano, D. M., Parano, E., Pavone, L. & Brzustowicz LM.et al. (1993) *Nature Genet.* **5**, 344-350.

10. Petrukhin, K., Lutsenko, S., Chernov, I., Ross, B. M., Kaplan, J. H. & Gilliam, T. C. (1994) *Human Molecular Genetics* **3**, 1647-1656.

11. La Fontaine, S. & Mercer, J. F. (2007) *Arch. Biochem. Biophys.* **463**, 149-167.

12. Forbes, J. R. & Cox, D. W. (2000) *Human Molecular Genetics* **9**, 1927-1935.

13. Thomas, G. R., Forbes, J. R., Roberts, E. A., Walshe, J. M. & Cox, D. W. (1995) *Nature Genet.* **9**, 210-216.

14. Firneisz G, Lakatos PL, Szalay F, Polli C, Glant T T & Ferenci P (2002) *Am J Med Genet.* **108**, 23-28.

15. Panagiotakaki E, Tzetis M, Manolaki N, Loudianos G, Papatheodorou A, Manesis E, Nousia-Arvanitakis S, Syriopoulou V & Kanavakis E (2004) *Am J Med Genet A.* **131(2)**, 168-173.

16. Cox DW, Prat L, Walshe JM, Heathcote J & Gaffney D (2005) *Hum Mutat.* **26(3)**, 280.

17. Folhoffer A, Ferenci P, Csak T, Horvath A, Hegedus D, Firneisz G, Osztovits J, Kosa JP, Willheim-Polli C, Szonyi L *et al.* (2007) *Eur J Gastroenterol Hepatol.* **19(2)**, 105-111.

18. Thomas, G. R., Roberts, E. A., Walshe, J. M. & Cox, D. W. (1995) *American Journal of Human Genetics* **56**, 1315-1319.

19. Okada, T., Shiono, Y., Hayashi, H., Satoh, H., Sawada, T., Suzuki, A., Takeda, Y., Yano, M., Michitaka, K., Onji, M. *et al.* (2000) *Human Mutation* **15**, 454-462.

20. Wu ZY, Wang N, Lin MT, Fang L, Murong SX & Yu L (2001) *Arch Neurol.* **58(6)**, 971-976.

21. Yoo HW (2002) *Genet Med.* **4**, 43-48.

22. Wan L, Tsai FJ, Lee CC, Hsu CM, Tsai Y & Tsai CH (2006) *Biochem Biophys Res Commun.* **345(2)**, 734-738.

23. Kenney, S. M. & Cox, D. W. (2007) *Human Mutation* **28**, 1171-1177.

24. Hung, I. H., Suzuki, M., Yamaguchi, Y., Yuan, D. S., Klausner, R. D. & Gitlin, J. D. (1997) *J. Biol. Chem.* **272**, 21461-21466.

25. Terada, K., Nakako, T., Yang, X.-L., Iida, M., Aiba, N., Minamiya, J., Nakai, M., Sakaki, T., Miura, N. & Sugiyama, T. (1998) *J. Biol. Chem.* **273**, 1815-1820.

26. Klomp, L. W., Lin, S. J., Yuan, D., Klausner, R. D., Culotta, V. C. & Gitlin, J. D. (1997) *J. Biol. Chem.* **272**, 9221-9226.

27. Hung, I. H., Casareno, R. L., Labesse, G., Mathews, F. S. & Gitlin, J. D. (1998) *J. Biol. Chem.* **273**, 1749-1754.

28. Schaefer, M., Hopkins, R. G., Failla, M. L. & Gitlin, J. D. (1999) *Am. J. Physiol.* **276**, G639-G646.

29. Petris, M. J., Mercer, J. F., Culvenor, J. G., Lockhart, P. & Camakaris, J. (1996) *EMBO J.* **15**, 6084-6095.

30. Petris, M. J., Camakaris, J., Greenough, M., LaFontaine, S. & Mercer, J. F. B. (1998) *Human Molecular Genetics* **7**, 2063-2071.

31. Monty, J. F., Llanos, R. M., Mercer, J. F. & Kramer, D. R. (2005) *J. Nutr.* **135**, 2762-2766.

32. Argüello JM, Eren E & González-Guerrero M (2007) *Biometals.* **20(3-4)**, 233-248.

33. Arnesano, F., Banci, L., Bertini, I., Ciofi-Baffoni, S., Molteni, E., Huffman, D. L. & O'Halloran, T. V. (2002) *Genome Res.* **12**, 255-271.

34. Tsivkovskii, R., Eisses, J. F., Kaplan, J. H. & Lutsenko, S. (2002) *Journal of Biological Chemistry* **277**, 976-983.

35. Voskoboinik, I., Mar, J., Strausak, D. & Camakaris, J. (2001) *J. Biol. Chem.* **276**, 28620-28627.

36. Tsivkovskii, R., Efremov, R. G. & Lutsenko, S. (2003) *J. Biol. Chem.* **278**, 13302-13308.

37. Barnes, N., Tsivkovskii, R., Tsivkovskaia, N. & Lutsenko, S. (2005) *Journal of Biological Chemistry* **280**, 9640-9645.

38. Hung YH, Layton MJ, Voskoboinik I, Mercer JF & Camakaris J (2007) *Biochem J.* **401(2)**, 569-579.

39. El Meskini, R., Culotta, V. C., Mains, R. E. & Eipper, B. A. (2003) *Journal of Biological Chemistry* **278**, 12278-12284.

40. Qin, Z. Y., Itoh, S., Jeney, V., Ushio-Fukai, M. & Fukai, T. (2005) *Faseb Journal* **19**, 334-+.

41. Lutsenko, S., Efremov, R. G., Tsivkovskii, R. & Walker, J. M. (2002) *J. Bioenerg. Biomemb.* **34**, 351-362.

42. Solioz, M. & Vulpe, C. D. (1996) *Trends Biochem. Sci.* **21**, 237-241.

43. Yuan, D. S., Stearman, R., Dancis, A., Dunn, T., Beeler, T. & Klausner, R. D. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 2632-2636.

44. Elam, J. S., Thomas, S. T., Holloway, S. P., Taylor, A. B. & Hart, P. J. (2002) *Copper-Containing Proteins* **60**, 151-219.

45. Wernimont, A. K., Huffman, D. L., Lamb, A. L., O'Halloran, T. V. & Rosenzweig, A. C. (2000) *Nat Struct Biol* **7**, 766-771.

46. Rosenzweig, A. C., Huffman, D. L., Hou, M. Y., Wernimont, A. K., Pufahl, R. A. & O'Halloran, T. V. (1999) *Structure Fold Des.* **7**, 605-617.

47. Arnesano, F., Banci, L., Bertini, I., Huffman, D. L. & O'Halloran, T. V. (2001) *Biochemistry* **40**, 1528-1539.

48. Wimmer, R., Herrmann, T., Solioz, M. & Wüthrich, K. (1999) *J. Biol. Chem.* **274**, 22597-22603.

49. Banci, L., Bertini, I., Del Conte, R., Markey, J. & Ruiz-Dueñas, F. J. (2001) *Biochemistry* **40**, 15660-15668.

50. Gitschier, J., Moffat, B., Reilly, D., Wood, W. I. & Fairbrother, W. J. (1998) *Nature Struct. Biol.* **5**, 47-54.

51. Banci, L., Bertini, I., Ciofi-Baffoni, S., Huffman, D. L. & O'Halloran, T. V. (2001) *J. Biol. Chem.* **276**, 8415-8426.

52. Lutsenko, S., Petrukhin, K., Cooper, M. J., Gilliam, C. T. & Kaplan, J. H. (1997) *J. Biol. Chem.* **272**, 18939-18944.

53. Ralle, M., Cooper, M. J., Lutsenko, S. & Blackburn, N. J. (1998) *J. Am. Chem. Soc.* **120**, 13525-13526.

54. DiDonato, M., Hsu, H. F., Narindrasorasak, S., Que, L. J. & Sarkar, B. (2000) *Biochemistry* **39**, 1890-1896.

55. Ralle, M., Lutsenko, S. & Blackburn, N. J. (2003) *J. Biol. Chem.* **278**, 23163-23170.

56. Iida, M., Terada, K., Sambongi, Y., Wakabayashi, T., Miura, N., Koyama, K., Futai, M. & Sugiyama, T. (1998) *FEBS Lett.* **428**, 281-285.

57. Forbes, J. R., Hsi, G. & Cox, D. W. (1999) *J. Biol. Chem.* **274**, 12408-12413.

58. Huster, D. & Lutsenko, S. (2003) *J. Biol. Chem.* **278**, 32212-32218.

59. Payne, A. S. & Gitlin, J. D. (1998) *J. Biol. Chem.* **273**, 3765-3770.

60. Banci, L., Bertini, I., Del Conte, R., D'Onofrio, M. & Rosato, A. (2004) *Biochemistry* **43**, 3396-3403.

61. Banci, L., Bertini, I., Cantini, F., Migliardi, M., Rosato, A. & Wang, S. (2005) *J. Mol. Biol.* **352**, 409-417.

62. Banci, L., Bertini, I., Cantini, F., Della Malva, N., Migliardi, M. & Rosato, A. (2007) *J. Biol. Chem.* **282**, 23140-23146.

63. Yatsunyk, L. A. & Rosenzweig, A. C. (2007) *J. Biol. Chem.* **282**, 8622-8631.

64. Bunce, J., Achila, D., Hetrick, E., Lesley, L. & Huffman, D. L. (2006) *Biochim. Biophys. Acta* **1760**, 907-912.

65. Larin, D., Mekios, C., Das, K., Ross, B., Yang, A. S. & Gilliam, C. T. (1999) *J. Biol. Chem.* **274**, 28497-28504.

66. van Dongen, E. M., Klomp, L. W. & Merkx, M. (2004) *Biochem. Biophys. Res. Commun.* **323**, 789-795.

67. Banci, L., Bertini, I., Cantini, F., Della Malva, N., Rosato, A., Herrmann, T. & Wüthrich, K. (2006) *J. Biol. Chem.* **281**, 29141-29147.

68. Strausak, D., Howie, M. K., Firth, S. D., Schlicksupp, A., Pipkorn, R., Multhaup, G. & Mercer, J. F. (2003) *J. Biol. Chem.* **278**, 20821-20827.

69. Achila, D., Banci, L., Bertini, I., Bunce, J., Ciofi-Baffoni, S. & Huffman, D. L. (2006) *Proc. Natl. Acad. Sci. USA* **103**, 5729-5734.

70. Banci, L., Bertini, I., Cantini, F., Rosenzweig, A. C. & Yatsunyk, L. A. (2008) *Biochemistry* **47**, 7423-7429.

71. Diaz, F. & Moraes, C. T. (2008) *Cell Calcium* **44**, 24-35.

72. DiMauro, S. & Schon, E. A. (2008) *Annual Review of Neuroscience* **31**, 91-123.

73. Grivell, L. A., rtal-Sanz, M., Hakkaart, G., de Jong, L., Nijtmans, L. G. J., van Oosterum, K., Siep, M. & van der Spek, H. (1999) *FEBS Letters* **452**, 57-60.

74. Tzagoloff, A. & Dieckmann, C. L. (1990) *Microbiol. Rev.* **54**, 211-225.

75. Chinnery, P. F. & Turnbull, D. M. (1999) *Lancet* **354**, SI17-SI21.

76. Wikstrom, M. (1998) *Curr. Opin. Struct. Biol.* **8**, 480-488.

77. Michel, H., Behr, J., Harrenga, A. & Kannt, A. (1998) *Annu. Rev. biophys. biomol. Struct.* **27**, 329-356.

78. Capaldi, R. A. (1990) *Annu Rev Biochem* **59**, 569-596.

79. Poyton, R. O. & Mcewen, J. E. (1996) *Annual Review of Biochemistry* **65**, 563-607.

80. Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. & Yoshikawa, S. (1995) *Science* **269**, 1069-1074.

81. Carr, H. S. & Winge, D. R. (2003) *Acc. Chem. Res.* **36**, 309-316.

82. Pecina, P., Gnaiger, E., Zeman, J., Pronicka, E. & Houstek, J. (2004) *American Journal of Physiology-Cell Physiology* **287**, C1384-C1388.

83. Cobine, P. A., Pierrel, F. & Winge, D. R. (2006) *Biochim. Biophys. Acta* **1763**, 759-772.

84. Clark, K. M., Taylor, R. W., Johnson, M. A., Chinnery, P. F., Chrzanowska-Lightowlers, Z. M. A., Andrews, R. M., Nelson, I. P., Wood, N. W., Lamont, P. J., Hanna, M. G. *et al.* (1999) *American Journal of Human Genetics* **64**, 1330-1339.

85. Schon, E. A., Bonilla, E. & Dimauro, S. (1997) *Journal of Bioenergetics and Biomembranes* **29**, 131-149.

86. Shoubridge, E. A. (2001) *American Journal of Medical Genetics* **106**, 46-52.

87. Massa, V., Fernandez-Vizarra, E., Alshahwan, S., Bakhsh, E., Goffrini, P., Ferrero, I., Mereghetti, P., D'Adamo, P., Gasparini, P. & Zeviani, M. (2008) *American Journal of Human Genetics* **82**, 1281-1289.

88. Diaz F (2009) *Biochim Biophys Acta.* **Epub ahead of print**.

89. Dahl, H. H. M. (1998) *American Journal of Human Genetics* **63**, 1594-1597.

90. Castro-Gago, M., Blanco-Barca, M. O., Campos-Gonzalez, Y., renas-Barbero, J., Pintos-Martinez, E. & Eiris-Punal, J. (2006) *Pediatric Neurology* **34**, 204-211.

91. Bernier, F. P., Boneh, A., Dennett, X., Chow, C. W., Cleary, M. A. & Thorburn, D. R. (2002) *Neurology* **59**, 1406-1411.

92. Leigh D (1951) *J Neurol Neurosurg Psychiatry.* **14(3)**, 216-221.

93. Tiranti, V., Hoertnagel, K., Carrozzo, R., Galimberti, C., Munaro, M., Granatiero, M., Zelante, L., Gasparini, P., Marzella, R., Rocchi, M. *et al.* (1998) *American Journal of Human Genetics* **63**, 1609-1621.

94. Poyau, A., Buchet, K. & Godinot, C. (1999) *FEBS Lett.* **462**, 416-420.

95. Williams, S. L., Valnot, I., Rustin, P. & Taanman, J. W. (2004) *Journal of Biological Chemistry* **279**, 7462-7469.

96. Tulinius, M., Moslemi, A. R., Darin, N., Westerberg, B., Wiklund, L. M., Holme, E. & Oldfors, A. (2003) *Neuropediatrics* **34**, 87-91.

97. Rahman, S., Blok, R. B., Dahl, H. H. M., Danks, D. M., Kirby, D. M., Chow, C. W., Christodoulou, J. & Thorburn, D. R. (1996) *Annals of Neurology* **39**, 343-351.

98. Merante, F., Petrovabenedict, R., MacKay, N., Mitchell, G., Lambert, M., Morin, C., Debraekeleer, M., Laframboise, R., Gagne, R. & Robinson, B. H. (1993) *American Journal of Human Genetics* **53**, 481-487.

99. Mootha, V. K., Lepage, P., Miller, K., Bunkenborg, J., Reich, M., Hjerrild, M., Delmonte, T., Villeneuve, A., Sladek, R., Xu, F. H. *et al.* (2003) *Proceedings of the National Academy of Sciences of the United States of America* **100**, 605-610.

100. Papadopoulou, L. C., Sue, C. M., Davidson, M. M., Tanji, K., Nishino, I., Sadlock, J. E., Krishna, S., Walker, W., Selby, J., Glerum, D. M. *et al.* (1999) *Nature Genet.* **23**, 333-337.

101. Stiburek, L., Hansikova, H., Tesarova, M., Cerna, L. & Zeman, J. (2006) *Physiol Res.* **55 Suppl 2**, S27-S41.

102. Valnot, I., Osmond, S., Gigarel, N., Mehaye, B., Amiel, J., Cormier-Daire, V., Munnich, A., Bonnefont, J. P., Rustin, P. & Rotig, A. (2000) *Am J Hum Genet* **67**, 1104-1109.

103. Jaksch, M., Ogilvie, I., Yao, J., Kortenhaus, G., Bresser, H. G., Gerbitz, K. D. & Shoubridge, E. A. (2000) *Hum. Mol. Genet.* **9**, 795-801.

104. Stiburek, L., Vesela, K., Hansikova, H., Hulkova, H. & Zeman, J. (2009) *American Journal of Physiology-Cell Physiology* **296**, C1218-C1226.

105. Stiburek, L., Vesela, K., Hansikova, H., Hulkova, H. & Zeman, J. (2009) *American Journal of Physiology-Cell Physiology* **296**, C1218-C1226.

106. Antonicka, H., Leary, S. C., Guercin, G. H., Agar, J. N., Horvath, R., Kennaway, N. G., Harding, C. O., Jaksch, M. & Shoubridge, E. A. (2003) *Hum Mol Genet* **12**, 2693-2702.

107. Coenen, M. J. H., van den Heuvel, L. P., Ugalde, C., ten Brinke, M., Nijtmans, L. G. J., Trijbels, F. J. M., Beblo, S., Maier, E. M., Muntau, A. C. & Smeitink, J. A. M. (2004) *Ann. Neurol.* **56**, 560-564.

108. Oquendo, C. E., Antonicka, H., Shoubridge, E. A., Reardon, W. & Brown, G. K. (2004) *Journal of Medical Genetics* **41**, 540-544.

109. Ghezzi, D., Saada, A., D'Adamo, P., Fernandez-Vizarra, E., Gasparini, P., Tiranti, V., Elpeleg, O. & Zeviani, M. (2008) *American Journal of Human Genetics* **83**, 415-423.

110. Sacconi, S., Salviati, L. & Trevisson, E. (2009) *Journal of Human Genetics* **54**, 419-421.

111. Weraarpachai, W., Antonicka, H., Sasarman, F., Seeger, J., Schrank, B., Kolesar, J. E., Lochmuller, H., Chevrette, M., Kaufman, B. A., Horvath, R. *et al.* (2009) *Nature Genetics* **41**, 833-U96.

112. Beers, J., Glerum, D. M. & Tzagoloff, A. (1997) *J. Biol. Chem.* **272**, 33191-33196.

113. Glerum, D. M., Shtanko, A. & Tzagoloff, A. (1996) *J. Biol. Chem.* **271**, 14504-14509.

114. Heaton, D. N., George, G. N., Garrison, G. & Winge, D. R. (2001) *Biochemistry* **40**, 743-751.

115. Maxfield, A. B., Heaton, D. N. & Winge, D. R. (2004) *J. Biol. Chem.* **279**, 5072-5080.

116. Cobine, P. A., Ojeda, L. D., Rigby, K. M. & Winge, D. R. (2004) *J. Biol. Chem.* **279**, 14447-14455.

117. Horng, Y. C., Cobine, P. A., Maxfield, A. B., Carr, H. S. & Winge, D. R. (2004) *J. Biol. Chem.* **279**, 35334-35340.

118. Tzagoloff, A., Capitanio, N., Nobrega, M. P. & Gatti, D. (1990) *EMBO J.* **9**, 2759-2764.

119. Hiser, L., Di Valentin, M., Hamer, A. G. & Hosler, J. P. (2000) *J. Biol. Chem.* **275**, 619-623.

120. Carr, H. S., George, G. N. & Winge, D. R. (2002) *J. Biol. Chem.* **277**, 31273-31242.

121. Banci, L., Bertini, I., Cantini, F., Ciofi-Baffoni, S., Gonnelli, L. & Mangani, S. (2004) *J. Biol. Chem.* **279**, 34833-34839.

122. Saveanu, C., Fromont-Racine, M., Harington, A., Ricard, F., Namane, A. & Jacquier, A. (2001) *Journal of Biological Chemistry* **276**, 15861-15867.

123. Khalimonchuk, O. & Rodel, G. (2005) *Mitochondrion* **5**, 363-388.

124. Lennard, A., Gaston, K. & Fried, M. (1994) *Dna and Cell Biology* **13**, 1117-1126.

125. Huxley C & Fried M (1990) *Mol Cell Biol.* **10(2)**, 605-614.

126. Mashkevich, G., Repetto, B., Glerum, D. M., Jin, C. & Tzagoloff, A. (1997) *Journal of Biological Chemistry* **272**, 14356-14364.

127. Yao, J. B. & Shoubridge, E. A. (1999) *Human Molecular Genetics* **8**, 2541-2549.

128. Barrientos, A., Barros, M. H., Valnot, I., Rotig, A., Rustin, P. & Tzagoloff, A. (2002) *Gene* **286**, 53-63.

129. Bundschuh, F. A., Hoffmeier, K. & Ludwig, B. (2008) *Biochimica et Biophysica Acta-Bioenergetics* **1777**, 1336-1343.

130. Smith, D., Gray, J., Mitchell, L., Antholine, W. E. & Hosler, J. P. (2005) *Journal of Biological Chemistry* **280**, 17652-17656.

131. Bundschuh, F. A., Hannappel, A., Anderka, O. & Ludwig, B. (2009) *Journal of Biological Chemistry* **284**, 25735-25741.

132. Bruijn, L. I., Miller, T. M. & Cleveland, D. W. (2004) *Annu. Rev. Neurosci.* **27**, 723-749.

133. Valentine, J. S., Doucette, P. A. & Potter, S. Z. (2005) *Annu Rev Biochem* **74**, 563-593.

134. Rosen, D. R., Siddique, T., Patterson, D., Figlewicz, D. A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J., Deng, H. X. *et al.* (1993) *Nature* **362**, 59-62.

135. McCord, J. M. & Fridovich, I. (1969) *J. Biol. Chem.* **244**, 6049-6055.

136. Andersen, P. M. (2001) *Amyotroph. Lateral Scler. Other Motor. Neuron. Disord.* **2 Suppl 1**, S37-S41.

137. Gaudette, M., Hirano, M. & Siddique, T. (2000) *Amyotroph Lateral Scler Other Motor Neuron Disord.* **1**, 83-89.

138. Cleveland, D. W. & Rothstein, J. D. (2001) *Nat. Rev. Neurosci.* **2**, 806-819.

139. Gurney, M. E., Pu, H., Chiu, A. Y., Dal Canto, M., Polchow, C. Y., Alexander, D. D., Caliendo, J., Hentati, A., Kwon, Y. W., Deng, H. X. *et al.* (1994) *Science* **264**, 1772-1775.

140. Okado-Matsumoto, A. & Fridovich, I. (2001) *J. Biol. Chem.* **276**, 38388-38393.

141. Sturtz, L. A., Diekert, K., Jensen, L. T., Lill, R. & Culotta, V. C. (2001) *J. Biol. Chem.* **276**, 38084-38089.

142. Bertini, I., Mangani, S. & Viezzoli, M. S. (1998) in *Advanced Inorganic Chemistry*, ed. Sykes, A. G. (Academic Press, San Diego, CA, USA), pp. 127-250.

143. Seetharaman SV, Prudencio M, Karch C, Holloway SP, Borchelt DR & Hart PJ (2009) *Exp Biol Med* 1140-1154.

144. Fridovich I (1989) *J Biol Chem.* **264(14)**, 7761-7764.

145. Reaume, A. G., Elliott, J. L., Hoffman, E. K., Kowall, N. W., Ferrante, R. J., Siwek, D. F., Wilcox, H. M., Flood, D. G., Beal, M. F., Brown, R. H., Jr. *et al.* (1996) *Nature Genet.* **13**, 43-47.

146. Wang J, Xu G, Li H, Gonzales V, Fromholt D, Karch C, Copeland NG, Jenkins NA & Borchelt DR (2005) *Hum Mol Genet.* **14(16)**, 2335-2347.

147. Bruijn LI, Becher MW, Lee MK, Anderson KL, Jenkins NA, Copeland NG, Sisodia SS, Rothstein JD, Borchelt DR, Price DL *et al.* (1997) *Neuron.* **18(2)**, 327-338.

148. Thayer, M. M., Ahern, H., Xing, D., Cunningham, R. P. & Tainer, J. A. (1995) *EMBO J.* **14**, 4108-4120.

149. Culotta, V. C., Klomp, L. W., Strain, J., Casareno, R. L., Krems, B. & Gitlin, J. D. (1997) *J. Biol. Chem.* **272**, 23469-23472.

150. Portnoy, M. E., Schmidt, P. J., Rogers, R. S. & Culotta, V. C. (2001) *Mol. Genet. Genomics* **265**, 873-882.

151. Gamonet, F. & Lauquin, G. J. M. (1998) *Eur. J. Biochem.* **251**, 716-723.

152. Wong, P. C., Waggoner, D., Subramaniam, J. R., Tessarollo, L., Bartnikas, T. B., Culotta, V. C., Price, D. L., Rothstein, J. & Gitlin, J. D. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 2886-2891.

153. Rothstein, J. D., Dykes-Hoberg, M., Corson, L. B., Becker, M., Cleveland, D. W., Price, D. L., Culotta, V. C. & Wong, P. C. (1999) *J. Neurochem.* **72**, 422-429.

154. Brown, N. M., Torres, A. S., Doan, P. E. & O'Halloran, T. V. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 5518-5523.

155. Furukawa, Y., Fu, R., Deng, H. X., Siddique, T. & O'Halloran, T. V. (2006) *Proc. Natl. Acad. Sci. USA* **103**, 7148-7153.

156. Chattopadhyay M, Durazo A, Sohn SH, Strong CD, Gralla EB, Whitelegge JP & Valentine JS (2008) *Proc Natl Acad Sci U S A.* **105**, 18663-18668.

157. Furukawa, Y Kaneko, K Yamanaka, K O'Halloran TV & Nukina N (2008) *J Biol Chem* **283**, 24167-24176.

158. Karch CM, Prudencio M, Winkler DD, Hart PJ & Borchelt DR (2009) *Proc Natl Acad Sci U S A* **(in press)**.

159. Moore, S. D. P., Chen, M. M. & Cox, D. W. (2000) *Cytogenetics and Cell Genetics* **88**, 35-37.

160. Bartnikas TB, Waggoner DJ, Casareno RLB, Gaedigk R, White RA & Gitlin JD (2000) *Mammalian Genome* **11**, 409-411.

161. Furukawa, Y. & O'Halloran, T. V. (2006) *Antioxidants & Redox Signaling* **8**, 847-867.

162. Lamb, A. L., Wernimont, A. K., Pufahl, R. A., Culotta, V. C., O'Halloran, T. V. & Rosenzweig, A. C. (1999) *Nature Struct. Biol.* **6**, 724-729.

163. Schmidt, P. J., Rae, T. D., Pufahl, R. A., Hamma, T., Strain, J., O'Halloran, T. V. & Culotta, V. C. (1999) *J. Biol. Chem.* **274**, 23719-23725.

164. Stasser JP, Siluvai GS, Barry AN & Blackburn NJ (2007) *Biochemistry* **46**, 11845-11856.

165. Lamb, A. L., Torres, A. S., O'Halloran, T. V. & Rosenzweig, A. C. (2001) *Nat Struct Biol* **8**, 751-755.

166. Doucette, P. A., Whitson, L. J., Cao, X., Schirf, V., Demeler, B., Valentine, J. S., Hansen, J. C. & Hart, P. J. (2004) *J. Biol. Chem.* **279**, 54558-54566.

Chapter 2

Materials and Methods

2.1 Genome browsing

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

A number of data banks are available and provide the scientific community with tools for searching gene banks, for the analysis of protein sequences, for the prediction of a variety of protein properties.

Primary databases contain information and annotations of DNA and protein sequences, DNA and protein structures and protein expression profiles. Some available databases for genome browsing are:

- **NCBI** (http://www.ncbi.nlm.nih.gov/sites/entrez/pubmed) - This web site integrates information from several databases (Swissprot, EMBL, all geneBank, etc...)
- **PDB** (http://www.rcsb.org/pdb/home/home.do) – A 3-D biological macromolecular structure database
- **Pfam** (http://pfam.sanger.ac.uk/) - A collection of different protein families organized in terms of different domains as obtained from multiple alignment.

Secondary or derived databases are so called because they contain the results of analysis of the primary resource including information on sequence patterns or motifs, variants and mutations and evolutionary relationships.

Once all biological data are stored, the requirement is to provide bioinformatics tools for extracting the meaningful information.

The most used programs for genome browsing are:

- **BLAST** (http://blast.ncbi.nlm.nih.gov/Blast.cgi): Standard BLAST (Basic Local Alignment Search Tool) is a set of similarity search programs designed to explore all of the available sequence databases regardless of whether the query is

protein or DNA. PHIBLAST is designed to search for proteins that contain a pattern specified by the user, and simultaneously are similar to the query sequence.

- **CLUSTALW** (http://www.ebi.ac.uk/Tools/clustalw2/index.html) is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylograms
- **PROSITE** (http://www.expasy.ch/prosite/) SCANPROSITE allows to scan a protein sequence, provided by the user, for the occurrence of patterns and profiles stored in the PROSITE database, or to search in protein databases all sequences with a user entered pattern.
- **STRING** (http://string.embl.de/) STRING is a database of predicted functional associations among genes/proteins. Genes of similar function tend to be maintained in close neighbourhood and tend to be present or absent together.

In our research group the attention has been focused on three main themes: (i) studying target proteins involved in rare diseases; (ii) finding new metal binding protein of unknown structure and function, with conserved consensus sequences for metal binding; (iii) discover in different organism metalloproteins homologues to the human ones. Once individuated a protein of particular interest, this is subjected to further bioinformatics investigations in order to predict important features like stability, solubility, hydrophobicity, secondary and tertiary structures.

2.2 Domain definition

Usually for each target protein several domains can be selected for expression, and of these domains various construct are cloned in order to increase the probability of obtaining a soluble and well-folded protein. Domain selection and construct design are performed considering several aspects.

The first important step is to perform the transmembrane regions prediction using dedicated bioinformatics programs. (e.g.TMHMM: www.cbs.dtu.dk/services/TMHMM-2.0) The result of this analysis make us know if we are dealing with a completely

soluble protein, with an integral transmembrane protein or with a protein mainly soluble with a small portion inserted into the phospholipidic bilayer. In the latter case it is possible to try to express a recombinant protein without the transmembrane tail. Using the program SIGNALP (www.cbs.dtu.dk/services/SignalP) it is possible to predict the presence and location of a signal-peptide sequence that enables the expression of the protein in the periplasmic area. Similarly using the program MITOPROT (http://ihg2.helmholtz-muenchen.de/ihg/mitoprot.html) it is possible to calculate the N-terminal protein region that can support a mitochondrial targeting sequence and the cleavage site.

Another important analysis to perform is the secondary structure prediction, e.g. exploiting the web server http://npsa-pbil.ibcp.fr/, in order to have an indication of the regions with tendency to form helices, beta-sheet, coil or turns. Combining these predictions with ClustalW alignments of the protein family, we can evaluate which are the domain borders, still maintaining the correct folding of the protein. It is also useful to search for the presence of intrinsically unstructured/disordered region using IUPred (http://iupred.enzim.hu/).

2.3 Cloning strategy

Proteins yield and solubility are highly dependent on the specific protein sequence, as well as on the vector, host cell, and culture conditions used. More is known about the characteristics of a protein, more easily it can be expressed, isolated and purified. For optimal efficiency, various combinations should be simultaneously screened, to determine the best conditions. The cloning strategy and the expression system are the first steps to be well designed since they will influence the expression protocol. The choice of the expression system depend on many factors, including cell growth characteristics, expression levels, intracellular and extracellular expression, posttranslational modifications, biological activity of the protein of interest (1, 2). For example, to express a protein of prokaryotic origin, the obvious choice is to use *E. coli* as host, but in case of eukaryotic proteins, different expression systems can be used and the choice will depend on many factors, since each system has its advantages and problems. Currently, many methodological improvement in non-prokaryotic hosts made more accessible and less expensive eukaryotic systems such as yeast, plants, filamentous fungi, insect or mammalian cells grown in culture and transgenic animals

(3-5). Also cell-free protein syntesis has a great potential for the expression of problematic proteins (6). However, especially for characterizations that require high amount of labelled samples, such as NMR, the prokaryotic and in particular the *E.coli* expression system is the most widely used. The choice of the expression vector depends on the expression system. For *E. coli*, a lot of expression plasmids are available for the screening of different expression conditions that can influence the yield of soluble recombinant protein. Plasmid vectors possess an origin of replication (ori), a gene for antibiotic resistance (usually Ampicillin), which allows for selection of cell clones carrying the plasmid, and a multicloning site, for the insertion of the target protein coding sequence. Classical cloning, using restriction enzymes, typically cannot be adapted to high-throughput approaches, due to the complication of selecting compatible and appropriate restriction enzymes for each cloning procedure and to its multistep process. High-throughput cloning therefore requires procedures which can help the screening of a broad range of conditions in less time, for these reasons new cloning technologies have been developed in recent years (**Fig. 1**).

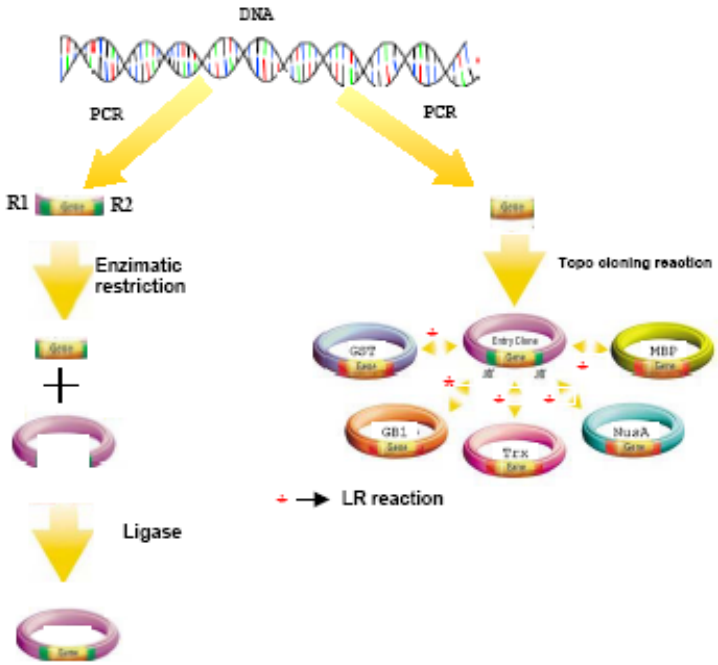


Fig. 1: *Schematic comparison between the Classic and the Gateway cloning technology. Theoretically parallel cloning is possible in both the cases but the Gateway system resulted faster due to the higher reaction efficiency and less sequencing demand since only the starting pENTR should be sequence verified.*

Landy and co-workers have found a universal cloning method based on the site-specific recombination (7). Gateway system (Invitrogen) is based on the bacteriophage lambda site specific recombination system which facilitates the integration of bacteriophage into the E. coli chromosome and the switch between the lytic and lysogenic pathways. Gateway system uses this machinery to clone a target gene into different expression vectors, without the time consuming reactions with restriction enzymes and ligase. PCR products made with primers containing terminal attB sites are substrates for the BP reaction. The result is an Entry Clone containing the PCR fragment(**Fig.2**). The entry vector can also be obtained via recombination reaction of the linearized pENTR/TOPO vector with our gene; the reaction is catalyzed by the topoisomerase I located on the pENTR vector. After isolation of the entry vector, the second step is to generate an expression construct by performing an LR recombination reaction between the entry clone and a Gateway destination vector of choice.

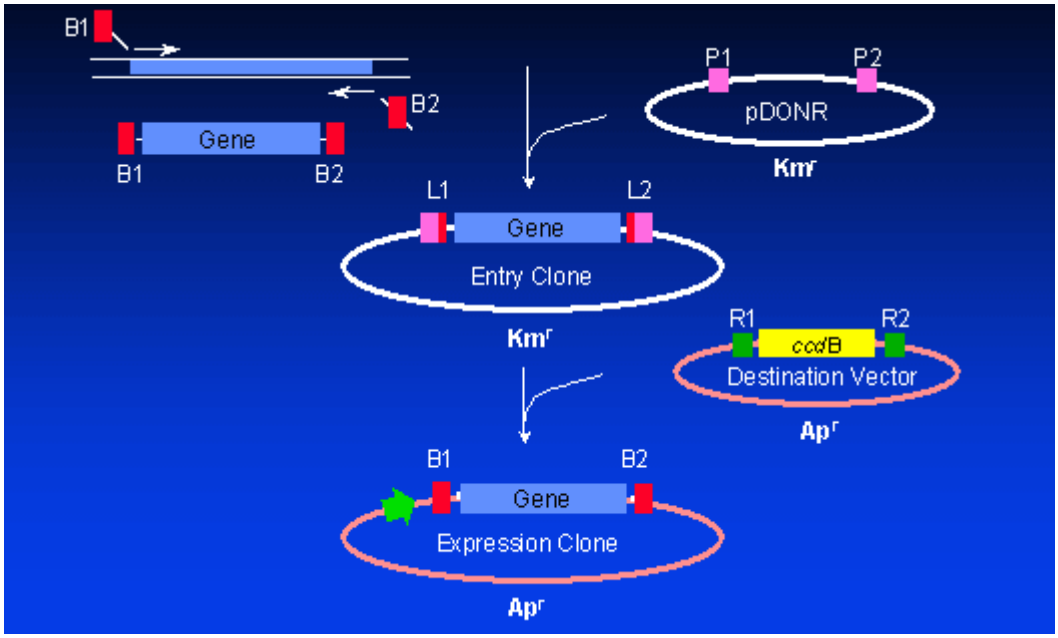


Fig. 2: *The Gateway cloning technology. PCR product can be transferred into an Entry Vector by the BP Reaction.Only plasmids without the ccdB gene that are also Kanamycin resistant (Kmr) will yield colonies. An Entry Clone, containing a gene flanked byrecombination sites, recombines with a Destination Vector by the LR Reaction to yield an Expression Clone and a by-product plasmid. The result is that a gene sequence in the Entry Clone is transferred into an Expression Vector with Ampicilline resistence (Apr), donated by the Destination Vector. The by-product plasmid contains the ccdB gene, and hence gives rise to no colonies when using standard strains of E. coli.*

Important elements of an expression plasmid that affect recombinant protein yield and solubility, are the promoter and the fusion tag. Many promoter are available for protein expression in *E. coli*, but only a few of them are commonly used. A useful promoter must be strong, tightly regulated to have a low basal expression level, the induction must be simple and cost-effective, and should be independent on the commonly used ingredients of growth media.

The most used promoter system for the bacterial expression of recombinant proteins is the T7/lac promoter (8). Genes under the control of T7/lac promoter can be transcribed by T7 RNA polymerase, in presence of lactose. Prokaryotic cells do not produce this type of RNA polymerase, and therefore only *E. coli* strains which has been genetically engineered to incorporate the gene for T7 RNA polymerase, the lac promoter and the lac operator in their genome, can be used for expression of target proteins. When lactose or lactose derivative or analogue to lactose, such as Isopropyl β-D-1-thiogalactopyranoside (IPTG), is added to the culture, it displaces the repressor from the lac operator. Since there are lac operators located upstream of the gene encoding the T7 RNA polymerase in the bacterial genome and of the gene encoding the target protein in the plasmid, IPTG activates both genes. The advantage of IPTG respect to lactose is that it is not a substrate of β-galactosidase. Therefore, since it can not be metabolized by *E. coli*, its concentration remains constant and the rate of expression of lac promoter controlled genes is not a variable during the experiment.

T7 RNA polymerase is so selective and active that, when fully induced, almost all of the cell's resources are converted to target gene expression and the desired product can comprise up to 50% of the total cell protein in few hours after induction. If the expression of the recombinant protein has to be reduced, as in case of toxic or membrane proteins, host strains containing the pLysS or pLysE vectors can be used. These vectors express the T7 lysozyme, a natural inhibitor of T7 RNA polymerase. The target protein can be expressed with different fusion partners, which has been developed to increase the expression yield and the solubility of the recombinant proteins, even if sometimes the expression of the native protein could be the best choice(2). Fusion tags can also be exploited to facilitate the purification procedures.

For example the ATP7B full leght (residues1- 633), ATP7B1-2 (residues 1-157) and the CCS1-2 (residues 1-238) entry vector are obtained via recombination reaction of the linearized pENTR/TOPO vector with the genes and then subcloned by LR reaction in

pETG-20A (His_6 tag) and pTH34 (GB1 tag) respectively. Instead SURF1 (reisdues 1-256), COX11 soluble part (reisdues 122-276) and CCS full length (reisdues 1-274) genes were cloned in pDONR 221 by BP reaction and then subcloned in pTH27 (COX11 and CCS full length) and pTH24 (SURF1) by LR reaction.

2.4 Site-directed mutagenesis

The functional analysis of a protein often requires the generation of a number of single or multiple amino acid variants of the wild type gene. Site-directed mutagenesis is a technique for carrying out vector modification and characterizing the dynamic, complex relationships between protein structure and function. The basic procedure utilizes a supercoiled double-stranded DNA vector with insert the gene of interest and two synthetic oligonucleotide primers, both containing the desired mutation. The primers, each complementary to opposite strands of the vector, are extended during PCR reaction by a high fidelity DNA polymerase (PfuUltra, Stratagene). Extension of the oligonucleotide primers generates a mutated plasmid. After PCR reaction the product is treated with an endonuclease specific for methylated and hemimethylated DNA, DpnI. This enzyme is used to digest the parental DNA template and thus select only the mutation-containing synthesized DNA; this happens because DNA isolated from almost all *E. coli* strains is dam methylated and therefore susceptible to endonuclease digestion. The vector containing the desired mutations is then transformed into XL1-Blue supercompetent cells and subsequently subjected to sequencing analysis.
To obtain Cystein 1,2,4 mutant of ATP7B protein I designed primers with both Cysteines of the conserved motif CxxC mutated in Alanines. The three mutagenesis reactions were performed one after the other, using as starting template the plasmid encoding for wild type ATP7B.

2.5 Protein expression

2.5.1 Expression systems

At the moment, several host systems are available for protein expression, including bacteria, yeast, plants, fungi, insect and mammalian cells. The choice is strongly

dependent upon the specific requirements for the final product, since it will affect not only the protein expression, but also the subsequent purification.

Often the first choice is the *E. coli* system, since it is the easiest, quickest and cheapest one. There are many commercial and non-commercial expression vectors available with different N- and C-terminal tags and many different bacterial strains which are optimized for special applications. Moreover one of the major advantages of *E. coli* is that the production of isotope labelled protein for NMR is very easy in compare to other systems. However, *E. coli* is a prokaryote and lacks intracellular organelles, such as the endoplasmic reticulum and the Golgi apparatus that are present in eukaryotes, which are responsible for post-translation modifications of the proteins being produced. Many eukaryotic proteins can be produced in *E. coli* but sometimes they are produced like a non-functional inclusion bodies protein, since glycosylation or post-translational modifications such as proteolytic processing, folding and disulfide bond formation do not occur. Therefore, researchers have recently turned to eukaryotic yeast and insect cell expression systems for protein production.

Yeast is a eukaryotic organism and has some advantages and disadvantages over *E. coli*. The major advantages are that yeast cultures can be grown to very high densities, which makes them especially useful for the production of isotope labelled protein for NMR, and the possibility to introduce some post-translational modification on the expressed protein (e.g. glycosylation). The main disadvantages are the higher prices of reagents required for isotope labelling and the very long time of growth (4-5 days instead of 1). The two most used yeast strains are *Saccharomyces cerevisiae* and the methanoltrophic yeast *Pichia pastoris*.

Insect cells are a higher eukaryotic system than yeast and are able to carry out more complex post-translational modifications than the other two systems. They also have a better machinery for the folding of mammalian proteins and, therefore, give you the best chance of obtaining soluble protein when you want to express a protein of mammalian origin. The disadvantages of insect cells are the higher costs and the longer duration before you get protein (usually 2 weeks).

All proteins studied in this project were expressed in *E. coli* cells. ATP7B was expressed in Rosetta (DE3) pLysS cells, while Surf1 and COX11 were expressed in C41(DE3) cells. CCS full length construct expression was detect in BL21(DE3) Codon Plus, while the short construct was expressed in Origami pLysS cells.

2.5.2 Expression in *E.coli*

In this section the methods used in our laboratory for protein expression optimization in *E. coli* are described. Other expression systems are not considered in detail since they have not been used in this thesis work.

As like gene cloning, also for protein expression the best solution is the use of a parallel expression strategy; depending on the needed production throughput the two possible choices are: (i) to express more genes in the same bacterial strain or (ii) to express a single gene in different strains. In our lab the second approach is more often used.

A preliminary expression test is performed in a small-volume scale using different strains, three expression temperatures (37-25-18°C), three inducer concentrations and different induction times (4h and 16h). Expression results are checked on Dot Blot in case of presence of an His-tag and on SDS polyacrilamide gel (SDS-PAGE). This kind of approach allows us to explore a large set of expression conditions and to evaluate which one gives the best yield of soluble protein. A second expression test is sometimes performed to better refine the expression conditions before starting a massive culture that is performed in flasks or fermentors to increase protein yield and have a better parameter control (i.e. pH, temperature, pO_2).

In case of negative or partially positive results some or all variables can be modified, in particular the more influencing ones like the choice of bacterial strain(9) and the kind of vectors. The majority of the target proteins are produced in the soluble fraction, but in case it is detected in the insoluble fraction, a possible approach is to try an in-vitro refolding screening with different additives in order to get a folded and soluble protein. The last choices are to redesign the expressed domains or to switch to other expression system.

Obviously when dealing with small bacterial proteins this procedure may be simplified, since the targets do not require a lot of efforts to be expressed in a soluble form, whereas bigger human proteins create more problems.

In dependence of the spectroscopic technique of choice, protein expression is performed in differently composed media. In fact, when large amounts of proteins must be isolated for techniques that do not require isotopic labelling, the culture is usually performed in a so-called rich or complex medium. Complex media contain water soluble extracts of plant or animal tissue (e.g., enzymatically digested animal proteins such as peptone and tryptone), and for this reason are rich in nutrients and minerals, assuring a fast bacterial

growth and a high expression level. Their exact composition is unknown and this can impair the reproducibility of cultures. Chemically defined (or minimal) media are composed of pure ingredients in measured concentrations, dissolved in milliQ water; this way the exact chemical composition of the medium is known, allowing high reproducibility of protein yields and grade and type of interferents. Typically, this class of media is composed of a buffering agent to maintain culture pH around physiological values, a carbon and energy source like a simple sugar (glucose) or glycerol, and an inorganic nitrogen source, usually an ammonium inorganic salt. In dependence of the bacterial strain and of the expressed proteins various mineral salts can be added and, if necessary, growth factors such as purified amino acids, vitamins, purines and pyrimidines. Chemically defined media are easier to isotopically enrich, simply using ^{15}N and ^{13}C enriched nitrogen and carbon sources in its composition, even if isotopically enriched complex media are also commercially available.

Uniform labelling of proteins is particularly important when dealing with 2D and 3D NMR. ^{15}N is very convenient because of the strategic locations of nitrogens in the backbone, and the absence of homonuclear couplings due to the intervening carbons. The replacement of all the carbons with ^{13}C is also possible, although in this case the spectroscopy has to deal with the network of couplings among bonded carbons. Partial or total deuteration (perdeuteration) attenuates the dipolar couplings among protons, reducing the relaxation rates of NMR-active nuclei, since the gyromagnetic ratio of ^{2}H is 6.5 times smaller than ^{1}H. This last kind of enrichment is the most difficult to achieve since incorporation of ^{2}H reduces the growth rate of organisms up to 50%, and decreases protein production as a consequence of the isotopic effect. For this reason, deuteration requires a gradual adaptation of bacteria to increasing concentration of D_2O.

2.6 Protein purification

The strategy of purification depends mainly upon the location of the expressed protein within the host; for example the protein can be secreted into the growth media, transported in the periplasmic space, located in the membrane or expressed like a soluble or insoluble (inclusion bodies) protein within the cytoplasm. In each of the cases the isolation has to be performed in different ways.

The majority of the proteins are usually expressed as soluble, while in few cases are expressed as insoluble proteins. Indeed, the most of the bacterial proteins are removed by different extraction steps with native buffer conditions, while the recombinant proteins are extracted from inclusion bodies with a denaturing buffer. Physical-chemical properties of the recombinant protein also drive the choice of purification protocols, thanks to peculiar properties of the recombinant protein.

Cell lysis is the first step in protein purification and the technique chosen for the disruption of cells must be compatible with the amount of material to be processed and the intended downstream applications. Many techniques are available, spanning from physical to detergent-based methods. Among the first class are mechanical disruption**,** liquid-based homogenization, such as the French press, where cells are lysed by forcing the suspension through a narrow space, and sonication, which uses pulsed, high frequency sound waves to agitate and lyse cells. Gentler ways of disrupting cells are freeze-thaw lysis, detergent lysis, enzymatic lysis and osmotic lysis.

The following step of purification procedure is composed of normally two or more rounds of chromatography, playing on the different physical chemical and biological characteristics of the protein.

In case of peptide fusion tags, a variety of affinity purification methods are available. The basic principle is that the protein tag specifically interacts with an affinity resin and binds the tagged protein whereas other polypeptides are not bound. This allows a simple purification by releasing the tagged protein with a suitable competing buffer. In this case, usually an intermediate purification step includes the cleavage of the fusion tag with proteases such as Enterokinase, Factor Xa, Thrombin or Tobacco Etch Virus (TEV) in dependence of the cleavage site. The protease recognition site is selected and cloned in the vector codifying for the protein sequence at the cloning step; pilot experiments should be done to find out suitable cleavage conditions. The fusion protein is then separated, either with a second round of affinity purification, or by size exclusion chromatography provided that the difference in size with the target polypeptide is large enough. There are a lot of different affinity tags and some of them, like Maltose Binding Protein (MBP) or Glutathione-S-transferase (GST), have also the role of facilitating soluble expression(10). Currently Immobilized Metal Ion Affinity Chromatography (IMAC) is the most used affinity technique; it exploits the interaction between chelated transition metal ions (generally Zn^{2+} or Ni^{2+}) and side-chains of specific amino acids (mainly histidine) added on the protein. In IMAC chromatography,

the target protein is usually washed from the impurities and eluted using increasing concentration of imidazole, which acts like a competitive agent. Histidine-tagged (His_6 tag) proteins can also be eluted from the column lowering the pH to decrease the binding affinity or stripping the metal ions from the column with EDTA. His_6 tag is removed as described. The procedure is illustrated in **Fig. 3**.

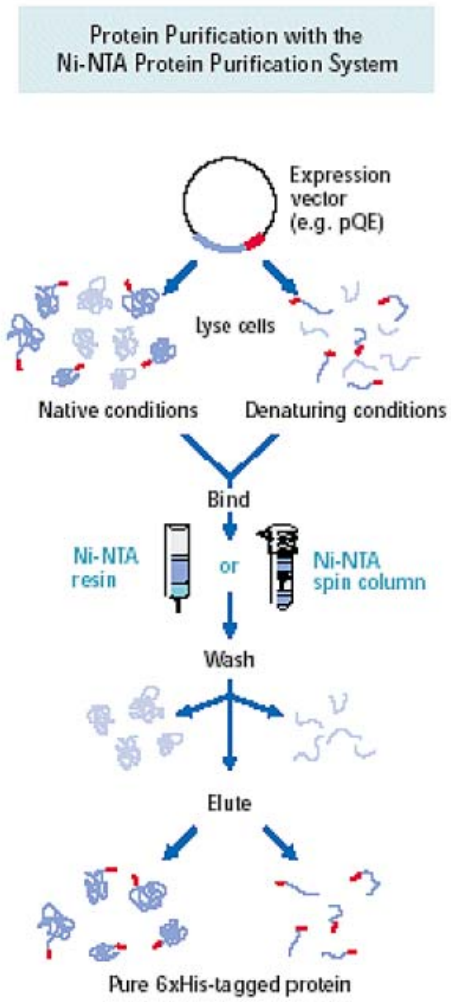


Fig. 3: *General purfication procedure for His-tag proteins.*

Gel filtration is the simplest and mildest of all the chromatographic techniques; the support for gel filtration chromatography is composed of beads which contain holes, called "pores," of given sizes. Larger molecules, which can't penetrate the pores, move around the beads and migrate through the spaces which separate the beads faster than the smaller molecules, which may penetrate the pores; in this technique the buffer composition does not play any role in the resolution. Size exclusion can be performed on a “rough” level, separating the components of a sample in major groups to remove

for example high or low molecular weight contaminants or to exchange buffers, while high resolution fractionation of biomolecules allows to isolate one or more components of a protein mixture, to separate monomers from aggregates and last but not least to determine molecular weights or to perform a molecular weight distribution analysis, provided that suitable standards are available.

For ATP7B, SURF1 and CCS proteins, all of them fused with His_6 tag, HisTrap chelating column was used for purification, while for COX11 protein gel filtration was performed after refolding.

2.6.1 Membrane protein purification

A particular protocol must be used for membrane proteins. In contrast to soluble proteins, the isolation of a peripheral or integral membrane protein requires extraction of the biological membrane containing the protein of interest prior to purification. Extraction of integral membrane proteins is commonly accomplished by solubilizing the protein-containing membrane fraction using a variety of detergents which are necessary to disrupts the lipid bilayer and brings the proteins into solution as protein-lipid-detergent complexes. **(Fig.4).**

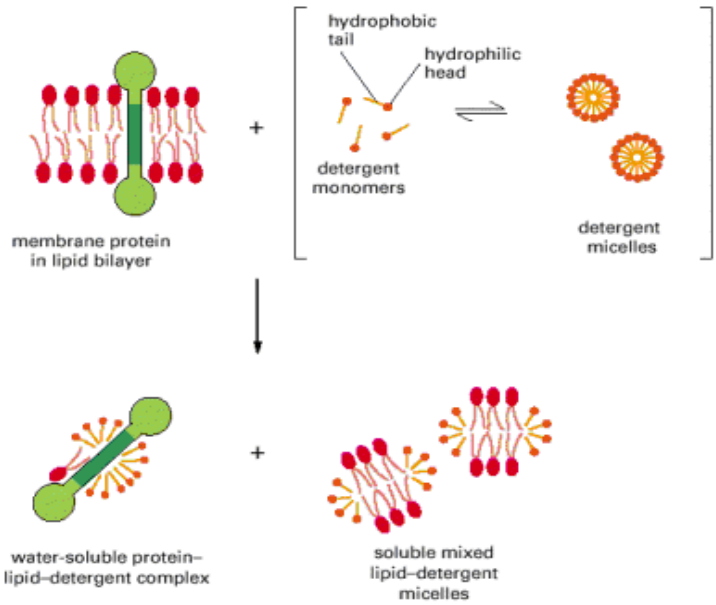


Fig. 4: *Solubilization of membrane protein by detergent with formation of a water-soluble protein-lipid-detergent complex.*

Detergents are amphiphilic molecules that mimic the lipid environment, which surrounds the protein in the membrane and keep the protein soluble in acqueous buffers. These molecules are composed of an hydrophilic head and an hydrophobic tail. There are several different types of detergents and are divided into groups according to the chemical property of the head part: ionic, non-ionic or zwitterionic. The head part of the ionic detergents can be either positive or negative, while zwitterionic detergents have both a positively and a negatively charged head group. One of the most harsh detergent is Fos-choline 12 (FC12) which is usually used to verify if a membrane protein is able to be solubized. For this reason FC12 was used to solubilised SURF1 protein from the membrane.

Following initial solubilization in high concentrations of detergent, intrinsic proteins can then be purified to homogeneity using a variety of biochemical techniques in the presence of relatively low concentrations of detergent. Solubilized membrane proteins from diverse sources and with different properties have successfully been purified by a combination of standard techniques such as gel filtration and different forms of the affinity chromatography method.

2.6.2 Refolding

Inclusion body formation is often an undesired event and a number of approaches can be utilized to obtain soluble protein. On the other hand, inclusion body formation does not necessarily have to be a dead-end process. Sometimes harvesting of proteins from inclusion bodies can prove to be very fruitful and may present a valid option for successful purification. New strains of *E.coli* and fusion partners are available, which offer a range of options to minimize, or maximize, inclusion body formation. Inclusion bodies are dense amorphous aggregates of misfolded protein. They are generally formed if a protein has a high propensity to misfold and aggregate, or if the cellular protein production machinery is unable to operate efficiently. The formation of inclusion bodies can sometimes be especially advantageous if the protein is toxic to the cell or susceptible to proteolytic attack. Successful renaturation depends on the purity of inclusion bodies, as contaminants can decrease the efficiency of renaturation, presumably as such contaminants can promote coaggregation. Inclusion bodies are therefore best washed and centrifuged a several times in buffer containing detergents such as Triton X-100. Following washing, inclusion bodies are then solubilized, usually

with GdnHCl or urea. It is important that during unfolding any disulfide bonds present are also reduced – this is usually achieved with β- mercaptoethanol or DTT. Upon solubilization of inclusion bodies, ample incubation time should be allowed for complete unfolding to occur. The renaturation process aims to effectively remove the denaturant and thiol reagents and allow the protein to refold. The refolding process is a competing reaction with misfolding and aggregation events and its success depends on a number of factors. Several different methods can be used to refold proteins. These include dialysis, dilution and column chromatography techniques. The method selected depends on the propensity of the protein to aggregate and the kinetics of refolding.

Refolding by dilution can be described as 'rapid' or 'slow'. In rapid dilution, the denatured protein is delivered to the refolding buffer such that in a very short time period, the concentrations of both the protein and denaturant decrease rapidly. The aim of dilution is to remove the denaturant, so its final concentration should be low. Rapid dilution has been used successfully for a range of proteins, however, the time period used dilution may be detrimental for some proteins, especially if they usually refold over a relatively long time period of minutes to hours. Forcing the protein to adopt its conformation in a short time frame may increase its chances of misfolding and aggregation. Slow dilution is an alternative method, as it is a more gentle approach and dramatically decreases the effective concentration of the refolding protein.

In refolding by dialysis, which is the protocol that was used to refold COX11 protein in my PhD project, the concentrated denatured protein is dialyzed against a refolding buffer, such that the concentration of denaturant decreases with buffer exchange. This slow removal of denaturant allows refolding to occur. Unfortunately, the slow removal of denaturant often results in the formation and exposure of long lived intermediate species over a time period and hence there may be an increased propensity for the protein to aggregate. A variation of one-step dialysis, where protein is dialyzed against refolding buffer containing no denaturant, is the use of a series of refolding buffers with decreasing denaturant concentrations.

On-column refolding has been applied to several proteins with success and offers an alternative where other methods may not be fruitful. If a protein has a His_6 tag, it can be immobilized on a Ni2+ affinity column in its denatured state. By applying buffers with a decreasing concentration of denaturant, either step-wise or by using a gradient, the protein can be refolded and then eluted(11).

2.7 Protein characterization

2.7.1 Mass Spectrometry

Mass spectrometry is an analytical tool used for measuring the molecular mass of a sample. For large samples such as biomolecules, molecular masses can be measured to within an accuracy of 0.01% of the total molecular mass of the sample. Mass Spectrometry can be used for different types of characterization:

− Accurate molecular weight measurements, to determine the purity of a sample, to verify amino acid substitutions, to detect post-translational modifications, to calculate the number of disulphide bridges;

− Reaction monitoring: monitor enzyme reactions, chemical modification, protein digestion;

− Amino acid sequencing: sequence confirmation, characterisation of peptides, identification of proteins by database from proteolytic fragmentation;

− Protein structure: protein folding monitored by H/D exchange, protein-ligand complex formation under physiological conditions, macromolecular structure determination.

Mass spectrometers can be divided into three fundamental parts: the ionisation source , the analyser and the detector. Sample molecules are ionised in the ionisation source, these ions are extracted into the analyser region of the mass spectrometer where they are separated according to their mass (m) -to-charge (z) ratios (m/z) . The separated ions are detected and this signal sent to a data system where the m/z ratios are stored together with their relative abundance for presentation in the format of a m/z spectrum.

Many ionisation methods are available, each with its own advantages and disadvantages, and the choice depend on the type of sample under investigation and the mass spectrometer available. The ionisation methods used for the majority of biochemical analyses are Electrospray Ionisation (ESI) and Matrix Assisted Laser Desorption Ionisation (MALDI). The main function of the mass analyser is to separate, or resolve, the ions formed in the ionisation source of the mass spectrometer according to their mass-to-charge (m/z) ratios. There are a number of mass analysers currently available, the better known of which include quadrupoles, time-of-flight (TOF) analysers, magnetic sectors, and both Fourier transform and quadrupole ion traps. These mass analysers have different features, including the m/z range that can be

covered, the mass accuracy, and the achievable resolution. The detector monitors the ion current, amplifies it and the signal is then transmitted to the data system where it is recorded as mass spectra, where the m/z values of the ions are plotted against their intensities to show the number of components in the sample, the molecular mass of each component, and the relative abundance of the various components in the sample.

MALDI is widely used in biochemical areas for the analysis of proteins, peptides, glycoproteins, oligosaccharides, and oligonucleotides and usually is performed in denaturing conditions. MALDI is based on the bombardment of sample molecules with a laser light to bring about sample ionisation. The sample is pre-mixed with a highly absorbing matrix compound which transforms the laser energy into excitation energy for the sample, which leads to sputtering of analyte and matrix ions from the surface of the mixture. The time-of-flight (TOF) analyser separates ions according to their mass(m)-to-charge(z) (m/z) ratios, by measuring the time it takes for ions to travel through a field free region known as the flight tube. The heavier ions are slower than the lighter ones. In negative ionisation mode, the deprotonated molecular ions (M-H-) are usually the most abundant species, accompanied by some salt adducts and possibly traces of dimeric or doubly charged materials. Negative ionisation can be used for the analysis of oligonucleotides and oligosaccharides. In positive ionisation mode, the protonated molecular ions (M+H+) are usually the dominant species, although they can be accompanied by salt adducts, a trace of the doubly charged molecular ion at approximately half the m/z value, and/or a trace of a dimeric species at approximately twice the m/z value. Positive ionisation is used in general for protein and peptide analyses.

2.7.2 Circular Dichroism

Circular Dichroism (CD) is an excellent method to analyze protein and nucleic acid secondary structure in solution and it can be used to follow the changes in folding as a function of temperature or denaturant. CD is a phenomenon occurring when asymmetrical molecules interact with circularly polarized light, thus absorbing left and right hand circularly polarized light with different absorption coefficients. In proteins the major optically active groups are the amide bonds of the peptide backbone, typically disposed in highly ordered arrays such as α-helices or β-pleated sheets. In dependence of the orientation of the peptide bonds in the arrays, given by the symmetry of its

disposition, optical transitions are differently split by exciton splitting, thus yielding characteristic spectral profiles for each of the three basic secondary structures of a polypeptide chain. A protein consisting of these elements will therefore display a spectrum that can be deconvoluted into the tree individual contributions; for this purpose several mathematical methods have been developed, all of them relying on the assumption that the spectrum of a protein can be represented by a linear combination of the spectra of its secondary structural elements, plus a noise term which includes the contribution of aromatic chromophores.

Circular dichroism spectroscopy is particularly good to:

· determine whether a protein is folded, and characterize its secondary structure content.

· compare the secondary structure elements content of proteins obtained from different sources (e.g. species or expression systems) or comparing structures for different mutants of the same protein.

· study the conformational stability of a protein upon the change of various condition (thermal stability, pH stability, and stability to denaturants) and how this stability is altered by buffer composition or addition of stabilizers.

· determine whether protein-protein interactions alter the conformation of protein.

2.7.3 Light Scattering

Static and dynamic light scattering represent an approach to studying protein complexes and oligomerization. In static light scattering, the scattering intensity is related to the molecular weight of the protein, in addition to its concentration, the scattering angle, and the wavelength. Dynamic light scattering is based on the auto-correlation of the time dependent fluctuations of scattered light intensity, which in turn depends upon the diffusion constant. This auto-correlation decays more slowly for slowly diffusing particles and thus, the diffusion constant is extracted from the value of the relaxation time of this function. In the case of ideal spherical particles, this provides a measure of the molecular weight. Light scattering is limited principally by sensitivity, with best results around 1 mg/ml, depending upon the size of the protein or the complex. While sensitivity limits preclude the determination of affinities and association or dissociation rate constants in most cases, light scattering is quite useful in characterizing the stoichiometry of complexes at high concentration.

2.8 Structural characterization

In order to understand completely the function of a protein the determination of the 3D structure is of basic importance. X-ray crystallography and NMR spectroscopy are the two main techniques that can provide structures of macromolecules at atomic resolution. Both techniques are well established and play a key role in structural biology as a basis for a detailed understanding of molecular functions. Their respective different advantages and disadvantages in terms of sample preparation and data collection and analysis make these two approaches complementary in Structural biology. Most (75%) of the NMR structures in the Protein Data Bank (PDB) do not have corresponding crystal structures, and many of these simply do not provide diffraction quality crystals. On the other hand structure determination of big proteins (more than 50 kDa) or of membrane protein is affordable only by X-ray spectroscopy.

2.8.1 Structural determination by NMR spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography can provide high-resolution structures of biological molecules such as proteins and nucleic acids and their complexes at atomic resolution. NMR can study molecules in solution, therefore, crystallization is not required, and crystal packing affects may not influence the structure, especially on the surface of a protein. Solution studies should be closer to native-like conditions found in the cell. Since crystals are not needed, protein folding studies can be monitored by NMR spectroscopy upon folding or denaturing of a protein in real time. More importantly, denatured states of a biomolecule, folding intermediates and even transition states can be characterized using NMR methods. NMR provides information about conformational or chemical exchanges, internal mobility and dynamics at timescales ranging from picoseconds to seconds, and is very efficient in determining ligand binding, and mapping interaction surfaces of protein/protein, protein/nucleic acid, protein/ligand or nucleic acid/ligand complexes and intramolecular interactions. Improvements in NMR hardware (magnetic field strength, cryoprobes) and NMR methodology, combined with the availability of molecular biology and biochemical methods for preparation and isotope labeling of recombinant proteins, have dramatically increased the use of NMR for the characterization of structure and dynamics of biological molecules in solution. Protein isotope labelling is necessary for

NMR analysis because not all atoms are magnetically active. The nuclei of naturally occurring atomic isotopes that constitute biological molecules have a nuclear spin determined by the spin quantum number (I) and its value depends on the composition of neutrons and protons in each nucleus. Because of the positive charge possessed, the nucleus rotation around its own axis generate a magnetic moment (m). Only atoms with $I \neq 0$ can be observed by NMR spectroscopy. For example, 12C is the most abundant isotopes in nature, but it has I=0, for this reason labelling with 13C is performed for biomolecules studied by NMR spectroscopy. Nuclei which have a spin of one-half, like 1H, or 13C, have two possible spin states: $m = ½$ or $m = -½$ (also referred to as α and β, respectively). The energies of these states are degenerated, hence the populations of the two states will be approximately equal at equilibrium. If a nucleus is placed in a magnetic field, there is interaction between the nuclear magnetic moment and the external magnetic field, then the nuclear spin state aligned with the external magnetic field will be more populated and the different states will have different energies. When a radiofrequency pulse is applied to match this energy difference, resonance absorption will occur and all nuclei of the same element would resonate at the same frequency. The resonance frequency is affected by the chemical environment of each nucleus, hence nuclei of the same element will have differences in the resonance frequency due to the interference of surrounding electrons, that decrease the magnitude of the effective magnetic field on the nucleus. These differences are called chemical shifts and are higher for more shielded nuclei. The chemical shift value depends also on the applied magnetic field and in order to have chemical shift values normalised on the static magnetic field strength, they are measured in parts per million (ppm).

In a protein, the resonance frequencies of each nucleus vary slightly due to chemical shifts, then a very short radiofrequency pulse is applied which inherently encodes a range of frequencies allowing to induce resonance for the whole frequency spectrum in one experiment (Fourier transform, FT NMR). Transient signals are detected as the system returns to equilibrium. The response obtained from a FT NMR experiment is a superposition of the frequencies of all spins in the molecule as a function of time. In order to obtain the corresponding spectrum as a function of frequency, a Fourier transformation is performed. Fourier transformation is a mathematical operation which translates a function in the time domain into the frequency domain. In large molecules, such as proteins, the number of resonances can be several thousand and a one-dimensional spectrum inevitably has overlaps. For this reason, proteins NMR spectra

cannot be resolved in a conventional one-dimensional spectra (1D) and multi-dimensional nuclear magnetic resonance spectroscopy is required to correlate the frequencies of different nuclei. There are different types of experiments that can detect through-bonds and through-space nucleusnucleus interactions.

The Heteronuclear Single Quantum Correlation (HSQC) spectrum is a 2D spectrum, where "heteronuclear" refers to nuclei other than ^{1}H. In theory the heteronuclear single quantum correlation has one peak for each H bound to a heteronucleus. Thus the ^{1}H ^{15}N-HSQC spectrum contains the signals of the HN protons in the protein backbone. Since there is only one backbone HN per amino acid, each HSQC signal represents one single amino acid, with the exception of proline, which has no amide-hydrogen due to the cyclic nature of its backbone. Moreover, this HSQC also contains signals from the NH_2 groups of the side chains of Asn and Gln and of the aromatic HN protons of Trp and His. The acquisition of NMR signals is performed during the relaxation process, which restore the equilibrium of the system. There are two types of relaxation:

− transverse relaxation, due to the interaction between different spins, and measured by the T2 time, which is directly proportional to the molecular weight;

− longitudinal relaxation, due to the interaction between spins and solvent molecules, and is measured by the T1 time.

The main problem in studies of biomolecules with molecular weights above 30 kDa is the fast decay of the NMR signal due to relaxation. Indeed, the line widths in the NMR spectra are inverse proportional to the relaxation rates. Therefore the signal-to-noise in NMR spectra of larger molecules is poor, due to poor resolution and sensitivity. There are NMR methods that can help the acquisition of spectra of large biomolecules, one of them is Transverse Relaxation Optimized Spectroscopy (TROSY). With these improvements high-resolution TROSY-HSQC spectra can be recorded of macromolecules with molecular weights up to several 100 000 Daltons. The exchange between two conformations, e.g. free and ligand bound forms of a protein, but also chemical exchange usually gives rise to two distinct NMR signals for a given spin due to different chemical environments in the two exchanging forms. If the exchange rate is slow on the chemical shift time scale, two sets of signals are observed, if the exchange rate is fast on the chemical shift time scale only one signal is observed at an average frequency corresponding to the populations of the two conformations. Intermediate exchange gives rise to very large line width.

Molecular interactions can be very efficiently characterized using very sensitive NMR experiments. Changes in the environment of a spin due to binding of a ligand give rise to chemical shift changes in the NMR spectrum. These changes are usually largest near the binding site. Therefore, the binding surface of a protein with a ligand can be mapped. In addition, from NMR titration experiments dissociation constants can be determined. Due to the relatively high sample concentration, even very weak interactions can be detected. Additional structural information, and long range interactions, as the relative orientation of two protein domains can be measured by the observation of Residual Dipolar Couplings (RDCs). Dipolar coupled spins are the result of spin/spin interactions through space and depend on the distance between the two spins and the orientation of the internuclear vector with respect to the static magnetic field. The chemical shift difference between 1H-15N (1H-15N J-coupling constant), is different in isotropic or anisotropic conditions.

2.8.2 X-ray and protein crystallization

X-ray crystallography is another technique used to characterize proteins and is essentially a form of very high resolution microscopy. It enables to visualize protein structures at the atomic level and enhances our understanding of protein function. With this technique it is possible to study how proteins interact with other molecules, how they undergo conformational changes, and how they perform catalysis in the case of enzymes. In all forms of microscopy, the amount of detail or the resolution is limited by the wavelength of the electro-magnetic radiation used. In order to see proteins in atomic detail we need to work with electro-magnetic radiation with a wavelength of around 0.1 nm or 1 Å.

The diffraction from a single molecule would be too weak to be measurable. So it is necessary to use an ordered three-dimensional array of molecules, in other words a crystal, to magnify the signal. Even a small protein crystal might contain a billion molecules. If the crystal is well ordered, then diffraction will be measurable at high angles or high resolution and a detailed structure should result. The X-rays are diffracted by the electrons in the structure and consequently the result of an X-ray experiment is a 3-dimensional map showing the distribution of electrons in the structure.

In order to crystallize a protein, the purified protein undergoes slow precipitation from an aqueous solution. As result, individual protein molecules align themselves in a repeating series of "unit cells" by adopting a consistent orientation. The production of good crystals is dependent upon a number of environmental factors, especially protein purity and pH conditions are very important.

In the past few years macromolecular crystallography has become a standard technique used by many pharmaceutical and biotechnology companies. This methodology offers details of protein-ligand interactions at levels of resolution virtually unmatched by any other technique, and this approach holds the promise of novel, more effective, safer and cheaper drugs. Although crystallography remains a laborious and rather expensive technique, remarkable advances in structure determination and structure based drug design (SBDD) have been made in recent years.

Reference List

1. Yokoyama, S. (2003) *Curr. Opin. Chem. Biol.* **7**, 39-43.

2. Terpe, K. (2006) *Applied Microbiology and Biotechnology* **72**, 211-222.

3. Junge, F., Schneider, B., Reckel, S., Schwarz, D., Dotsch, V. & Bernhard, F. (2008) *Cellular and Molecular Life Sciences* **65**, 1729-1755.

4. Lico, C., Chen, Q. & Santi, L. (2008) *Journal of Cellular Physiology* **216**, 366-377.

5. Condreay, J. P. & Kost, T. A. (2007) *Current Drug Targets* **8**, 1126-1131.

6. Klammt, C., Schwarz, D., Lohr, F., Schneider, B., Dotsch, V. & Bernhard, F. (2006) *Febs Journal* **273**, 4141-4153.

7. Landy, A. Dynamic, Structural, and Regulatory Aspects of Lambda Site-specific Recombination. Ann.Rev.Biochem. 58, 913-949. 1989.
   Ref Type: Generic

8. Dubendorff, J. W. & Studier, F. W. (1991) *J. Mol. Biol.* **219**, 45-59.

9. (1996) *J. Mol. Biol.* **260**, 298.

10. Kapust, R. B. & Waugh, D. S. (1999) *Protein Sci.* **8**, 1668-1674.

11. D'Alatri, L., Di Massimo, A. M., Anastasi, A. M., Pacilli, A., Novelli, S., Saccinto, M. P., De Santis, R., Mele, A. & Parente, D. (1998) *Anticancer Research* **18**, 3369-3373.

Chapter 3

Wilson Disease

Expression and characterization of

human Wilson protein

AnNMRStudy of the Interaction of the N-terminal Cytoplasmic Tail of the Wilson Disease Protein with Copper(I)-HAH1

Lucia Banci[‡§¶], Ivano Bertini[‡§1], Francesca Cantini[‡§], **Chiara Massagni**[‡¶], Manuele Migliardi[‡], and Antonio Rosato[‡§]

From the [‡]Magnetic Resonance Center, University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy, the [§]Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy, and the [¶]FIORGEN Foundation,Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 284, NO. 14, pp. 9354–9360, April 3, 2009

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 284, NO. 14, pp. 9354–9360, April 3, 2009

An NMR Study of the Interaction of the N-terminal Cytoplasmic Tail of the Wilson Disease Protein with Copper(I)-HAH1*[S]

Received for publication, August 4, 2008, and in revised form, January 30, 2009 Published, JBC Papers in Press, January 30, 2009, DOI 10.1074/jbc.M805981200

Lucia Banci[‡§¶], **Ivano Bertini**[‡§1], **Francesca Cantini**[‡§], **Chiara Massagni**[‡¶], **Manuele Migliardi**[‡], **and Antonio Rosato**[‡§]

From the [‡]*Magnetic Resonance Center, University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy, the* [§]*Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy, and the* [¶]*FIORGEN Foundation, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy*

ATP7B is a human P_{1B}-type ATPase that has a crucial role in maintaining copper(I) homeostasis. Mutations in the corresponding gene are the cause of Wilson disease. Among its various distinguishing features is a long (~630 amino acids) N-terminal cytosolic tail containing six domains that are individually folded and capable of binding one copper(I) ion each. We expressed the entire tail as a single construct in *Escherichia coli* and investigated its interaction with its copper chaperone (*i.e.* HAH1) by solution NMR spectroscopy. We observed that all six of the metal-binding domains were metallated by Cu(I)-HAH1, with the first, the second, and the fourth domains forming an adduct with it. This behavior is different from that of the highly similar human ATPase ATP7A, in which only two domains form such an adduct. The distinct behaviors of the different domains were analyzed in terms of the energetics of Cu(I) transfer, hinting at a specific role of the interaction with copper(I)-HAH1 in the overall functional process.

Human ATP7B is a P_{1B}-type ATPase that, like the other human copper(I)-transporting ATPase, ATP7A, can translocate copper(I) to the *trans*-Golgi network, where the metal is incorporated into copper-dependent enzymes. Copper stimulation results in the redistribution of both ATP7A and ATP7B to the plasma membrane through intracellular vesicles in order to export excess copper out of the cytosol (1–4). Mutations in the ATP7B autosomal gene produce the copper accumulation in tissues that characterizes Wilson disease (5–7). Mutations in ATP7A are instead responsible for Menkes disease (8, 9). ATP7A and ATP7B are often referred to as the MNK and WLN proteins, respectively.

P_{1B}-type ATPases contain four major regions/domains (the N-terminal copper-binding tail, the transmembrane domain, the ATP-binding domain, and the phosphatase domain) as well as a short C-terminal tail (10). The N-terminal copper-binding cytosolic tails of WLN and MNK are both ~630 amino acids long. They contain six 70-amino acid, independently folded metal-binding domains (MBDs),[2] which are relatively similar in sequence and structure (11–15). Each MBD harbors the conserved sequence motif GM*X*C*XX*C, through which it can bind one equivalent of copper(I) (16). The N-terminal tail has an important role in tuning the activity of the enzyme (17–19) and in modulating the intracellular trafficking rates of ATP7A and ATP7B (20–22), an aspect of the cellular function of these ATPases that is lacking in less complex organisms, including yeast. The presence of either the intact fifth or intact sixth metal-binding domain in the human proteins is sufficient to support both WLN/MNK activity and intracellular trafficking at essentially normal levels (21, 23–26). This has stimulated researchers to investigate the cellular role of the additional four domains that are present in the two human ATPases as well as in several other mammalian homologues. The K_{Cu} binding affinity of the six MBDs of both the MNK (13, 27, 28) and WLN (29, 30) proteins range between 1- and 5-fold the binding affinity of the copper(I) metallochaperone HAH1. Instead, the two MBDs of the yeast Ccc2 ATPase have essentially the same affinity as the Atx1 metallochaperone (31). Copper(I) is transferred from the chaperone to the MBDs of the ATPases via a metal-bridged intermediate (32–35), which may or may not be present at detectable levels in solution. All the six MBDs of the N-terminal tail of MNK are metallated by the metallochaperone. Additionally, two of them form, at relatively high levels, an intermolecular copper(I)-bridged adduct with HAH1 (28).

In this work, we investigated the interaction of copper(I)-HAH1 with the whole N-terminal tail of WLN to characterize the different behavior of its six MBDs metal-binding domains and compare it to the properties of the N-terminal tail of MNK. The variations between these two systems may be relevant for the details of their different functioning in tissues where both proteins are expressed (such as kidney and placenta). The features of the interactions of the MBDs of P_{1B}-type ATPases with their partners are rationalized on the basis of a specific energetic profile of the copper transfer reaction.

EXPERIMENTAL PROCEDURES

Preparation of Protein Samples—A DNA segment corresponding to residues 1–633 of WLN (WLN1-6 hereafter) was amplified by PCR and cloned in the Gateway Entry vector

* This work was supported by Ministero dell'Istruzione, dell'Università e della Ricerca (Fondo per gli Investimenti della Ricerca di Base project RBLA032ZM7), 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"), and the European Commission (Project SPINE2-COMPLEXES-031220).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Fig. S1 and Tables S1–S3.

[1] To whom correspondence should be addressed: Magnetic Resonance Center, University of Florence, Via L. Sacconi, 6, 50019 Sesto Fiorentino, Italy. Tel.: 39-055-4574272; Fax: 39-055-4574271; E-mail: ivanobertini@cerm.unifi.it.

[2] The abbreviations used are: MBD, metal-binding domain; HSQC, heteronuclear single quantum coherence.

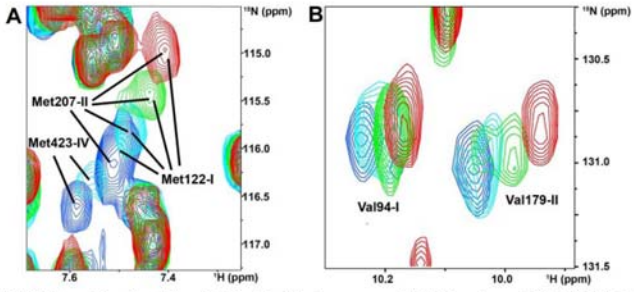


FIGURE 1. **Formation of metal-mediated adducts between copper(I)-HAH1 and some MBDs of WLN1-6.** In panel *A* and *B* selected peaks from the first, second, and fourth MBDs of WLN1-6 are shown as a function of the copper(I)-HAH1/WLN1-6 ratio. MBDs numbers are indicated in roman numerals. *Blue*, 0:1; *cyan*, 1.5:1; *green*, 3:1; *red*, 6:1 copper(I)-HAH1/WLN1-6 ratio.

pENTR/TEV/D-TOPO (Invitrogen) to include the tobacco etch virus protease cleavage site at the N-terminal end. This segment was then subcloned into the pETG-20A EMBL (36) through the Gateway LR reaction, yielding a plasmid expressing the protein fused with thioredoxin A and a His tag at the N terminus. Mutants were generated using the QuikChange mutagenesis kit (Stratagene).

Wild type and mutant WLN1-6 were expressed in *Escherichia coli* Rosetta(DE3) cells (Novagen) in minimal medium cultures and purified using HisTrap chelating FF columns (GE Healthcare). For isotope enrichment, $(^{15}NH_4)_2SO_4$ and ^{13}C-labeled glucose were used. The His tag and the fusion protein were cleaved with AcTEV (Invitrogen). Protein purity was checked by SDS-PAGE and matrix-assisted laser desorption ionization time-of-flight mass spectra. Copper(I) was added to samples in an inert atmosphere chamber (Coy Lab) by adding an acetonitrile solution of tetrakis(acetonitrile)copper(I) hexafluorophosphate as copper source in the presence of 1 mM dithiothreitol. All samples contained around 0.3 mM protein in 50 mM phosphate, 50 mM arginine, 50 mM glutamate at pH 7.0. The addition of arginine and glutamate to the buffer extended the lifetime of the samples (37), which, however, remained in the range of 2–3 days.

HAH1 samples were prepared as already reported, always without a poly-His tag (38). In titration experiments, we added copper(I)-HAH1 and apo-WLN1-6 directly in the NMR tube under N_2 atmosphere, using the same procedure previously followed for MNK1-6 (28).

NMR Spectroscopy—NMR experiments were acquired using Bruker Avance spectrometers operating at proton frequencies of 500, 700, and 900 MHz, all equipped with cryogenically cooled probes. Resonance assignments of WLN1-6 were performed through conventional multidimensional NMR techniques based on triple resonance experiments (39). Titrations of apo-WLN1-6 with copper(I)-HAH1 were followed through ^{1}H-^{15}N SOFAST-heteronuclear multiple quantum coherence experiments (40). The backbone dynamics of copper(I)-HAH1 in the presence of apo-WLN1-6 was investigated through the analysis of ^{15}N R_1, R_2 relaxation rates (41). More than one sample was needed to complete NMR data collection. A table summarizing the NMR experiments performed is given in the supplemental material (Table S1). All of the figures reporting NMR spectra were generated with the program CARA (42).

RESULTS

The first 633 amino acids of WLN (WLN1-6, with the wild-type sequence), comprising the entire cytosolic tail and in particular the six MBDs, were expressed in *E. coli* as a single construct; the protein could be enriched in the ^{15}N and ^{13}C stable isotopes, permitting its characterization by NMR spectroscopy. Protein solubility was only about 300 μM at pH 7.0; in addition, the samples were poorly stable, displaying significant degradation after 2–3 days. Fig. S1 reports a two-dimensional ^{1}H-^{15}N HSQC spectrum of apo-WLN1-6. The particularly favorable line widths of the spectra are due to the presence of long linker segments between the various domains, which allow them to tumble regardless of one another. We assigned signals from all six MBDs using triple resonance experiments; the backbone resonance assignments both for the apo and copper(I) forms are reported in Tables S2 and S3. In the other regions of the polypeptide chain, we could observe and assign some signals from the linker connecting domains 1 and 2, domains 3 and 4, and domains 5 and 6. The present assignment was more extensive than that achieved for MNK1-6 (28). Using a labile inorganic copper(I) complex, we could fully metallate the protein. The HSQC spectra of metallated WLN1-6 were used to readily identify metal transfer from copper(I)-HAH1 to the various MBDs during titrations of apo-WLN1-6 with HAH1 (see below).

As mentioned, the major focus of this work was on the analysis of the interaction between apo-WLN1-6 and copper(I)-HAH1. This was investigated by adding the latter to WLN1-6, with only one of the two proteins enriched in ^{15}N so that only its signals appeared in ^{1}H-^{15}N HSQC spectra. The samples contained 1 mM dithiothreitol. The interaction with copper(I)-HAH1 caused a variation of the chemical shifts of the signals from the backbone amide moieties of domains 1, 2, and 4 (Fig. 1), whose magnitude increased upon increasing concentration of the titrant. Instead, for domains 3, 5, and 6, we observed slow or intermediate exchange with the copper(I)-bound species (Fig. 2), whose signals increased in intensity upon increasing the partner concentration. These features progressed to the same extent throughout the titration (Figs. 1 and 2). The linkers connecting the various metal-binding domains instead did not experience appreciable chemical shift changes calculated through the following equation (Fig. 3).

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

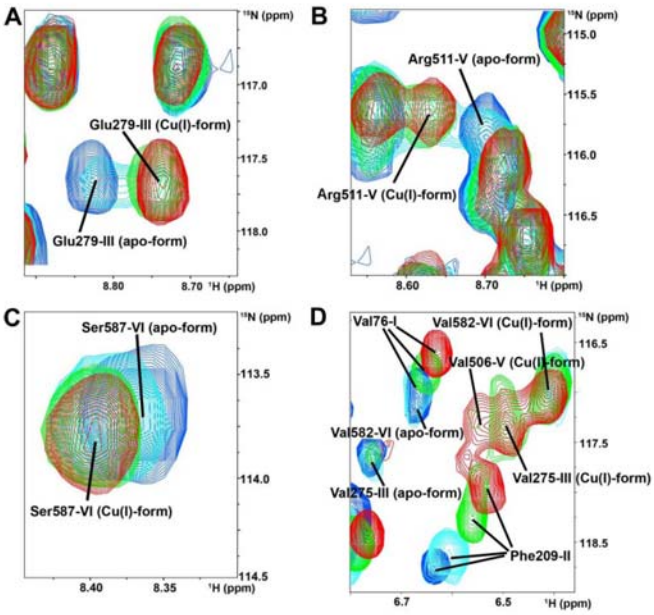


FIGURE 2. **Metallation of the third, fifth, and sixth domains of WLN1-6 by copper(I)-HAH1.** Panels *A–D* show an overlay of a selected region of the ^{1}H-^{15}N HSQC spectra of apo-WLN1-6 (*blue*) and of apo-WLN1-6 in the presence of 1.5 (*cyan*), 3.0 (*green*), and 6.0 (*red*) equivalents of unlabeled Cu(I)-HAH1. MBDs numbers are indicated in roman numerals.

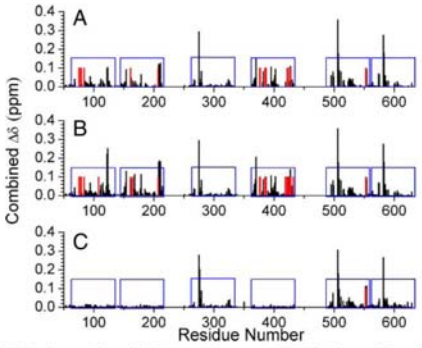


FIGURE 3. **Interaction of WLN1-6 with copper(I)-HAH1 (viewed from the side of WLN1-6).** Variations in the position of signals between WLN1-6 alone and in the presence of 3.0 eq (*A*) and 6.0 eq (*B*) of copper(I)-HAH1. *C*, variations in the position of signals between C1,2,4WLN1-6 mutant alone and in the presence of 6.0 eq of copper(I)-HAH1. These variations are quantitated through the combined chemical shift (52), which is calculated from the experimental ^{1}H and ^{15}N chemical shift changes ($\Delta\delta(^{1}H)$ and $\Delta\delta(^{15}N)$, respectively) measured between corresponding peaks in the two forms, through Equation 1. The *red bars* indicate the peaks that became broad beyond detection upon interaction. The boundaries of the six MBDs along the sequence are shown by *blue boxes*.

The signals of domains 1, 2, and 4 always appeared in a single form, and the corresponding peaks had a position in the spectrum that depended on the copper(I)-HAH1/WLN1-6 ratio (Figs. 1 and 2). The chemical shift changes observed for corresponding residues in these domains were similar throughout the titration (Fig. 3). The perturbed signals of domains 1, 2, and 4 belonged to both the metal-binding loop and the loop between the second α-helix and the fourth β-strand of each domain (loop 5). Loop 5 is located at the interface of the metal-bridged adduct formed by the homologous yeast proteins (35). Finally, some signals of the residues of the MBD of domain 4 experienced significant line broadening (*e.g.* Met-423 in Fig. 1), which led to them becoming undetectable. Eventually, signals from the copper(I) form could be detected.

For domains 3, 5, and 6, the chemical shift changes were mainly limited to the metal-binding loop, and loop 5 was not affected by interaction with the partner (Fig. 3). The signals of the apo form disappeared with a similar dependence on the concentration of copper(I)-HAH1. However, not all of the corresponding signals of the copper(I) form reappeared at the same protein ratio (see the signals from Val-275, Val-506, and Val-582 in Fig. 2). Domains 3, 5, and 6 were essentially fully metallated at the end of the titration (6:1 copper(I)-HAH1/WLN1-6 ratio).

Looking at ^{15}N-enriched copper(I)-HAH1 when interacting with apo-WLN1-6, it was observed that the interaction caused a change in the position of various peaks in the ^{1}H-^{15}N HSQC. In addition, essentially all of the signals of the residues around the metal-binding loop (Fig. 4) became broad beyond detection. In order to evaluate whether this was the result of the formation of an intermolecular adduct involving HAH1 and WLN1-6, we measured the correlation time (τ_R) for reorientation in solution of copper(I)-HAH1 alone and in the presence of equimolar WLN1-6. An increase of the τ_R of HAH1 in the latter system up to 7.4 ns from 4.9 ns of the isolated protein (38) was determined, which resulted from the generalized increase of ^{15}N R_2 relaxation rates and the concomitant decrease of ^{15}N R_1 relaxation rates. This increase in the τ_R of HAH1 indicates that the latter protein formed an intermolecular adduct with one or more of the domains of WLN1-6.

The overall behavior indicates for WLN1-6 a clear cut differentiation between the signals from domains 1, 2, and 4 on the one hand and 3, 5, and 6 on the other. From all of these data and also on the basis of the similar behavior of other systems, such

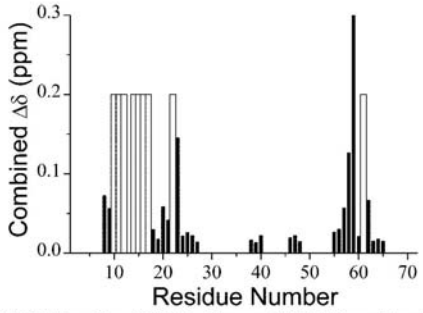


FIGURE 4. **Interaction of WLN1-6 with copper(I)-HAH1 (viewed from the side of HAH1).** Variations in the position of signals between copper(I)-HAH1 alone and in the presence of 1 eq of WLN1-6 are quantitated through the combined chemical shift variation (see legend to Fig. 3). The *white bars* indicate the peaks that became broad beyond detection upon interaction.

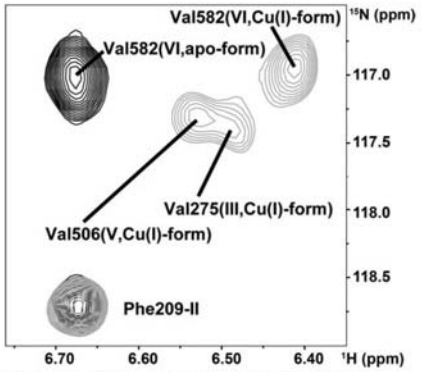


FIGURE 5. **Copper(I) transfer from HAH1 to C1,2,4WLN1-6.** Overlay of a selected region of the ^{1}H-^{15}N HSQC spectra of apo-C1,2,4WLN1-6 (*black*) and of apo-C1,2,4WLN1-6 in the presence of 3 eq of unlabeled Cu(I)-HAH1 (*light gray*).

as human Menkes and yeast Ccc2, it can be concluded that (i) all of the six domains of WLN1-6 were metallated by copper(I)-HAH1; (ii) the latter protein formed metal-bridged adducts with domains 1, 2, and 4 at levels detectable by NMR spectroscopy, which were in fast (on the NMR time scale) exchange with the single species; and (iii) the copper(I) form of domains 3, 5, and 6 exchanged with the apo form in an intermediate to slow regime.

To discriminate whether domains 3, 5, and 6 were directly metallated by copper(I)-HAH1 or if an intramolecular copper transfer occurred, we studied the interaction of copper(I)-HAH1 with a mutant of WLN1-6, where the metal-binding cysteines of domains 1, 2, and 4 had been mutated to alanine (C1,2,4WLN1-6), thus making these domains unable to bind copper. The ^{1}H-^{15}N HSQC spectrum of C1,2,4WLN1-6 was essentially superimposable to that of wild type WLN1-6 recorded under the same experimental conditions, with the exception of the mutated amino acids and of their neighboring residues in sequence. Upon the addition of copper(I)-HAH1 to the mutant, we observed that domains 3, 5, and 6 acquired copper(I) from HAH1, since they experienced the same spectral changes that were observed in native WLN1-6 (Fig. 3). The new sets of signals of the metallated domains were in slow exchange with the apo form (Fig. 5). On the contrary, the signals of residues from domains 1, 2, and 4 did not experience significant chemical shift changes, indicating that no adducts were formed with the partner (Fig. 3).

DISCUSSION

When the WLN1-6 construct was presented with copper(I)-HAH1, the behavior of its six MBDs differentiated. In fact, the NMR signals of the backbone amide moieties from three domains (the first, second, and fourth) experienced a variation of their position that was dependent on the copper(I)-HAH1/WLN1-6 ratio (Fig. 1). Instead, for the other three domains (the third, fifth, and sixth), the addition of copper(I)-HAH1 gave rise to another set of signals in slow or intermediate exchange with the apo form (Fig. 2).

It has been extensively demonstrated in the available literature for related systems that the present behavior of the signals of the amino acids in domains 1, 2, and 4 results from the formation of an intermolecular, copper(I)-bridged adduct with HAH1 (14, 15, 28, 35, 43). In the present work, the formation of this adduct was experimentally shown by the increase of the τ_R of HAH1. Instead, the behavior shown by domains 3, 5, and 6 is attributable to the occurrence of metal transfer from the metallochaperone without accumulation of the protein-protein adduct. Indeed, in a closely related study of a two-domain construct of WLN that comprised only domains 3 and 4, it was demonstrated that the addition of copper(I)-HAH1 to apo-WLN3-4 induced a significant increase of the tumbling rate of domain 4, consistent with formation of an intermolecular copper(I)-bridged adduct, but did not affect the tumbling rate of domain 3 (15).

Within each group of domains (1, 2, and 4 on one hand, 3, 5, and 6 on the other), the signals from corresponding residues of the various domains did not all behave in the same way. Regarding domains 1, 2, and 4, the features of the signals of the fourth MBD suggested that the complex between this domain and copper(I)-HAH1 had a slower dissociation rate than in the case of the first and second MBDs. For domains 3, 5, and 6, the signals of the copper(I) form became detectable at different copper(I)-HAH1/WLN1-6 ratios, despite the similar affinity for the metal ion of the three domains (29). This is presumably due to the diverse dynamics of this loop in each MBD, related to conformational equilibria involving the apo and copper(I) forms, resulting in different broadening of the signals.

We then investigated whether the metallation of domains 3, 5, and 6 could be carried out by HAH1 alone or required the prior formation of an adduct involving one among domains 1, 2, or 4. As shown in Figs. 3 and 5, the C1,2,4WLN1-6 mutant could be metallated by copper(I)-HAH1. In addition, there was no significant perturbation of the backbone signals of domains 1, 2, and 4 (Fig. 3). The presence of the metal-binding cysteines

is thus required for the formation of the intermolecular adduct. This confirms that the adduct is metal-mediated, analogously to the yeast Atx1/Ccc2a system (35). Furthermore, it appears that the metal transfer process from copper(I)-HAH1 to the third, fifth, or sixth MBDs can take place even without any prior intermolecular interaction with the other domains. In a previously reported study (14), a double domain construct comprising the fifth and sixth metal-binding domains of the WLN protein was unable to receive directly the metal from copper(I)-HAH1. This difference from our results is most likely linked to the different protein as well as to the different experimental conditions employed in the two studies and is not unprecedented. Indeed, for the third domain of the MNK protein, a difference in metal-binding capability has been reported when comparing the isolated domain to the six-domain construct (28, 44). In addition, the isolated second WLN MBD required somewhat higher concentrations of the partner metallochaperone than the fourth to afford comparable chemical shift variations (14), whereas in the present WLN1-6 construct, the first, second, and fourth MBDs experienced similar chemical shift variations along the titration. Considering all of the available experimental data on the wild type and mutant WLN and MNK systems (28), it appears that copper(I) can be transferred directly by the metallochaperone to any of the six MBDs. A mechanism in which the metal ion is first transferred to a domain and then relocates to another domain (to be ultimately pumped across the membrane) cannot be ruled out.

The different interaction of the MBDs of the MNK (28) and WLN proteins with copper(I)-HAH1 reflects different thermodynamic properties of the various species involved in the metal transfer pathway. In all known cases, including those of the yeast and bacterial homologues of the present proteins, it can be assumed that copper(I) transfer to a soluble MBD of the ATPase occurs according to the following scheme (35),

$$\begin{array}{c}\text{Cu(I)-Chaperone} + \text{apo-ATPase}\\ \rightleftarrows\\ \text{Chaperone-Cu(I)-ATPase}\\ \rightleftarrows\\ \text{apo-chaperone} + \text{Cu(I)-ATPase}\end{array}$$

SCHEME 1

where the metal ion bridges the proteins in the chaperone-Cu(I)-ATPase species, as it is coordinated by two cysteines of one protein and one cysteine of the other. For the yeast system, there are two forms of the chaperone-Cu(I)-ATPase species in equilibrium (35). The two forms differ in the identity of the protein providing the two cysteines, the form with the ATPase providing two cysteines being more stable (35). Scheme 1 is in agreement with a body of experimental evidence based on the combination of site-directed mutagenesis data and various spectroscopic techniques (32–35).

Within the above general reaction scheme, the various systems experimentally investigated up to now feature a remarkable differentiation of the energetic and kinetic aspects of the process. In yeast, the affinity for copper(I) of the metallochaperone (Atx1) and of the first MBD of the Ccc2 ATPase is nearly the same (31). When 1 mM copper(I)-Atx1 is mixed with 1 mM

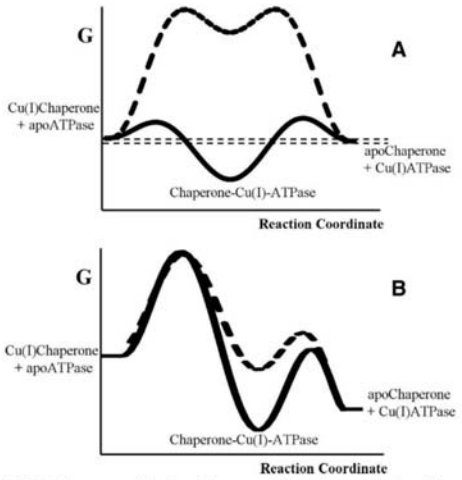


FIGURE 6. **Energy profile along the copper(I) transfer reaction for different kinds of metallochaperone/ATPase pairs.** *A*, the isolated proteins have very similar affinity for copper(I) and may (*continuous line*) or may not (*dashed line*) form a stable complex; *B*, the isolated MBD of the ATPase has higher affinity for copper(I) than the metallochaperone. The two proteins may (*continuous line*; *e.g.* MBD1, MBD2, and MBD4 of WLN1-6) or may not (*dashed line*; *e.g.* MBD3, MBD5, and MBD6 of WLN1-6) form a stable complex. In both cases, the complex and the isolated proteins may be in fast or slow (on the NMR time scale) exchange, depending on the height of the energetic barriers separating the various species. The two equilibria of Equation 1 do not need to be equally rapid; indeed, for the homologous system of *Synechocystis* PCC 6803, one of the two equilibria was fast, whereas the other was slow (53). G stands for Gibbs energy.

Ccc2a, a copper(I)-bridged adduct forms, involving ~70% of the protein molecules (45). The adduct is in fast exchange with the isolated proteins. The various MBDs of MNK (27, 46) and WLN (29) have a higher affinity for copper(I) than HAH1. In WLN, three MBDs form an adduct at detectable amounts with copper(I)-HAH1, whereas for the other three MBDs, the complex is not detectable, and only the final metallated form can be observed. In MNK, only two domains (the first and the fourth) form an adduct (28) at similar protein concentrations; the second MBD of MNK is metallated by copper(I)-HAH1 without the formation of detectable amounts of a protein-protein adduct. This behavior is observed both in the context of the entire tail (28) and for the isolated domain (46). The different behaviors described above correspond to different energy profiles for the copper transfer process, which are sketched in Fig. 6, for the reaction of Scheme 1. In Ccc2, the final state has a free energy only slightly lower than the initial state (Fig. 6*A*), whereas this difference is somewhat higher in MNK and WLN (Fig. 6*B*). Indeed, the HAH1-Cu(I)-WLN complex is more similar to the final product than the Atx1-Cu(I)-Ccc2 complex. In both cases, the occurrence of detectable amounts of the intermediate adduct is possible if its free energy is close to that of the final state of Scheme 1 but is independent of the relative affinity for copper(I) of the metallochaperone and the ATPase. When the adduct is formed, it is in fast equilibrium with the free pro-

teins, implying low activation barriers. On the other hand, for those MBDs that do not form detectable amounts of adduct, the equilibrium between the apo and copper(I) forms is slow, regardless of the presence of HAH1. The latter, therefore, does not act as a catalyst in lowering the activation barrier between the two forms of the metal-binding domains (Fig. 6*B*).

The N-terminal tail of P_{1B}-type ATPases may have a regulatory role in the enzyme activity and/or trafficking (17, 20, 21). In particular, the tail could have an autoinhibitory role in the absence of copper(I) (18, 24). The structural model of CopA, a P_{1B}-type ATPase from *Archaeoglobus fulgidus*, has suggested that in the absence of copper(I), the tail is in close contact with the other cytoplasmic domains (47). Both for WLN and MNK, the regulation of the ATPase activity and/or trafficking by the N-terminal tail could depend upon the interaction with copper(I)-HAH1, leading to a variation of the contacts of the tail with the other domains of the enzyme. Indeed, the present data, as well as those on MNK1-6 (28), suggested that the aforementioned rearrangements are linked to the formation, and presumably accumulation at high intracellular copper(I) concentrations, of adducts between the MBDs of the tail and copper(I)-HAH1. This view would also provide a rationale for the observed dependence of ATPase trafficking on the expression levels of HAH1 (48). Even when the equilibrium leading to the formation of the adduct with the metallochaperone or to the transfer of copper(I) is not fully shifted toward the products, every time one of these molecular events takes place, it can trigger the conformational rearrangements of the ATPase that ultimately lead to ATP hydrolysis. The latter reaction is effectively irreversible and thus pushes forward the catalytic cycle. The effective protein concentration in the cell is not known; the relatively high equilibrium constants for the protein-protein adducts observed here imply that the processes described above may occur also at micromolar concentrations.

The available literature points to the MBD or to the two MBDs closest to the transmembrane region of MNK and WLN as playing a role different than the most N-terminal four domains (49, 50). In this respect, it is noteworthy that for both human proteins, the fifth and sixth MBDs are metallated by the metallochaperone without forming detectable amounts of complex. The metallation of these two domains does not require the formation of copper(I)-mediated adducts with other metal-binding domains within the cytosolic tail, as shown by mutagenesis data for both MNK (28) and WLN (Fig. 3). This is at variance with what observed for yeast Ccc2. This ATPase has only two MBDs, of which at least one forms a complex with copper(I)-Atx1.

Of the first four MBDs, three domains in WLN and two in MNK can form detectable amounts of an adduct. The above differences are the consequence of subtle mutations of their surface properties, which define the energetic profile of the interaction, with copper(I)-HAH1 among those depicted in Fig. 6. This differentiation could contribute to tuning the regulation of the activity and/or trafficking of the ATPases. Specifically, for WLN and MNK, this could be a means to discriminate their properties in the tissues where they are coexpressed (51).

Acknowledgment—We thank David Huffman for many insightful discussions.

REFERENCES

1. Petris, M. J., Mercer, J. F., Culvenor, J. G., Lockhart, P., and Camakaris, J. (1996) *EMBO J.* **15,** 6084–6095
2. Schaefer, M., Hopkins, R. G., Failla, M. L., and Gitlin, J. D. (1999) *Am. J. Physiol.* **276,** G639–G646
3. Harrison, M. D., Jones, C. E., Solioz, M., and Dameron, C. T. (2000) *Trends Biochem. Sci.* **25,** 29–32
4. Puig, S., and Thiele, D. J. (2002) *Curr. Opin. Chem. Biol.* **6,** 171–180
5. Bull, P. C., Thomas, G. R., Rommens, J. M., Forbes, J. R., and Cox, D. W. (1993) *Nat. Genet.* **5,** 327–337
6. Petrukhin, K., Fischer, S. G., Pirastu, M., Tanzi, R. E., Chernov, I., Devoto, M., Brzustowicz, L. M., Cavanis, E., Vitale, E., and Russo, J. J. (1993) *Nat. Genet.* **5,** 338–343
7. Tanzi, R. E., Petrukhin, K., Chernov, I., Pellequer, J. L., Wasco W., Ross, B., Romano, D. M., Parano, E., Pavone, L., and Brzustowicz, L. M., Devoto, M., Peppercorn, J., Bush, A. I., Sternlieb, I., Pirastu, M., Gusella, J. F., Evgrafov, O., Penchaszadeh, G. K., Honig, B., Edelman, I. S., Soares, M. B., Scheinberg, I. H., and Gilliam, T. C. (1993) *Nat. Genet.* **5,** 344–350
8. Mercer, J. F., Livingston, J., Hall, B., Paynter, J. A., Begy, C., Chandrasekharappa, S., Lockhart, P., Grimes, A., Bhave, M., and Siemieniak, D. (1993) *Nat. Genet.* **3,** 20–25
9. Vulpe, C. D., Levinson, B., Whitney, S., Packman, S., and Gitschier, J. (1993) *Nat. Genet.* **3,** 7–13
10. Solioz, M., and Vulpe, C. (1996) *Trends Biochem. Sci.* **21,** 237–241
11. Gitschier, J., Moffat, B., Reilly, D., Wood, W. I., and Fairbrother, W. J. (1998) *Nat. Struct. Biol.* **5,** 47–54
12. Jones, C. E., Daly, N. L., Cobine, P. A., Craik, D. J., and Dameron, C. T. (2003) *J. Struct. Biol.* **143,** 209–218
13. Banci, L., Bertini, I., Del Conte, R., D'Onofrio, M., and Rosato, A. (2004) *Biochemistry* **43,** 3396–3403
14. Achila, D., Banci, L., Bertini, I., Bunce, J., Ciofi-Baffoni, S., and Huffman, D. L. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103,** 5729–5734
15. Banci, L., Bertini, I., Cantini, F., Rosenzweig, A. C., and Yatsunyk, L. A. (2008) *Biochemistry* **47,** 7423–7429
16. Lutsenko, S., Petrukhin, K., Cooper, M. J., Gilliam, C. T., and Kaplan, J. H. (1997) *J. Biol. Chem.* **272,** 18939–18944
17. Rice, W. J., Kovashilin, A., and Stokes, D. L. (2006) *Biochem. Biophys. Res. Commun.* **348,** 124–131
18. Hatori, Y., Hirata, A., Toyoshima, C., Lewis, D., Pilankatta, R., and Inesi, G. (2008) *J. Biol. Chem.* **283,** 22541–22549
19. DiDonato, M., Hsu, H. F., Narindrasorasak, S., Que, L. J., and Sarkar, B. (2000) *Biochemistry* **39,** 1890–1896
20. Cater, M. A., La Fontaine, S., and Mercer, J. F. (2007) *Biochem. J.* **401,** 143–153
21. Cater, M. A., Forbes, J. R., La Fontaine, S., Cox, D., and Mercer, J. F. (2004) *Biochem. J.* **380,** 805–813
22. Voskoboinik, I., Mar, J., Strausak, D., and Camakaris, J. (2001) *J. Biol. Chem.* **276,** 28620–28627
23. Payne, A. S., and Gitlin, J. D. (1998) *J. Biol. Chem.* **273,** 3765–3770
24. Huster, D., and Lutsenko, S. (2003) *J. Biol. Chem.* **278,** 32212–32218
25. Strausak, D., La Fontaine, S., Hill, J., Firth, S. D., Lockhart, P. J., and Mercer, J. F. (1999) *J. Biol. Chem.* **274,** 11170–11177
26. Voskoboinik, I., Strausak, D., Greenough, M., Brooks, H., Petris, M., Smith, S., Mercer, J. F., and Camakaris, J. (1999) *J. Biol. Chem.* **274,** 22008–22012
27. Banci, L., Bertini, I., Cantini, F., Migliardi, M., Rosato, A., and Wang, S. (2005) *J. Mol. Biol.* **352,** 409–417
28. Banci, L., Bertini, I., Cantini, F., Della Malva, N., Migliardi, M., and Rosato, A. (2007) *J. Biol. Chem.* **282,** 23140–23146
29. Yatsunyk, L. A., and Rosenzweig, A. C. (2007) *J. Biol. Chem.* **282,** 8622–8631
30. Bunce, J., Achila, D., Hetrick, E., Lesley, L., and Huffman, D. L. (2006) *Biochim. Biophys. Acta* **1760,** 907–912

31. Huffman, D. L., and O'Halloran, T. V. (2000) *J. Biol. Chem.* **275,** 18611–18614
32. Larin, D., Mekios, C., Das, K., Ross, B., Yang, A. S., and Gilliam, C. T. (1999) *J. Biol. Chem.* **274,** 28497–28504
33. Strausak, D., Howie, M. K., Firth, S. D., Schlicksupp, A., Pipkorn, R., Multhaup, G., and Mercer, J. F. (2003) *J. Biol. Chem.* **278,** 20821–20827
34. van Dongen, E. M., Klomp, L. W., and Merkx, M. (2004) *Biochem. Biophys. Res. Commun.* **323,** 789–795
35. Banci, L., Bertini, I., Cantini, F., Felli, I. C., Gonnelli, L., Hadjiliadis, N., Pierattelli, R., Rosato, A., and Voulgaris, P. (2006) *Nat. Chem. Biol.* **2,** 367–368
36. Alzari, P. M., Berglund, H., Berrow, N. S., Blagova, E., Busso, D., Cambillau, C., Campanacci, V., Christodoulou, E., Eiler, S., Fogg, M. J., Folkers, G., Geerlof, A., Hart, D., Haouz, A., Herman, M. D., Macieira, S., Nordlund, P., Perrakis, A., Quevillon-Cheruel, S., Tarandeau, F., Van Tilbeurgh, H., Unger, T., Luna-Vargas, M. P., Velarde, M., Willmanns, M., and Owens, R. J. (2006) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62,** 1103–1113
37. Golovanov, A. P., Hautbergue, G. M., Wilson, S. A., and Lian, L. Y. (2004) *J. Am. Chem. Soc.* **126,** 8933–8939
38. Anastassopoulou, J., Banci, L., Bertini, I., Cantini, F., Katsari, E., and Rosato, A. (2004) *Biochemistry* **43,** 13046–13053
39. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, and Skelton, N. J. (1996) *Protein NMR Spectroscopy: Principles and Practice*, pp. 478–528, Academic Press, Inc., San Diego
40. Schanda, P., Kupce, E., and Brutscher, B. (2005) *J. Am. Chem. Soc.* **33,** 199–211
41. Ishima, R., and Torchia, D. A. (2000) *Nat. Struct. Biol.* **7,** 740–743
42. Keller, R. (2004) *The Computer Aided Resonance Assignment Tutorial*, CANTINA Verlag, Goldau, Switzerland
43. Banci, L., Bertini, I., Cantini, F., Chasapis, C., Hadjiliadis, N., and Rosato, A. (2005) *J. Biol. Chem.* **280,** 38259–38263
44. Banci, L., Bertini, I., Cantini, F., Della Malva, N., Rosato, A., Herrmann, T., and Wüthrich, K. (2006) *J. Biol. Chem.* **281,** 29141–29147
45. Arnesano, F., Banci, L., Bertini, I., Cantini, F., Ciofi-Baffoni, S., Huffman, D. L., and O'Halloran, T. V. (2001) *J. Biol. Chem.* **276,** 41365–41376
46. Banci, L., Bertini, I., Chasapis, C., Ciofi-Baffoni, S., Hadjiliadis, N., and Rosato, A. (2005) *FEBS J.* **272,** 865–871
47. Wu, C. C., Rice, W. J., and Stokes, D. L. (2008) *Structure* **16,** 976–985
48. Hamza, I., Prohaska, J., and Gitlin, J. D. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 1215–1220
49. Tsivkovskii, R., MacArthur, B. C., and Lutsenko, S. (2001) *J. Biol. Chem.* **276,** 2234–2242
50. Mercer, J. F., Barnes, N., Stevenson, J., Strausak, D., and Llanos, R. M. (2003) *Biometals* **16,** 175–184
51. Linz, R., and Lutsenko, S. (2007) *J. Bioenerg. Biomembr.* **39,** 403–407
52. Garrett, D. S., Seok, Y. J., Peterkofsky, A., Clore, G. M., and Gronenborn, A. M. (1997) *Biochemistry* **36,** 4393–4398
53. Banci, L., Bertini, I., Ciofi-Baffoni, S., Kandias, N. G., Spyroulias, G. A., Su, X. C., Robinson, N. J., and Vanarotti, M. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103,** 8325

Chapter 4

Cytochrome c Oxidase Deficiency

Expression and characterization of human SURF1 and human COX11 proteins

4.1 Introduction

Cytochrome c Oxidase (CcO) is the terminal electron acceptor of the mitochondrial respiratory chain. In mammals, this protein consists of 13 individual subunits, 3 of which, subunits I–III (COX1, COX2, COX3), are encoded in the mitochondrial genome and serve as the catalytic centers binding copper and *heme*. In particular CcO has two copper binding sites (CuA and CuB), two *heme*s (*a* and *a*3), a magnesium and a zinc ion. The binuclear CuA site is located in COX2 and the CuB site is located in COX1 togheter with the two *heme*s. Electrons flow from reduced cytochrome c to the binuclear CuA site, then to *heme-a* before entering the oxygen binding *heme-a*3-CuB site (1).

Biogenesis of mitochondrial CcO relies on a large number of assembly factors. SURF1 and COX11 proteins are both involved in this process and associated to CcO deficiency. The loss of human SURF1 membrane protein function is associated with Leigh syndrome, an autosomal recessive inherited neurodegenerative disorder characterized by focal, bilateral lesions in one or more areas of the central nervous system.

Human *SURF1* is the first gene of the *SURFEIT* gene locus on chromosome 9, encodes a 30-kDa protein (2, 3). Located in the inner mitochondrial membrane, SURF1 is an amphiphilic integral membrane protein ubiquitously expressed. It presents a mitochondrial signal peptide (MSP) at N-terminus and it is predicted to form two transmembrane helices connected by a long loop facing the intermembrane space (4), (5). Sequence alignment confirm the presence of SURF1 homologs in many eukaryotes and prokaryotes (6). Even if is known that mutations in the *SURF1* gene leads to Leigh syndrome, the role of the protein in this disease remains unclear. It was suggested that the apo SURF1 protein interacts with *heme-a* synthase to receive the *heme-a* group and then probably interacts directly with the nascent CcO subunit I polypeptide emerging from the ribosome (7).

During my PhD I try to obtain the whole 256aa SURF1 protein segment without the MSP(**Fig.1**) in order to characterize and solve the structure by NMR and to study its role in the CcO assembly.

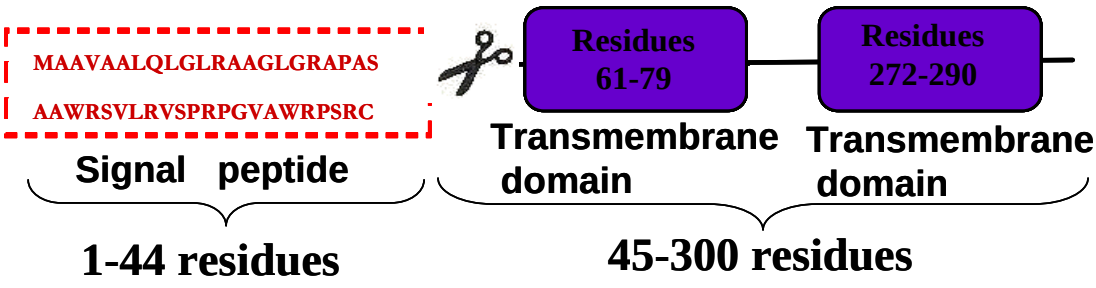


Fig.1: *A scheme of SURF1 protein. In my study I removed the MSP in order to obtain only the segments with the two transmembrane domains.*

COX11 is a copper-binding protein originally identified in yeast as essential for CcO activity and *heme-a* stability in COX1 (8). Actually we know that it is implicated in the assembly of the two copper centers of CcO, indeed it presents two conserved Cys residues which are candidate ligands for the Cu(I) ion. For this reason it is hypothesized that it is involved in CcO deficiency (9).

COX11 is a 28KDa protein ubiquitously expressed and it is located in the inner mitochondrial membrane. It is anchored to the membrane by a single transmembrane segment and presents a soluble C-terminal globular domain protruding into the intermembrane space. Even if the precise role of COX11 in CcO assembly remain unknown, it is possible that copper transfer to and from COX11 occurs through heterodimeric interaction of a single COX11 subunit with copper chaperone COX17 and COX1 CcO subunit respectively. Moreover is unclear if COX11 mediated copper incorporation occurs after CcO assembly or co-translationally, prior to the association of COX1 with other subunits. As COX11 is implicated in the assembly of the CuB site of CcO in yeast and a human homologue has been previously identified(10), this protein is a reasonable candidate for human CcO deficiency.

For this reason I focus my work on COX11 in order to obtain the 155aa C-terminal soluble domain of human COX11(**Fig.2**) and to characterize and solve the solution structure by NMR.

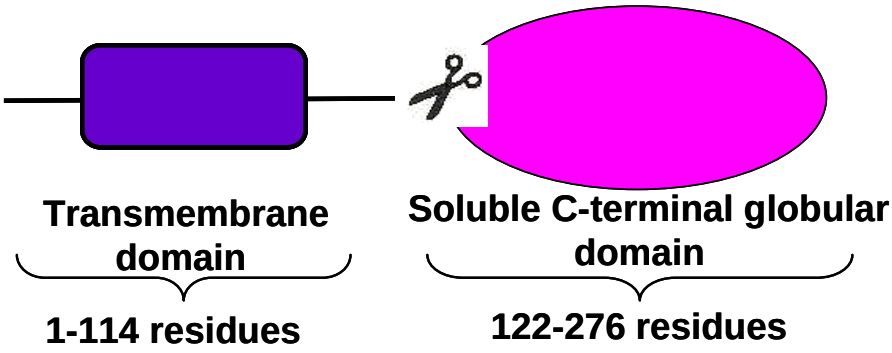


Fig.2: *COX11 protein. A scheme of the protein. In my work I try to obtain the soluble C-terminal globular domain excluding the transmembrane domain.*

4.2 Materials and methods

The DNA segment without the MSP of *SURF1* (256aa) and the one of C-terminal soluble domain of human *COX11* (155aa) were amplified from cDNA by PCR. Two primers combination (with and without TEV site in the primers) were used to amplify the segments. Different polymerases and conditions were used to obtain the amplification:

- Phusion Hot Start (BioLabs), +/- $MgCl_2$, 60°C/64°C
- KOD Polymerase(Novagen), +/- $MgCl_2$, 60°C
- Accuprime Pfx(Invitrogen), +/- $MgSO_4$, +/- DMSO, 60°C

The *SURF1* gene amplification was obtained with KOD Polymerase, while the *COX11* gene amplification was obtained with Accuprime Pfx + $MgSO_4$ + DMSO.
Genes were cloned in the Gateway Entry vector pDONR 221. SURF1 and COX11 were then subcloned successfully into destination vectors through the Gateway LR reaction. Three different pDEST were used: pTH27, PMM (N-terminal 6-His tag) and pTH24 (C-terminal 6-His tag) which were prepared in Karolinska Institutet laboratories. Proteins were transformed in three different Escherichia coli expression cells : Rosetta (DE3), BL21(DE3) and C41(DE3).
All clones were grown simultaneously at two different temperatures (25°C and 37°C) and were lisated both with a native lysis buffer (Tris 20mM pH 8.0 + NaCl 100 mM + Lysozyme 1 mg/ml + Benzonase 10 U/ml) and with a buffer supplemented with FOS-

CHOLINE-12 (FC12- Anatrace) 1% as detergent. Proteins purification were done by ultracentrifugation using Nickel sepharose resin (Amersham) and the protein expression was checked by Dot-Blot with His-Probe-HRP(Pierce) and SDS-PAGE.
SURF1 was solubilised using FC12 1% as detergent and lysate was purified using HisTrap chelating FF column. The His-tag was cleaved with TEV protease and removed using HisTrap chelating FF column again.
The NMR sample were prepared using Phosphate buffer solution 20mM +NaCl 100mM+FC12 0,1% at pH7.00.
COX11 protein was expressed in inclusion bodies and pellet washed with Triton 1% were solubilised with Guanidine 6M. Dialysis with different concentration of Urea (4M, 2M, 0M) and Arginine (0,5M, 0M) are performed follow one after the other, and then the protein was gel filtrated using Akta Prime plus with Superdex 75(GE Healthcare). The His-tag was cleaved with TEV protease and removed by gel filtration again. The NMR samples were prepared under reducing atmosphere in different buffer conditions:

- TrisHCl 50mM pH 8.0 + NaCl 100mM + DTT 10mM
- Phosphate Buffer 50mM pH 7.0 + NaCl 100mM + DTT 10mM
- BisTris 50mM pH 6.5 + NaCl 100mM + DTT 10mM

The last condition was the best one even if it presents precipitation.

4.3 Results and discussion

The first part of this project was carried out in Pär Nordlund's Lab at the Karolinska Institutet in Stockholm, in order to perform High Throughput detection of different constructs and strains for both SURF1 and COX11 proteins. Moreover in Pär Nordlund's research group I was able to use new modified plasmids (pTH27 and pTH24) together with the commercial one in order to increase the probability to obtain proteins expression . We first amplify genes from a cDNA and clone them in different plasmids. After sequencing the constructs were transformed in different strains of *E.Coli* expression cells. We then proceed with the expression test. The use of 96-well plates allowed us to grow all the strains simultaneously and to perform an High Throughput screening. This kind of analysis is very useful in order to exploit an

elevated number of different conditions (presence/absence of TEV cleavage site, different plasmids, three different strains, two different growth temperatures and two lysis buffer with/ without the addiction of detergent) for the expression of both COX11 and SURF1 proteins. To detect the proteins expression, the cells were lysated both with a native lysis buffer and with a buffer with addiction of FC12 detergent. The lysates were successfully cleared from cell debris by ultracentrifugation of the 96-well plates. The next step was the His-tag purification made by adding the Nickel sepharose resin to the lysates and by washing and eluting the proteins with different imidazol concentrations buffers. The presence of a 6-His tag in all constructs allowed us to detect the presence of both proteins by Dot-Blot using an anti His_6 tag probe (**Fig.3**).

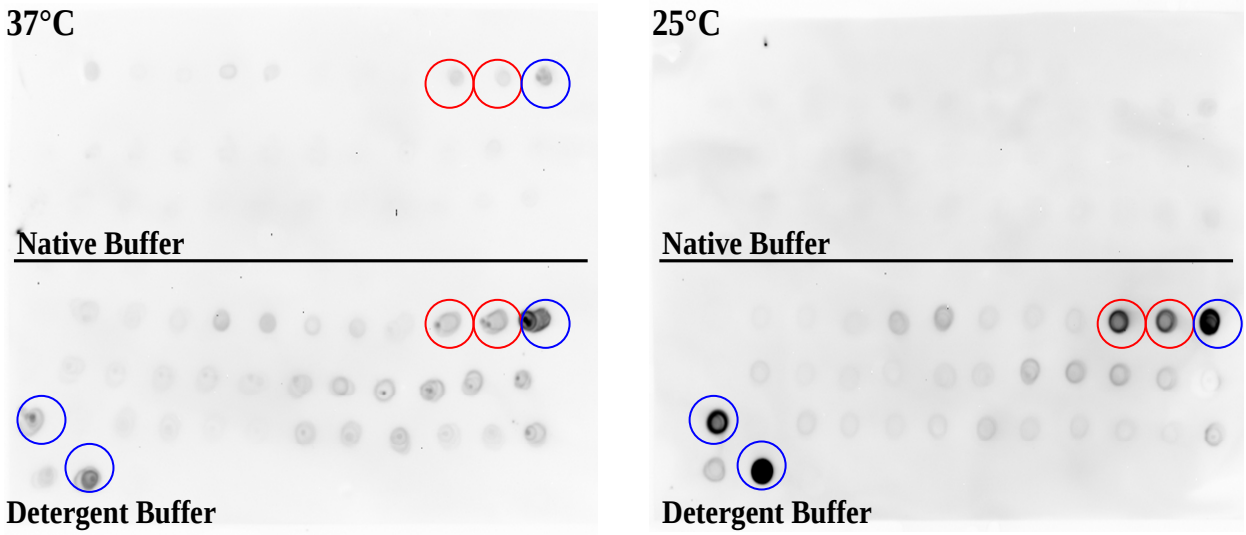


Fig.3: *Dot-Blot screening. Highlighted circles show the best expression for SURF1 in blue and for COX11 in red.*

The Dot-Blot analysis showed that both COX11 and Surf1 proteins expressed at 37°C in the native and in the FC12 detergent buffers. At the contrary, at room temperature, only in FC12 detergent buffer we have protein expression.

In particular, the best expression conditions (more intense blot) for SURF1 protein were:

- 37°C, Native buffer, no TEV site, pTH24, C41 cells
- 37°C, Detergent buffer, with TEV site, pTH24, C41/BL21/Rosetta cells
- 25°C, Detergent buffer, with TEV site, pTH24, C41/BL21/Rosetta cells

The best expression conditions for COX11 protein were:

- 37°C, Native buffer, with TEV site, pTH24, C41 cells
- 37°C, Detergent buffer, with TEV site, pTH27, C41/Rosetta cells
- 25°C, Detergent buffer, with TEV site, pTH27, C41/Rosetta cells

The minor intensity of some blots showed in **Fig.3** is often due to a less absorbance of the drop on the membrane rather than to differences in the protein expression yield. For this reason a more detailed analysis have to be carried out by SDS-PAGE of each blots (**Fig.4**). The latter analysis allowed to further restrict the number of expression conditions.

We conclude that for SURF1 protein the best expression conditions were:

- 37°C, Detergent Buffer, pTH24, C41/BL21/Rosetta cells with TEV site
- 37°C, Detergent Buffer, pTH27, C41/Rosetta cells without TEV site
- 25°C, Detergent Buffer, pTH24, C41/BL21/Rosetta cells without TEV site

For COX11 protein the best expression conditions were:

- 37°C, Detergent Buffer, pTH24, C41 cells, with TEV site
- 37°C, Detergent Buffer, pTH27 C41/Rosetta cells, with TEV site
- 25°C, Detergent Buffer, pTH24, C41 cells, with/without TEV site

The above conditions were subsequently used for the scale up of the coltures and expression optimization and purification were then carried out for both proteins.

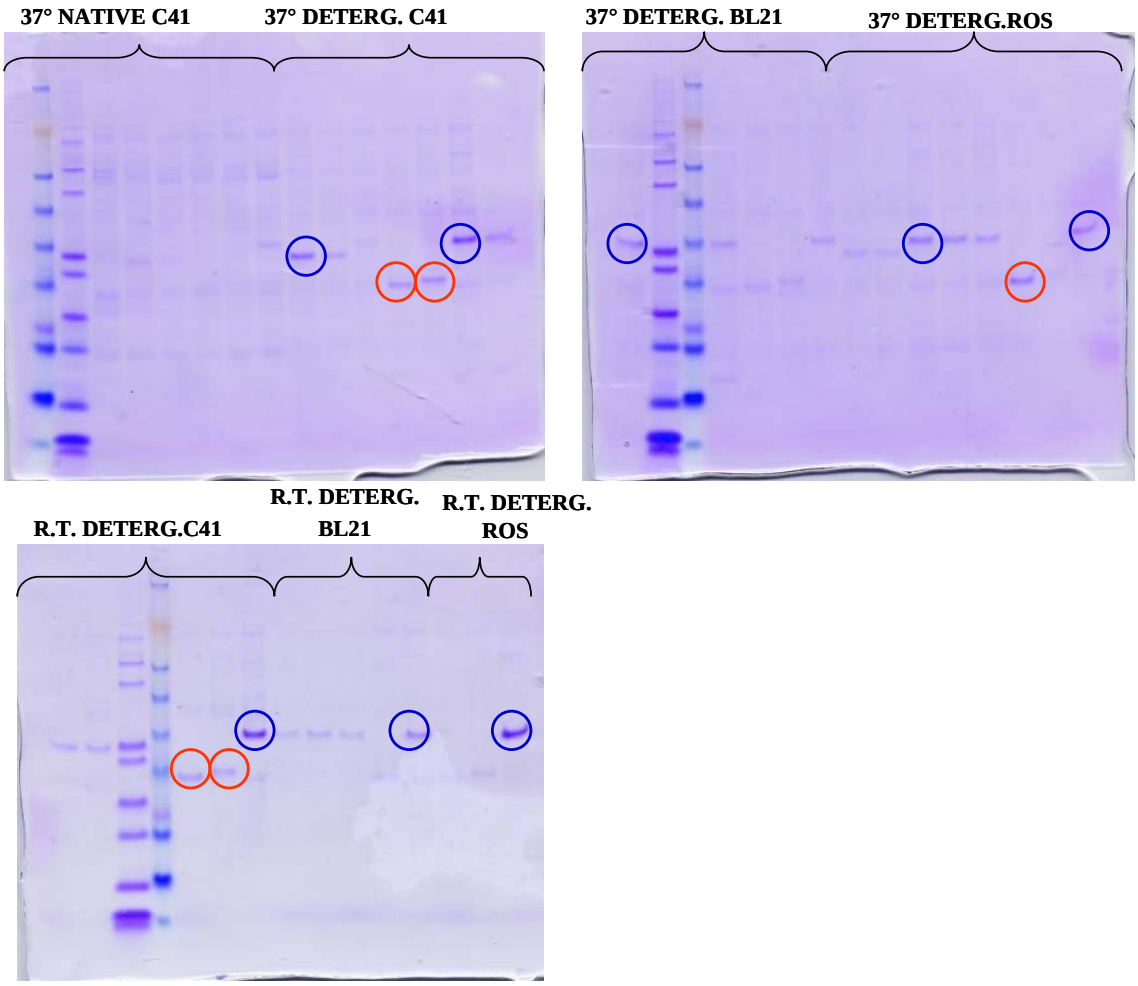


Fig.4: *SDS-PAGE. Gels show best expression conditions for SURF1 protein(in blue) and for COX11 protein (in red).*

4.3.1 SURF1 protein

Scale up was done growing one liter colture rich medium in flasks and the best construct results the pTH24 in C41 expression cells. The expression growing colture has been performed at 18°C obtaining the higher yield (**Fig.5**).

The cell debris was obtained by centrifugation of the colture and then cells were disrupted by freeze-thaw lysis and sonication. Then, a first centrifugation (10 000 rpm, 10 min) is necessary to remove the inclusion bodies and a second one (40 000 rpm, 45 min) to isolate the membrane were SURF1 protein is located. 1% Fos-choline12 (FC12) detergent buffer was used to solubilise the protein from the membrane with an

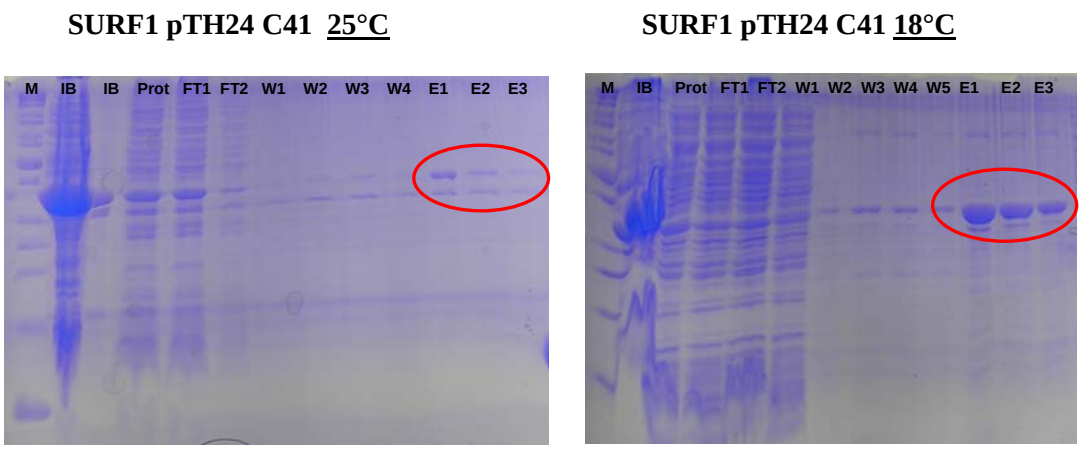


Fig. 5: *Expression optimization. The gels show the increasing of expression from colture growth at 25°C and colture growth at 18°C.*

over night incubation at 4°C. The solubilised protein is then purified using HisTrap chelating column eluting the proteins with different imidazol concentrations buffers and 0.1% FC12. After cleavage with TEV enzyme, the protein was purified again with His-Trap column keeping always 0.1%FC12 concentration. MALDI was then performed on a protein sample obtained directly cutting from an SDS-PAGE gel the band corresponding to the protein. The MALDI positive outcome showing 58.4% coverage of the peptides obtained from the sample by Trypsin digestion (**Fig.6**).

In order to better characterize the protein and to check if the tertiary structure is well folded a ^{15}N labelled sample was prepared in phosphate buffer 20mM pH7.00 + NaCl 100mM + FC12 0.1%. The ^{1}H-^{15}N HSQC spectrum acquired at 900MHz spectrometer at 298K (**Fig.7 A**), shows the peaks are clustered in the range of ^{1}H and ^{15}N frequencies typical of unfolded protein. Unfortunately the increasing of temperature up to 310K did not improve the quality of ^{1}H-^{15}N HSQC (**Fig.7 B**). However both spectra shows the resonance typical of the side chain $NH_{\varepsilon 1}$ of the Tryptophan (high lightened in red in **Fig.7**) suggesting that the protein is not completed unfolded and that we can improve the quality of the spectrum increasing the amount of surfactant. The presence of these peaks suggest that some regions of the protein are folded. The detergent is necessary for the protein to be stable in solution, and up to now FC12 at 0.1% was the only detergent and concentration that we tested, but it appears evident that a detailed screening of different detergents concentrations is necessary. FC12 is indeed a very harsh detergent which can be useful for proteins solubilisation from the membranes, but it should be not the best detergent for NMR sample preparation.

In order to obtain a good folded SURF1 protein it will be interesting to screen different detergents (DHPC, DDM, LDAO, OG…) that will be tested at different levels of protein purification. First new detergents can replace FC12 completely starting from the beginning of the purification protocol and can be used to solubilise SURF1 protein from the membrane. Anyway in case no other detergents will be able to solubilise the protein, FC12 can be used only during solubilisation and than other detergents should be tested during the purification protocol in the elution buffers. Moreover, it will be necessary to try different concentration of detergents in sample preparation preferably higher than 0.1%.

The right concentration should be tested easily adding the detergent directly in the NMR tube with SURF1 protein sample and acquiring an HSQC after each addition. Finally it will be important to test also different pHs for buffer.

When the conditions will be optimized it will be very interesting to obtain the solution structure of SURF1 protein by NMR to better understand the relationship between the structure and the function of this membrane protein. Another important step should be also build SURF1 pathogenic mutants to better understand their role in CcO assembly process and how the mutations are involved in CcO-deficient Leigh syndrome.

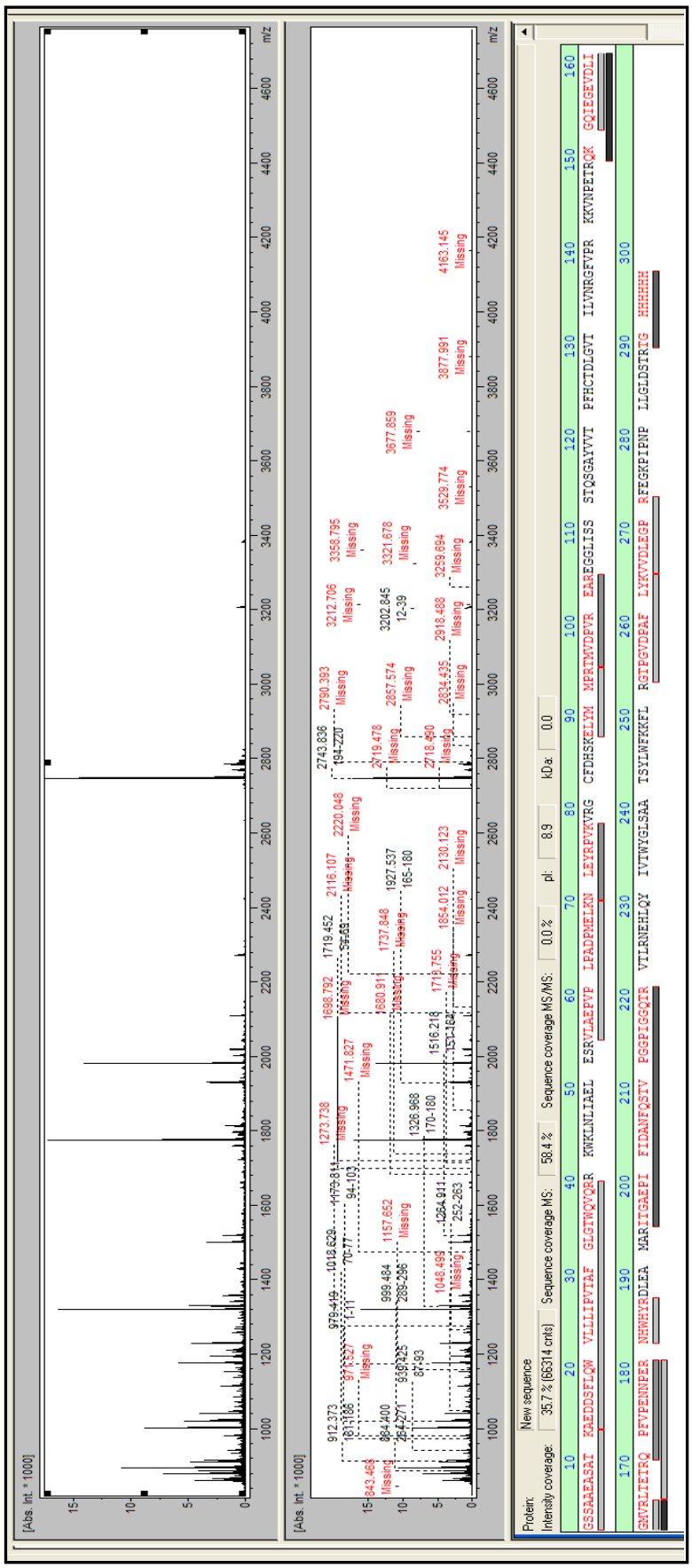


Fig.6: *MALDI. Here is shown the coverage between peptides and amino acidic sequence of SURF1 protein.*

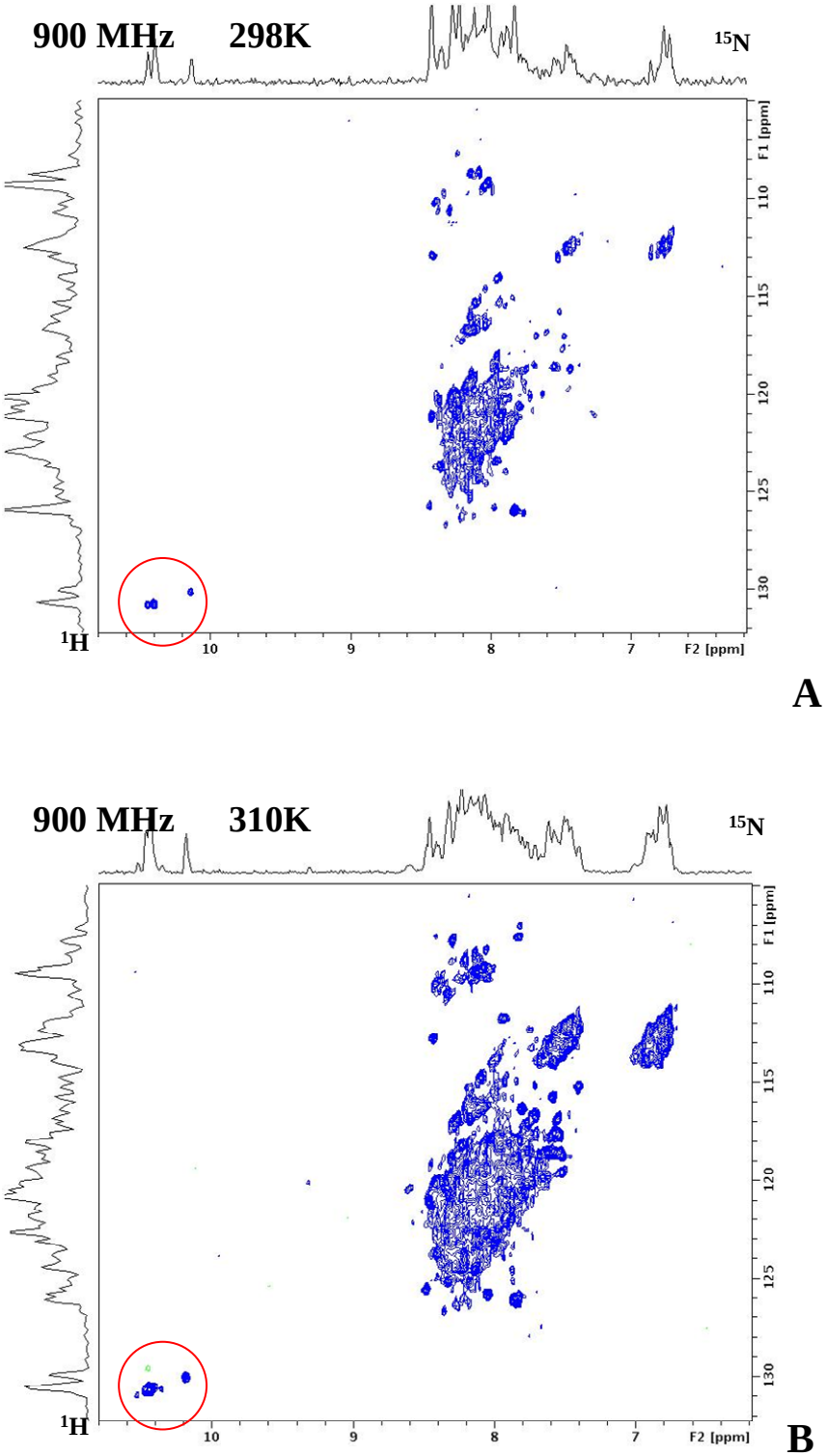


Fig. 7: *^{1}H-^{15}N HSQC spectra at 900MHz of SURF1 protein. HSQC was done at 298K (A)and 310K(B) on SURF1 sample in Phosphate buffer 20mM pH7.00 + NaCl 100mM + FC12 0,1%. $NH_{\varepsilon 1}$ of tryptophan are high lightened in red.*

4.3.2 COX11 protein

The scale up was done growing the constructs resulted positive in SDS-PAGE in one litre rich medium coltures in flasks. Cell debris obtained by centrifugation was disrupted by freeze-thaw lysis and sonication. Then the lisate was centrifuged at 40 000 rpm x 45' in order to isolate inclusion bodies from soluble part. The soluble part was then purified using affinity His tag column and eluted with buffer with increasing imidazol concentration. SDS-PAGE gels showed that in all constructs, growth both at 37°C and 25°C, COX11 was usually expressed in the inclusion bodies and no protein was detected in soluble part (In **Fig.8A** and **Fig. 8B** is shown the case of COX11 pTH27 construct in C41 cells). In order to obtain a soluble expression of the protein the coltures were growth at lower temperature, exactly at 18°C and 4°C. Also in this case the protein was detected only in inclusion bodies as shown in SDS-PAGE gels in **Fig. 8C** and **Fig. 8D** for COX11 pTH27 in C41 cells.

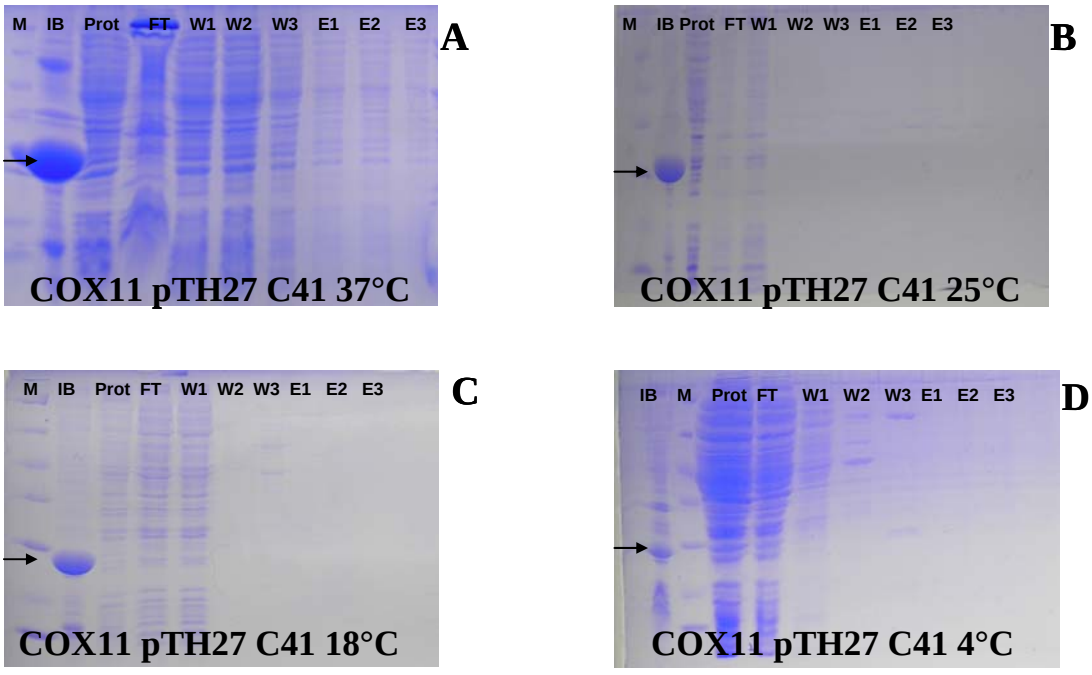


Fig.8: *SDS-PAGE gels show that COX11 is detected only in inclusion bodies at all tested temperatures (37°C in A; 25°C in B; 18°C in C; 4°C in D).*

In order to extract the protein from inclusion bodies, refolding was performed. The inclusion bodies were first washed with Triton 1% and then solubilised with Guandine HCl 6M. The first approach was refolding by dilution, which is a fast kind of refolding based on the rapid dilution of the concentrated protein. The solubilised concentrated protein was dropped slowly in a big amount of buffer Tris HCl 200mM pH 8.0 + EDTA 10mM + DTT 10mM + L-Arginine 1M. The protein diluted was then concentrated and cleaved with TEV enzyme. After purification by His tag affinity column the Circular Dycroism (CD) was performed (**Fig. 9**)

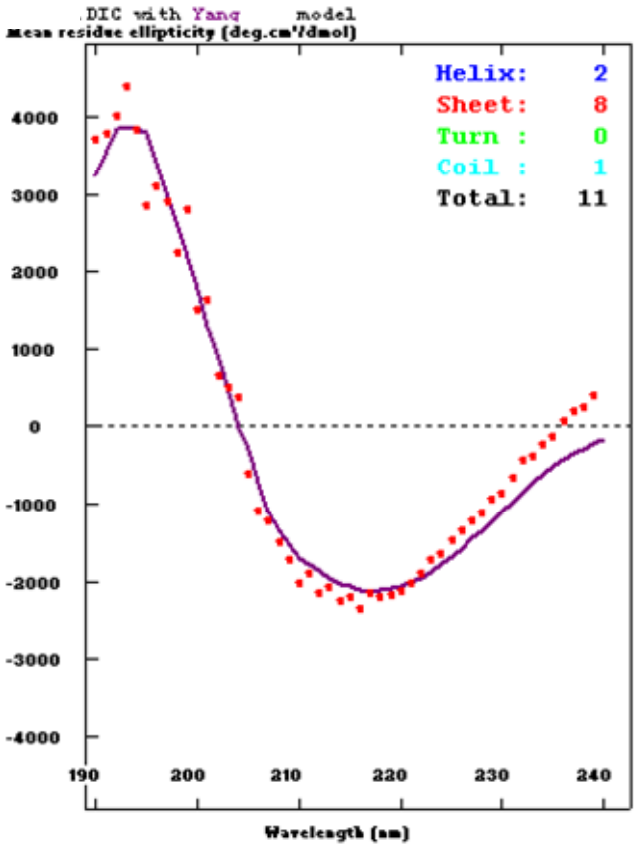


Fig. 9: *CD spectra of COX11 protein refolded by dilution.*

The Yang model shows a good fitting with 73% of β sheet, 18% of α helix and 9% of random coil. This data suggest that the protein presents an high content of secondary structure elements. In order to better characterized the protein, a COX11 sample was

prepared using phosphate buffer solution 50mM pH7.00 + NaCl 100mM + DTT 5mM and 1D spectrum was acquired at 700MHz spectrometer at 298K (**Fig. 10**).

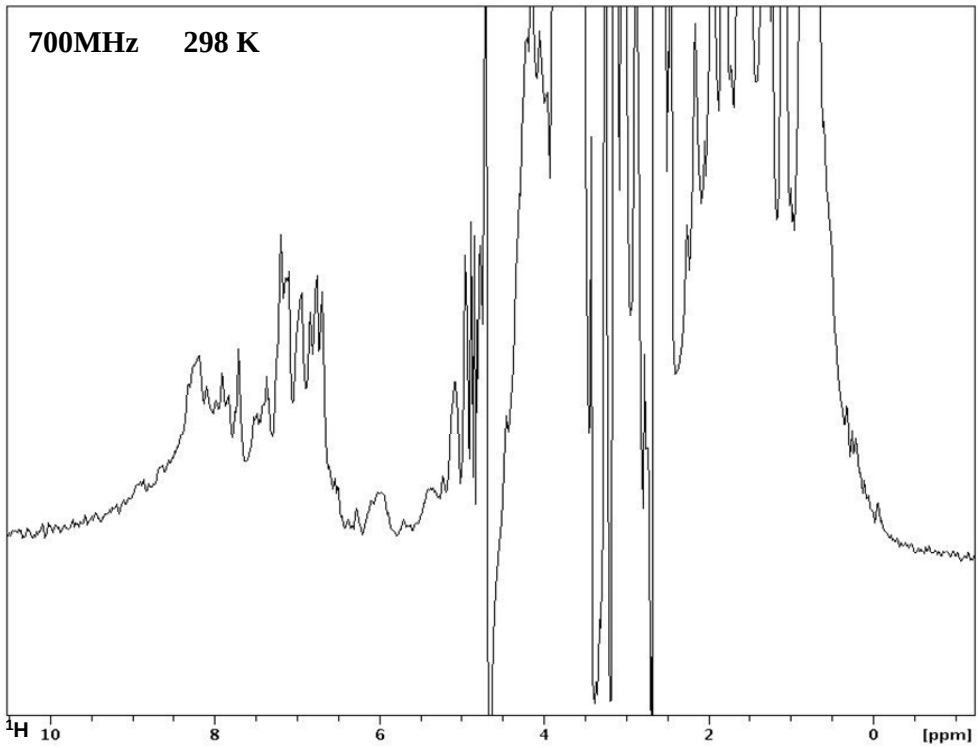


Fig. 10: *1D spectrum of COX11 refolded by dilution.*

The spectrum showed a reduced signal dispersion suggesting that the tertiary structure of the protein is not well folded. Moreover the sample presents high precipitation indicating that the protein is not stable.

The second approach consisted in a slower refolding protocol where many dialysis were performed. In this case the solubilised inclusion bodies were dialysed several times in buffers with decreasing concentration of urea (4M, 2M and 0M) and increasing concentration of L-Arginine (0.3M and 0.5M) which is necessary for protein refolding. During dialysis we observed slightly protein precipitation. Therefore a gel filtration was necessary to remove impurities and not folded protein (**Fig. 11**). The chromatogram shows a big amount of protein (higher peak) isolated from the impurity (lower peaks). The protein was cleaved subsequently with TEV enzyme and gel filtrated again to remove the His_6 tag. We carry out a CD spectrum which gives 66 % of β sheet, 13% of α helix and 21% of random coil suggesting an high content of secondary structure elements (**Fig.12**). ^{15}N labelled sample was prepared in TrisHCl 50mM pH 8.0 + NaCl 100mM + DTT 10mM and ^{1}H-^{15}N HSQC spectrum was acquired at 700MHz

spectrometer and 298K(**Fig.13**). The spectrum shows that the protein was folded but we counted a lower number of amide signals than expected.

Different samples conditions were tried to improve signal dispersion and to increase the number of signals in the ^{1}H-^{15}N HSQC spectrum(**Fig.14**, **Fig.15**). We observed that the best results has been obtained with BisTris 50mM pH 6.5 + NaCl 100mM + DTT 10mM (**Fig.16**). In spite of the high protein instability we acquired the triple resonances experiments necessary for backbone assignment.

Therefore we are now working to improve sample stability and we are trying to produce other two constructs of COX11 protein: one without the His_6 tag, which is no more necessary in case of refolding protocol; the other construct is a short one modelled on the sequence of COX11 homolog from *Sinorhizobium meliloti,* which presents 47.5% of identity as calculated with LALIGN program (http://www.ch.embnet.org/software/LALIGN_form.html) (**Fig.17**).

After the optimization of sample stability it will be interesting to solve the solution structure of COX11 because of it will help to better understand protein function. Interesting perspectives will be build COX11 cysteines mutants to understand how it binds Cu(I) ions and, consequently, how it is involved in the assembly of copper centre of CcO protein. Moreover it will be useful to study the interactions between COX11 and COX17 to understand how latter protein release Cu(I) ions to COX11 protein.

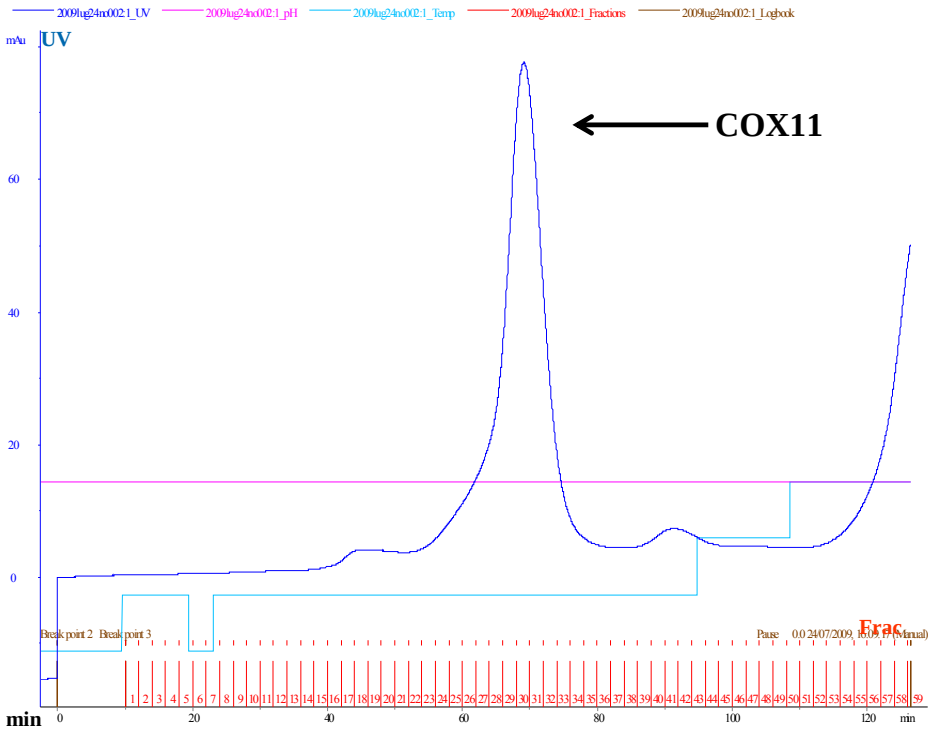


Fig.11: *Gel filtration chromatogram of COX11 refolded by dialysis.*

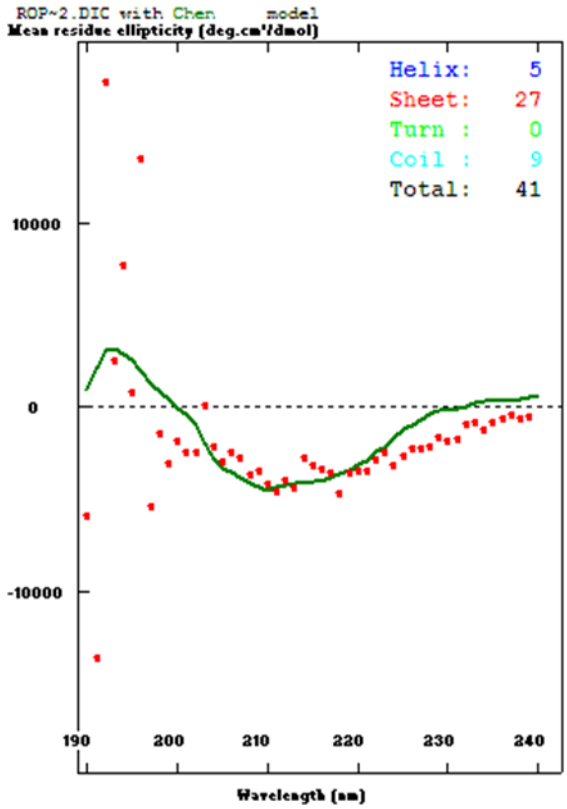


Fig.12: *CD spectrum of COX11 protein refolded by dialysis.*

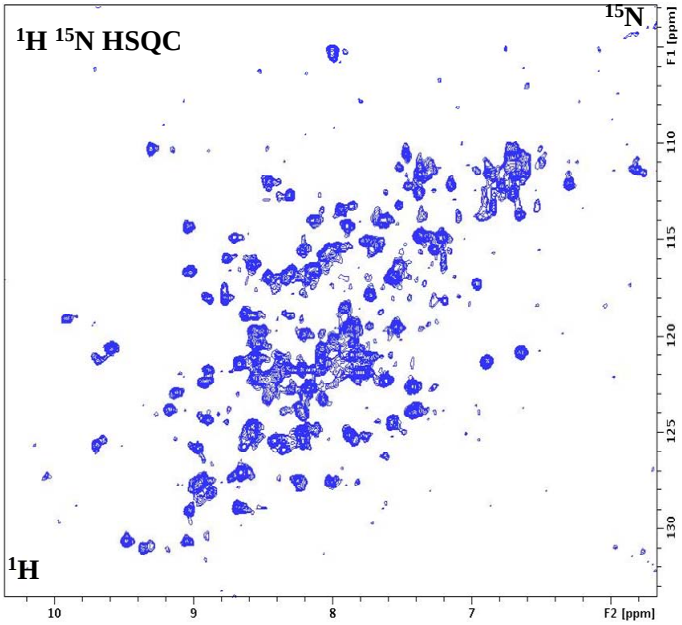


Fig.13: *^{1}H-^{15}N HSQC spectrum at 700MHz, 298K of COX11 protein. The sample is prepared in TrisHCl 50mM* ***pH 8.0*** *+ NaCl 100mM + DTT 10mM.*

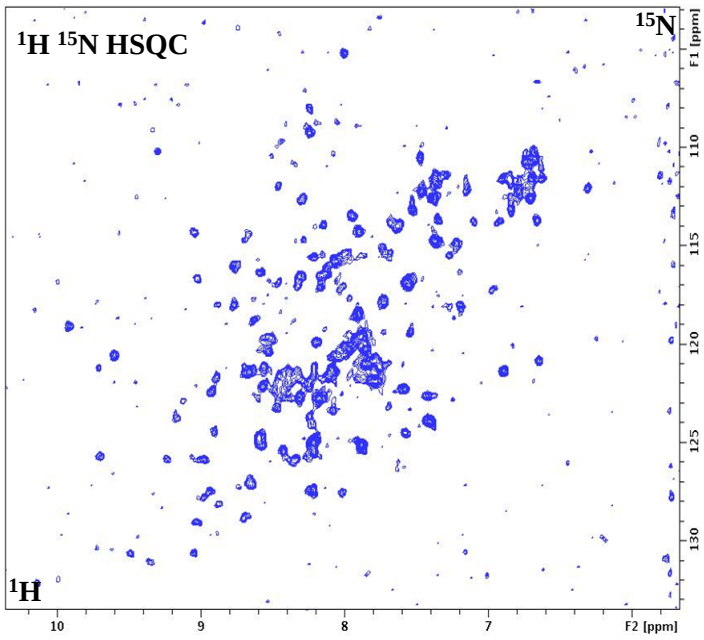


Fig.14: *^{1}H-^{15}N HSQC spectrum at 700MHz, 298K of COX11 protein. The sample is prepared in Phosphate Buffer 50mM* ***pH 7.0*** *+ NaCl 100mM + DTT 10mM.*

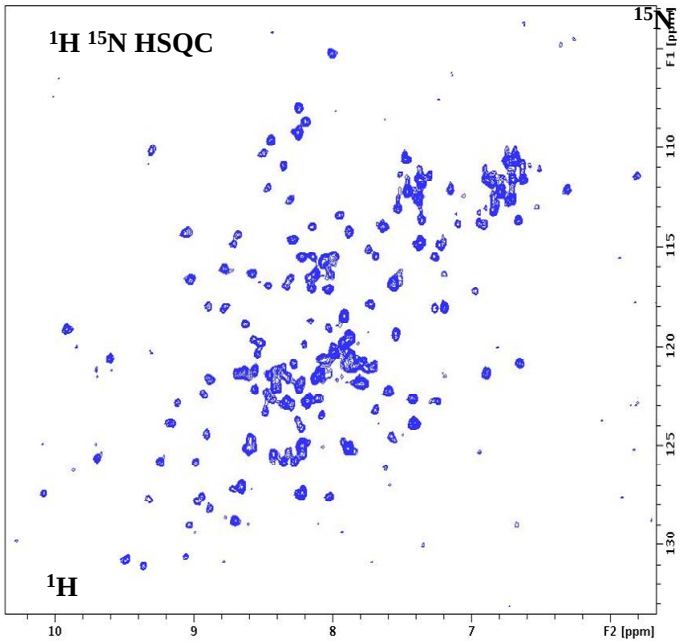


Fig.15: *^{1}H-^{15}N HSQC spectrum at 700MHz, 298K of COX11 protein. The sample is prepared in BisTris 50mM **pH 6.5** + NaCl 100mM + DTT 10mM.*

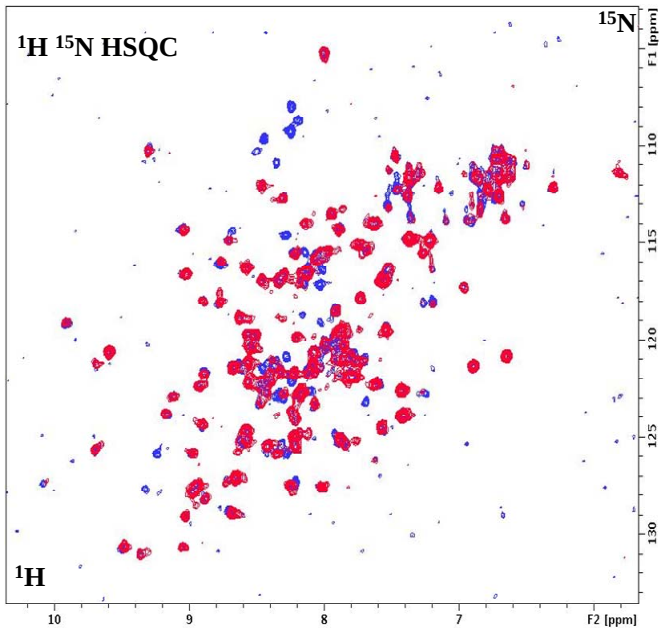
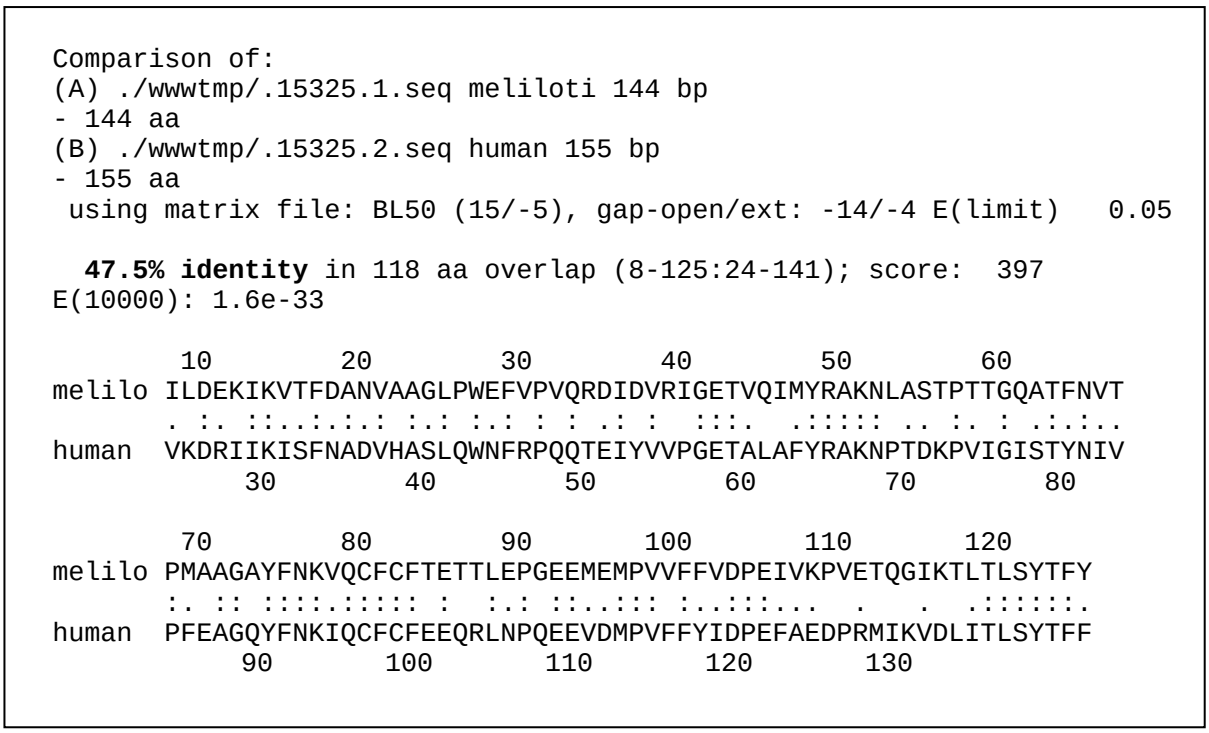


Fig.16: *Overlapping of ^{1}H-^{15}N HSQC spectrum of COX11 protein in TrisHCl 50mM **pH 8.0** + NaCl 100mM +DTT 10mM (red) and spectrum in BisTris 50mM **pH 6.5** + NaCl 100mM + DTT 10mM (blue). Both spectra are acquired at 700MHz, 298K.*



```
Comparison of:
(A) ./wwwtmp/.15325.1.seq meliloti 144 bp
- 144 aa
(B) ./wwwtmp/.15325.2.seq human 155 bp
- 155 aa
 using matrix file: BL50 (15/-5), gap-open/ext: -14/-4 E(limit)   0.05

  47.5% identity in 118 aa overlap (8-125:24-141); score:  397
E(10000): 1.6e-33

        10        20        30        40        50        60
melilo ILDEKIKVTFDANVAAGLPWEFVPVQRDIDVRIGETVQIMYRAKNLASTPTTGQATFNVT
       . :. ::.:.:.: :.: :.: : : .: :  :::.  .:::::: .. :. : .:.:..
human  VKDRIIKISFNADVHASLQWNFRPQQTEIYVVPGETALAFYRAKNPTDKPVIGISTYNIV
            30        40        50        60        70        80

        70        80        90       100       110       120
melilo PMAAGAYFNKVQCFCFTETTLEPGEEMEMPVVFFVDPEIVKPVETQGIKTLTLSYTFY
       :. :: ::::.::::: :  :.: ::..::: :..:::...  .   .  .:::::::.
human  PFEAGQYFNKIQCFCFEEQRLNPQEEVDMPVFFYIDPEFAEDPRMIKVDLITLSYTFF
            90       100       110       120       130
```

Fig.17: *comparison of Sinorhizobium meliloti COX11 sequence and Human COX11 sequence looking for local similarities using LALIGN program.*

Reference List

1. Capaldi, R. A. (1990) *Annu Rev Biochem* **59**, 569-596.

2. Tiranti, V., Hoertnagel, K., Carrozzo, R., Galimberti, C., Munaro, M., Granatiero, M., Zelante, L., Gasparini, P., Marzella, R., Rocchi, M. *et al.* (1998) *American Journal of Human Genetics* **63**, 1609-1621.

3. Zhu, Z., Yao, J., Johns, T., Fu, K., De Bie, I., Macmillan, C., Cuthbert, A. P., Newbold, R. F., Wang, J., Chevrette, M. *et al.* (1998) *Nature Genet.* **20**, 337-343.

4. Mashkevich, G., Repetto, B., Glerum, D. M., Jin, C. & Tzagoloff, A. (1997) *Journal of Biological Chemistry* **272**, 14356-14364.

5. Yao, J. B. & Shoubridge, E. A. (1999) *Human Molecular Genetics* **8**, 2541-2549.

6. Poyau, A., Buchet, K. & Godinot, C. (1999) *FEBS Lett.* **462**, 416-420.

7. Bundschuh, F. A., Hannappel, A., Anderka, O. & Ludwig, B. (2009) *Journal of Biological Chemistry* **284**, 25735-25741.

8. Tzagoloff, A., Capitanio, N., Nobrega, M. P. & Gatti, D. (1990) *EMBO J.* **9**, 2759-2764.

9. Carr, H. S., George, G. N. & Winge, D. R. (2002) *J. Biol. Chem.* **277**, 31273-31242.

10. Petruzzella, V., Tiranti, V., Fernandez, P., Ianna, P., Carrozzo, R. & Zeviani, M. (1998) *Genomics* **54**, 494-504.

Chapter 5

Familial Amyotrophic Lateral Sclerosis

Expression and characterization of human CCS protein

5.1 Introduction

Familial Amyotrophic Lateral Sclerosis (FALS) is a motor neuron disease characterized by a progressive muscle weakness and atrophy. 5 to 10% of all cases of Amyotrophic Lateral Sclerosis (ALS) are constituted by familial cases (FALS) which are inherited in a dominant manner (1, 2). Each FALS type is characterized by mutation in a specific gene. In particular, type 1, is caused by autosomal dominant mutations in the gene *SOD1*, that encodes copper zinc superoxide dismutase protein (SOD1)(3). The latter is an antioxidant enzyme found in the cytosol, nucleus, peroxisomes, and mitochondrial intermembrane space of eukaryotic cells (4-6). The human enzyme is a 32-kDa homodimer, with one copper- and one zinc-binding site per 153-amino acid subunit and an intrasubunit disulfide bond. The dimer interface is stabilized by numerous main-chain to main-chain hydrogen bonds, water-mediated hydrogen bonds, and hydrophobic contacts (7, 8).

The metallation of the SOD1 polypeptide is aided by CCS (copper chaperone for SOD1) (9, 10). CCS, which is able to bind copper ions and one zinc ion, is implicated in the insertion of the copper co-factor and in the formation of the disulfide bond during insertion of copper into SOD1 in yeast(11). CCS protein is ubiquitously expressed in all human tissues and presents three functional distinct domains (**Fig.1**): the N-terminal domain is homologous to ATX1 copper chaperone with a MXCXXC copper binding motif (12); the second domain has a high homology to SOD1 and is able to recognize the SOD1 protein allowing copper transfer (13). The last domain presents a CXC motif able to bind copper(14). This domain, that results disordered in the crystal structure of yeast CCS(13), is crucial for SOD1 activation. It has been hypothesized that domain III lies near domain I and together then insert copper in SOD's site (14).

Yeast mutants lacking CCS express an apo form of SOD1 protein without copper ions (15) and mice with disruptions in CCS gene exhibit marked reductions in SOD1 activity (16). Moreover CCS has been implicated in SOD1-related FALS disease showing *SOD1* gene missense mutation along all protein sequence. Indeed an altered maturation of SOD1 mutants can be due to incorrect copper insertion or disulfide formation CCS-mediated processes.

The aim of this work was to obtain the full length CCS protein in order to characterize the solution structure, to analyze its copper binding properties and the mechanism of copper transfer to copper depleted ZnSOD1.

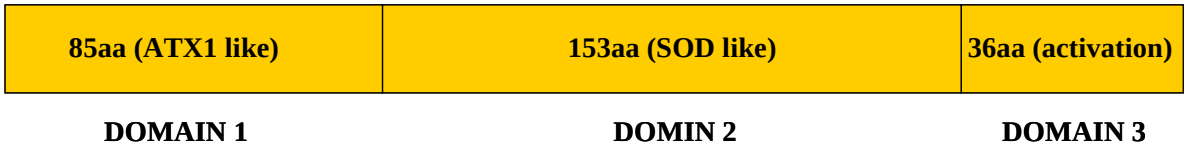


Fig.1: *A scheme of the three domains of CCS protein.*

5.2 Materials and methods

The DNA segment of full length human CCS (274aa) were amplified from cDNA by PCR. Primers with C-terminal TEV site or with N-terminal TEV site or without TEV site were used with different kind of polymerase :

- Phusion Hot Start (BioLabs), +/- $MgCl_2$, 60°C/64°C
- KOD Polymerase(Novagen), +/- $MgCl_2$, 60°C
- Accuprime Pfx(Invitrogen), +/- $MgSO_4$, +/- DMSO, 60°C

Using the latter with $MgSO_4$ and the Touchdown Polymerase chain reaction was obtained a good amplification of the CCS DNA segment. The latter was then cloned in the Gateway Entry vector pDONR 221 by BP Gateway reaction and then subcloned with LR Gateway reaction in pTH27, PMM (N-terminal 6-His tag) and pTH24 (C-terminal 6-His tag) plasmids. The best expression was observed with the construct containing the TEV site in pTH27 plasmid transformed in BL21(DE3) Codon Plus expression cells. Scale up colture was done using fermentors (Inforce) at 25°C. The expression was induced by 0.7mM IPTG and $ZnSO_4$ 1mM was added after one hour directly in the colture. Cells were disrupted by sonication and the protein was purified using Nickel sepharose resin (Amersham). The His-tag was cleaved with TEV protease and removed using HisTrap chelating column.

NMR sample was prepared with different buffers in order to optimize sample conditions:

- Phosphate Buffer 50mM pH 7.0
- Bis Tris 50mM pH 6.5
- Bis Tris 50mM+ 5mM EDTA pH 6.5

The DNA segment of the first two domains of CCS protein (238aa) was also amplified by PCR using TAQ polymerase (Invitrogen). The obtained fragments were then cloned in the Gateway Entry vector pENTR /TEV/D-TOPO to include the TEV protease cleavage site at the N-terminal end. The segments were finally subcloned into destination vector pTH34 (GB1 tag) through the Gateway LR reaction and plasmid expressing the protein His-tagged at the N-terminus in Origami pLysS expression cells were selected. The scale up was done growing colture in rich medium flasks at 37°C. 0.5mM IPTG was used for induction and $ZnSO_4$ 1mM was added after one hour directly in the colture. Cell debris was disrupted by sonication and the protein was purified using Nickel sepharose resin (Amersham). The protein was cleaved with TEV enzyme to remove His_6 and GB1 tag and purified with HisTrap chelating column. ^{15}N labelled samples of full length and two domains constructs were prepared growing colture in minimum medium with the addiction of glycerol 50%. Bis Tris 50mM pH 6.5 and Bis Tris 50mM+ 5mM EDTA pH 6.5 were the selected buffer for sample preparation.

5.3 Results and discussion

Amplification of the DNA segment of the full length CCS protein was obtained by PCR. Different kind of polymerase were tested in order to obtain the amplification because of the length of the primers necessary to perform the BP Gateway reaction. The gene was amplified with Accuprime Pfx (Invitrogen) with $MgSO_4$ using a particular kind of PCR, Touchdown Polymerase chain reaction, during which the annealing temperature is decreased in increments for every subsequent set of cycles, in order to start with not very specific amplification and to obtain at the end of the PCR only specific amplifications. These amplified genes were cloned in the Gateway Entry vector pDONR 221 by BP Gateway reaction, in order to subclone all in different destination vectors (pTH27, PMM and pTH24) which were prepared in Karolinska Institutet laboratories and kindly made available from Pär Nordlund's group. Different kind of expression cells were then tested resulting pTH27 construct in BL21(DE3) Codon Plus cells as the one with high expression levels of CCS protein. The scale up was done growing an over night colture in fermentors at 25°C. Cells obtained by centrifugation were disrupted by sonication and the soluble part was isolated from inclusion bodies by ultracentrifugation. The protein was binded to Nickel sepharose resin (Amersham) by one hour agitation in cold room and then eluted from the packed resin using different

imidazol concentrations buffers. After cleavage with TEV enzyme the His_6 tag was removed using HisTrap chelating column (**Fig.2**). A gel filtration was performed using Superdex 75 to further purification of the sample and the chromatogram shows the presence of one form of CCS protein (**Fig.3**).

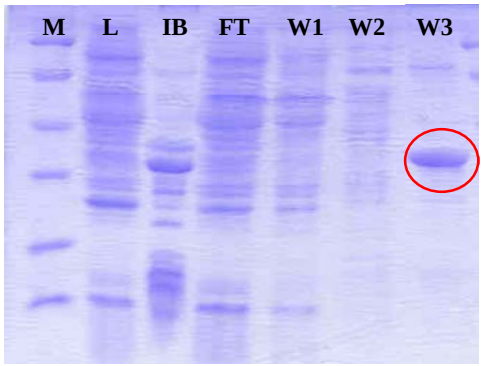


Fig. 2: *SDS-PAGE of CCS protein purified after TEV cleavage.*

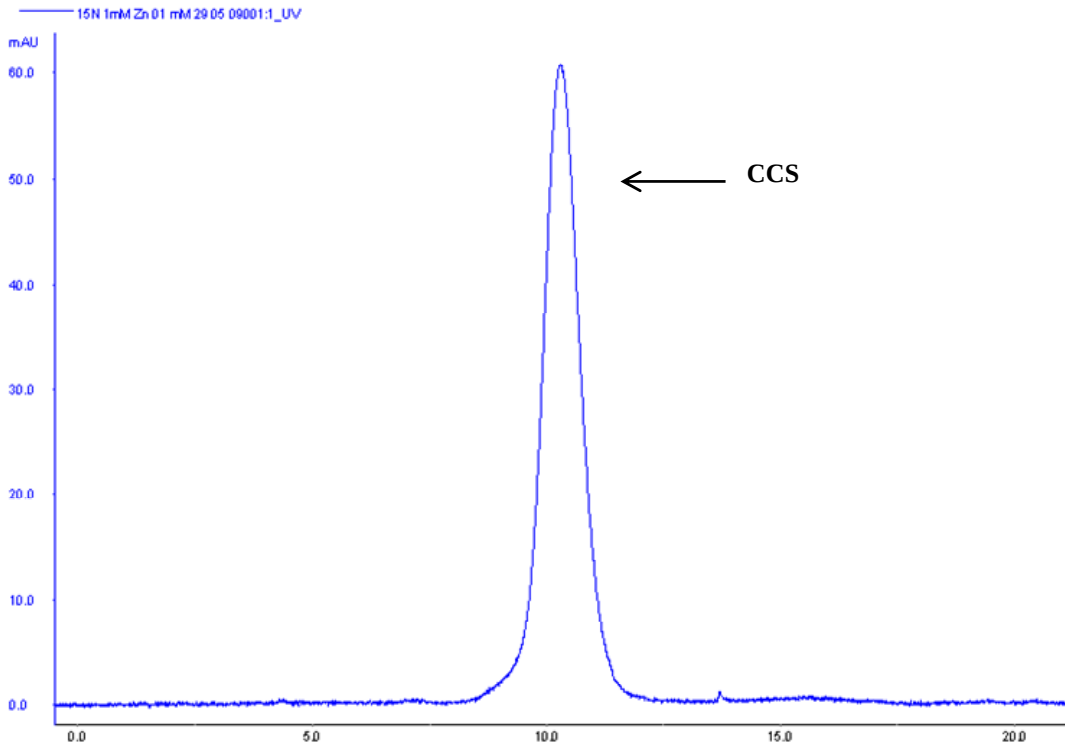


Fig.3: *Gel filtration chromatogram of CCS full length protein.*

To detect if it is a monomer (29040 Da) or a dimer (58080 Da) Light Scattering was performed clearly demonstrating that the CCS protein is in the dimeric form (**Fig.4**).

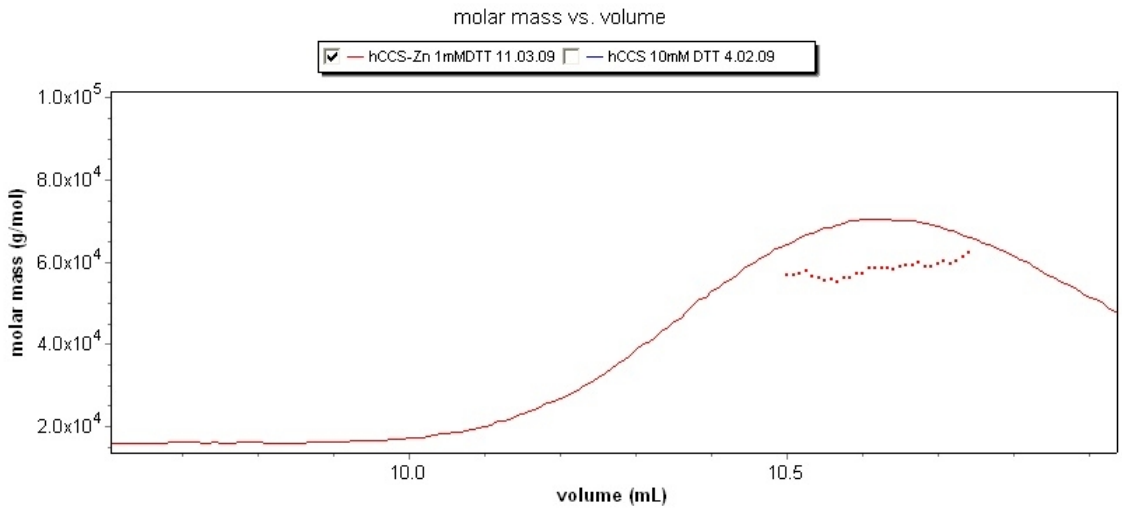


Fig.4: *Light Scattering spectrum shows that CCS is in a dimeric form.*

Atomic Absorption spectroscopy was then performed for determining the concentration of metals in a sample, with the result of 100% Zn and less than 5% Cu.

ESI mass spectrometry also shows a main peak of dimeric forms and a minor peak of the monomeric form with major mass, corresponding to CCS (theoretical) + 1Zn + 1Cu.

ESI mass was also done in acidic conditions showing two or more protein forms, with different molecular masses, where the major form is CCS - 1 Zn - 1 Cu. This result shows that two metal ions (one Zn and one Cu) are bound to each monomer subunit of CCS in the sample. Addition of Cu(I) does not lead to appearance of additional metal ion binding by ESI mass spectrometry, which suggests that protein is saturated with metal ions.

Titration of CCS + Cu sample with DTT reveals dissociation of one Cu ion at around 7 mM DTT, thus resulting an affinity of 2.4 fM. Only Cu(I) ion is dissociated in these conditions, whereas the Zn ion remains to be bound to CCS (**Fig.5**).

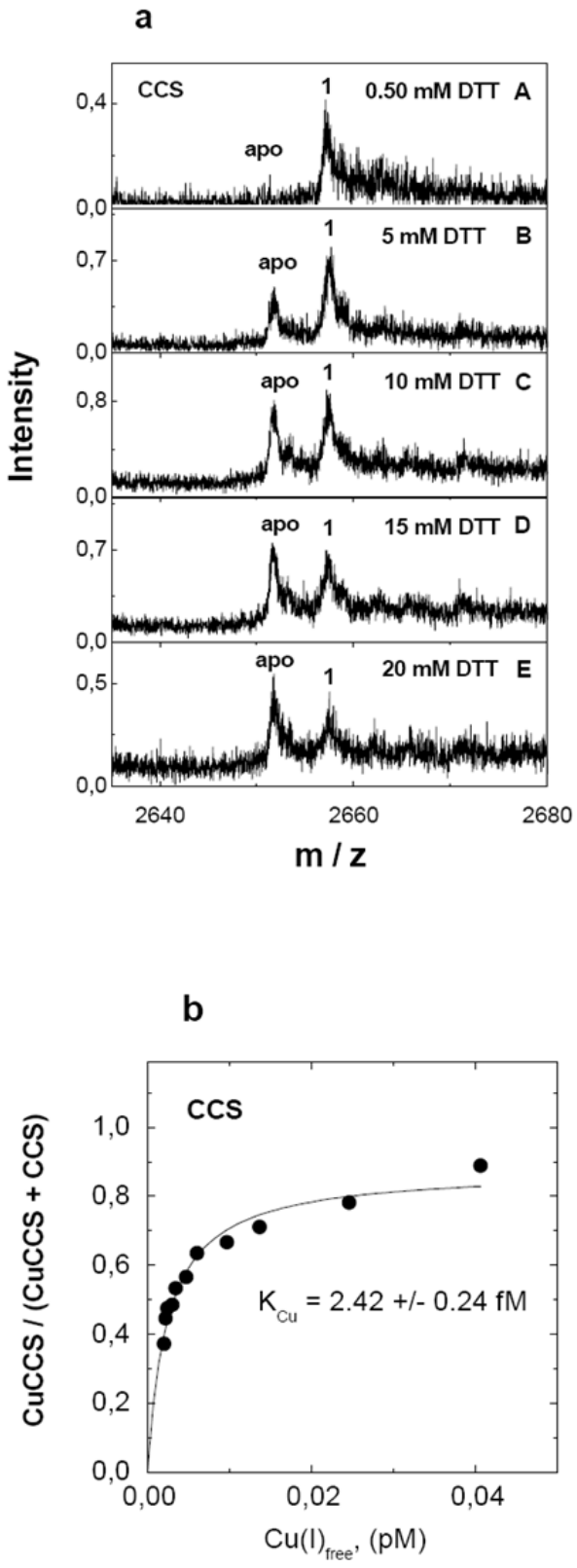


Fig.5: *a) ESI mass spectra repeated during titration of CCS full length sample with DTT show the increasing amount of apo form of CCS. b) Fitting of the affinity affinity between CCS and Cu(I). The graph shows the relation between the CCs-Cu(I) complex and the formation of free Cu(I) due to the presence of DTT.*

Apparently it needs higher concentration of DTT but at higher than 30 mM concentrations of DTT peaks, which are not very intensive, disappear. CCS has high molecular weight and it is not seen as proteins with lower molecular weights. From the data anyway it does not follow from which domain of CCS metal is dissociated.

To better characterize the structural properties of CCS protein a ^{15}N labelled sample was prepared in Phosphate Buffer 50mM pH 7.0 and ^{1}H-^{15}N HSQC spectrum was acquired at 500MHz spectrometer at 310 K (**Fig.6**). Several cross-peaks have resonances typical of folded protein, but their number are less than expected for a protein of 238aa. Moreover several cross-peaks are found in the typical spectral region of unfolded proteins. Therefore we try to optimize the sample conditions changing pH and buffer condition. The NMR sample was prepared in Bis Tris 50mM pH 6.5 and the ^{1}H-^{15}N HSQC was acquired at 900MHz spectrometer both at 298K (**Fig. 7**) and 310K (**Fig.8**). The spectra show that cross-peaks are sharper at pH 6.5 and at 310K thus obtaining an improvement in the number of peaks. In particular we can detect a cross-peak corresponding to the typical SOD1 domain Glycine residue. A futher improvement was obtained adding 5mM EDTA directly in the NMR tube of CCS sample. Indeed, ^{1}H-^{15}N HSQC performed at 900MHz and at 310K shows a large increment in cross-peaks number (**Fig.9**).

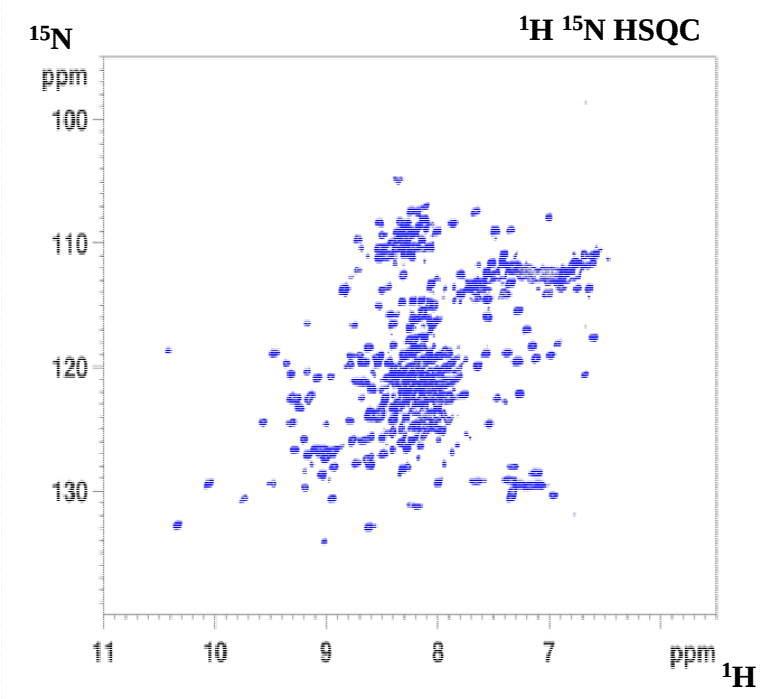


Fig.6: *^{1}H-^{15}N HSQC spectrum at 500MHz, 310K of CCS protein. The sample is prepared in Phosphate Buffer 50mM pH 7.0.*

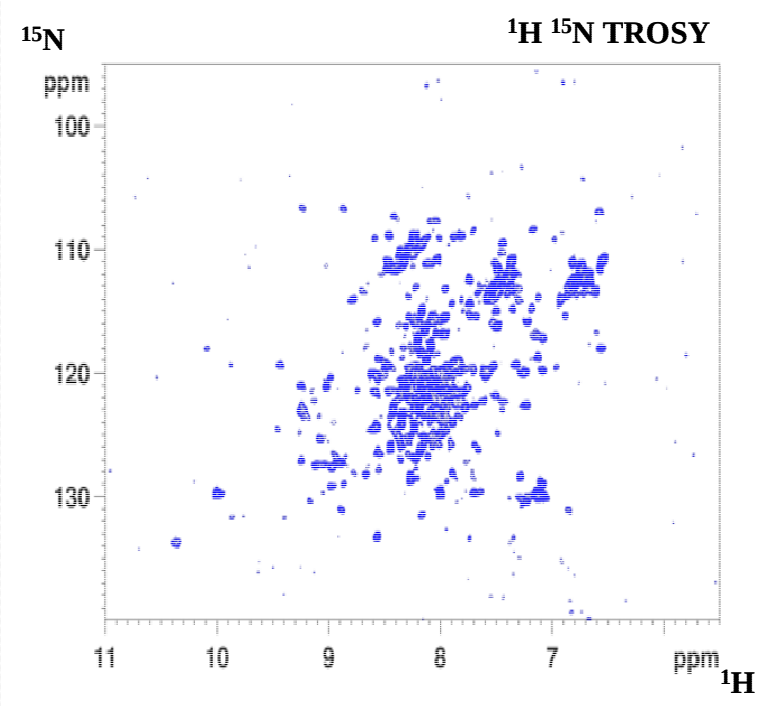


Fig.7: *^{1}H-^{15}N HSQC spectrum at 900MHz, 298K of CCS protein. The sample is prepared in Bis Tris 50mM pH 6.5.*

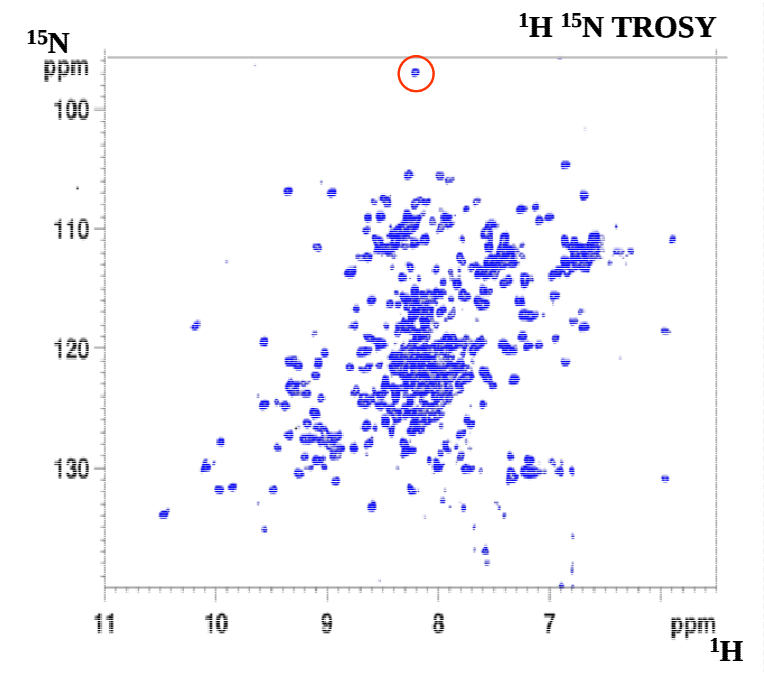


Fig.8: *^{1}H-^{15}N HSQC spectrum at 900MHz, 310K of CCS protein. The sample is prepared in Bis Tris 50mM pH 6.5. In red is high lightened a Glycine typical of SOD spectrum.*

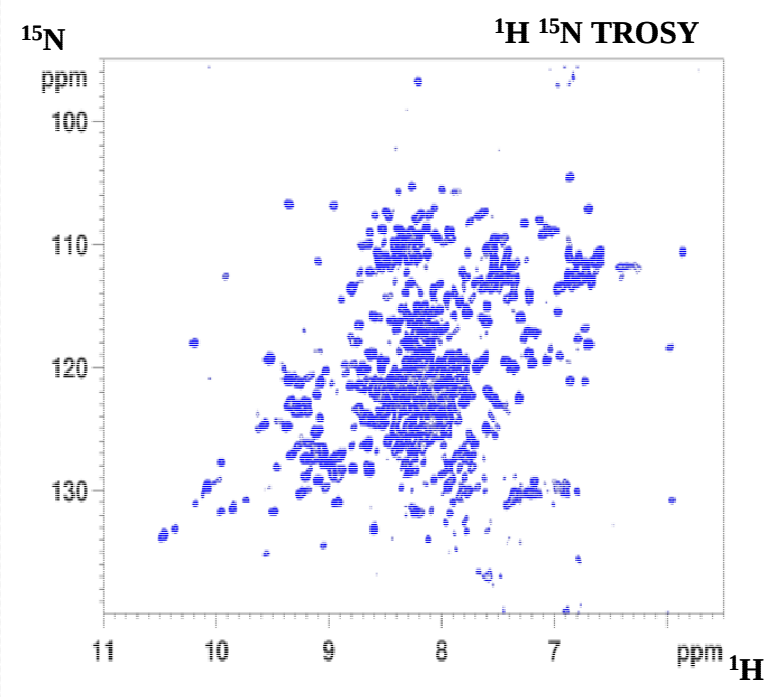


Fig. 9: *^{1}H-^{15}N HSQC spectrum at 900MHz, 310K of CCS protein. The sample is prepared in Bis Tris 50mM pH 6.5+5mM EDTA*

Anyway a part of the protein still remains unfolded. Since it is known that in the crystal structure of CCS the third domain is disordered, we hypothesized that the unfolded part observed in the ^{1}H-^{15}N HSQC map should belong to this domain. For that reason we study a short construct of CCS protein comprehensive only the first two folded domains (CCS1-2). The 238aa long gene segment was amplified by PCR and cloned with TOPO-TEV reaction in pENTR/TEV/D-TOPO plasmid. Using the LR gateway reaction the gene was then subcloned in different destination plasmids and transformed in different strains. CCS1-2 construct in pTH34 expression plasmid (characterized by a GB1 tag) transformed in Origami pLysS expression cells was the only one showing high soluble expression levels. Cells were disrupted by sonication and purification was performed using Nickel sepharose resin. Protein eluted from the packed resin using different imidazol concentrations was cleavage with TEV enzyme and the protein was finally purified using HisTrap chelating column. ESI Mass Spectrometry on the protein gave the right weight of the protein (25114Da). Gel filtration analysis show only one peak corresponding to the dimeric protein state as proved by light scattering analysis.

^{1}H-^{15}N HSQC spectra in BisTris 50mM pH6.5 were performed at 900MHz spectrometer at different temperatures: 285K(**Fig.10**), 298K(**Fig.11**) and 310K(**Fig.12**).

Frome the analysis of the spectra we can conclude that at higher temperatures the number of cross-peaks improves a lot; at 285K only few cross-peaks belonging to unstructured protein segment can be detected, wile the cross-peaks number increases at 298K and even more in the one at 310K, in which again the typical SOD domain like Gycine was observed(**Fig.13)**. As already tested for the full length construct we obtained a further improvement by adding 5mM EDTA in the sample. The ^{1}H-^{15}N HSQC spectrum performed at 900MHz spectrometer at 310K is overlapped with the one without EDTA and it clearly shows an increasing number of amide signals (**Fig.14).**

After sample optimization of both construct we are now producing $^{2}H^{13}C^{15}N$ labelled sample for acquisition of triple resonance experiments for the assignment with the final goal of solving the full length solution structure of human CCS protein. The triple labelled sample of CCS 1-2 will be done to facilitate the backbone resonance assignment. It will be interesting to study the interaction between CCS and copper and to build cysteines residues mutants to understand in detail the functions of the three domains. Moreover these data will be necessary to study the interaction of CCS with SOD1 protein to characterize at the molecular level how CCS is involved in SOD1 maturation.

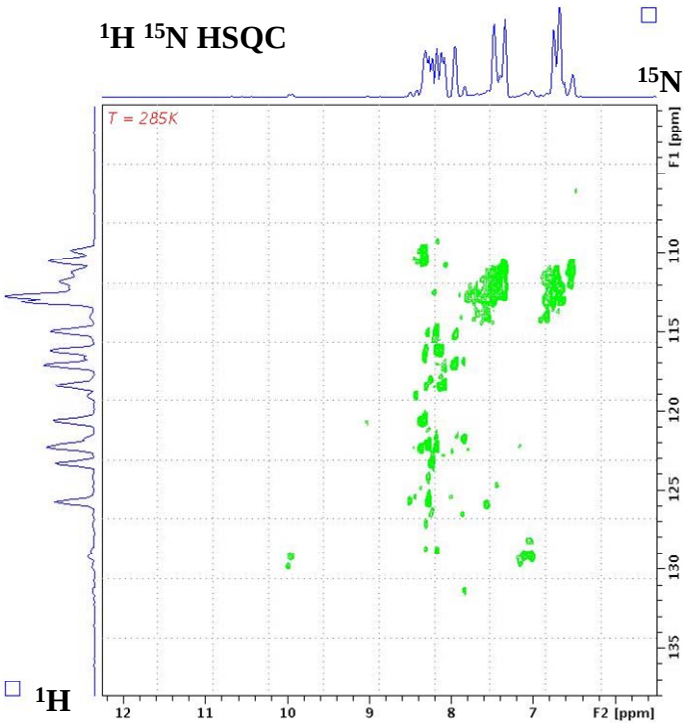


Fig.10: *^{1}H-^{15}N HSQC spectrum at 900MHz, 285K of CCS1-2 protein. The sample is prepared in Bis Tris 50mM pH 6.5.*

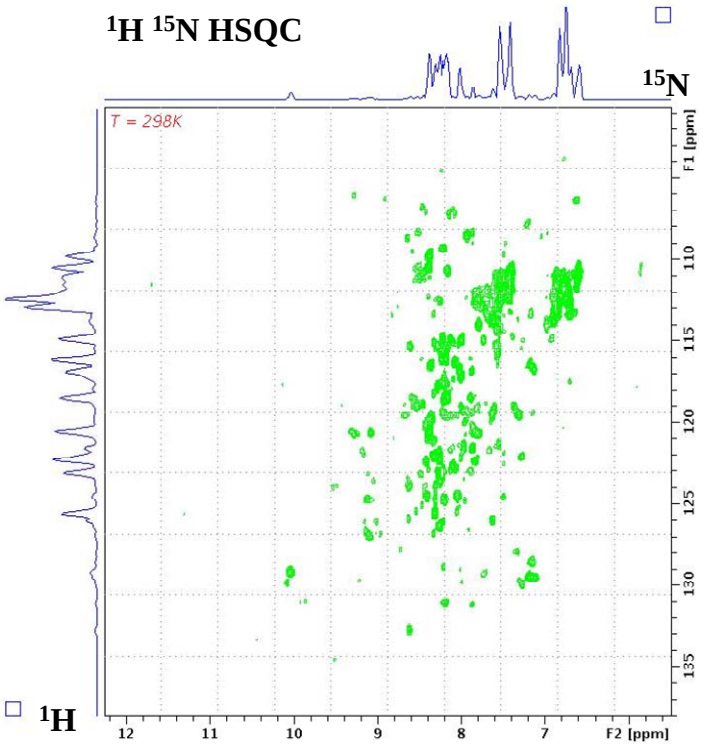


Fig.11: *^{1}H-^{15}N HSQC spectrum at 900MHz, 298K of CCS1-2 protein. The sample is prepared in Bis Tris 50mM pH 6.5.*

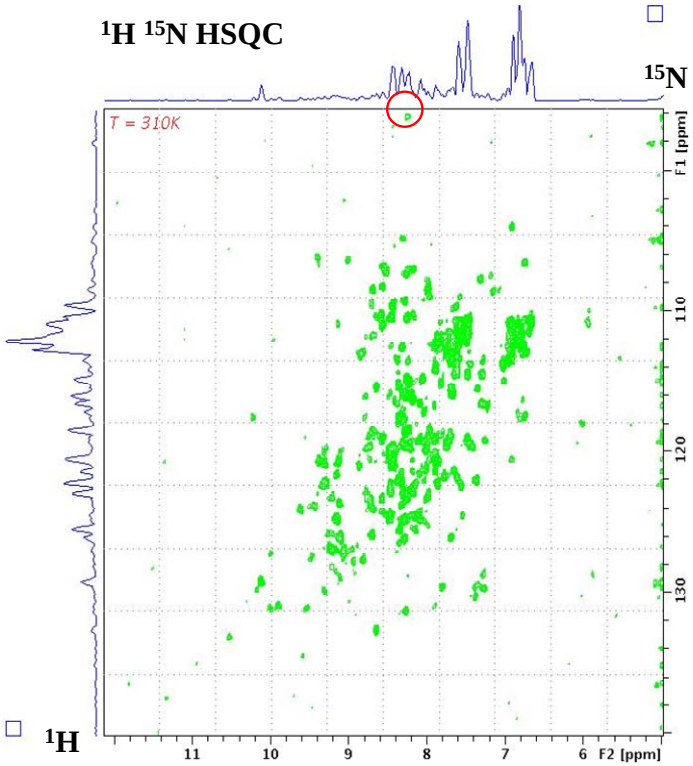


Fig.12: *^{1}H-^{15}N HSQC spectrum at 900MHz, 310K of CCS1-2 protein. The sample is prepared in Bis Tris 50mM pH 6.5. In red is high lightened a Glycine typical of SOD spectrum.*

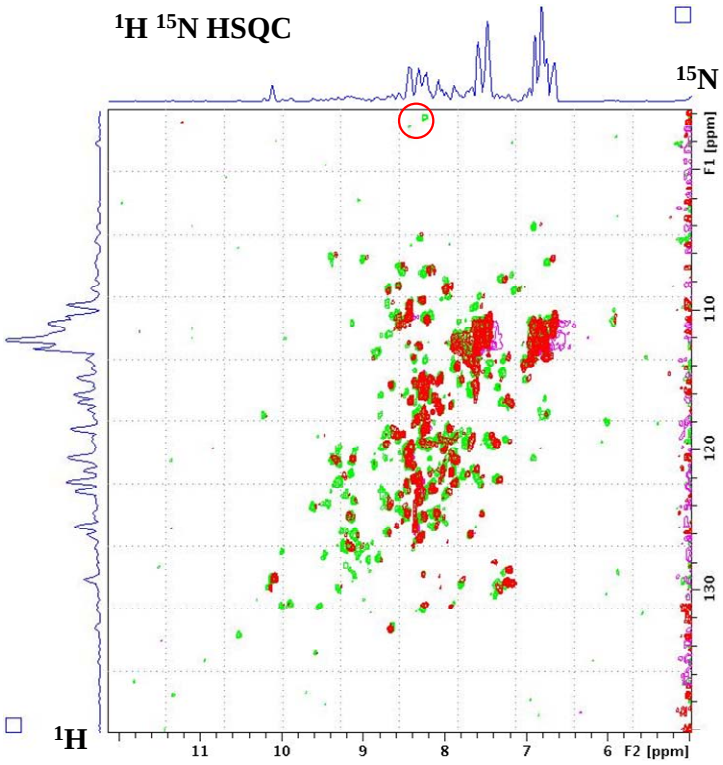


Fig.13: *Overlapping of ^{1}H-^{15}N HSQC spectrum of CCS1-2 protein in BisTris 50mM pH 6.5 at 298K (red) and the one at 310K (green). Both spectra are acquired at 900MHz spectrometer.*

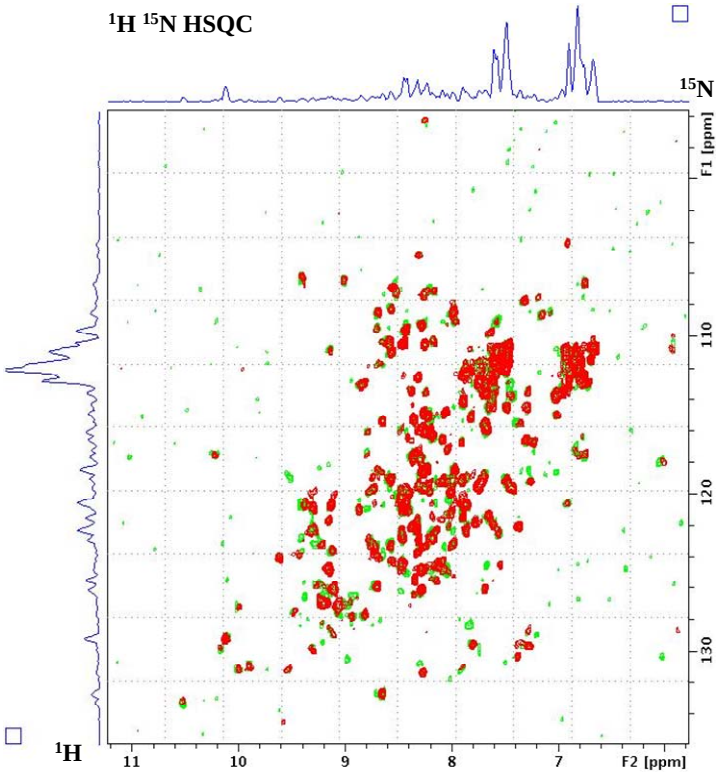


Fig.14: *Overlapping of ^{1}H-^{15}N HSQC spectrum of CCS1-2 protein in BisTris 50mM pH 6.5 (red) and the spectrum of the sample with the addiction of 5mM EDTA (green). Both spectra are acquired at 900MHz spectrometer, at 310K.*

Reference List

1. Bruijn, L. I., Miller, T. M. & Cleveland, D. W. (2004) *Annu. Rev. Neurosci.* **27**, 723-749.

2. Valentine, J. S., Doucette, P. A. & Potter, S. Z. (2005) *Annu Rev Biochem* **74**, 563-593.

3. McCord, J. M. & Fridovich, I. (1969) *J. Biol. Chem.* **244**, 6056-6063.

4. Fridovich, I. (1997) *J. Biol. Chem.* **272**, 18515-18517.

5. Okado-Matsumoto, A. & Fridovich, I. (2001) *J. Biol. Chem.* **276**, 38388-38393.

6. Sturtz, L. A., Diekert, K., Jensen, L. T., Lill, R. & Culotta, V. C. (2001) *J. Biol. Chem.* **276**, 38084-38089.

7. Bertini, I., Mangani, S. & Viezzoli, M. S. (1998) in *Advanced Inorganic Chemistry*, ed. Sykes, A. G. (Academic Press, San Diego, CA, USA), pp. 127-250.

8. Hough, M. A., Grossmann, J. G., Antonyuk, S. V., Strange, R. W., Doucette, P. A., Rodriguez, J. A., Whitson, L. J., Hart, P. J., Hayward, L. J., Valentine, J. S. *et al.* (2004) *Proc. Natl. Acad. Sci. USA* **101**, 5976-5981.

9. O'Halloran, T. V. & Culotta, V. C. (2000) *J. Biol. Chem.* **275**, 25057-25060.

10. Elam, J. S., Taylor, A. B., Strange, R., Antonyuk, S., Doucette, P. A., Rodriguez, J. A., Hasnain, S. S., Hayward, L. J., Valentine, J. S., Yeates, T. O. *et al.* (2003) *Nature Struct. Biol.* **10**, 461-467.

11. Furukawa, Y., Torres, A. S. & O'Halloran, T. V. (2004) *EMBO J.* **23**, 2872-2881.

12. Cozzolino, M., Amori, I., Pesaresi, M. G., Ferri, A., Nencini, M. & Carri, M. T. (2008) *J. Biol. Chem.* **283**, 866-874.

13. Lamb, A. L., Wernimont, A. K., Pufahl, R. A., Culotta, V. C., O'Halloran, T. V. & Rosenzweig, A. C. (1999) *Nature Struct. Biol.* **6**, 724-729.

14. Schmidt, P. J., Rae, T. D., Pufahl, R. A., Hamma, T., Strain, J., O'Halloran, T. V. & Culotta, V. C. (1999) *J. Biol. Chem.* **274**, 23719-23725.

15. Gamonet, F. & Lauquin, G. J. M. (1998) *Eur. J. Biochem.* **251**, 716-723.

16. Wong, P. C., Waggoner, D., Subramaniam, J. R., Tessarollo, L., Bartnikas, T. B., Culotta, V. C., Price, D. L., Rothstein, J. & Gitlin, J. D. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 2886-2891.

Chapter 6

Summary and perspectives

Summary and perspectives

Information about the three dimensional structure of biological macromolecules as proteins is vital for understanding their function at the molecular level. The role of structural biology has become increasingly important in many fields of the life science and has gained great importance also in drug design. NMR and X-ray crystallography are the methods used to obtain 3D structures of macromolecules. In particular biomolecular NMR spectroscopy is the most important method to obtain information at the atomic level in solution, providing not only structural data but also information about conformational dynamics and exchange processes, internal mobility and dynamics of biomolecules at timescales ranging from picoseconds to seconds. NMR is very efficient in mapping interactions with other molecules, e.g. protein/protein, protein/nucleic acid, protein/metal ions or ligand interactions. Furthermore it should be kept in mind that a large majority of the functional processes in the cell occurs through weak transient interactions among proteins (interactome). Unfortunately, the characterization at the atomic level of the weak interactions is difficult, due to their dynamic nature. NMR is the more appropriate technique to characterize these weak protein-protein interaction processes and it is very efficient in mapping interaction surfaces of protein/protein complexes.

Anyway, biophysical investigations of macromolecules usually rely on the availability of a purified and stable protein samples. In fact, an inhomogeneous preparation and/or aggregation of the protein target may severely compromise its structural characterization. The first step to perform an NMR study therefore involves optimization of the measurement conditions as pH, ionic strength, and temperature that can more appropriately be adjusted to mimic physiological conditions. Furthermore, in the case of membrane proteins, which I have produced during my PhD work, the choice of detergent and the path by which the protein is transferred from organic solvents into detergent micelles can greatly affect the structure and aggregation state of the transmembrane domains, and hence their suitability for NMR structure determination. During my PhD studies, I acquired knowledge on the production and purification of membrane proteins; in particular, my stay at the Karolinska Institutet in Stockholm allowed me to approach advanced high-throughput techniques in gene cloning and in the handling of membrane proteins.

In this PhD work I was involved in the production of four different proteins implicated in rare diseases. Different protocols were used on the basis of their features in order to obtain proper quantity and purity necessary for NMR characterization of them.

The first protein I have studied is the Wilson protein (ATP7B), an important human copper(I)-transporting P-type ATPase involved in copper-regulated trafficking between the trans-Golgi network (TGN) and plasma membranes. Mutation in the gene *ATP7B* give rise to Wilson disease (WD) which is an autosomal recessive copper toxicosis condition.

The N-terminal tail of ATP7B protein was obtained to study the behaviour of the six metal binding domains in presence of Cu(I) and Cu(I)-HAH1 complex in order to understand the functional role of the N-terminal tail in copper trafficking and in Wilson disease. Moreover a short construct comprehensive of the first two domains of the protein and a cysteines mutated ATP7B protein were obtained for further analysis by which we conclude that domains 1, 2 and 4 form an intermolecular Cu(I)-bridged adduct with HAH1, while domains 3, 5 and 6 remove Cu(I) from the metallochaperone HAH1 independently from the adduct formation.

SURF1 and COX11 proteins were studied in order to understand their role in CcO assembly mechanism. SURF1 membrane protein was solubilised with FC12 detergent and high purification was obtained even if NMR spectrum suggest that screening of different detergents and concentrations are necessary to obtained a folded protein. The resolution of SURF1 protein structure should be the starting point to study how it is involved in CcO assembly process. Moreover build SURF1 mutants should help to obtain information about their role in Leigh syndrome.

High yield of C-terminal soluble part of COX11 protein was obtained using refolding protocol and nice NMR spectra were obtained. The solution structure of COX11 will be then solved and its copper binding properties characterized in order to understand how COX11 is involved in assembly of copper centers of CcO protein. In this frame interaction studies by NMR between COX11 and COX17 chaperone will be carried out.

The last project was focussed on the role of the human copper chaperone of superoxide dismutase(CCS) in FALS disease. Full length and short constructs of CCS were obtained and characterized by CD and NMR. Both protein constructs are nicely folded and we are therefore preparing triple labelled NMR samples in order to solve the solution structure and to use it as starting point to study interactions with SOD1 protein and with Cu(I) ions.

Production and structural characterization of these proteins will be the starting point to understand the mechanisms by which all these proteins act and are involved in different cellular mechanisms and in the respective diseases. Moreover, the syndromes here studied are rare diseases and these data should be the hint for new studies in drug discovery.