University of Florence

International Doctorate in Structural Biology

Cycle XXIII (2008-2010)

Structural and functional aspect of proteins involved in iron-sulfur cluster biogenesis and metal transfer

Ph.D. thesis of

Maciej Mikolajczyk

Tutor	**Coordinator**
Prof. Lucia Banci	Prof. Ivano Bertini

This thesis has been approved by the University of Florence,

the University of Frankfurt and the Utrecht University

Table of contents

Chapter 1- INTRODUCTION 4

1.1. The role of iron in biological systems 5

1.2. Iron metabolism and cellular uptake 7

1.3. The role of iron in mitochondrial processes 9

1.4. The role and types of ISCs 11

1.5. Biogenesis of ISCs 12

1.5.1. Mitochondrial ISC assembly machinery 13

1.5.2. Mitochondrial ISC export machinery 16

1.5.3. CIA - cytosolic iron-sulfur cluster assembly machinery 17

1.6 Biogenesis of ISC and human diseases 18

1.7 Wilson and Menkes – diseases linked with copper homeostasis ... 20

1.8 Aims and topics of the research 22

References 25

Chapter 2- MATERIALS AND METHODS 32

2.1. Genome browsing 33

2.2. Construct design 34

2.3. Gene cloning 35

2.4. Protein expression 38

2.5. Protein purification 39

2.6. Protein refolding 41

2.7. Protein characterization 43

2.7.1. Metals content determination by ICP 43

2.6.2. Electron Paramagnetic Resonance 43

2.6.3. Mass spectrometry 44

2.6.4. Circular dichroism 47

2.6.5. NMR – protein structural characterization 48

Chapter 3- RESULTS..... 53

3.1. Anamorsin is a 2Fe2S cluster-containing substrate of the Mia40-dependent mitochondrial protein trapping machinery 54

3.2. Characterization of glutaredoxin5, a protein involved in Fe-S cluster biogenesis in mitochondria 88

3.3. Wilson and Menkes proteins: cloning, expression and purification of ATP-binding and P-domain..... 107

Chapter 4- CONCLUSIONS AND PERSPECTIVES.....123

Acknowledgments.....126

Chapter 1- INTRODUCTION

1.1. The role of iron in biological systems

Out of the more than 100 chemical elements known to scientists today, only a relatively small number of these elements are found in the human body. In fact, only 24 different elements are thought to be essential for humans. The largest elemental components of the body, by mass, are oxygen (65%), carbon (18%), hydrogen (10%), and nitrogen (3%).[1] The other elements present in the body are known to physiologists as trace elements. Some of the trace elements are metal ions like copper, zinc, iron and manganese which are vital to the body's proper functioning. They must be present at the proper level, and they must also be available to react with other elements to form the molecules in the functional state.

Metals are involved in a wide range of biochemical processes in all living organisms. They are ''key'' players in electron transfer, oxygen transport and numerous catalytic reactions. The use of metals in living organisms is mainly related to red-ox reactions and to the ability to catalyze numerous essential functions such as hydrolysis. Metalloprotein is a generic term for a protein that contains a metal ion cofactor. Indeed, about one quarter to one third of all proteins require metals to carry out their functions.[2,3,4] The metal ion is usually coordinated by nitrogen, oxygen or sulfur atoms belonging to amino acids in the polypeptide chain and/or a macrocyclic ligand incorporated into the protein. The presence of the metal ion allows metalloproteins to perform functions that cannot efficiently be performed by the limited set of functional groups found in amino acids. Numerous proteins bind one or more metal ions which can be involved either in functional process or necessary to obtain/maintain the protein's tertiary or quaternary structure or both. Also in the case of RNA metal ions are very important for function and structure. [5,6]

One of the most abundant essential trace elements in the human body is iron. Healthy adults possess from 4 to 6 g of iron, which approximately comprises only 0,008% of the body mass (75 kg).[7,8] Nevertheless it is necessary for our survival. The bulk of this iron is bound to hemoglobin in the red blood cells, where it is necessary for oxygen transport in the blood. In humans, the hemoglobin molecule is an assembly of four globular protein subunits. Each subunit is composed of a protein chain tightly associated with a non-protein heme group. Iron is the central atom of the heme group, a metal complex that binds

molecular oxygen (O_2) in the lungs and carries it to all of the other cells in the body that need oxygen to perform their activities (e.g., muscle cells). In addition to hemoglobin**,** other important proteins in the body that contain heme groups (and therefore contain iron) include myoglobin**,** which takes oxygen from hemoglobin and allows the oxygen to diffuse throughout the muscle cells, and the cytochromes**,** which transfer electrons.[7,9] Other class of iron bound proteins are those binding iron-sulfur clusters (**Fig. 1**). These proteins are characterized by the presence of iron-sulfur (Fe-S) clusters containing sulfide-linked di-, tri-, and tetra-iron centers in variable oxidation states. Fe-S clusters are found in a variety of metalloproteins, such as the ferredoxins, as well as NADH dehydrogenase, hydrogenases, Coenzyme Q - cytochrome c reductase, succinate - coenzyme Q reductase and nitrogenase. Fe-S clusters are best known for their role in the oxidation-reduction reactions of mitochondrial electron transport. Both Complex I and Complex II of oxidative phosphorylation have multiple Fe-S clusters. They have many other functions including catalysis as illustrated by aconitase, generation of radicals as illustrated by SAM-dependent enzymes, and as sulfur donors in the biosynthesis of lipoic acid and biotin. Additionally some Fe-S proteins regulate gene expression.

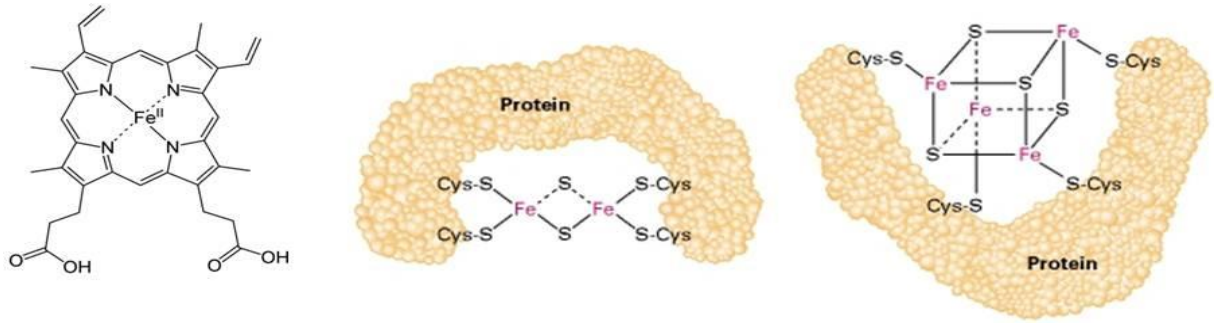


Figure 1. Examples of protein cofactors containing iron.

Other proteins, such as those needed for DNA synthesis and cell division, also rely on iron. Furthermore, iron contributes to produce the connective tissues in our body, some of the neurotransmitters in our brain, and to maintain the immune system.

While iron is necessary for allowing the respirating cells to obtain O_2, for supplying the body with a reliable source of energy, and for maintaining several other important structures and systems in the body, yet its properties also make it potentially deleterious. The biological importance of iron is largely attributable to its chemical properties as a transition metal. It readily engages in one-electron oxidation-reduction reactions between its ferric (+3) and ferrous (+2) states. However, the same chemical property explains why

an excess of "free," reactive iron is toxic. In the cytoplasm, a significant fraction of iron is reduced and can participate in "Fenton-type" redox chemistry: ferrous iron reacts with hydrogen peroxide (H_2O_2) or lipid peroxides to generate ferric iron, OH^-, and the highly reactive hydroxyl radical (OH•) or lipid radicals such as LO• and LOO•. These radicals damage lipid membranes, proteins, and nucleic acids.

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH\cdot + OH^-$$

$$Fe^{3+} + H_2O_2 \rightarrow Fe^{2+} + OOH\cdot + H^+$$

To avoid the above harmful redox reactions, iron homeostasis is achieved through iron transport, storage and regulatory proteins. That allows the cells to use the benefits of iron, but also limits its ability to do harm.[10,11]

1.2. Iron metabolism and cellular uptake

Iron uptake is tightly regulated by the human body, which do not have pathways for iron excretion. Only small amounts of iron are lost daily due to mucosal and skin epithelial cell sloughing, desquamation of skin and urinary cells, blood loss, or sweat. To compensate this loss only about 1-2 mg of dietary iron per day is absorbed by duodenal enterocytes. This amount of iron compared to the amount ingested is typically low. The efficiency with which iron is absorbed varies largely depending on the source and generally the best absorbed forms of iron come from products of animal origin.[8,12] Duodenal enterocytes have special proteins that allow them to move iron into the body. Iron can be absorbed in its ferrous Fe^{2+} form. A ferric reductase enzyme on the enterocytes brush border, reduces ferric Fe^{3+} to ferrous Fe^{2+}, and in this form iron is imported into enterocyte via divalent metal transporter 1 (DMT1). Iron is released directly into the blood stream by ferroportin which is located at the basolateral membrane of duodenal enterocytes. It is oxidized to the Fe^{3+} form and binds to transferrin (Tf). Tf-$2Fe^{3+}$ complexes circulate in the plasma, and deliver iron to all the tissues. Most of the iron is incorporated into hemoglobin, in erythroid precursors and mature red blood cells (approx. 1800 mg). Approximately 400 mg is present

in the other tissues (in enzymes and cytochromes) and muscle fibers (in myoglobin). Liver (parenchymal cells) and reticuloendothelial macrophages are responsible for iron storage (approx. 1600 mg). The macrophages provide most of the usable iron by degrading hemoglobin in aged or damaged erythrocytes and reloading ferric iron onto transferrin for delivery to cells, because the amount of plasma iron is just around 10% (approx. 3 mg) of the amount used daily, which means that plasma iron is turned over many times each day (**Fig. 2**). All cells use some iron, and must get it from the plasma. Since iron (in the ferric state) is tightly bound to transferrin (Tf), cells throughout the body have receptors (TfR1) for Tf-2Fe^{3+} complexes on their surfaces. These receptors engulf and form endosomes containing both the protein and the iron attached to it. A proton pump decreases the pH within the endosomes, leading to conformational changes in the transferin that result in the release of iron. The iron is reduced to the ferrous state and transported by DMT1 across the endosomal membrane, to enter the cytoplasm. Intracellular iron can be utilized for direct incorporation into iron proteins or delivered to mitochondria, where the iron-dependent step of heme synthesis and critical steps for Fe-S cluster biogenesis are localized. The fraction of the cytosolic iron that is not utilized for metalation reactions can be stored in ferritin in a nontoxic form or exported via ferroportin, which works together with ferroxidases for iron loading onto transferrin in plasma. Meanwhile, transferrin (Apo-Tf) and transferrin receptor are recycled to the cell surface, where each can be used for further cycles of iron binding and iron uptake.[5,10,13]

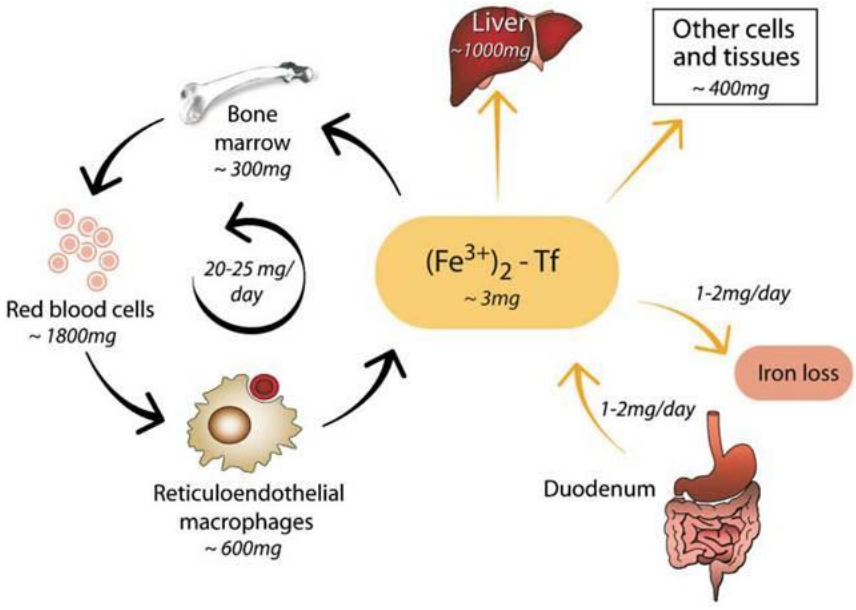


Figure 2. Circulation of iron in human body. Reprinted from Hentze et al. *Cell* (2004).

Disruptions in iron homeostasis from both iron deficiency and overload account for some of the most common human diseases (e.g. hemoglobin and iron deficiency anemia or hemochromatosis). That is why cells developed mechanisms for sensing their own need of iron. Iron metabolism is maintained by two regulatory systems:

- the first one functions systemically and relies on the hormone hepcidin that directly inhibits ferroportin. By inhibiting ferroportin, hepcidin prevents enterocytes of the intestine and macrophages of the reticuloendothelial system from secreting iron into the blood stream, thereby functionally reducing the iron level.

- the second one predominantly controls the cellular iron metabolism through iron-regulatory proteins (IRP1 and IRP2). When the iron level in the cell is low IRPs bind to iron-responsive elements in regulated messenger RNAs of the transferrin receptors and ferritin. The binding of IRPs to iron responsive elements (IREs) in the 5'-untranslated regions (UTRs) of mRNA blocks the translation (e.g. of ferritin), while IRP binding to the IRE in the 3'-UTR stabilizes the mRNA (e.g. of transferrin receptor). When the iron concentration in the cell is high IRP1 binds an [4Fe-4S] cluster (and acts like aconitase) and IRP2 degrades with the result that they can no longer bind to mRNA. This liberates the IRE in the 5'-untranslated region of Ferritin mRNA and directs the cell to produce more storage molecules. On the other hand the free IRE at the 3'-UTR of transferrin receptors accelerates its mRNA degradation resulting in decreased protein expression.[14,15]

1.3. The role of iron in mitochondrial processes

The mitochondrion is an eukaryotic cellular organelle that contains outer and inner membranes composed of phospholipid bilayers and proteins.[20] Because of this double-membraned organization, there are five distinct compartments within the mitochondrion (**Fig. 3**):

- the outer mitochondrial membrane,
- the intermembrane space (the space between the outer and inner membranes),
- the inner mitochondrial membrane,
- the crista space (formed by infoldings of the inner membrane),
- and the matrix (space within the inner membrane). Main place for iron utilization.

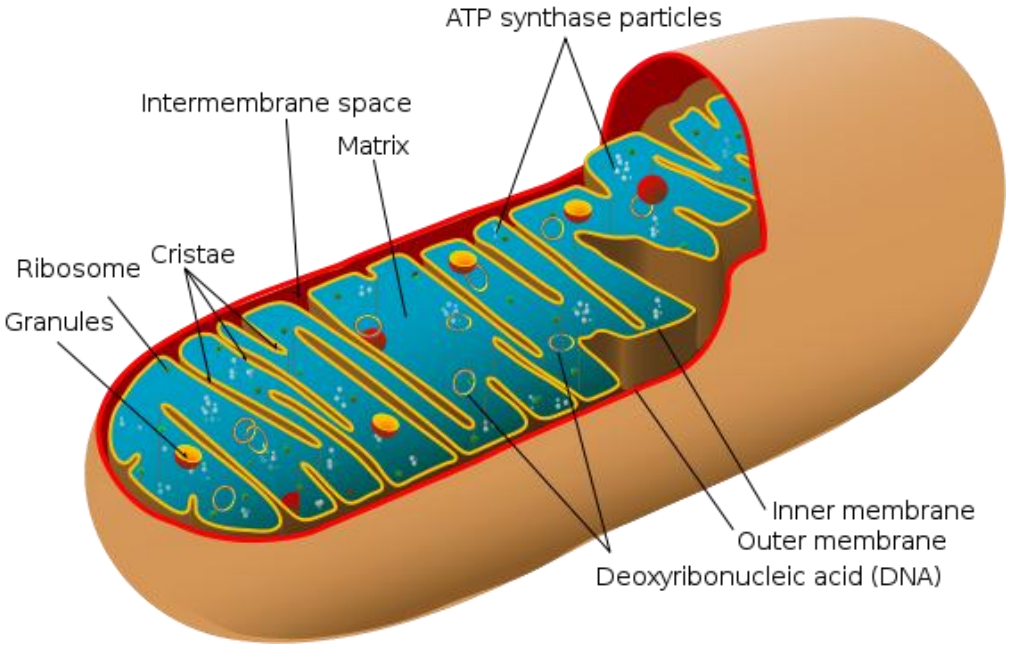


Figure 3. Schematic representation of the mitochondrion structure.

Mitochondria are well known for their role in ATP synthesis. The overall process of oxidizing glucose to carbon dioxide is known as cellular respiration. ATP can be produced via three main pathways: the glycolysis and the citric acid cycle/oxidative phosphorylation, both components of cellular respiration, and the beta-oxidation. The majority of this ATP production by a non-photosynthetic aerobic eukaryote takes place in the mitochondria, which can make up nearly 25% of the total volume of a typical cell. Iron is directly involved in the ATP synthesis, because its reversible oxidation states allow electron transfer processes that are necessary for respiratory chain via its heme and iron-sulfur clusters (ISCs) containing complexes. Additionally, the mitochondrion is also the main site of heme synthesis and a major generator of ISCs, both of which are present in mitochondria and cytosol. Moreover, the synthesis of ISCs is the only essential biosynthetic process performed by these organelles.[16,17,18,19]

1.4. The role and types of ISCs

ISCs are ubiquitous protein cofactors highly conserved from bacteria to human. Practically, they play a crucial and essential role in every living cell. In the earliest forms of life, when the earth's atmosphere was a reducing and anaerobic environment (few billion years ago), Fe-S cluster proteins were used to perform fundamental metabolic processes (amino acid metabolism). After transformation of the atmosphere into an oxidative and aerobic environment (which does not support the formation, neither the stability of ISCs) the organism evolved a whole set of proteins to coordinate a safe, efficient and specific ISC biogenesis. Another reason why biosynthesis of ISCs has to be strictly controlled is their potential toxicity. Improper biogenesis of the clusters may fuel the Fenton chemistry, which produces highly toxic reactive oxygen species (ROS).[19,21,22] ISCs are indispensable for life because of the ability of the iron to accept or donate electrons, formally altering between the ferrous (Fe2+) and ferric (Fe3+) oxidation states. In fact Fe-S proteins have a wide range of reduction potentials (–700 mV to 400 mV) and are among the most important electron carriers in nature.[23] They play a central role in three processes: respiration, photosynthesis and nitrogen fixation which are crucial for life on earth. Fe-S clusters might serve to transfer electrons, perform catalysis, stabilize three-dimensional protein structures or have regulatory function. Enzymatic functions of Fe-S proteins include Lewis acid catalysis in aconitase-type proteins, DNA and RNA synthesis and repair, and iron and heme metabolism.[24]

Fe-S clusters are formally composed of ferrous (Fe^{2+}) and/or ferric (Fe^{3+}) iron, and inorganic sulfide (S^{2-}) ions. However a class of low-molecular-weight iron-containing proteins called rubredoxins, sometimes classified as iron-sulfur proteins, do not contain inorganic sulfide but coordinates iron ion by the sulfurs of four cysteine residues. With exception of rubredoxins, the most common Fe-S clusters present in living organisms are the [2Fe–2S] and [4Fe–4S] clusters (**Fig. 4**). [2Fe–2S] is the simplest polymetallic system constituted by two iron ions bridged by two sulfides. The cofactor is usually coordinated by four cysteine residues or two cysteines and two histidines (in Rieske proteins). The [2Fe–2S] cluster exists in two oxidation states; the oxidized proteins contain two Fe^{3+} ions, whereas the reduced proteins contain one Fe^{3+} and one Fe^{2+} ion.[25,26] In [4Fe–4S] clusters four iron ions and four sulfide ions are placed at the vertices of a cubane-type structure. It is formed by the fusion of two [2Fe–2S] clusters. Proteins coordinating [4Fe–4S] clusters

can be subdivided in two groups; the low and high potential depending on the pair of the commonly accessible oxidation states. In high potential proteins the cluster shuttles between [$2Fe^{3+}$, $2Fe^{2+}$] ($Fe_4S_4^{2+}$) and [$3Fe^{3+}$, Fe^{2+}] ($Fe_4S_4^{3+}$), whereas in low potential [Fe^{3+}, $3Fe^{2+}$] ($Fe_4S_4^{+}$) and [$2Fe^{3+}$, $2Fe^{2+}$] ($Fe_4S_4^{2+}$). The third type of cluster coordinated by proteins is the [3Fe–4S] cluster, where three sulfide ions bridge two iron ions each, while the fourth sulfide bridges three iron ions. Their formal oxidation states may vary from $[Fe_3S_4]^{+}$ (all-Fe^{3+} form) to $[Fe_3S_4]^{2-}$ (all-Fe^{2+} form). In a number of Fe-S proteins, the $[Fe_4S_4]$ cluster can be reversibly converted by oxidation and loss of one iron ion to a $[Fe_3S_4]$ cluster. For example the inactive form of aconitase possesses an $[Fe_3S_4]$ and is activated by addition of Fe^{2+} and reductant.[21,22,23] Despite the simple structure and easy incorporation of the cluster into purified apo-proteins *in vitro*, the mechanisms of cluster assembly *in vivo* requires dedicated machineries.

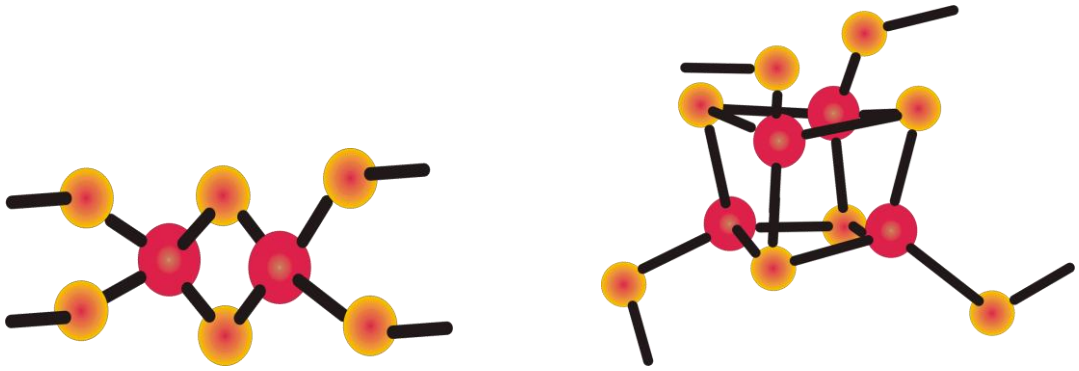


Figure 4. Schematic representation of [2Fe-2S] and [4Fe-4S] clusters. Iron and sulfur atoms are represented as red and yellow balls respectively.

1.5. Biogenesis of ISCs

The biosynthesis of the Fe-S clusters raised quite interest in the past years.[27,28,29] The most extensive studies were focused on bacteria (*Escherichia coli* and *Azetobacter vinelandii*) and yeast *Saccharomyces cerevisiae*. Three different systems were identified for the biogenesis of bacterial Fe-S proteins: the NIF system, for specific maturation of nitrogenase in azototrophic bacteria; the ISC assembly and SUF systems, for the generation of housekeeping Fe-S proteins under normal and oxidative-stress conditions, respectively. The yeast *S. cerevisiae* has served as an excellent model organism to establish the first details of the complex biosynthesis pathways in eukaryotes.

Recent investigations in human cell culture and other model systems have demonstrated that the entire process is highly conserved from yeast to human.

Three distinct protein machineries are required in the (non-plant) eukaryotic cells for the biogenesis of the Fe-S clusters; the ISC assembly machinery in mitochondria (inherited from bacteria during evolution), the ISC export machinery located in the mitochondrial inter membrane space, and the cytosolic ISC assembly machinery (CIA)(**Fig. 5**). ISC export and CIA machineries are specifically involved in the maturation of cytosolic and nuclear ISC proteins, whereas the mitochondrial ISC assembly machinery is required for the generation of all cellular Fe-S proteins.[30,31,32]

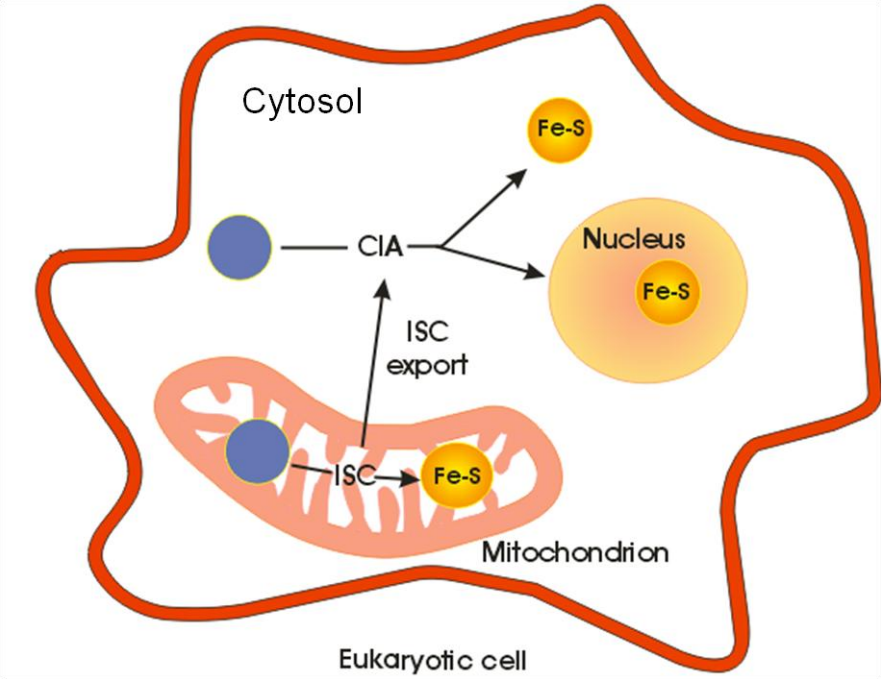


Figure 5. Schematic representation of ISC assembly in eukaryotic cell

1.5.1. Mitochondrial ISC assembly machinery

The overall biogenesis process in mitochondria can be split into two main steps: the *de novo* assembly of an Fe-S cluster on a scaffold protein, and the transfer of the Fe-S cluster from the scaffold to target apo-proteins. Each of these steps involves the participation of several proteins and cofactors, which execute specific biosynthetic partial reactions.[33] Indeed over 15 proteins are known to play a role in this complicated biosynthetic process. The mitochondrial components required for Fe-S cluster formation in

higher organisms are similar to the bacterial ISC proteins and apparently were inherited from a prokaryotic progenitor.[34] The central component of the mitochondrial ISC assembly machinery is the scaffold protein IscU on which *de novo* assembly of a Fe-S cluster takes place. IscU is one of the most conserved proteins in evolution and is found in many bacteria, and virtually all eukaryotes. It has three cysteine residues which are necessary for the cluster synthesis. Both [2Fe-2S] and [4Fe-4S] clusters can be assembled on IscU.[35,36,37] To accomplish *de novo* cluster formation, sources of iron and sulfur are necessary. Iron is imported from the cytosol into mitochondria through the innermembrane transporters Mitoferrin1/2 (Mrfn1/2) in its ferrous (Fe^{2+}) form, and is delivered to the scaffold by Frataxin which was shown to interact with IscU *in vivo* and *in vitro* in an iron stimulated fashion.[38,39,40] The free cysteine is the source of the sulfur, the latter is released and delivered to IscU by the complex of cysteine desulfurase IscS-Isd11. A persulfide is formed on a conserved cysteine residue of the enzyme, and is then transferred to the scaffold protein. Although isolated IscS has the enzymatic activity (as a cysteine desulfurase) *in vitro*, the IscS–Isd11 complex is the functional entity for sulfur transfer from IscS to IscU *in vivo*.[41,42,] This reaction is aided by direct interaction between IscS and IscU. Before the assembly reaction on IscU is accomplished the sulfur is reduced by the electron transfer chain composed of ferredoxin and ferredoxin reductase which receives the electron from NADH (**Fig. 6**).

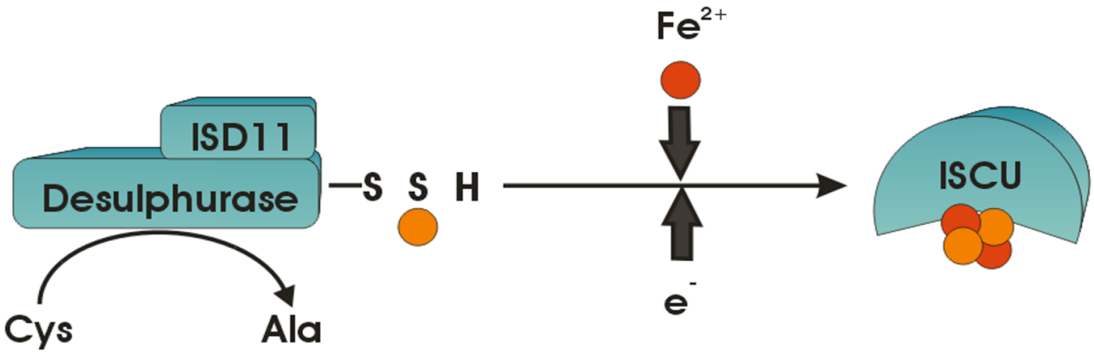


Figure 6. First step of ISC biosynthesis process in mitochondria - assembly of [Fe-S] cluster on ISCU.

The second major step performed by late components of the ISC assembly machinery is the release of the Fe-S cluster from IscU and its transfer and incorporation into recipient apo-proteins. The interaction of scaffold protein with co-chaperone HSCB and HSPA9 promotes the conformational changes of IscU and dislocates the cluster.[43] Afterwards the delivery of the assembled Fe-S cluster to the mitochondrial recipients is performed by a group of chaperones, e.g. Glutaredoxin5, IscA1/2, C1orf69 (specifically required for the gaining activity of aconitase-type and SAM-dependent Fe-S enzymes) or recently reported Ind1 (necessary for cluster incorporation into respiratory Complex I of oxidative phosphorylation in mitochondria).[44,45,46,47] Mitochondrial and cytosolic Fe-S cluster protein maturation strictly depends on the function of the ISC machinery, which produces a still unidentified compound X that is later exported from mitochondria by the ISC export machinery and used for cytosolic Fe-S cluster formation (**Fig. 7**).

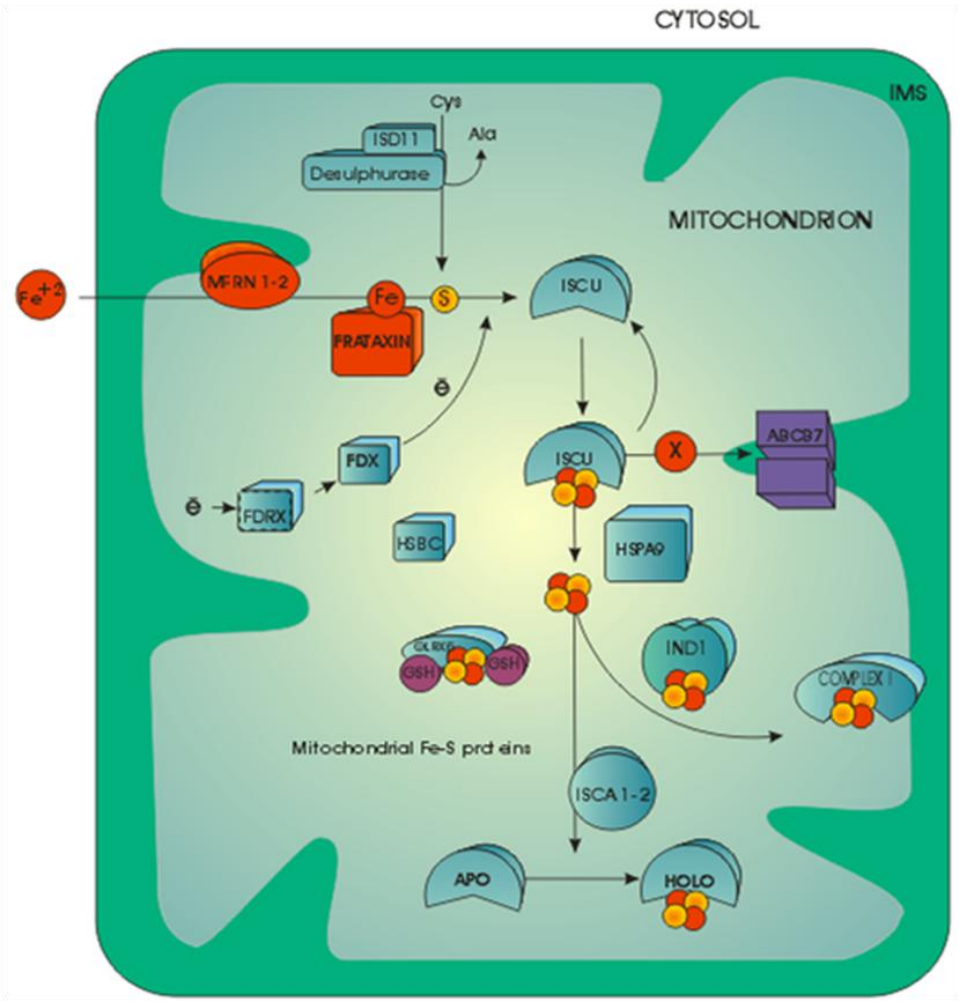


Figure 7. Mitochondrial ISC assembly machinery components and pathway.

1.5.2. Mitochondrial ISC export machinery

The necessity of mitochondrial ISC assembly for cytosolic Fe-S protein maturation was previously explained by the export of ready, preassembled clusters. This idea was abandoned when the ISC export system was discovered. The export machinery consists of three components; the inner mitochondrial membrane ABC transporter - ABCB7, the intermembrane space protein - ALR and three peptide glutathione – GSH (**Fig. 8**). Downregulation or depletion experiments of latter proteins in yeast and human cell lines lead to defective cytosolic and nucleus Fe-S proteins due to improper maturation and increased iron uptake in the cell and mitochondrial iron overload.[48,49] It was shown *in vitro* that ATPase activity of the yeast homologue of ABCB7 (the yeast homologue is called Atm1) is stimulated by the sulfhydryl groups (e.g. cysteine residues).[50] Because the ATPase of ABC transporters is usually stimulated by their physiological substrates, this may indicate that the substrate X transported from the mitochondria contains a sulfhydryl group in a peptidic environment. Moreover, a further step of the export process is performed by the ALR, which has sulfhydryl oxidase activity and is located in the mitochondrial intermembrane space. ALR has also been shown to assist incorporation of disulfide bonds by Mia40 into substrate proteins during protein import and trapping in the IMS.[51,52,53] The last component of the ISC export machinery – GSH is a reducing agent, existing at a concentration of approximately 10 mM in animal cells. In summary, all known components of the ISC export machinery deal with sulfhydryl chemistry, but identification of their specific function will depend on the identification of the exported moiety.[19,26]

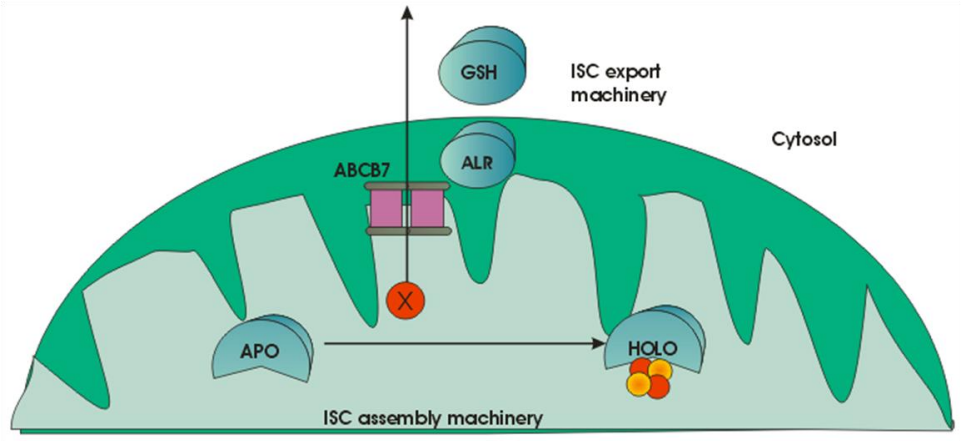


Figure 8. Mitochondrial ISC export machinery.

1.5.3. CIA - cytosolic iron-sulfur cluster assembly machinery

The maturation of cytosolic and nucleus Fe-S proteins requires not only the functionality of the mitochondrial ISC assembly machinery but also a dedicated cytoplasmic system (CIA). The overall process can be subdivided into two steps; first the cluster is assembled on scaffold proteins and later transferred to apo-proteins (**Fig. 9**). Till date six proteins were identified to play a crucial role in this process;

- NUBP2 and NUBP1 (both P-loop NTPases) that form a hetero complex where *de novo* cluster formation takes place,
- IOP1 required for cluster transfer to the apo-proteins,
- CIAO1 that mediates protein-protein interaction between NUBP1 and IOP1,
- CIAPIN1 (also called anamorsin) whose yeast homologue Dre2 was shown to be partially localized in the mitochondrial intermembrane space,
- and Ndor1 that was shown to transfer electrons from NADH to the Fe-S cluster of anamorsin.[22,26,48,54,55,56]

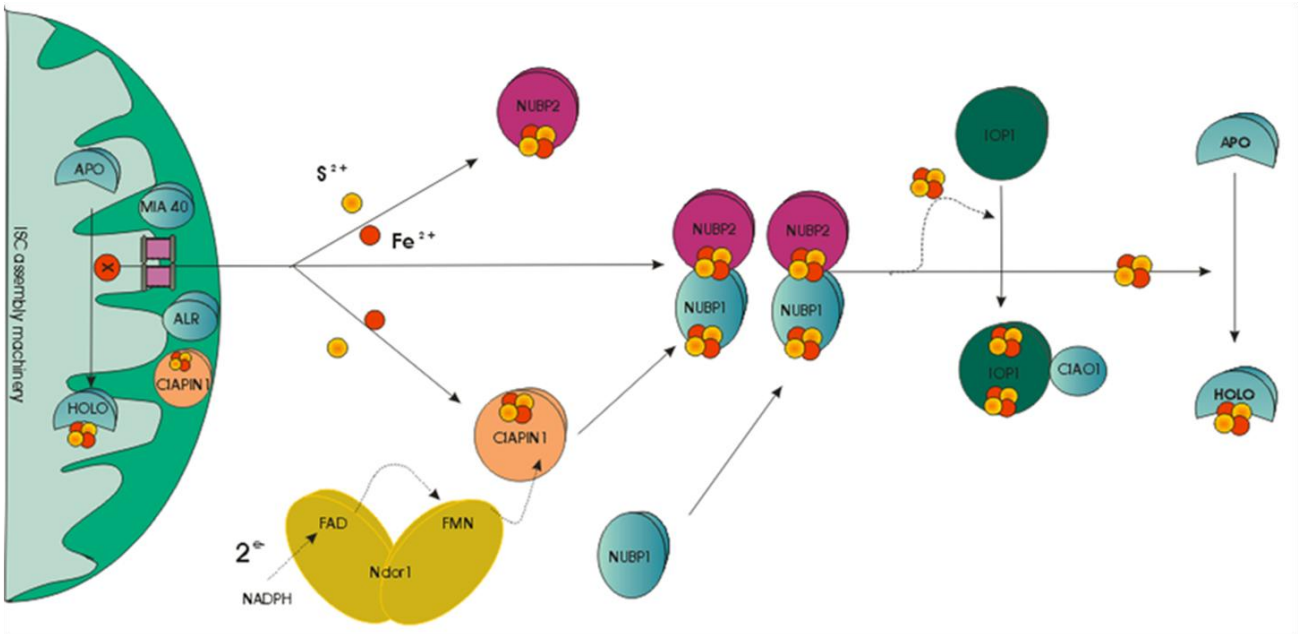


Figure 9. Components of cytosolic ISC assembly machinery.

Interestingly, the incorporation of anamorsin's Fe-S cluster is independent from the CIA machinery. Considering anamorsin's double localization, and it's CIA-independent Fe-S assembly, this protein can link the mitochondrial and the cytosolic assembly machineries.

All the components of the CIA machinery are defined by having a primary location in the cytoplasm and they act just in the cytosolic (not the mitochondrial) Fe-S cluster protein maturation. This distinguishes CIA proteins from ISC factors that are important for both mitochondrial and cytosolic Fe-S cluster biogenesis and ISC export proteins that are required for cytosolic cluster biogenesis but are located only within the mitochondria. The cytosolic Fe-S maturation machinery is functionally conserved from yeast to humans. Complementation tests performed on yeast strains depleted for the Cia1 and Dre2 (homologues of CIAO1 and Anamorsin, respectively) showed that human proteins can functionally replace them *in vivo*. Depletion experiments of single CIA factors in yeast and human cell lines lead to a massive reduction in growth rates, malfunction of iron homeostasis (impaired maturation of IRP1) and decrease of cytosolic Fe-S protein activities, with preserved activity of mitochondrial Fe-S proteins. This suggests that the cytosolic ISC assembly system works downstream of the ISC assembly machinery in mitochondria.[57,58] Identification of factor X, necessary for cytosolic cluster formation, produced by the ISC assembly machinery and exported by the ISC export machinery will give a new insight into the overall cellular maturation process, and will help to fully understand the connection between these three distinct machineries.

1.6 Biogenesis of ISC and human diseases

Fe-S clusters are essential cofactors of many proteins and the disruption of their biosynthesis leads to dysfunction of many fundamental cellular processes. Indeed the mutations in numerous genes encoding the members of ISC and CIA biogenesis pathways were connected with human diseases. The two classical examples of Fe-S protein-linked diseases are Friedreich's ataxia (FRDA) and X-Linked Sideroblastic Anemia and Ataxia (XLSA/A). The first one described was FRDA, which is the most common autosomal recessive ataxia, affecting about 1 person in 50,000.[59,60] The disease is caused by mutations in the gene encoding Frataxin, a protein that is essential for ISC biogenesis in

mitochondria. In most cases, the mutant gene contains expanded GAA triplet repeats in the first intron which causes the gene silencing.[61] Patients show signs of mitochondrial Fe-S protein defects (e.g. complexes I–III of the respiratory chain and aconitase) and accumulate iron in the heart and neuronal tissues. These cause progressive damage to the nervous system resulting in symptoms ranging from gait disturbance and speech problems to heart malfunction, decreased life span, and tumor growth.[62,63] XLSA/A is caused by mutations in the human gene encoding the ATP-binding cassette of the ISC export machinery - ABCB7.[64] The phenotype developed by ABCB7 gene inactivation in man represents Fe-S protein defects in the cytosol (but not in mitochondria), heme defects and iron accumulation in mitochondria that forms so-called ring sideroblasts (i.e., iron-loaded ring-shaped tubules, which are concentrated around the nucleus).[65,66] The illness is rare, and only few patient families have been described. Recently, several other diseases have been connected with defects in ISC assembly. A malfunction of the human mitochondrial monothiol glutaredoxin5 (Glrx5) was found in a patient who had microcytic anemia. In GLRX5-deficient cells, ISC biosynthesis was impaired which lead to high activity of IRP1, increased levels of IRP2, cytosolic iron depletion, and mitochondrial iron overload.[67] A myopathy with exercise intolerance, intracellular iron overload, and deficiencies in succinate dehydrogenase and aconitase activity were connected with the splicing mutation in the gene encoding the central component of the ISC assembly machinery in mitochondria – the scaffold protein IscU.[68]

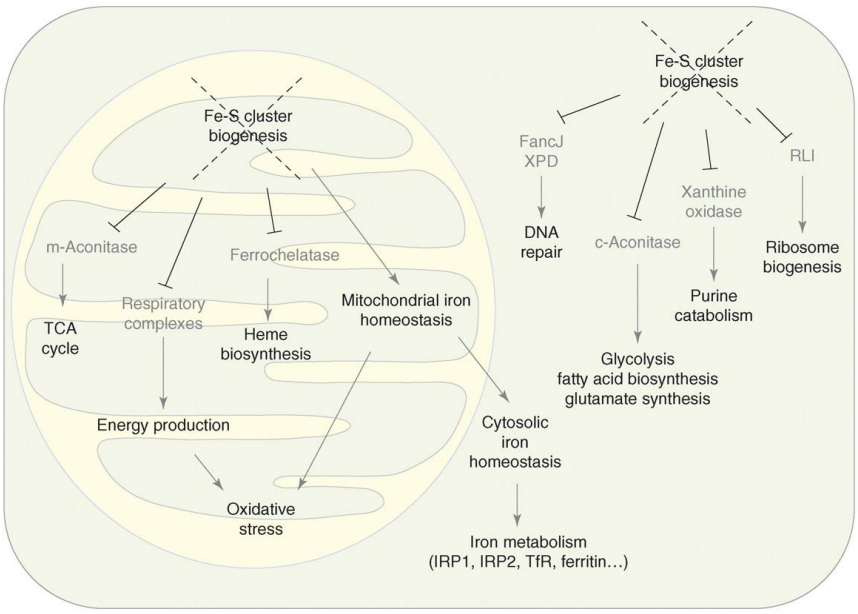


Figure 10. Cellular processes affected upon malfunction of mitochondrial and cytosolic ISC biogenesis. Reprinted from Rouault et al., *Trends Genet* (2008).

In summary, the disruption in cellular ISC biogenesis lead to malfunction of essential processes, e.g. TCA cycle, energy production, heme biosynthesis, cellular iron homeostasis, DNA repair or glycolysis (fig.). [69,70] Syndromes of the connected diseases can appear in the central nervous system and heart (FRDA), in developing red blood cells (GLRX5 deficiency) or in skeletal muscles (myopathy with deficiencies in IscU). Iron-sulfur clusters are cofactors of many proteins that play crucial roles in essential metabolic pathways common to all tissues. The fact that their dysfunction can cause such diverse disease presentations makes it very probable that mutations in other conserved Fe-S cluster biogenesis components lead to diseases for which the cause is presently unknown.[25]

1.7 Wilson and Menkes – diseases linked with copper homeostasis

Copper (as iron) is an essential element that participates in biological processes in eukaryotic and prokaryotic cells, because of its ability to change the redox state from Cu^{+} to Cu^{2+}. However the copper homeostasis has to be strictly controlled, because of copper potential toxicity.[71] Indeed, excessive tissue accumulation of redox-active transition metals, as copper, result in an array of cellular disturbances characterized by oxidative stress and increased free radical production. Alterations of copper homeostasis lead to errors of metabolism and damage to cells, resulting in pathological conditions.[72] Central regulators of cellular copper metabolism are two homologues copper-transporting P-type ATPases ATP7A (MNK) and ATP7B (WLN).[73,74] The proteins differ in their tissue distribution of expression: WLN is predominantly expressed in the liver and is responsible for delivery of Cu to ceruloplasmin (the main plasma form of Cu) and biliary efflux of Cu, whereas MNK is present in most non-hepatic tissue, especially skeletal muscle, fibroblast and brain.[75,76] MNK has two important cellular functions: to facilitate the export of copper from non-hepatic tissues, its absorption into circulation, and to deliver copper to the secretory pathway for incorporation in copper-dependent enzymes such as lysyl oxidase.[77] Both ATP7A and ATP7B are integral membrane proteins, respectively of 1500 and 1476 amino acids, with various soluble domains. The Wilson protein has 67% amino acid identity to the Menkes protein[78] and the same structural characteristics; six soluble metal

binding domains, and the core unit composed of eight transmembrane helices and two soluble domains, the Actuator-domain (A) and the ATP-binding domain. The latter one is composed by two sub-domains: the Nucleotide-domain (N) and the Phosphorylation-domain (P) (**Fig. 11**).[79,80]

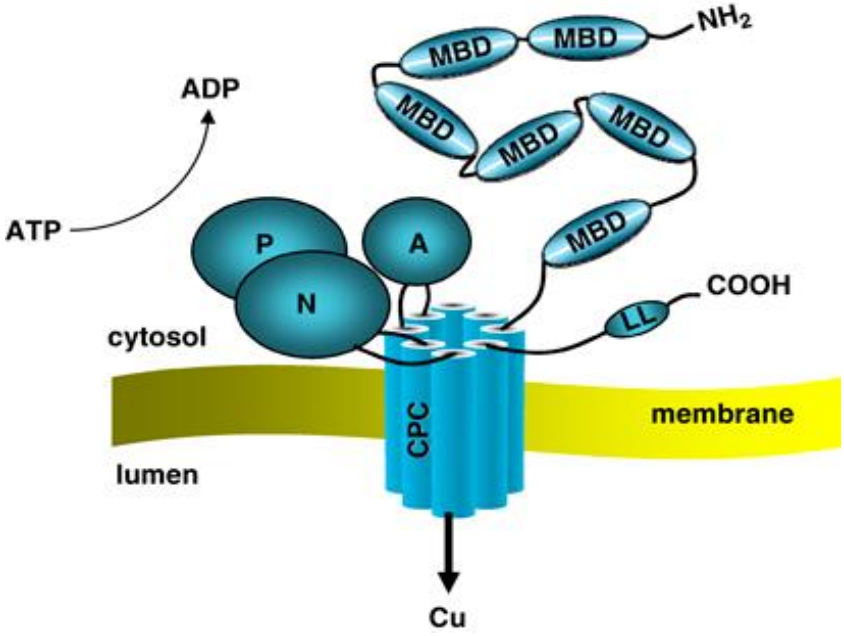


Figure 11. Schematic representation of WLN and MNK proteins. Reprinted from Tumer et al., *Eur J Hum Genet* (2010)

Mutations in ATP7A or ATP7B genes disrupt the homeostatic copper balance, resulting in copper deficiency (Menkes disease) or copper overload (Wilson disease), respectively.[81] Menkes disease is a fatal X-linked copper deficiency disorder; affected individuals die in early childhood and have multiple abnormalities that can be related to deficiencies in cuproenzymes. Differently in the Wilson disease very high concentrations of copper are accumulated in the liver because the biliary excretion of the metal ion is defective. If this disease is not treated, death can result from liver failure.[79,82] The overall view of copper uptake, efflux and distribution in various tissues is shown in **Fig. 12**.

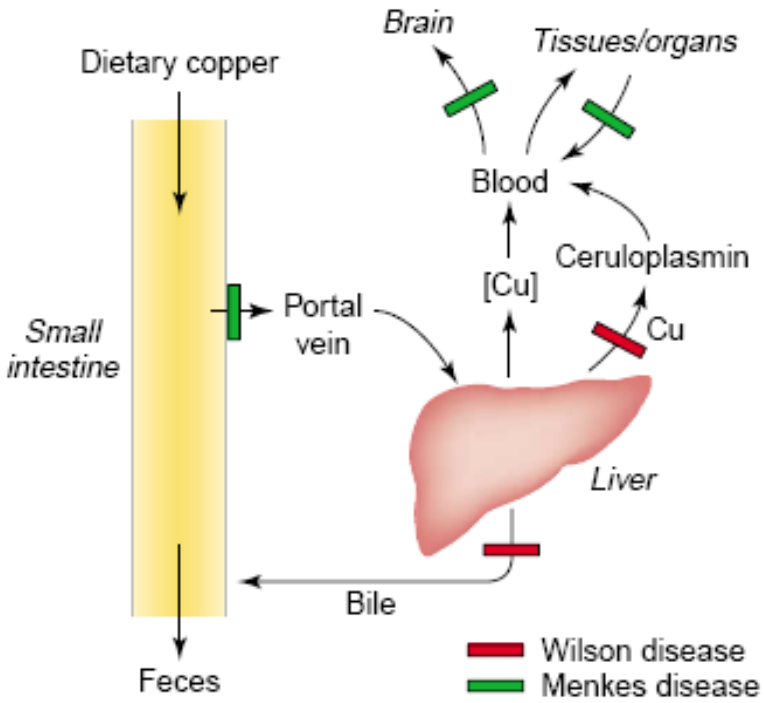


Figure 12. Pathway of copper in the body, and the blocks in Menkes and Wilson diseases. Reprinted from Mercer, *TRENDS in Molecular Medicine* (2001).

1.8 Aims and topics of the research

During the three years of my PhD course my research was focused on expression, purification, biochemical and structural characterization of recombinant human proteins involved in iron-sulfur cluster biogenesis and copper homeostasis. A multidisciplinary approach integrating several biophysical techniques has been used.

Several target proteins involved in ISC assembly in the cytosol and the mitochondria were chosen in order to understand at the molecular level their functional role in the pathways responsible for the Fe-S cluster incorporation in iron-sulfur dependent enzymes.

1. The cytokine-induced anti-apoptotic molecule (CIAPIN1) had been found to be a differentially-expressed gene involved in a variety of cancers, and it was also considered as a candidate tumor suppressor gene in gastric, renal and liver cancer. Owing to its prognostic value for human tumors and its involvement in cancer progression and tumor cell resistance to anticancer agents, CIAPIN1 has been proposed as an attractive target for

new anticancer interventions.[83,84,85] Recently its yeast homologue called Dre2 was reported to be a member of the cytosolic iron-sulfur cluster assembly machinery, and was found both in the cytosol and the mitochondrial intermembrane space. Moreover CIAPIN1 was able to substitute Dre2 *in vivo*.[54] My work in this project was devoted to the structural characterization of CIAPIN1, and the characterization of its ISC-binding properties as well as its interaction with the potential physiological partner Mia40, which is responsible for protein entrapment in the IMS. Full length protein and several constructs were cloned, expressed and purified. These proteins were characterized by NMR, EPR and mass spectroscopy techniques in order to give new hints on the function of this extremely important molecule in the iron-sulfur cluster biogenesis process.

2. Glutaredoxins (GRX) are redox proteins which use glutathione as a cofactor. They are divided into two classes, monothiol and dithiol and in each class, several GRX have been shown to form [2Fe-2S] cluster-coordinating homodimers or homotetramers. Glutaredoxin 5 is a monothiol glutaredoxin and was found to play a crucial role in the ISC assembly process in mitochondria. Absence of GRLX5 causes iron accumulation in the cell, which in turn could promote oxidative damage, and inactivation of enzymes requiring Fe-S clusters for their activities.[44] The protein was recombinantly expressed in *E.coli*, purified and Fe-S cluster binding properties were characterized. Using ^{1}H-paramagnetic NMR and UV-visible spectroscopy we indeed observed the binding of a [2Fe-2S] cluster. Interestingly, our NMR relaxation data showed that the cofactor is bound by the monomeric protein, at variance with what has been found in the crystal structure of holo-GLRX5. This result suggests that this monomeric form is the physiological one acting in the mitochondrial matrix. GLRX5 is indeed proposed to be involved in Fe-S cluster delivery to ISCA1 and ISCA2 protein partners. The monomeric form can more easily be the active state in the transfer process with respect to the tetramer in which Fe-S cluster is buried between the subunits. Protein-protein interactions and Fe-S cluster transfer between this monomeric form of holo-GLRX5 and apo- ISCA1 and ISCA2 are now in progress.

3. The project that I was involved in at the beginning of my PhD course was devoted to the Menkes and Wilson proteins, which play essential roles in the copper homeostasis in the human body. Within the frame of this major project at CERM, where several domains of both the Menkes (MNK) and Wilson (WLN) protein were produced and characterized in terms of their structural and functional properties. I have attempted the production of the ATP-domains of both proteins and the P-domain of the MNK protein

with the final goal of understanding the molecular interactions between these soluble domains and other previously characterized domains that drives the copper pumping machinery.

References

1. J. Emsley *The Element, 3rd ed.* (Oxford: Clarendon Press: 1998).
2. Bertini, I.;Sigel, A.; Sigel, H. Handbook on Metalloproteins. *Journal of the American Chemical Society* **123**, 12748 (2001).
3. Finney, L.A. & O'Halloran, T.V. Transition Metal Speciation in the Cell: Insights from the Chemistry of Metal Ion Receptors. *Science* **300**, 931-936 (2003).
4. Bidlack, W.R. Metal Ions in Biological System, Volume 36: Interrelations Between Free Radicals and Metal Ions in Life Processes. *J Am Coll Nutr* **18**, 368-369 (1999).
5. Hentze, M.W., Muckenthaler, M.U., Galy, B. & Camaschella, C. Two to Tango: Regulation of Mammalian Iron Metabolism. *Cell* **142**, 24-38 (2010).
6. Bidlack, W.R. Metal Ions in Biological Systems. Vol. 35: Iron Transport and Storage in Microorganisms, Plants and Animals. *J Am Coll Nutr* **18**, 95 (1999).
7. Andrews, N.C. Disorders of iron metabolism. *N. Engl. J. Med* **341**, 1986-1995 (1999).
8. Beard, J.L., Dawson, H. & Piñero, D.J. Iron Metabolism: A Comprehensive Review. *Nutrition Reviews* **54**, 295-317 (1996).
9. Lippard, S.J. & Berg, J.M. *Principles of bioinorganic chemistry*. (University Science Books: 1994).
10. Hentze, M.W., Muckenthaler, M.U. & Andrews, N.C. Balancing Acts: Molecular Control of Mammalian Iron Metabolism. *Cell* **117**, 285-297 (2004).
11. Letelier, M.E., Sánchez-Jofré, S., Peredo-Silva, L., Cortés-Troncoso, J. & Aracena-Parks, P. Mechanisms underlying iron and copper ions toxicity in biological systems: Pro-oxidant activity and protein-binding effects. *Chem Biol Interact* (2010).doi:10.1016/j.cbi.2010.06.013
12. Dunn, L., Rahmanto, Y. & Richardson, D. Iron uptake and metabolism in the new millennium. *Trends in Cell Biology* **17**, 93-100 (2007).
13. Andrews, N.C. Forging a field: the golden age of iron biology. *Blood* **112**, 219-230 (2008).
14. Rouault, T.A. The role of iron regulatory proteins in mammalian iron homeostasis and disease. *Nat Chem Biol* **2**, 406-414 (2006).
15. Cairo, G. & Recalcati, S. Iron-regulatory proteins: molecular biology and pathophysiological implications. *Expert Rev Mol Med* **9**, 1-13 (2007).
16. Lill, R. et al. The essential role of mitochondria in the biogenesis of cellular iron-sulfur proteins. *Biol. Chem* **380**, 1157-1166 (1999).

17. McBride, H.M., Neuspiel, M. & Wasiak, S. Mitochondria: More Than Just a Powerhouse. *Current Biology* **16**, R551-R560 (2006).

18. Richardson, D.R. et al. Mitochondrial iron trafficking and the integration of iron metabolism between the mitochondrion and cytosol. *Proceedings of the National Academy of Sciences* **107**, 10775 -10782 (2010).

19. Lill, R. & Mühlenhoff, U. Maturation of iron-sulfur proteins in eukaryotes: mechanisms, connected processes, and diseases. *Annu. Rev. Biochem* **77**, 669-700 (2008).

20. Henze, K. & Martin, W. Evolutionary biology: essence of mitochondria. *Nature* **426**, 127-128 (2003).

21. Py, B. & Barras, F. Building Fe–S proteins: bacterial strategies. *Nat Rev Micro* **8**, 436-446 (2010).

22. Xu, X.M. & Møller, S.G. Iron-Sulfur Clusters: Biogenesis, Molecular Mechanisms and Their Functional Significance. *Antioxid Redox Signal* (2010).doi:10.1089/ars.2010.3259

23. Rouault, T.A. & Tong, W. Iron-sulphur cluster biogenesis and mitochondrial iron homeostasis. *Nat Rev Mol Cell Biol* **6**, 345-351 (2005).

24. Rudolf, J., Makrantoni, V., Ingledew, W.J., Stark, M.J.R. & White, M.F. The DNA repair helicases XPD and FancJ have essential iron-sulfur domains. *Mol. Cell* **23**, 801-808 (2006).

25. Rouault, T.A. & Tong, W.H. Iron-sulfur cluster biogenesis and human disease. *Trends Genet* **24**, 398-407 (2008).

26. Lill, R. et al. Mechanisms of iron-sulfur protein maturation in mitochondria, cytosol and nucleus of eukaryotes. *Biochim. Biophys. Acta* **1763**, 652-667 (2006).

27. Johnson, D.C., Dean, D.R., Smith, A.D. & Johnson, M.K. STRUCTURE, FUNCTION, AND FORMATION OF BIOLOGICAL IRON-SULFUR CLUSTERS. *Annu. Rev. Biochem.* **74**, 247-281 (2005).

28. Johnson, M.K. & Smith, A.D. Iron-Sulfur Proteins. *Encyclopedia of Inorganic Chemistry* (2006).at <http://onlinelibrary.wiley.com/doi/10.1002/0470862106.ia116/full>

29. Lill, R. & Mühlenhoff, U. Iron-sulfur-protein biogenesis in eukaryotes. *Trends Biochem. Sci* **30**, 133-141 (2005).

30. Fosset, C. et al. RNA silencing of mitochondrial m-Nfs1 reduces Fe-S enzyme activity both in mitochondria and cytosol of mammalian cells. *J. Biol. Chem* **281**, 25398-25406 (2006).

31. Hausmann, A., Samans, B., Lill, R. & Mühlenhoff, U. Cellular and Mitochondrial

Remodeling upon Defects in Iron-Sulfur Protein Biogenesis. *Journal of Biological Chemistry* **283**, 8318 -8330 (2008).

32. Martelli, A. et al. Frataxin is essential for extramitochondrial Fe–S cluster proteins in mammalian tissues. *Human Molecular Genetics* **16**, 2651 -2658 (2007).

33. Lill, R. Function and biogenesis of iron-sulphur proteins. *Nature* **460**, 831-838 (2009).

34. Tokumoto, U., Kitamura, S., Fukuyama, K. & Takahashi, Y. Interchangeability and distinct properties of bacterial Fe-S cluster assembly systems: functional replacement of the isc and suf operons in Escherichia coli with the nifSU-like operon from Helicobacter pylori. *J. Biochem* **136**, 199-209 (2004).

35. Agar, J.N., Zheng, L., Cash, V.L., Dean, D.R. & Johnson, M.K. Role of the IscU Protein in Iron−Sulfur Cluster Biosynthesis: IscS-mediated Assembly of a [Fe2S2] Cluster in IscU. *Journal of the American Chemical Society* **122**, 2136-2137 (2000).

36. Chandramouli, K. et al. Formation and Properties of [4Fe-4S] Clusters on the IscU Scaffold Protein†. *Biochemistry* **46**, 6804-6811 (2007).

37. Agar, J.N. et al. IscU as a scaffold for iron-sulfur cluster biosynthesis: sequential assembly of [2Fe-2S] and [4Fe-4S] clusters in IscU. *Biochemistry* **39**, 7856-7862 (2000).

38. Gerber, J., Mühlenhoff, U. & Lill, R. An interaction between frataxin and Isu1/Nfs1 that is crucial for Fe-S cluster synthesis on Isu1. *EMBO Rep* **4**, 906-911 (2003).

39. Leidgens, S., De Smet, S. & Foury, F. Frataxin interacts with Isu1 through a conserved tryptophan in its beta-sheet. *Hum. Mol. Genet* **19**, 276-286 (2010).

40. Foury, F., Pastore, A. & Trincal, M. Acidic residues of yeast frataxin have an essential role in Fe-S cluster assembly. *EMBO Rep* **8**, 194-199 (2007).

41. Adam, A.C., Bornhovd, C., Prokisch, H., Neupert, W. & Hell, K. The Nfs1 interacting protein Isd11 has an essential role in Fe-S cluster biogenesis in mitochondria. *EMBO J* **25**, 174-183 (2006).

42. Biederbick, A. et al. Role of human mitochondrial Nfs1 in cytosolic iron-sulfur protein biogenesis and iron regulation. *Mol. Cell. Biol* **26**, 5675-5687 (2006).

43. Bonomi, F., Iametti, S., Morleo, A., Ta, D. & Vickery, L.E. Studies on the mechanism of catalysis of iron-sulfur cluster transfer from IscU[2Fe2S] by HscA/HscB chaperones. *Biochemistry* **47**, 12795-12801 (2008).

44. Rodríguez-Manzaneque, M.T., Tamarit, J., Bellí, G., Ros, J. & Herrero, E. Grx5 is a mitochondrial glutaredoxin required for the activity of iron/sulfur enzymes. *Mol. Biol. Cell* **13**, 1109-1121 (2002).

45. Dutkiewicz, R. et al. Sequence-specific interaction between mitochondrial Fe-S scaffold protein Isu and Hsp70 Ssq1 is essential for their in vivo function. *J. Biol.*

Chem **279**, 29167-29174 (2004).

46. Andrew, A.J., Dutkiewicz, R., Knieszner, H., Craig, E.A. & Marszalek, J. Characterization of the interaction between the J-protein Jac1p and the scaffold for Fe-S cluster biogenesis, Isu1p. *J. Biol. Chem* **281**, 14580-14587 (2006).

47. Sheftel, A.D. et al. Human ind1, an iron-sulfur cluster assembly factor for respiratory complex I. *Mol. Cell. Biol* **29**, 6059-6073 (2009).

48. Sharma, A.K., Pallesen, L.J., Spang, R.J. & Walden, W.E. Cytosolic iron-sulfur cluster assembly (CIA) system: factors, mechanism, and relevance to cellular iron regulation. *J. Biol. Chem* **285**, 26745-26751 (2010).

49. Cavadini, P. et al. RNA silencing of the mitochondrial ABCB7 transporter in HeLa cells causes an iron-deficient phenotype with mitochondrial iron overload. *Blood* **109**, 3552-3559 (2007).

50. Kuhnke, G., Neumann, K., Mühlenhoff, U. & Lill, R. Stimulation of the ATPase activity of the yeast mitochondrial ABC transporter Atm1p by thiol compounds. *Mol. Membr. Biol* **23**, 173-184 (2006).

51. Hell, K. The Erv1-Mia40 disulfide relay system in the intermembrane space of mitochondria. *Biochim. Biophys. Acta* **1783**, 601-609 (2008).

52. Bihlmaier, K., Mesecke, N., Kloeppel, C. & Herrmann, J.M. The disulfide relay of the intermembrane space of mitochondria: an oxygen-sensing system? *Ann. N. Y. Acad. Sci* **1147**, 293-302 (2008).

53. Banci, L. et al. MIA40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria. *Nat. Struct. Mol. Biol* **16**, 198-206 (2009).

54. Zhang, Y. et al. Dre2, a conserved eukaryotic Fe-S cluster protein, functions in cytosolic Fe-S protein biogenesis. *Mol. Cell. Biol* **28**, 5569-5582 (2008).

55. Vernis, L. et al. A newly identified essential complex, Dre2-Tah18, controls mitochondria integrity and cell death after oxidative stress in yeast. *PLoS ONE* **4**, e4376 (2009).

56. Netz, D.J.A. et al. Tah18 transfers electrons to Dre2 in cytosolic iron-sulfur protein biogenesis. *Nat Chem Biol* (2010).doi:10.1038/nchembio.432

57. Song, D. & Lee, F.S. A Role for IOP1 in Mammalian Cytosolic Iron-Sulfur Protein Biogenesis. *Journal of Biological Chemistry* **283**, 9231 -9238 (2008).

58. Stehling, O. et al. Human Nbp35 Is Essential for both Cytosolic Iron-Sulfur Protein Assembly and Iron Homeostasis. *Mol. Cell. Biol.* **28**, 5517-5528 (2008).

59. Puccio, H. & Koenig, M. Friedreich ataxia: a paradigm for mitochondrial diseases. *Curr. Opin. Genet. Dev* **12**, 272-277 (2002).

60. Babady, N.E. et al. Advancements in the pathophysiology of Friedreich's Ataxia and new prospects for treatments. *Molecular Genetics and Metabolism* **92**, 23-35

61. Campuzano, V. et al. Friedreich's Ataxia: Autosomal Recessive Disease Caused by an Intronic GAA Triplet Repeat Expansion. *Science* **271**, 1423-1427 (1996).

62. Condò, I. et al. In vivo maturation of human frataxin. *Human Molecular Genetics* **16**, 1534 -1540 (2007).

63. Pandolfo, M. Iron and Friedreich ataxia. *J. Neural Transm. Suppl* 143-146 (2006).

64. Allikmets, R. et al. Mutation of a Putative Mitochondrial Iron Transporter Gene (ABC7) in X-Linked Sideroblastic Anemia and Ataxia (XLSA/A). *Human Molecular Genetics* **8**, 743 -749 (1999).

65. Pondarré, C. et al. The mitochondrial ATP-binding cassette transporter Abcb7 is essential in mice and participates in cytosolic iron-sulfur cluster biogenesis. *Hum. Mol. Genet* **15**, 953-964 (2006).

66. Pondarre, C. et al. Abcb7, the gene responsible for X-linked sideroblastic anemia with ataxia, is essential for hematopoiesis. *Blood* **109**, 3567-3569 (2007).

67. Ye, H. et al. Glutaredoxin 5 deficiency causes sideroblastic anemia by specifically impairing heme biosynthesis and depleting cytosolic iron in human erythroblasts. *J. Clin. Invest* **120**, 1749-1761 (2010).

68. Mochel, F. et al. Splice Mutation in the Iron-Sulfur Cluster Scaffold Protein ISCU Causes Myopathy with Exercise Intolerance. *Am J Hum Genet* **82**, 652-660 (2008).

69. Gottlieb, E., King, A. & Selak, M.A. Succinate dehydrogenase and fumarate hydratase: linking mitochondrial dysfunction and cancer. (2006).at <http://eprints.gla.ac.uk/23549/>

70. Rudolf, J., Makrantoni, V., Ingledew, W.J., Stark, M.J. & White, M.F. The DNA Repair Helicases XPD and FancJ Have Essential Iron-Sulfur Domains. *Molecular Cell* **23**, 801-808 (2006).

71. Banci, L. et al. Affinity gradients drive copper to cellular destinations. *Nature* **465**, 645-648 (2010).

72. Brewer, G.J. Copper in medicine. *Current Opinion in Chemical Biology* **7**, 207-212 (2003).

73. Payne, A.S., Kelly, E.J. & Gitlin, J.D. Functional expression of the Wilson disease protein reveals mislocalization and impaired copper-dependent trafficking of the common H1069Q mutation. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 10854 -10859 (1998).

74. Fontaine, S.L. et al. Correction of the Copper Transport Defect of Menkes Patient Fibroblasts by Expression of the Menkes and Wilson ATPases. *Journal of Biological*

Chemistry **273**, 31375 -31380 (1998).

75. Hung, I.H. et al. Biochemical Characterization of the Wilson Disease Protein and Functional Expression in the Yeast Saccharomyces cerevisiae. *Journal of Biological Chemistry* **272**, 21461 -21466 (1997).

76. Paynter, J.A., Grimes, A., Lockhart, P. & Mercer, J.F. Expression of the Menkes gene homologue in mouse tissues lack of effect of copper on the mRNA levels. *FEBS Letters* **351**, 186-190 (1994).

77. Voskoboinik, I. & Camakaris, J. Menkes copper-translocating P-type ATPase (ATP7A): biochemical and cell biology properties, and role in Menkes disease. *J. Bioenerg. Biomembr* **34**, 363-371 (2002).

78. Bull, P.C., Thomas, G.R., Rommens, J.M., Forbes, J.R. & Cox, D.W. The Wilson disease gene is a putative copper transporting P-type ATPase similar to the Menkes gene. *Nat. Genet* **5**, 327-337 (1993).

79. Tumer, Z. & Moller, L.B. Menkes disease. *Eur J Hum Genet* **18**, 511-518 (2010).

80. Barry, A.N., Shinde, U. & Lutsenko, S. Structural organization of human Cu-transporting ATPases: learning from building blocks. *J. Biol. Inorg. Chem* **15**, 47-59 (2010).

81. de Bie, P., Muller, P., Wijmenga, C. & Klomp, L.W.J. Molecular pathogenesis of Wilson and Menkes disease: correlation of mutations with molecular defects and disease phenotypes. *Journal of Medical Genetics* **44**, 673 -688 (2007).

82. Mercer, J.F.B. The molecular basis of copper-transport diseases. *Trends in Molecular Medicine* **7**, 64-69 (2001).

83. Li, X. et al. A new apoptosis inhibitor, CIAPIN1 (cytokine-induced apoptosis inhibitor 1), mediates multidrug resistance in leukemia cells by regulating MDR-1, Bcl-2, and Bax. *Biochem. Cell Biol* **85**, 741-750 (2007).

84. Wang, Y., Qi, H., Li, X., Chen, X. & Liu, J. [CIAPIN1 expression in human lung cancer tissues and inhibitory effects of the gene on human pulmonary carcinoma NCI-H446 cells]. *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi* **24**, 434-437 (2008).

85. Shizusawa, T. et al. The expression of anamorsin in diffuse large B cell lymphoma: possible prognostic biomarker for low IPI patients. *Leuk. Lymphoma* **49**, 113-121 (2008).

86. Kim, K., Chung, W., Kim, H., Lee, K. & Roe, J. Monothiol glutaredoxin Grx5 interacts with Fe-S scaffold proteins Isa1 and Isa2 and supports Fe-S assembly and DNA integrity in mitochondria of fission yeast. *Biochem. Biophys. Res. Commun* **392**, 467-472 (2010).

87. Stemmler, T.L., Lesuisse, E., Pain, D. & Dancis, A. Frataxin and Mitochondrial FeS

Cluster Biogenesis. *Journal of Biological Chemistry* **285**, 26737 -26743 (2010).

88. Wiedemann, N. et al. Essential role of Isd11 in mitochondrial iron-sulfur cluster synthesis on Isu scaffold proteins. *EMBO J* **25**, 184-195 (2006).

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 have become an integral part of research and development in the biomedical sciences. Bioinformatics now has an essential role both in deciphering genomic, transcriptomic and proteomic data, generated by high-throughput experimental technologies, and in organizing information gathered from traditional biology. It is an interdisciplinary research area at the interface between the biological and computational science. The ultimate goal of bioinformatics is to uncover the biological information hidden in the mass of data and obtain a clearer insight into the fundamental biology of organisms.

Genome browsing is the first and crucial step for the expression and characterization of a recombinant protein. In this step, bioinformatic tools are necessary to analyse the nucleotide and amino acid sequences and obtain information useful for the choice of the protein target. A number of data banks are available and provide the scientific community 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** (www.ncbi.nlm.nih.gov/Entrez/) - This web site integrates information from several databases (Swissprot, EMBL, all geneBank, etc...)
- **PDB** (www.rcsb.org/pdb) - 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 alignments.[1]

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 is stored, the requirement is to provide bioinformatic tools for extracting the meaningful information. The most used programs for genome browsing are:

- **BLAST** (www.ncbi.nlm.nih.gov/BLAST/): Standard BLAST (Basic Local Alignment Search Tool) is a set of similarity search programs designed to explore all of the available sequence databases regardless of whether the query is protein or DNA. PHIBLAST is designed to search for proteins that contain a pattern specified by the user, and simultaneously are similar to the query sequence.
- **CLUSTALW** (www.ebi.ac.uk/clustalw/) is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylograms
- **PROSITE** (www.expasy.org/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.[2]

2.2. Construct design

Once the target protein is chosen, it is subjected to further bioinformatics investigations in order to predict important features like stability, solubility, hydrophobicity, secondary and tertiary structures. In this crucial step other bioinformatic tools can be used.

- Several tools can be used in order to predict if the target is a fully soluble, transmembrane protein, or a protein containing both soluble and transmembrane domains;
  http://www.sbc.su.se/~miklos/DAS/,
  http://www.cbs.dtu.dk/services/TMHMM-2.0/,

http://www.ch.embnet.org/software/TMPRED/form.html;

- The presence in the sequence of subcellular targeting signals can be confirmed by the tools from servers like: SignalP, TargetP and PSORT.org;[3,4,5]
  http://www.cbs.dtu.dk/services/TargetP/,
  http://www.cbs.dtu.dk/services/SignalP/,
  http://www.psort.org/;
- Topological and structural predictions can help in the identification of intrinsically unstructured regions and the prediction of secondary and tertiary structures; [6,7,8,9]
  http://iupred.enzim.hu/,
  http://bip.weizmann.ac.il/fldbin/findex/,
  http://www.sbg.bio.ic.ac.uk/~3dpssm/,
  http://pbil.univ-lyon1.fr/;
- The construct should be designed with respect to the N-end rules;[10]

These tools applied for each target protein, domain or construct selected for expression, can increase the probability of obtaining a soluble and well-folded protein.

2.3. Gene cloning

In order to get high yield of soluble proteins, many factors have to be taken in consideration, such as the choice of the vector, of the cloning strategy, and of culture conditions. Of course, the knowledge of the protein characteristics will help in this choice, thus increasing the chance of success. In the last years several eukaryotic expression systems were optimized, such as mammalian, yeast and insect cell expression. Cell-free protein synthesis has also a great potential, in particular with membrane proteins. [11,12,13] However, especially for NMR which requires high yields of labelled ^{15}N,^{13}C and ^{2}H samples, the *E.coli* expression system is nowadays the most widely used. That is why many efforts were focused on the optimization of this expression system. Gene fusion technology can facilitate purification, enhance protein expression and solubility, chaperone proper folding, reduce protein degradation, and in some cases, generate proteins with a native N-terminus. Factors such as reduced temperature, changes in the *E.coli* expression strain, different promoters or induction conditions and co-expression of molecular chaperones and folding modulators have all been examined and, in some specific cases,

they lead to enhancements of soluble protein production. Nevertheless, protein expression remains an arduous task that involves a complex decision tree. To date, no technology or reagent is a panacea. Thus, establishing tools and optimal conditions for each protein remains an empirical exercise.[14] The best way to maximize the probability of obtaining a soluble and correctly folded target protein is to proceed with a parallel cloning and expression of it with a high number of fusion partners. 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 to screen a broad range of conditions in less time, for these reasons new cloning technologies have been developed in recent years. Landy and coworkers have described a universal cloning method (Gateway technology) based on the site-specific recombination of bacteriophage lambda, which facilitates the integration of lambda into *E.coli* chromosome and the switch between the lytic and lysogenic cycle. [15] By this technology (developed by Invitrogen) it is possible to clone a target gene into different expression vectors eliminating the time consuming work with restriction enzymes and ligase (**Fig. 1**).

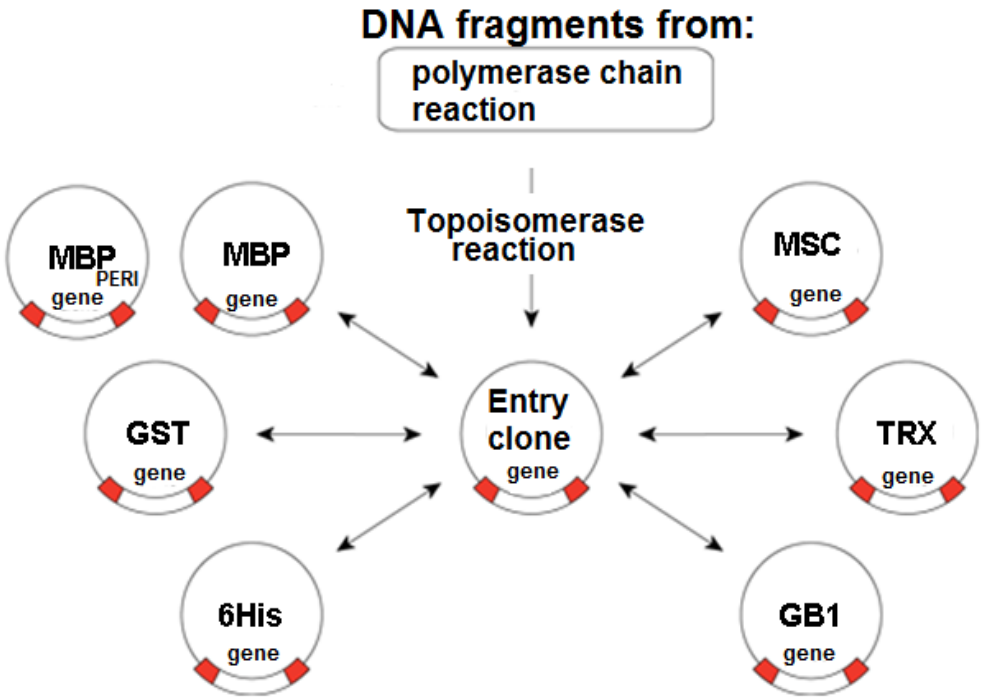


Figure 1. Gateway technology allows one-step cloning in the pENTR vector, and by site-specific recombination the target gene can be inserted in different expression vectors.

The entry vector is obtained via recombination reaction of the pENTR/TOPO vector with our gene (PCR fragment); the reaction is catalyzed by the Topoisomerase I. 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. This cloning method is faster due to the higher reaction efficiency and to the fact that only sequencing of the entry clone is required, and most of all, the destination vector are compatible with one single entry clone making the parallel approach easier. Moreover, using different destination vectors allows expressing a chosen protein with different fusions or tags, which can increase solubility, yield and make the purification of the target much easier. However, the main function of an expression vector is to yield the product of a gene (usually the more product the better). Therefore, expression vectors are ordinarily equipped with very strong promoters; the rationale is that the more mRNA that is produced, the more protein product will be made. The ideal promoter is strong, has a low basal expression level, is easy to induce and is independent from the common components of culturing media. The most used promoter system for the bacterial expression of recombinant proteins is the T7/lac promoter.[16] Genes under the control of T7/lac promoter can be transcribed by T7 RNA polymerase, in presence of lactose. Since *E.coli* cells do not produce this type of RNA polymerase, in order to use this promoter system they have to be genetically modified by incorporation of the genes encoding the T7 RNA polymerase, the lac promoter and the lac operator into their genome. The repressor is displaced from the lac operator when lactose or a similar molecule like Isopropyl β-D-1-thiogalactopyranoside (IPTG), is added to the culture. IPTG activates genes encoding the T7 RNA polymerase and the target protein in the plasmid, because lac operators are located upstream in the bacterial genome. Because of the high selectivity of the T7 RNA polymerase, almost all of the cell's resources are converted to target gene expression and the amount of desired product can reach up to 50% of the total cell protein in few hours after induction. In some cases (e.g. toxic or membrane proteins) the basal expression of the recombinant protein has to be reduced. In order to achieve this, host strains containing the pLysS or pLysE vectors that express T7 lysozyme (natural inhibitor of T7 RNA polymerase) should be used.

2.4. Protein expression

After bio-informatic characterization and choice of the cloning strategy several conditions for the target protein expression have to be tested in order to obtain a high yield of soluble protein. The variables which affect the expression of a recombinant protein are: host strain, growth medium and induction parameters (temperature, IPTG concentration and duration of induction step). A preliminary expression test is performed in a small-volume scale using at least:

- three different *E.coli* strains (e.g.: Bl21(DE3)pLysS a protease deficient strain, Rosetta(DE3) for rare codons containing genes and Origami(DE3) for disulphide containing proteins);
- three different expression vectors (containing different tags and/or fusion partners);
- three expression temperatures (37-25-17°C);
- three inducer (IPTG) concentrations and different induction times (4-6-16h).

Expression results are checked on SDS polyacrylamide gels (SDS-PAGE). This kind of approach allows 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 the scale-up. On the basis of these preliminary results, the expression protocol can be optimised and, in case of negative results, it is possible to try the expression of mutants, change the cloning strategy, the construct or the expression system. With such an approach it is possible to find good expression conditions for many proteins. Anyway some proteins can be difficult to obtain, since *E.coli* is a prokaryote and lacks intracellular organelles, such as the endoplasmic reticulum and the Golgi apparatus, which are responsible for post-translation modifications of the produced proteins.

In case of negative results, variables like bacterial strain, induction time, the kind of vectors and expression promoters can be modified. If the main fraction of the protein is produced in the insoluble fraction, another 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.[17]

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; in 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.

2.5. Protein purification

Basically, protein purification is a series of processes intended to isolate a single type of protein from a complex mixture. Protein purification is vital for the characterization of the function, structure and interactions of the protein of interest. Separation of one protein from all others is typically the most laborious aspect of protein purification. Separation steps exploit differences in protein size, physico-chemical properties and binding affinity.

The location of expressed protein within the host will affect the choice of methods for its isolation and purification. Bacterial host may secrete the protein into the growth media, transport it to the periplasmic space, express a cytosolic protein or store it as

insoluble inclusion bodies within the cytoplasm. For insoluble proteins, the first purification step is the extraction from inclusion bodies. Indeed, most of the bacterial proteins are removed by different extraction steps with native buffer conditions, while the recombinant protein is extracted from inclusion bodies with a denaturing buffer.

One of the most popular techniques of purification, which can be adopted for denatured and native state proteins, is immobilized metal ion affinity chromatography (IMAC). It is based on the specific coordinate covalent bond of amino acids, particularly histidine, to metals. This common technique involves engineering the sequence in such a way that 6 to 12 histidines are added to the N- or C-terminus of the protein. The polyhistidine binds strongly to divalent metal ions such as nickel. The protein can be passed through a column containing immobilized nickel ions, which binds the polyhistidine tag. All untagged proteins pass through the column. The protein can be eluted with imidazole, which competes with the polyhistidine tag for binding to the column, or by a decrease in pH (typically to 4.5), which decreases the affinity of the tag for the resin (**Fig. 2**).

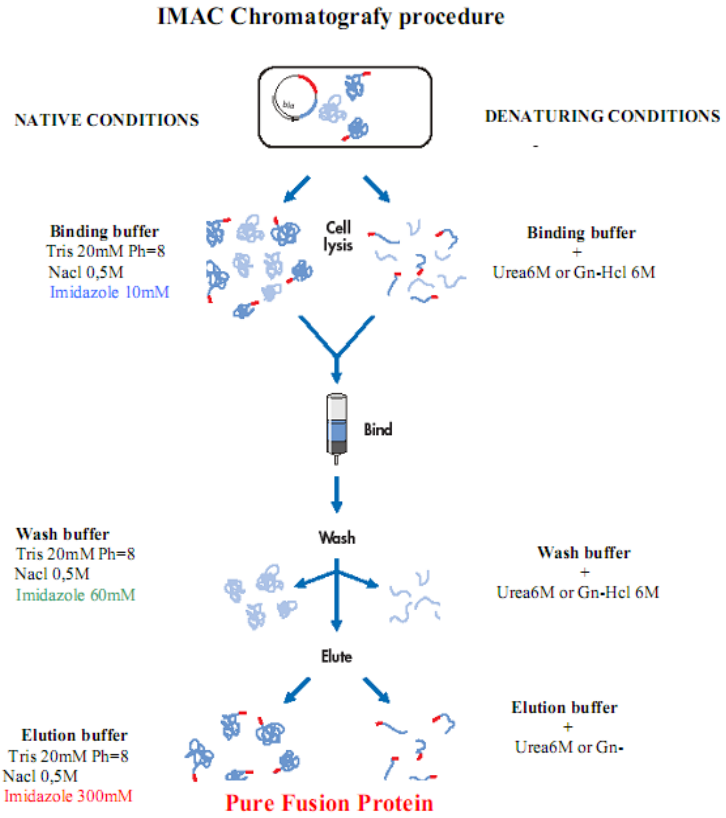


Figure 2. General purification procedure of a typical His-tagged proteins

(Reprinted from Qiagen handbook for expression and purification of His-tagged proteins)

After the affinity purification, the fusion-tag must be removed from the recombinant protein. Indeed many expression vectors are engineered to express a protease clevage site between the fusion-tag and the protein. Tobacco Etch Virus (TEV), Factor Xa, Thrombin, Prescission Protease, recombinant Enterokinase are some examples of proteases that are normally used for the cleavage of tags. A second IMAC is generally performed in order to separate the fusion from the target native protein. However, if the fused protein is expressed in inclusion bodies, it must be refolded before performing the tag cleavage. This is not always possible, since fusion tags may interfere with protein folding.

Gel filtration (GF) chromatography is the most common used purification step if the protein after affinity chromatography is not pure enough. GF 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. 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 bio-molecules 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.

2.6. Protein refolding

Heterologous expression of foreign genes in *E. coli* often leads to production of the expressed proteins in IBs. IBs must then be solubilized and refolded into an active conformation. Refolding of IBs is not a straightforward process, often requiring an extensive trial-and-error approach. There are two important issues in recovering active proteins from IBs, i.e., solubilization and refolding. Solubilization must result in monomolecular dispersion and minimum non-native intra- or inter-chain interactions. Choice of solubilizing agents, e.g., urea, guanidine HCl, or detergents, plays a key role in

solubilization efficiency, in the structure of the proteins in denatured state, and in subsequent refolding.[18,19]

Refolding is initiated by reducing concentration of denaturant used to solubilize IBs. Protein refolding is not a single reaction and competes with other reactions, such as misfolding and aggregation, leading to inactive proteins. Rate of refolding and other reactions is determined both by the procedure to reduce denaturant concentration and the solvent condition. The procedures to reduce the denaturant will be shortly described below:

- One-step dialysis: Denatured, unfolded protein samples in concentrated denaturant solution are dialyzed against a refolding buffer, and hence, exposed to descending concentration of the denaturant.
- Step-wise dialysis: This protocol uses descending concentration of denaturant for dialysis and has been successfully used for refolding antibodies. Unfolded protein sample is initially brought to equilibrium with high denaturant concentration, then with middle concentration, and with low concentration.
- Buffer-exchange by gel filtration: Gel filtration column is equilibrated with the final refolding buffer. Unfolded protein sample in denaturant is applied to the column and run through it with the refolding buffer. Use of desalting column will separate proteins from denaturant, while use of protein-sizing column will fractionate protein species.
- Dilution: Protein samples at high denaturant concentration are delivered into a large volume of refolding buffer. Dilution brings the unfolded sample into a rapid collapse, whereby bypassing the intermediate denaturant concentration.
- Solid phase refolding: Denatured protein is initially non-covalently bound to solid matrix such as Ni-resin or ion-exchange resin in the presence of denaturant. Denaturant concentration is then decreased to initiate refolding. Since protein molecules are bound to resin, this procedure minimizes aggregation of unfolded protein or folding intermediates.

2.7. Protein characterization

2.7.1. Metals content determination by ICP

Inductively Coupled Plasma (ICP) is an analytical technique used for the detection of trace metals and allows multi-elemental, simultaneous analysis of most elements in the periodic table, with excellent sensitivity (ng/l or ppt) and high sample throughput. The analytical principle used is optical emission spectroscopy, triggered by plasma ionization. The liquid sample is introduced in the pre-chamber, usually by means of a peristaltic pump, then nebulized and entrained in the flow of plasma support gas, which is typically Argon. The analytes are excited into a state of radiated light emission by the plasma vapour and the emitted radiation is converted to an electrical signal that can be measured quantitatively by resolving the light into its component radiation (almost always by means of a diffraction grating). The light intensity is measured, at the specific wavelength for each element line, with a CCD (*charge coupled device*, basically an array of tiny, light-sensitive diodes) which converts it to electrical signals. The electron signal intensity is then compared to previously measured intensities of standard solutions of the target elements, and a concentration is computed. This spectroscopic technique is very suitable for the protein's metal content determination, since it requires very small amounts of sample, and is not perturbed by the polypeptidic matrix.

2.6.2. Electron Paramagnetic Resonance

Electron paramagnetic resonance (EPR) or electron spin resonance (ESR) spectroscopy is a technique for studying chemical species that have one or more unpaired electrons, such as organic and inorganic free radicals or inorganic complexes possessing a transition metal ion. The basic physical concepts of EPR are analogous to those of nuclear magnetic resonance (NMR), but electron spins are excited instead of spins of atomic nuclei. Because most stable molecules have all their electrons paired, the EPR technique is less widely used than NMR. However, this limitation to paramagnetic species also means that the EPR technique is of great specificity, since ordinary chemical solvents and matrices do not give

rise to EPR spectra. In a typical experiment the sample is placed in a resonant cavity which has a high quality factor Q (high Q leads to high sensitivity of the EPR method). At a fixed microwave frequency, the magnetic field B_o is varied until resonance occurs at the value B_e given by

$$B_e = hv = g\beta B_o$$

In this equation g is the "g-factor" of the unpaired electron and β is the unit magnetic moment of the spinning electron (Bohr magneton). Free radicals have values that are fairly close to g = 2.0023 which characterizes a free electron, while some transition ions, such as cobalt (Co^{2+}) and copper (Cu^{2+}), have g-values which differ significantly form that theoretical value.

The resonant absorption is not infinitely narrow, since unpaired electrons do not only interact with the externally applied magnetic field, but also with the magnetic fields in their neighbourhood. By observing spectral line width and intensity, it is possible to obtain information about the spin environment. Electron spin exchange between identical and non-identical molecules, chemical exchange between the paramagnetic molecule and its environment, and the interaction of nearby molecules having unpaired spins are some examples of environmental effects which can influence line width and intensity in the EPR spectrum. Moreover, an observed spectrum can split in several lines referred to as hyperfine structure, arising from the electrons interacting with surrounding nuclear spins. This last property becomes very useful when investigating biological molecules that contain transition metal ions in their structure, since it can be exploited to characterize their coordination environment in dependence of their intrinsic geometry.

2.6.3. 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 with 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 databases 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, including quadrupoles, time-of-flight (TOF) analysers**,** magnetic sectors, both Fourier transform and quadrupole ion traps, and Orbitrap. 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. 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 it is usually 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.[20]

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.[21]

In Orbitrap ions are electrostatically trapped in an orbit around a central, spindle shaped electrode. The electrode confines the ions so that they both orbit around the central electrode and oscillate back and forth along the central electrode's long axis. This oscillation generates an image current in the detector plates which is recorded by the instrument. The frequencies of these image currents depend on the mass to charge ratios of the ions. Mass spectra are obtained by Fourier transformation of the recorded image currents. Orbitraps have a high mass accuracy, high sensitivity and a good dynamic range.[22]

2.6.4. 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 if so characterizing its secondary structure;
- Compare the structures of a protein 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 under stress (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 the protein.

2.6.5. NMR – protein structural characterization

Discovering the tertiary structure of a protein, or the quaternary structure of its complexes, can provide important clues about how the protein performs its function. Nuclear Magnetic Resonance (NMR) is an experimental method of structure determination that can produce information at atomic resolution. NMR experiments are able to provide information from which a subset of distances between pairs of atoms can be estimated, and the final possible conformations for a protein are determined by solving a distance geometry problem. Moreover NMR spectroscopy is unique among the methods available for three-dimensional structure determination of proteins at atomic resolution, since the NMR data can be recorded in solution. Considering that body fluids such as blood, stomach liquid and saliva are protein solutions where these molecules perform their physiological functions, knowledge of the molecular structures in solution is highly relevant. Furthermore, in addition to protein structure determination, NMR applications include investigations of dynamic features of the molecular structures, as well as studies of structural, thermodynamic and kinetic aspects of interactions between proteins and other solution components, which may either be other proteins or low molecular weight ligands.[23,24]

The preparation of the protein sample is a crucial step of this process, since a highly purified protein sample is required. Inhomogeneous preparation and/or aggregation of the protein as well as low molecular weight impurities may severely harm the structure determination. The NMR technique exploits the property that magnetic nuclei, with an uneven spin quantum number such as the isotopes ^{1}H, ^{2}H, ^{13}C, ^{15}N, have in a magnetic field an energy splitting. Therefore, protein isotope labeling is necessary for NMR analysis because not all atoms are magnetically active. An applied electromagnetic (EM) pulse causes the nuclei to absorb energy, transient time domain signals are detected as the system returns to equilibrium. Fourier transformation of the transient signal into a frequency domain yields a one-dimensional NMR spectrum, which is a series of resonances from the various nuclei at different frequencies, or chemical shifts. The chemical shift of an atom depends on the electronic environment of its nucleus. Ideally, each distinct nucleus in the molecule experiences a distinct chemical environment and thus has a distinct chemical shift by which it can be recognized. However, 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, protein 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. Multi-dimensional NMR spectra provide both, increased resolution and correlations which are easy to analyse.

There are different types of experiments that can detect through-bonds and through-space nucleus/nucleus interactions. The ^{1}H-^{15}N HSQC (Heteronuclear Single Quantum Correlation) experiment is probably the most frequently recorded experiment in protein NMR. Each residue of the protein (except proline) has an amide proton attached to nitrogen in the peptide bond. If the protein is folded, the peaks are usually well dispersed, and most of the individual peaks can be distinguished. Being a relatively cheap and quick experiment, the HSQC is useful to screen candidates for structure determination by NMR. The number of peaks in the spectrum should match the number of residues in the protein (though side chains with nitrogen-bound protons will add additional peaks). It will probably be difficult to solve the structure of the protein if this is not the case. The labour-intensive process of structure determination is usually not undertaken until a good HSQC spectrum is obtained. In order to analyze the nuclear magnetic resonance data, it is important to get a resonance assignment for the protein. That is to find out which chemical shift corresponds to which atom. This is typically achieved by sequential walking using information derived from several different types of NMR experiment.

For small proteins (less than 10 kDa), it is not required to label the sample with ^{13}C or ^{15}N. With unlabelled protein the usual procedure is to record a set of two dimensional homonuclear NMR experiments through correlation spectroscopy (COSY), of which several types include conventional correlation spectroscopy, total correlation spectroscopy (TOCSY) and nuclear Overhauser effect spectroscopy (NOESY). TOCSY gives information about different nuclei *via J* coupling and through-bond correlations, whereas NOESY gives information about correlations through space.[25,26] Homonuclear NMR is usually restricted to small proteins or peptides because of the signal overlap problem in the case of bigger proteins. This occurs when different protons have the same or very similar chemical shifts, and becomes greater as the protein becomes larger.

This barrier can be overcome with 3D NMR techniques and uniformly ^{13}C and ^{15}N labelled proteins. In the case of labelled proteins it is possible to record an experiment that transfers magnetization over the peptide bond, and thus connects different spin systems through bonds. The 3D experiments exclusively correlate the resonances of the protein

backbone, and the experiments used are: HNCA, HNCACB, HN(CO)CA, HN(CO)CACB, HNCO and HN(CA)CO.

In the case of proteins with a molecular weight larger than 30 kDa the use of TROSY (Transverse Relaxation Optimized Spectroscopy) experiments is necessary. The TROSY technique benefits from a variety of triple resonance NMR experiments as the 3D HNCA and HN(CO)CA. 3D H(C)CH-TOCSY and (H)CCH-TOCSY experiments are then used to link the side chain spin systems to the backbone assignments.[27,28]

Protein NMR techniques are continually being used and developed. 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.

References

1. Finn, R.D. et al. The Pfam protein families database. *Nucleic Acids Research* **38**, D211-D222 (2009).
2. Jensen, L.J. et al. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37**, D412-416 (2009).
3. Schneider, G. & Fechner, U. Advances in the prediction of protein targeting signals. *Proteomics* **4**, 1571-1580 (2004).
4. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protocols* **2**, 953-971 (2007).
5. Emanuelsson, O. Predicting protein subcellular localisation from amino acid sequence information. *Brief. Bioinformatics* **3**, 361-376 (2002).
6. Dosztányi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433-3434 (2005).
7. Prilusky, J. et al. FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435 -3438
8. Kelley, L.A. & Sternberg, M.J.E. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protocols* **4**, 363-371 (2009).
9. Perrière, G. et al. Integrated databanks access and sequence/structure analysis services at the PBIL. *Nucleic Acids Research* **31**, 3393 -3399 (2003).
10. Varshavsky, A. The N-end rule: functions, mysteries, uses. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 12142 -12149 (1996).
11. Boettner, M., Prinz, B., Holz, C., Stahl, U. & Lang, C. High-throughput screening for expression of heterologous proteins in the yeast Pichia pastoris. *Journal of Biotechnology* **99**, 51-62 (2002).
12. Ozawa, K., Wu, P.S.C., Dixon, N.E. & Otting, G. N-Labelled proteins by cell-free protein synthesis. Strategies for high-throughput NMR studies of proteins and protein-ligand complexes. *FEBS J* **273**, 4154-4159 (2006).
13. Ozawa, K., Dixon, N.E. & Otting, G. Cell-free synthesis of 15N-labeled proteins for NMR studies. *IUBMB Life* **57**, 615-622 (2005).
14. John P. Hall Applying Fusion Protein Technology to E. Coli. (2006).a <http://biopharminternational.findpharma.com/biopharm/article/articleDetail.jsp?id=301977&sk=&date=&pageID=5>
15. Landy, A. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem* **58**, 913-949 (1989).
16. Dubendorff, J.W. & Studier, F.W. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol* **219**, 45-59 (1991).
17. Miroux, B. & Walker, J.E. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol* **260**, 289-298 (1996).
18. Tsumoto, K., Ejima, D., Kumagai, I. & Arakawa, T. Practical considerations in refolding proteins from inclusion bodies. *Protein Expression and Purification* **28**, 1-8 (2003).
19. Tsumoto, K. et al. Role of Arginine in Protein Refolding, Solubilization, and Purification. *Biotechnol. Prog.* **20**, 1301-1308 (2004).
20. Tanaka, K. et al. Protein and polymer analyses up tom/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151-153 (1988).

21. Wollnik, H. Time-of-flight mass analyzers. *Mass Spectrom. Rev.* **12**, 89-114 (1993).
22. Hu, Q. et al. The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **40**, 430-443 (2005).
23. Dyson, H.J. & Wright, P.E. Insights into protein folding from NMR. *Annu Rev Phys Chem* **47**, 369-395 (1996).
24. Wüthrich, K. NMR Studies of Structure and Function of Biological Macromolecules. *Biosci Rep* **23**, 119-168 (2003).
25. Wider, G., Macura, S., Kumar, A., Ernst, R.R. & Wüthrich, K. Homonuclear two-dimensional 1H NMR of proteins. Experimental procedures. *Journal of Magnetic Resonance (1969)* **56**, 207-234 (1984).
26. Bax, A. & Lerner, L. Two-dimensional nuclear magnetic resonance spectroscopy. *Science* **232**, 960-967 (1986).
27. Bax, A. & Ikura, M. An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins. *J. Biomol. NMR* **1**, 99-104 (1991).
28. Kay, L.E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **89**, 496-514 (1990).

Chapter 3- RESULTS

3.1. Anamorsin is a 2Fe2S cluster-containing substrate of the Mia40-dependent mitochondrial protein trapping machinery

Lucia Banci [1,2,*], Ivano Bertini [1,2,*], Simone Ciofi-Baffoni [1,2], Francesca Boscaro [3], Maciej Mikolajczyk [1,2], Julia Winkelmann [1,2]

[1] Magnetic Resonance Center CERM, University of Florence, Via Luigi Sacconi 6, 50019, Sesto Fiorentino, Florence, Italy.
[2] Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Florence, Italy.
[3] Centro Interdipartimentale di Spettrometria di Massa (CISM), University of Florence, Viale G. Pieraccini 6, 50139 Florence, Italy

[*] Correspondence:
Prof. Ivano Bertini; Phone: +39 055 4574272; Fax: +39 055 4574271; E-mail: bertini@cerm.unifi.it
Prof. Lucia Banci; Phone: +39 055 4574263; Fax: +39 055 4574253; E-mail: banci@cerm.unifi.it

Running Title: Structure and Mia40-interaction of 2Fe2S-anamorsin

I.B. and L.B. planned the research and coordinated the writing of the text, to which all the co-authors contributed; M.M. coordinated and performed protein production and characterization; S.C.-B. guided the flow of experiments, planned and recorded NMR spectra; J.W. solved the solution structure of N-terminal domain of anamorsin and analyzed all NMR experimental data; F.B. performed and analyzed ESI- and MALDI-MS. Manuele Migliardi is also acknowledged for performing EPR measurements.

Introduction

Beside their well-established function in energy supply for all cell compartments, mitochondria are also the primary source of iron-sulfur (Fe/S) clusters that are utilized by proteins throughout the cell.[1] Such Fe/S cluster proteins have essential functions in metabolism, electron transport and regulation of gene expression. Despite the relative simplicity of Fe/S clusters in terms of structure and composition, their synthesis and assembly is a highly complex and coordinated process which yet has not been understood in a comprehensive way.[2,3,4] Defects in human Fe/S protein biogenesis are responsible of numerous diseases like e.g. Friedreich's ataxia or sideroblastic anemia.[5,6] The cluster biogenesis in eukaryotes involves three distinct machineries localized in two different cellular compartments: i) the iron-sulfur cluster (ISC) assembly machinery and ii) the ISC export machinery, both in mitochondria, and iii) the cytosolic iron-sulfur cluster assembly (CIA) machinery in the cytoplasm.[7] The components of the ISC assembly machinery are similar to the proteins mediating Fe/S cluster assembly in bacteria and apparently were inherited from the prokaryotic ancestor of mitochondria.[8]

Generation of extra-mitochondrial (both cytosolic and nuclear) Fe/S proteins requires a yet unidentified product of the ISC assembly pathway in mitochondria. This compound is exported to the cytosol by the ISC export machinery whose central component is an ABC transporter in the mitochondrial inner membrane. The export reaction is assisted by the sulfhydryl oxidase Erv1/ALR and glutathione in the IMS.[4,9,10]

In yeast, a protein, Dre2, has been recently suggested to be implicated in the cytosolic Fe/S protein maturation.[11] Dre2 is a Fe/S cluster protein and the phenotypes of dre2 mutants or of Dre2-depleted cells are similar to those caused by alteration of CIA components. It has recently also been found that the diflavin reductase Tah18, which is also part of the yeast CIA machinery, forms a complex with Dre2 in the cytosol. This complex is part of an electron transfer chain functioning in an early step of the cytosolic Fe/S protein biogenesis.[12] The pathway is conserved in mammals as the human counterparts, Ndor1 and CIAPIN1, are able to functionally replace the yeast Tah18 and Dre2 proteins in yeast and to form a stable complex like the yeast homologues. However, at variance with all the other CIA protein components, for which no mitochondrial localisations have ever been reported, a fraction of Dre2 is localized in the mitochondrial IMS.[11] Additionally, it has been reported that Tah18 relocalises to the mitochondria to act as a molecular sensor for high oxidative stress levels that involves mitochondria and accurately

controls an oxidative induced cell death pathway in yeast.[13]

How the dual mitochondrial/cytoplasmic localization of Dre2 and Tah18 is accomplished is still unknown. Since Dre2 has two CX2C motifs, similar to some other IMS proteins, it has been proposed that the dual cellular protein distribution might be explained by a yet unknown process of folding of the Dre2 precursor outside mitochondria which competes with the oxidative folding (and trapping) inside the IMS performed by Mia40 and Erv1.[11] The latter proteins trap in the IMS a class of mitochondrial proteins having characteristic cysteine motifs (twin CX3C or CX9C, or CX2C). Specifically, the mitochondrial protein can be directed to the IMS by targeting signals that allow passage through the import translocase; subsequently, it can be retained within this compartment by interaction with the Mia40 protein partner which catalyses the formation of two disulphides between the twin cysteine motifs of the substrate.[14,15] The human homologue of yeast Dre2, called anamorsin or CIAPIN1 (hereafter named as anamorsin) has not only been associated with Fe/S cluster assembly, but it has been also identified as a necessary molecule for hematopoiesis which mediates antiapoptotic effects of various cytokines (Cytokine-Induced Apoptosis Inhibitor 1), without sharing any homology to known apoptosis regulatory molecules such as Bcl-2 family, caspase family, or signal transduction molecules.[16] Additionally, anamorsin was reported to be involved in the development of multidrug resistance in gastric and leukemia cells.[17] Several preclinical studies have also demonstrated that anamorsin is associated with increased tumor recurrence and shorter patient survival in different human tumor models, making anti-anamorsin therapy a new cancer treatment strategy.[18] Anamorsin is therefore an extremely important molecule playing a crucial role in cellular survival. To contribute to the comprehension of the possible functional role of anamorsin in the IMS, we have investigated in this work the structural and metal binding properties of anamorsin as well as its interaction with Mia40. From this molecular analysis we conclude that anamorsin is a Mia40-substrate able to bind a 2Fe2S cluster before and after the Mia40-driven Cys-oxidation, in agreement with a possible role of anamorsin in the Fe/S cluster biogenesis also once localized in the intermembrane space of mitochondria.

Materials and methods

Bioinformatic analysis

Homologous sequences to anamorsin were searched via BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) in the database of non-redundant protein sequences using Blastp (protein-protein BLAST). All the sequences were then divided into domains using the Pfam/SMART databases.[19] Sequence alignments were then performed on the full-length sequences and on the identified domains, using the ClustalW program and default parameters.[20] The folding probability was predicted by the algorithm FoldIndex©.[21] Homologous structures to the N-terminal domain of anamorsin were identified via the DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server/).[22]

Molecular cloning, expression and purification of full-length, N-terminal domain and C-terminally truncated (1-270aa) anamorsin

The cDNA coding for the human protein anamorsin (UniProtKB/Swiss-Prot: Q6FI81, Genbank Accession No. BC024196) was commercially acquired from Imagenes (ORF shuttle Clone IOH4051) and after amplification by PCR inserted into the Gateway® pEntr-TEV-d-Topo vector (InvitrogenTM). For recombinant expression it was cloned via the Gateway®technology (InvitrogenTM) into the expression vector pCold-DEST. This vector is based on the pColdI (Takara Bio Inc, Japan) and has been modified to make it suitable for the Gateway® system. According to the vector construction the protein is expressed with an N-terminal His12 tag and for the release from it a sequence coding for the specific recognition site of the TEV protease was located at the 5' end of the anamorsin. All the recombinant proteins contain four additional amino acids, GSFT, at the N-terminus.

BL21-CodonPlus (DE3)-RIPL Competent Cells (Stratagene, La Jolla, CA) were transformed with the anamorsin/pCOLD-DEST construct. The cultures were grown at 37°C under vigorous shaking overnight, diluted 1:100 into LB medium or minimal medium containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol and grown further at 37°C until OD600 reached 0.8-1. The culture was transferred to 25°C and protein expression was induced by addition of IPTG to a final concentration of 0.2 mM and incubation was continued for 3 h (rich media) or 4 h (minimal media). The full-length anamorsin and C-terminally truncated construct were expressed in the presence of 125 μM

FeCl3. Subsequently, the cells were harvested by centrifugation (6000 x g, 15 min, 4°C) and the pellets were stored at -20°C. The frozen cell pellets were re-suspended in degassed lysis buffer (Tris-HCl 50mM, NaCl 500mM, glycerol 5%, imidazole 20mM, pH 8) containing lysozyme (0,25mg/ml), DNA-se I (0.01mg/ml) and MgSO4 (40mM). Cells were incubated on ice for 30 min with gentle shaking and subsequently, disrupted on ice by sonication alternating 30 seconds of sonication and 5 minutes of resting for 10 times. The resulting cell lysate was centrifuged (20000 x g, 40 min, 4°C). Employing the N-terminal His tag, all the proteins were purified by immobilized metal ion affinity chromatography, under anaerobic conditions in a glove box for the full-length protein and the C-terminally truncated construct and under aerobic conditions for the N-terminal domain. The supernatant was applied to a nickel-charged HiTrap chelating HP column (Amersham Pharmacia Biosciences) and anamorsin constructs were eluted with degassed elution buffer (50mM Tris-HCl, 500mM NaCl, 500 mM imidazole, 5% glycerol, pH 8) and concentrated in presence of 3mM DTT with an Amicon Ultra-15 Centrifugal Filter Units with a MWC of 10 kDa (Millipore). The His-tag was cleaved from the recombinant proteins by incubation with TEV protease (3µl TEV/1mg anamorsin) and used for reverse Ni(II) affinity chromatography in order to remove the cleaved His-Tag. Size exclusion chromatography was performed as the final purification step using a HiLoad 16/60 Superdex 75 pg column (GE Healthcare) and degassed 50mM Tris-HCl, 500mM NaCl, 2mM DTT, 5% glycerol, pH 8 as a running buffer. The protein was used directly for analysis or frozen in liquid nitrogen and stored at -80°C.

Degradation of the full-length anamorsin was followed by MALDI-MS. 1 µl of protein solution was mixed with a 1 µl of matrix solution (SA 10 mg/ml in 70 % acetonitrile/30% water, 0.1% TFA) and analysed on a MALDI TOF/ TOF mass spectrometer Ultraflex III (Bruker Daltonics, Bremen, Germany) by using Flex Control 3.0 as data acquisition software in positive linear mode. The instrument was externally calibrated prior to analysis using the Bruker Protein I calibrant standard kit (5-17 kDa).

Reconstitution of 2Fe2S-anamorsin

Purified full-length anamorsin or C-terminally truncated construct was dialyzed (MWC:10kDa) against the 500-fold volume of degassed reconstitution buffer (50 mM Tris-HCl, 200 mM NaCl, 5 mM DTT, pH 7.5) containing 10-fold molar excess of FeCl3 and Na2S for 8h at 25°C. To remove unbound ions, the mixture was exchanged with fresh reconstitution buffer (without FeCl3 and Na2S) using PD-10 Desalting column (Amersham Biosciences) and then the reconstituted protein was further dialysed at 4°C for 12h. All the reconstitution experiments were performed under anaerobic conditions and in dark. The apo-form of anamorsin was prepared according to the procedure published by Kennedy et al.[23]

Iron and sulfide quantification

Iron concentration and sulfide content of purified, reconstituted anamorsin were determined by colorimetric assays. To estimate the iron content a standard procedure has been followed with few modifications.[11] 350µL of each sample dilution were treated with 50µl of 5% SDS for 5 minutes and then reduced with 50µl of 10mM sodium dithionite followed by addition of 50µl of 10mM bathophenanthrolinedisulfonic acid disodium salt to chelate iron(II). Samples were incubated in the dark for 20 minutes, mixed and centrifuged for 5 minutes at 12000 x g. Absorbance at 515nm was measured and the concentration was estimated through standard curve analysis using $NH_4Fe(SO_4)_2$. To estimate the sulfide content, a standard procedure has been followed with few modifications.[24] Samples (600µl) were treated with 50µl of 6% NaOH and incubated for 5 minutes. Samples were mixed with 125µl of 0.1% DPD (N,N-dimethyl-p-phenylenediamine dihydrochloride, Sigma-Aldrich) followed by addition of 50µl of 11.5 mM $FeCl_3$ and incubated for 30 minutes at room temperature. Sulfide concentration was estimated at 670nm through standard curve analysis using sodium sulfide Na_2S.

UV/Vis, EPR and NMR spectroscopy

UV-visible spectra of oxidized and reduced 2Fe2S-anamorsin in degassed 50mM phosphate buffer at pH 7.0 were performed on a Cary 50 Eclipse spectrophotometer. EPR spectra of oxidized and reduced 2Fe2S-anamorsin in 50mM Tris-HCl pH 8.0, 500mM NaCl or 50mM phosphate buffer at pH 7.0 and 10% glycerol was performed on a Bruker Elexsys E500 spectrometer equipped with a X-band microwave bridge (microwave frequency, 9.45 GHz) and an ER 4131 VT (unit for temperature control). EPR parameters were the following: sample temperature 45K, microwave frequency 9.45 GHz ; microwave power 5 mW; modulation frequency 9,387691 GHz; modulation amplitude 2500 G, time constant 167 ms. To reduce the cluster, 1 to 10 mM dithionite was added under anaerobic conditions and the sample was immediately frozen. ^{1}H paramagnetic NMR spectra of 2 mM oxidized 2Fe2S-anamorsin (50 mM phosphate buffer pH 7.0 in D2O, 1mM DTT) were recorded on a Bruker Avance 600 MHz spectrometer equipped with a selective ^{1}H probe head (5 mm SEL ^{1}H) at 298 K. All 1D spectra were taken by using a super WEFT pulse sequence with a recycle delay of 54 ms.[25] ^{15}N R_1, R_2, and steady-state heteronuclear NOE measurements were performed at 14.1 T (600MHz), 298K, using the pulse sequences proposed by Farrow et al. on ^{15}N-labeled samples.[26,27] The overall rotational correlation time values were estimated from the R_2/R_1 ratio using the program QUADRATIC_DIFFUSION (Lee, et al., 1997).[28] The relaxation data for those NHs having an exchange contribution to the R_2 value or exhibiting large-amplitude internal motions, as monitored by low NOE values, were excluded from the analysis.

Interaction of anamorsin with Mia40

The cysteine mutant (C4S) of the Mia40 protein was recombinantly expressed and purified as described previously in Banci et al. (Banci, et al., 2009)with the exception that the BL21-Gold(DE3) *E. coli* strain (Stratagene, La Jolla, CA) was used for protein expression. The protein was oxidized with 10 mM $K_3[Fe(CN)_6]$ (Sigma-Aldrich) for 75 min at 4°C in order to assure the formation of all three disulfide bonds. $K_3[Fe(CN)_6]$ was then removed by a PD-10 column.

To monitor the interaction by NMR, titrations of ^{15}N-labeled apo- or 2Fe2S-anamorsin with unlabeled Mia40 and of unlabeled apo- or 2Fe2S-anamorsin with ^{15}N-labeled Mia40 were performed in anaerobic conditions. Chemical shift changes were followed by ^{1}H-^{15}N HSQC spectra after addition of increasing amounts of the unlabeled partner.

To detect Mia40-induced disulfide formation in anamorsin by MALDI-MS, a full-length 2Fe2S-anamorsin sample was divided into two aliquots I and II (50 µl each). Sample II was mixed with 12.5 µl of 700 µM Mia403S-S (phosphate buffer 50 mM pH 7.0, 0.5 mM EDTA) and incubated at room temperature for 5 minutes. Both samples were then precipitated with 20% TCA, subsequently washed with acetone and re-suspended in 50µl of degassed buffer (Tris-HCl 50 mM, Urea 8M pH 7.5). 10µl of freshly prepared 100 mM IAM were added to each sample and the mixture was incubated in anaerobic conditions at 25°C in the dark for 1h. Reacted samples were applied on a SDS-PA gel and run without addition of DTT, so that all the cysteine residues that are not involved in disulfide bonds have to carry the IAM modification. Bands corresponding to anamorsin from both sample I and II were excised from the gel and individually transferred to a 1.5 ml microcentrifuge tube. Spots were washed three times for about 10 min in 40 µl of acetonitrile and then in 40 µl of a 0.1 M ammonium bicarbonate water solution. The solution was removed and digested either with trypsin or chymotrypsin. 30 µl of a 1 ng/µl of modified trypsin (Promega, Madison, WI) in 10 mM ammonium bicarbonate was added to each spot. Samples were kept at 4 °C for 30 min, then the solution was removed and replaced with 25 µl of 10 mM ammonium bicarbonate. After overnight digestion at 37 °C, the supernatant was recovered and the reaction blocked by addition of 1 µl of 10% TFA. For chymotrypsin digestion, 10 µl of a 100 ng/ µl of the enzyme dissolved in AMBIC 25 mM, CaCl2 10 mM were added to the gel pieces. Digestion was proceeded overnight at 37°C and stopped by addition of 1 µl of 10% TFA and analysed by MALDI-MS. 0.3 µl of matrix solution (HCCA 10 mg/ml in 70 % acetonitrile/30% water, 0.1% TFA) were mixed with 0.3 µl of the digested proteins on an anchorchip target (600 µm diameter). Spectra were acquired in positive Reflecton mode linear mode over the m/z range 800-5,000. The instrument was externally calibrated prior to analysis using the Bruker Peptide calibrant standard kit (1046-3147Da).

ESI-MS data were acquired in positive ion mode, setting the spray voltage at 4 kV, the capillary voltage and temperature respectively at 45 V and 220 °C, and the tube lens at

230V. For acquisition, Xcalibur 2.0 software (Thermo) was used and monoisotopic and average deconvoluted masses were obtained by using the integrated Xtract tool. For spectra acquisition a nominal resolution (at m/z 400) of 100.000 was used.

NMR spectroscopy and structure calculations

All NMR spectra were recorded at 298K using Bruker Avance 500, 600 and 900 MHz spectrometers, processed using the standard Bruker software Topspin and analyzed by the program CARA (Keller, 2004).[29] The NMR experiments used for resonance assignment and structure calculations were performed on 0.4-1 mM ^{13}C, ^{15}N-labeled and ^{15}N-labeled samples of the N-terminal domain of anamorsin (aa 1-172) in 50 mM phosphate buffer pH 7, 2 mM DTT containing 10% (v/v) D2O. The ^{1}H, ^{13}C and ^{15}N backbone resonance assignment was performed using standard triple-resonance NMR experiments and side chain assignment using TOCSY- and NOESY-based NMR experiments. Completeness of all atom assignment: 92.56%, completeness of proton atom assignment: 94.34%, completeness of heavy atom assignment: 90.37%.

Structure calculations were performed with the software package UNIO (ATNOS/CANDID/CYANA)[30,31,32], using as input the amino acid sequence, the chemical shift lists, four [^{1}H,^{1}H]-NOE experiments (two-dimensional NOESY, three-dimensional ^{15}N-resolved NOESY and two different three-dimensional ^{13}C-resolved NOESY), ψ and φ backbone torsion angle constraints, derived from ^{1}H, ^{13}C and ^{15}N chemical shifts through the program TALOS+[33] were also used in the structure calculations. The protein hydrogen-bonding database potential (HBDB)[34] was also used at the structure refinement stage. The 30 conformers with the lowest residual target function values were subjected to restrained energy minimization in explicit water with the program AMBER 10.[35] The quality of the structures was evaluated by the programs PROCHECK, PROCHECK-NMR, WHAT IF, PSVS (Bhattacharya, et al., 2007) and iCing (http://nmr.cmbi.ru.nl/cing/iCing.html).[36,37,38] The conformational and energetic analyses of the final REM family of 30 structures are reported in **Table S2**. The Ramachandran plot of the 30 conformers (aa 2-165) shows that 88.7% of the residues lie in the most favorable region of the plot, 10.2% in the allowed region, 1% in the generously allowed and 0 % in the disallowed regions.

Results

Anamorsin has a CIAPIN1 domain potentially responsible of its IMS trapping through the mitochondrial Mia40-Erv1 machinery

Bioinformatic analysis shows that a CIAPIN1 domain is present in the anamorsin amino acid sequence. The CIAPIN1 domain contains a highly conserved Cys-rich motif, i.e. $CX_{5-14}CX_2CXCX_nCX_2CX_7CX_2C$ (**Fig. 1A**), and is present in 182 sequences found in 103 eukaryotic species (average length of the domain 85 a.a.). In these proteins the CIAPIN1 domain is typically located at the C-terminus and can be fused to other functional domains. In anamorsin, the CIAPIN1 domain is fused at the N-terminus to a methyltransferase-like domain. Secondary structure and folding predictions reveal that these two domains are separated by an unstructured amino acid linker of ~ 50 residues (**Fig. 1B**) and that a large part of the CIAPIN1 domain, in particular its initial segment (residues 222 to 266), is structurally highly disordered, while the final segment is predicted to be more structured (Fig. 1B). The latter region of the CIAPIN1 domain contains a twin CX2C motif while the initial segment of the CIAPIN1 domain contains the additional four cysteines of the Cys-rich motif (Fig. 1B). Since it has been shown that the highly homologous yeast protein, Dre2, is able to bind a Fe/S cofactor, it is likely that cysteines from this Cys-rich motif in anamorsin are involved in iron binding.[11]

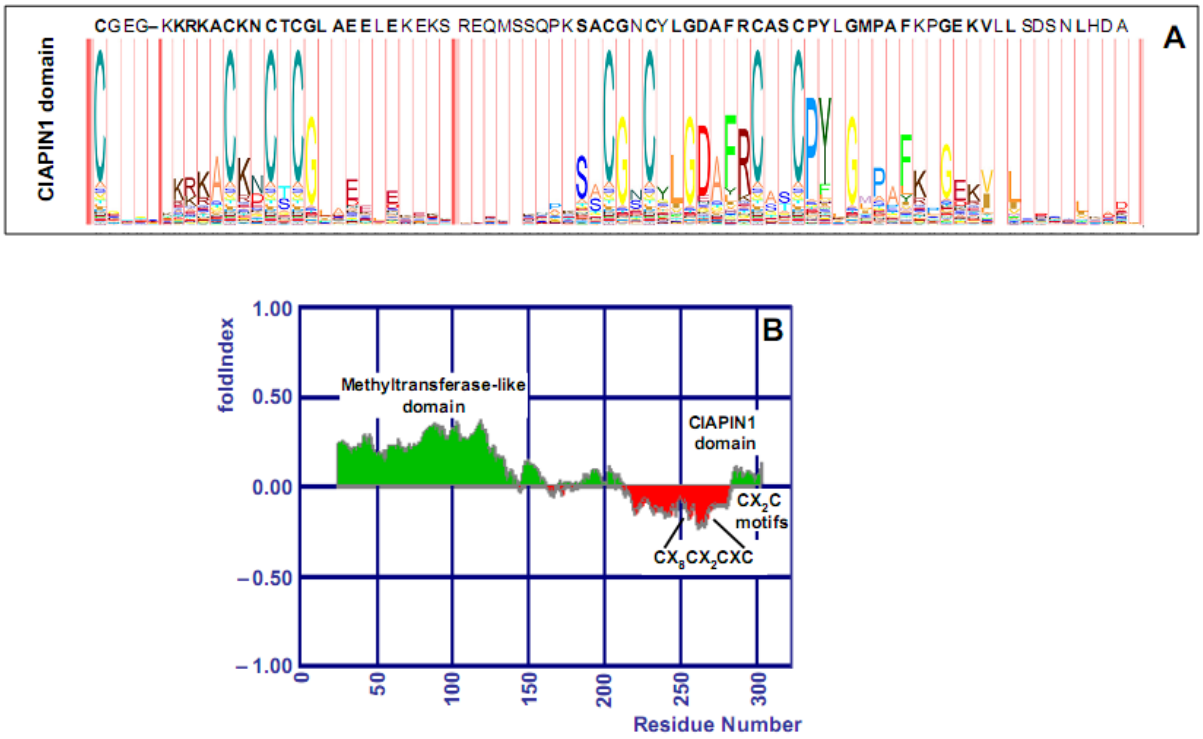


Figure 1. Bioinformatic analysis of anamorsin. (A) Sequence alignment of the CIAPIN1 domain conserved in 182 sequences by using LogoMat-M visualization tool.[39] The relative size of a letter expresses its probability to occur at this position in the sequence alignment. Pink column indicates an insert region of a number of residues with no conservation in the sequence alignment. (B) Folding prediction of anamorsin derived from FoldIndex program.[21] Regions shaded in green are predicted to be in a folded state, while red shaded regions are predicted to be unfolded.

In cell fractionation studies, anamorsin was found in the cytoplasm, nucleus, and nucleolus, while mitochondria were not specifically examined.[40] Accordingly, bioinformatic predictors show the presence of a putative nuclear localization signal and a putative leucine-rich nuclear export signal, while the typical mitochondrial matrix targeting signals are not bioinfomatically predicted.[40] However, the yeast Dre2 homologue has been shown to be localized in the IMS and it has been proposed to be a partner of the Mia40-Erv1 IMS-trapping machinery.[11] Mia40 is involved in the recognition of IMS-imported proteins typically containing a twin CX9C or CX3C motif or CX2C motif.[41,42,43] Recently, a hydrophobic IMS-targeting signal (ITS) located close to these twin CX9C or CX3C motifs in Mia40 substrates has been identified.[44,45] This ITS sequence is able to promote the docking of the substrate into a hydrophobic cleft of Mia40 which is then able to catalyze the formation of disulfide bond(s) in the substrate, thus trapping Mia40 substrates in the IMS. Anamorsin has a twin CX2C motif at the C-terminus of the CIAPIN1 domain

that could potentially be recognized and oxidized by Mia40. According to this proposal, sequence comparative analysis shows that the CIAPIN1 domain has highly conserved hydrophobic residues following the last cysteine of the twin CX2C motif, strongly suggesting that it could act as a possible ITS recognizable by Mia40. In conclusion, this analysis suggests that anamorsin could potentially be trapped in the
IMS through the Mia40-Erv1 machinery.

Anamorsin contains a well-folded N-terminal domain and an unstructured, flexible C-terminal part carrying a Fe/S cluster

Recombinant expression of anamorsin in the presence of iron and subsequent purification results in a reddish-brown protein solution indicating the presence of a Fe/S cluster. The protein is monomeric in solution and degrades within 7-10 days at room temperature under anaerobic conditions resulting into a stable fragment of 19 kDa (**Fig. 2**). During degradation the color of the solution is lost and mass spectrometry analysis revealed that the stable fragment is constituted by the N-terminal part (aa 1-172) of the protein. This part does not contain the CIAPIN1 domain and, accordingly, neither the cysteines potentially involved in the coordination of the Fe/S cluster. The ^{1}H-^{15}N HSQC spectra of the full-length protein (**Fig. 2A**) shows highly crowded cross-peaks in the spectral region between 8 and 8.5 ppm indicating that the protein contains an unstructured region, in addition to several signals spread over a larger spectral range, indicative of folded conformations. After degradation, the ^{1}H-^{15}N HSQC spectrum of the resulting fragment (**Fig. 2B**) shows the disappearance of many of the signals in the crowded, central region, while the well-resolved resonances are still present and do not show significant chemical shift changes.

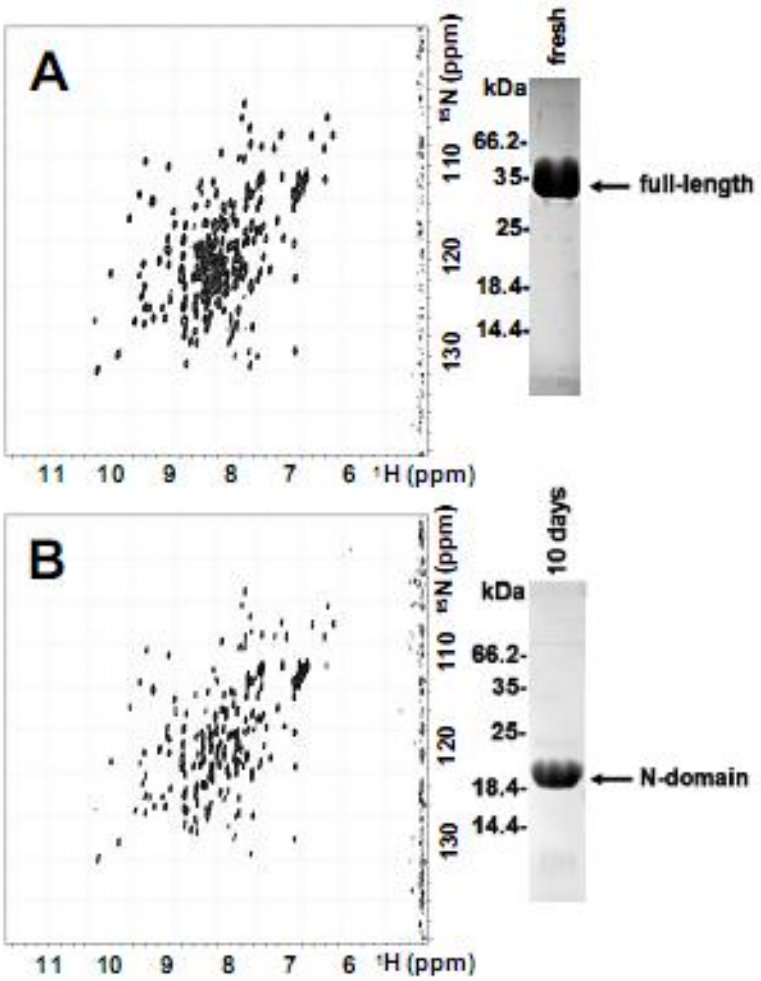


Figure 2. Structural properties of anamorsin. (A) ^{1}H-^{15}N HSQC spectrum at 500 MHz and 308K of a freshly purified 2Fe2S-anamorsin. SDS PAGE of the corresponding sample. (B) ^{1}H-^{15}N HSQC spectrum at 500 MHz and 308K of anamorsin recorded after 10 days of incubation under anaerobic conditions. SDS PAGE of the corresponding sample after 10 days. Anamorsin degraded into a stable fragment of ~19 kDa. PD10 column was performed under anaerobic conditions before recording the NMR spectrum to remove smaller degradation fragments. NH signals clustered in the central region of ^{1}H-^{15}N HSQC spectrum of the freshly purified anamorsin disappear after its degradation. (see also **Fig. S1**).

^{15}N{^{1}H} NOEs experiments on the full-length anamorsin showed that the residues not belonging to the N-terminal domain (which are in the highly crowded spectral region between 8 and 8.5 ppm) exhibited negative or very low NOEs, indicative of a high degree of flexibility in the C-terminal part of the protein (**Fig. S1**). We can conclude that the N-terminal part is well-folded while the C-terminal part is largely unstructured and flexible, and therefore, highly prone to degradation. Comparison of the signals that can be unambiguously assigned in both, the full-length and the N-terminal domain ^{1}H-^{15}N HSQC

spectra (70 spread over the entire sequence), shows similar chemical shifts with average weighted Δ_{HN} values being smaller than 0.05 ppm (**Fig. S1**) (except for two residues at the beginning and the last residues of the N-terminal domain). This result indicates that there are no interactions between the N- and C-terminal domains in the full-length protein and that the cysteines involved in Fe/S coordination are all part of the CIAPIN1 domain.

Anamorsin coordinates a 2Fe2S cluster in the CIAPIN1 domain

The binding motifs for the different Fe/S clusters in proteins are not highly conserved, yet some consensus motifs have been recognized.[7,46] In particular, the binding motifs for the different 2Fe2S cluster proteins were divided into three main groups.[2,47] However, comparison of the anamorsin cysteine motifs with these groups gave no clear hint on the kind of cluster.

Chemical reconstitution of full-length anamorsin with a 10-fold access of iron and sulfur in the presence of DTT resulted in an enhanced intensity of its reddish-brown color. After in vitro binding of the cluster to the protein, the absorbance ratio A_{420}/A_{280} (which is an indicator of the structural integrity of the cluster) is around 0.38, in good agreement with that of spinach ferredoxin, and shows that nearly all protein molecules are loaded with the cluster.[48] At a protein concentration of 76 μM the amount of iron and sulfur is determined to be 150 μM and 157 μM, respectively. This ratio of Fe:S:protein of 2:2:1 after cluster reconstitution has been achieved in all preparations.

In the UV/vis spectra (**Fig. 3A**) an absorption peak at 333 nm and weaker, broad peaks at around 420 nm and 460 nm are observed, that are typical absorption features of 2Fe2S clusters.[49] These absorbances are quenched after reduction of the cluster with 1 mM dithionite (**Fig. 3A inset**). In a typical 2Fe2S cluster, absorption from 420 nm to 460 nm is attributed to the a charge transfer transition between iron and bridging inorganic sulfur, whereas absorption at 330 nm is a result of a charge transfer transition between the iron and cysteinyl sulfur of the polypeptide backbone.[49] The protein with the oxidized cluster is EPR silent consistent with the presence of a S=0 $[2Fe2S]^{2+}$ cluster. Anaerobic addition of dithionite results in cluster reduction and the resulting spectra measured at 45K (**Fig. 3B**) exhibits a rhombic pattern with g values of g_z 2.00, g_y 1.96, g_x 1.92, consistent with a S=1/2 $[2Fe2S]^{+}$ cluster. The shape and the g-values are similar to those of plant-like ferredoxins and identical g-values have been reported for the yeast homologue Dre2 under these conditions.[11,50]

The 600 MHz ^{1}H paramagnetic NMR spectrum (**Fig. 3C**) of the oxidized protein at 298K shows a broad peak around 34 ppm (CH_{β}-Cys protons) and a sharp signal at 10.5 ppm (CH_{α}-Cys protons). These signals are typical of proteins with 2Fe2S cluster.[51,52,53]

In conclusion, all these results indicate that anamorsin coordinates one 2Fe2S cluster per protein molecule.

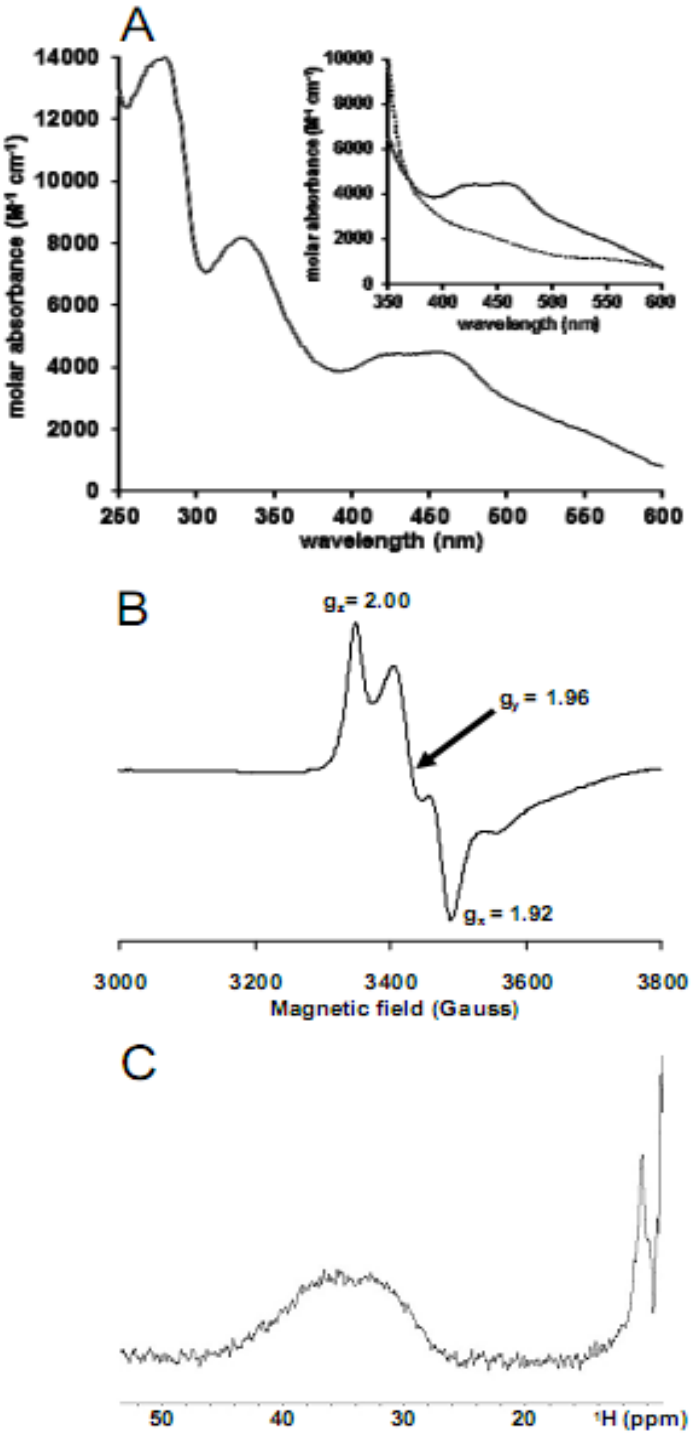


Figure 3. Anamorsin binds a 2Fe2S cluster. (A) UV-visible spectrum of 2Fe2S-anamorsin in 50mM phosphate buffer, pH 7.0. In the inset, UV-visible spectra of anamorsin without (solid line) or with 1mM sodium dithionite (dashed line) are shown. The peaks at around 330, 420 and 460 nm of the spectrum of the oxidized cluster disappear upon reduction with dithionite. (B) EPR spectrum of 2Fe2S-anamorsin (200 μM) in 50mM Tris-HCl pH 8.0, 500mM NaCl and 10% glycerol in the presence of 1mM dithionite. EPR parameters were the following: sample temperature 45 K, microwave frequency 9.45 GHz ; microwave power 5 mW; modulation frequency 9,387691 GHz; modulation amplitude 2500 G, time constant 167 ms. (C) Paramagnetic 1H NMR spectrum of 2Fe2S-anamorsin (2 mM) in 50mM phosphate buffer pH 7.0 at 600 MHz and 298K. (**see also Fig. S2**).

Anamorsin is a Mia40 substrate

Following the suggestion of our bioinformatic sequence analysis and the localization results on the yeast homologue, we predict that anamorsin can be a putative substrate of Mia40 in the intermembrane space of mitochondria.[11] Therefore, we have investigated the interaction of these two proteins in vitro.

The fully oxidized Mia40 containing three disulfide bonds ($Mia40_{3S\text{-}S}$) was incubated with anamorsin from which the Fe/S cluster was removed (apo-anamorsin) and which therefore has ten free cysteines (two in the N-terminal folded domain and eight in the CIAPIN1 domain). The analysis of the redox state of the two proteins after their incubation, performed by reaction with 4-acetamido-4′-maleimidylstilbene-2,2′-disulfonic acid (AMS) and followed by SDS-PAGE, showed that the band of Mia40 is shifted with respect to the band of the fully oxidized $Mia40_{3S\text{-}S}$ before reaction (**Fig. 4A**). This indicates that its cysteines were partially reduced upon incubation with anamorsin and therefore, they were able to bind two AMS molecules (0.5 kDa/thiol moiety). Anamorsin in the completely reduced state (before Mia40-incubation) binds ten AMS molecules resulting in a larger shift of the band (**Fig. 4A**). After incubation with $Mia40_{3S\text{-}S}$ this shift is, however, decreased indicating that fewer AMS molecules were bound to anamorsin, as a consequence of the oxidation of some of its cysteine residues.

This redox process between apo-anamorsin and Mia40 was then followed by NMR. Upon addition of increasing amounts of unlabelled reduced apo-anamorsin to ^{15}N labeled $Mia40_{3S\text{-}S}$, NH chemical shift changes were detected in the ^{1}H-^{15}N HSQC spectra for Mia40 residues of the catalytic CPC motif and its surroundings, indicating its reduction ($Mia40_{2S\text{-}S}$, **Fig. 4B**). The observed chemical shift variations are fully comparable with those observed upon oxidation of typical IMS Mia40-substrates with twin CX_9C motifs.[14] The reaction was complete at a Mia40:apo-anamorsin ratio of about 1:1 (**Fig. 4C**). The same behavior was observed when the NMR experiment was performed using anamorsin carrying the Fe/S cluster (2Fe2S-anamorsin hereafter). Moreover, the UV/vis and EPR spectra of 2Fe2S-anamorsin are identical after or before reaction with Mia40 (**Fig. S2**). All these data indicate that the coordination of the cluster is not affected by the formation of the Mia40-induced disulfide bond(s).

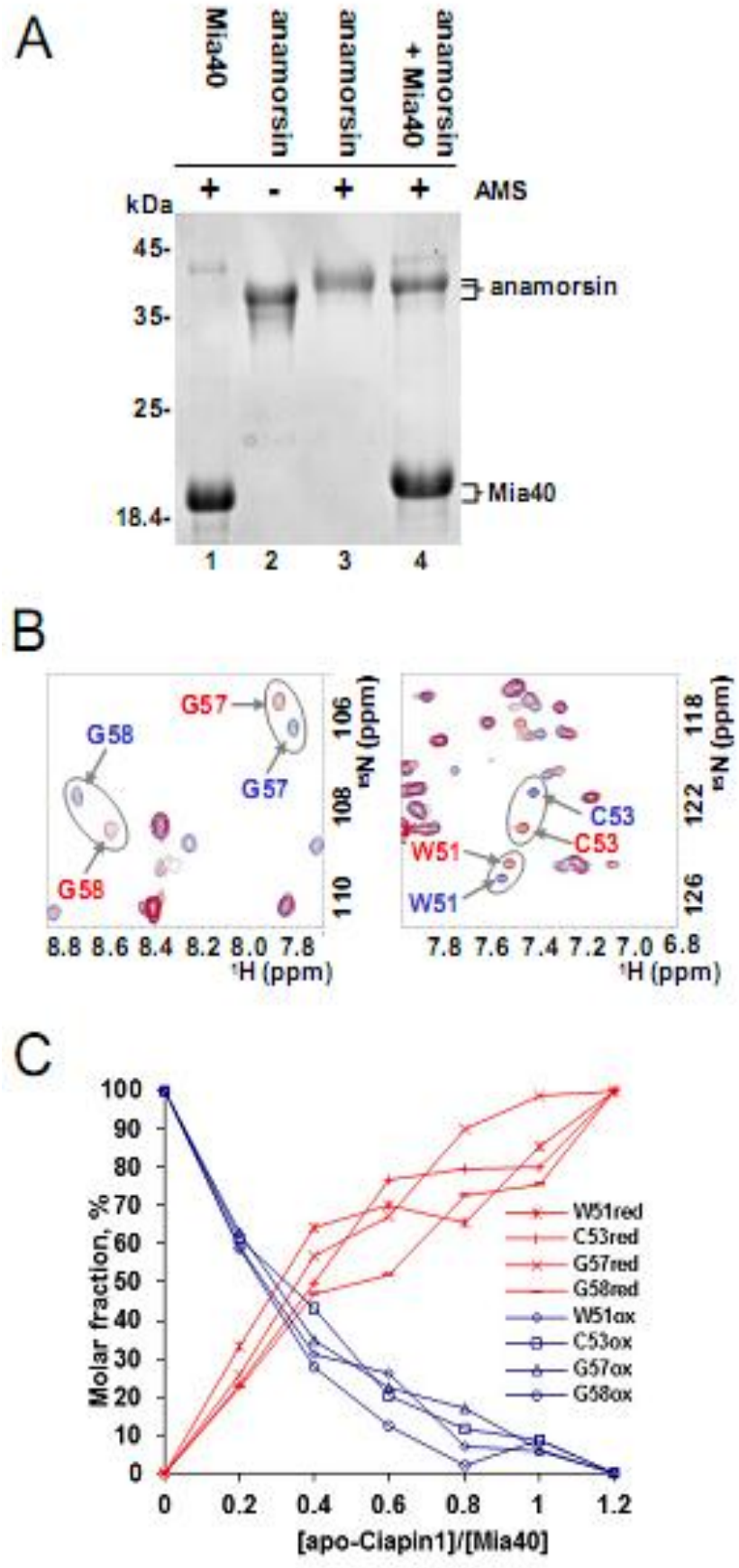


Figure 4. Interaction of anamorsin with Mia40. (A) 17% SDS PAGE showing the redox state of anamorsin and Mia40 alone or after their incubation in the presence (+) or absence (-) of the cysteine-alkylating agent AMS. After the incubation of the two proteins (lane 4) the band corresponding to Mia40 is shifted compared to the fully oxidized protein (lane 1) indicating a partial reduction of the protein. The shift of the anamorsin band in line 4 is not as strong as without incubation with Mia40 (lanes 2 and 3) due to a partial oxidation of anamorsin. (B) Overlay of two selected regions of the ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled Mia40 in the presence of 0 (blue, $Mia40_{3S-S}$) and 1.2 equivalents (red, $Mia40_{2S-S}$) of unlabeled apo-anamorsin. (C) The plot reports $Mia40_{3S-S}$/$Mia40_{2S-S}$ molar fraction as a function of the anmorsin/Mia40 molar ratio. The NH cross-peaks of residues Trp51, Cys53, Gly57 and Gly58, whose 1H- and 15N chemical shifts change substantially depending on the redox state of Mia40, have been selected to calculate the $Mia40_{3S-S}$/$Mia40_{2S-S}$ molar fraction. (see also **Fig. S3**).

When detecting ^{15}N labeled 2Fe2S-anamorsin reacting with unlabeled $Mia40_{3S\text{-}S}$, no significant chemical shift variations were observed in the ^{1}H-^{15}N HSQC maps for the residues of the folded N-terminal domain, indicating that it is not involved in the reaction (**Fig. S3**). On the contrary, chemical shift variations are observed for NH signals in the crowded, central spectral region, indicating a structural rearrangement in the unstructured C-terminal region of anamorsin upon disulfide formation (**Fig. S3**). Mia40-induced oxidation of anamorsin also produces a dramatically enhanced stability as shown by SDS-PAGE analysis (**Fig. S3**).

Carboxyamidomethylation with iodoacetamide (IAM) of the cysteine residues of 2Fe2S-anamorsin was performed before and after incubation with Mia40 in order to investigate how many and which are the cysteine residues involved in the formation of the disulfide bond(s). After SDS-PAGE protein separation and proteolytic trypsin or chymotrypsin digestion, the resulting peptide fragments were analyzed by MALDI-MS. For anamorsin not treated with Mia40, all the ten cysteines reacted with IAM (**Table S1**), while in the sample incubated with Mia40 the last four cysteine residues (C274, C277, C285 and C288) did not react with IAM while the remaining six cysteine residues were modified (**Table S1, Fig. 5**). This means that the last four cysteine residues are involved in two disulfide bonds and consequently the other four upstream cysteines are coordinating the Fe/S cluster (C237, C246, C249 and C251), as the remaining two IAM-modified cysteines C92 and C116, located within the well-folded N-terminal domain, are not involved in Fe/S cluster binding according to the not-observed interaction between the N-terminal and CIAPIN1 domain (see second section of Results).

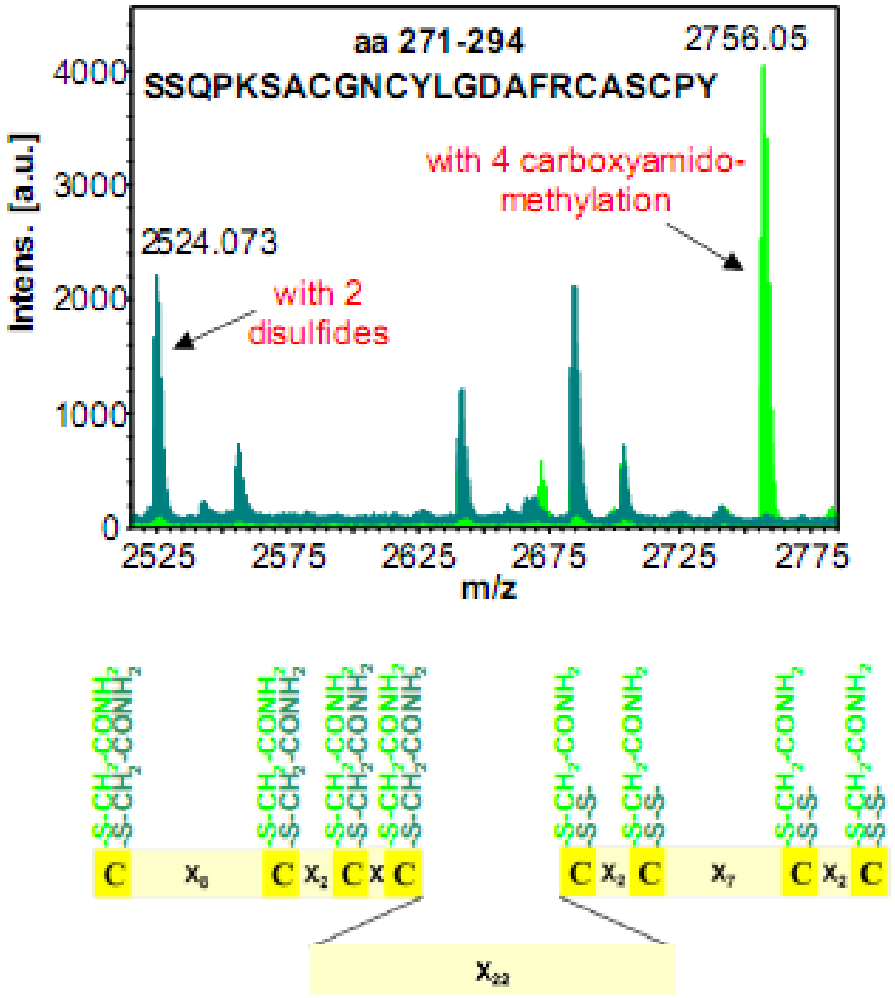


Figure 5. Identification of the cysteine residues involved in the Mia40-induced oxidation. Overlay of MALDI-MS spectra of anamorsin before (light green) and after (dark green) reaction with Mia40. The cysteine residues were carboxyamidomethylated with IAM before proteins were subjected to SDS-PAGE and in-gel digested by chymotrypsin. Schematic representation of the alkylation pattern of the cysteine residues of the CIAPIN1 domain before (light green) and after (dark green) reaction with Mia40. (**see also Table S1**).

To confirm this finding on the coordination mode of the 2Fe2S cluster, we produced a construct without the last four cysteine residues which are oxidized by Mia40. The UV/vis and EPR spectra of this construct showed the same features of the full-length anamorsin (**Fig. 6A and 6B**), in agreement with the participation of the CX_8CX_2CXC motif in 2Fe2S cluster binding as shown by the MALDI-MS analysis. In this construct, the presence of a 2Fe2S cluster was also confirmed through ESI-MS analysis. The molecular weight difference between 2Fe2S-anamorsin and apo-anamorsin is 175.78 which corresponds with the mass of 2Fe and 2S atoms (**Fig. 6C**).

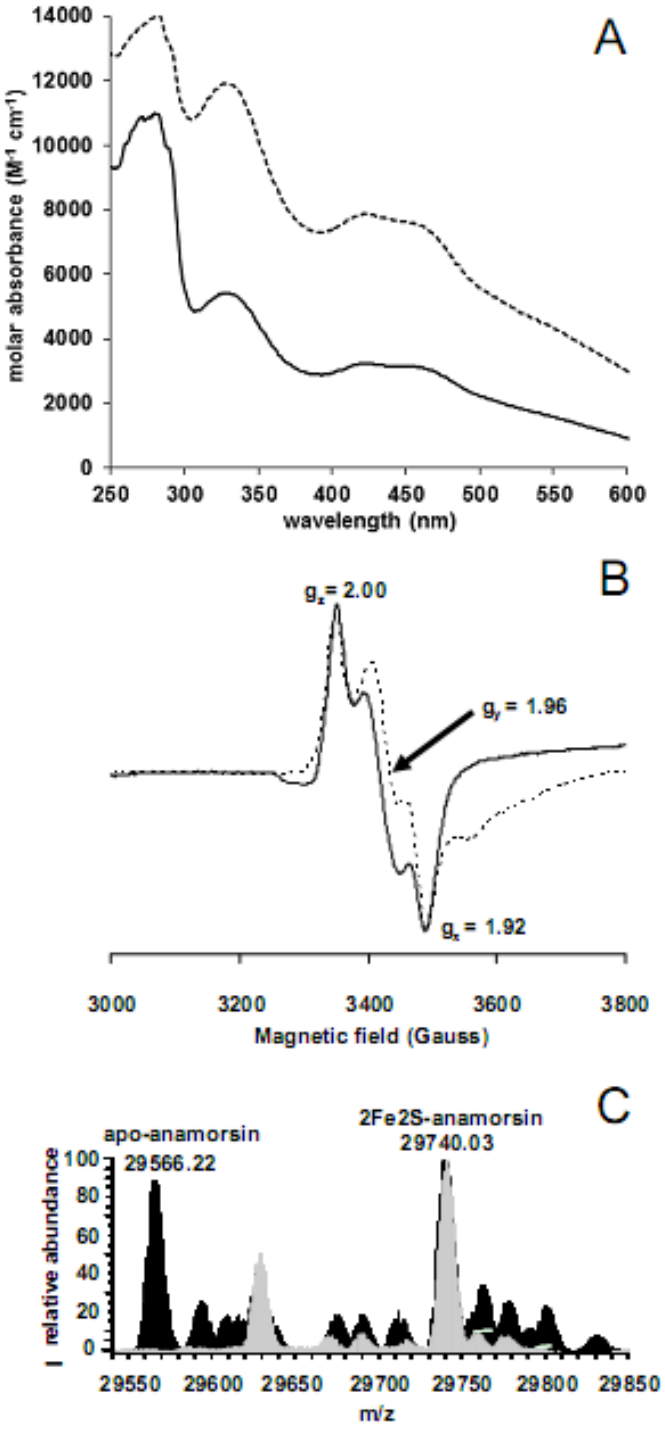


Figure 6. Spectral features of the Fe/S cluster of the full-length and C-terminally truncated anamorsin are similar. (A) UV-visible (50 mM Tris pH 8, 500 mM NaCl, 5% Glycerol) and (B) EPR spectra (200 µM, 50 mM Tris pH 8, 500 mM NaCl, 10% glycerol) of the C-terminally truncated (solid line) and full-length 2Fe2S-anamorsin (dashed line). EPR conditions where the same like in Fig. 3. (C) Overlay of the ESI-MS spectra of C-terminally truncated 2Fe2S-anamorsin (grey shaded) and after addition of 0.025% formic acid (black). The C-terminally truncated apo form (MW 29566.22) was formed by acidification. The relative abundance was scaled with respect to the peak corresponding to the MW of C-terminally truncated 2Fe2S-anamorsin.

Solution structure of the N-terminal domain of anamorsin

The construct comprising the well-folded N-terminal domain (aa 1-172) features essentially homogeneous heteronuclear relaxation properties, with a few exception. The correlation time for overall protein tumbling, as estimated from the R_2/R_1 ratio, is 10.82 ± 0.77 ns, as expected for a protein of this size in a monomeric state. Average ^{15}N R1, R2, and $^{15}N\{^1H\}$ NOEs values of 1.22 ± 0.13 s^{-1}, 14.70 ± 1.98 s^{-1} , and 0.76 ± 0.09 are obtained, respectively, excluding residue 1-2 and 161-172, which have low or negative $^{15}N\{^1H\}$ NOEs values and are characterized by fast motions on the ns-ps timescale. Also the region between residues 102-107 shows a certain degree of motions on the same timescale as shown by values lower than the mean R_2/R_1 ratio (**Fig. S4**).

The solution structure of the above construct, solved by NMR, shows five α-helical and six β-strand alternated structural elements (**Fig. 7**). The β-strands are all parallel, except for the last which is antiparallel, forming a sheet around which the α-helices are arranged (**Fig. 7**). This fold resembles the core fold of S-adenosyl-L-methionine (SAM)-dependent methyltransferases. The order of the β-strands in these proteins is generally 3214576 which is also found in the N-terminal domain of anamorsin with the exception that the third β-strand is absent. In addition, the fold of anamorsin diverges from SAM-dependent methyltransferases in the lack of the second α-helix and in the insertion of an α-helix between the two last β-strands. The glycine-rich motif in the loop between the first β-strand and first α-helix, which is typically responsible for the binding of the co-factor SAM in such methyltransferases is not conserved in the N-terminal domain of anamorsin (**Fig. 7**).[54,55,56] These structural differences suggest that the N-terminal domain of anamorsin is not involved in SAM-binding, as indeed experimentally verified through ^1H-^{15}N HSQC spectra by incubating the ^{15}N-labeled N-terminal domain with a large excess of SAM molecule, but it is probably involved in partner recognition.

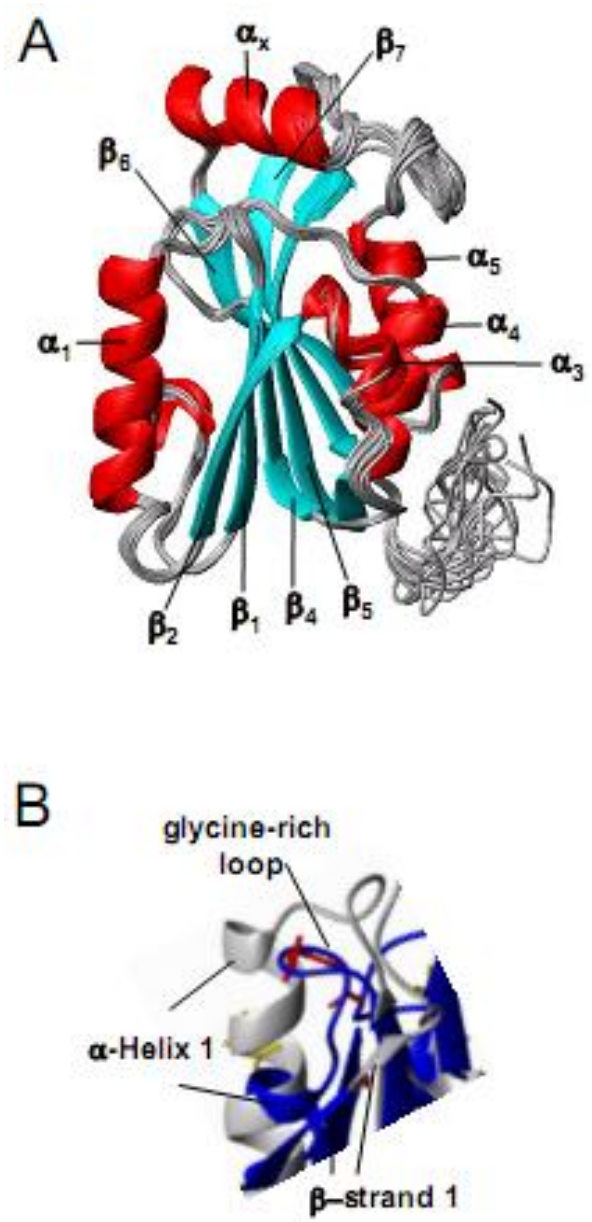


Figure 7. Solution structure of the N-terminal domain of anamorsin. (A) Ribbon representation of the 20 conformers with the lowest energy. α-helical structural elements are colored in red while β-strands are colored in cyan. The numbering of the secondary structure elements follows the arrangement in the typical SAM-dependent methyltransferase fold (α_x: alpha-helix insertion in the N-terminal domain of anamorsin). (B) Section of the overlay between the structures of the N-terminal domain of anamorsin (grey) and the structurally closest methyltransferase (blue, BH2331 of *Bacillus halodurans*). The typical glycine-rich loop involved in SAM binding in methyltransferases is shown (glycine residues of *B. halodurans'* methyltransferase in red). The corresponding loop of anamorsin does not have any analogous glycine residues, but only glycine residues in helix α1 and beta-strand β4 in the shown section (in yellow). (see also **Fig. S4** and **Table S2**).

Discussion

The human anamorsin protein is formed by a SAM-dependent methyltransferase-like structural domain but with no co-factor SAM-binding activity, and a long unstructured and flexible Cys-rich region. Bioinformatic sequence analysis shows that the latter motif is highly conserved and part of a large eukaryotic protein family (with members ranging from basal protozoa to fungi, higher plants and animals, up to humans). This Cys-rich motif of anamorsin is involved in the binding of a 2Fe2S cluster through the first four cysteines of the motif, i.e. CX_{5}-$_{14}CX_2CXC$, defining a novel cysteine-binding motif for Fe/S proteins, as shown by an extensive characterization performed through NMR, UV-vis, EPR and both ESI-MS and MALDI-MS analysis. This feature is different from the yeast homologue Dre2 that has been proposed to bind a 4Fe4S cluster in addition to a 2Fe2S cluster, based, however, only on EPR data, which, as also mentioned by the authors, need further detailed investigation to be fully interpreted.[11,57]

It is being increasingly recognized that the intermembrane space plays a pivotal role in the coordination of mitochondrial activities with other cellular processes.[57] Proteins of the IMS are particularly relevant for the coordinated exchange of components/signals into and out the mitochondria. One of these processes is represented by the connection of the Fe/S protein biogenesis machinery present in the mitochondrial matrix with the cytosolic one, but still not all of the involved proteins are known neither the exact pathways. Up to now, only one component, an ABC membrane transporter, located in the mitochondrial inner membrane, is found to be involved in exporting from the matrix a yet unidentified product of the ISC assembly pathway which is essential for the extra-mitochondrial Fe/S protein production.[58,59,60] The orientation of the ABC transporter in the mitochondrial inner membrane, with its C-terminal nucleotide-binding domain in the matrix, predicts that it functions in the export of substrate from the matrix to the intermembranous space.[61] This export reaction is further facilitated by the presence of the sulfhydryl oxidase Erv1/ALR in the IMS and glutathione.[9,10] On the basis of its protein IMS localization, the yeast homologue of anamorsin, Dre2, has been proposed to play a functional role in the assembly process of cytosolic Fe/S clusters once imported in the IMS, i.e. to generate a component or signal that is permissive for the subsequent assembly steps in the cytoplasm.[11] In addition, it has recently been suggested a working model which involves Dre2 as a cytosolic electron donor necessary for the Fe/S cluster incorporation into cytosolic CIA proteins, by receiving the electrons from the diflavin reductase Tah18.[12]

The cytosolic Dre2-Tah18 protein interaction has also been proposed to be implicated in controlling mitochondrial integrity and cell death after oxidative stress through a still unknown mechanism which, however, implies the mitochondrial relocalization of Tah18.[13] The cytosolic electron transfer pathway involving Dre2 and Tah18 seems also to be conserved in humans as the human homologues, anamorsin/CIAPIN1 and Ndor1, can functionally replace their yeast counterparts and interact with each other.[12]

Our data show that anamorsin interacts with a well-known mitochondrial oxidase, Mia40. In particular, Mia40 is able to specifically catalyze the formation of two disulfide bonds which involve the four cysteines of the twin CX_2C motif. The latter is indeed not involved in iron-binding and is located at the C-terminus subsequent to the Cys-motif involved in Fe/S cluster-binding. Typically, mitochondrial proteins are directed to the IMS by internal cysteine-containing targeting signals that allow passage through the import TOM translocase.[15,45] The cysteine residues in these signals (usually CX_3C, CX_9C motifs but also CX_2C motifs) are recognized by the IMS-localized oxidoreductase Mia40 which serves as an intramitochondrial import receptor and traps substrate proteins within this compartment by forming disulfides and inducing folding.[15] These IMS-targeted proteins are indeed unfolded before import and interaction with Mia40. Our interaction data with Mia40 suggests that also anamorsin could in this way be transported into and retained in the IMS. Consistent with this proposal we have also found that i) the CIAPIN1 domain is completely unfolded, similarly to the typical CX_3C and CX_9C Mia40-substrates, and therefore is structurally appropriate to be imported into the IMS through the TOM channel and be recognized by Mia40, ii) the Mia40-dependent disulfide bond formation does not interfere with the cluster formation, indicating that the insertion and the presence of the Fe/S cluster within the IMS could be possible in order to perform an electron transfer functional role. The latter feature is also found in the mitochondrial copper chaperone Cox17, which is involved in the assembly of the copper centers of cytochrome c oxidase.[62] Cox17, similarly to anamorsin, undergoes indeed formation of two disulfide bonds between a twin CX_9C motif by Mia40, while two additional cysteines, involved in copper binding and positioned upstream of the twin CX_9C motif, as found for the Cys motif involved in Fe/S cluster-binding, are not oxidized by Mia40, thus being capable to coordinate the copper(I) ion.[63] The fact that the cluster assembly in yeast Dre2 is independent of the cytosolic CIA components is also in agreement with a possible cluster maturation reaction in Dre2 being operative not only in the cytoplasm but also in other compartments such as the IMS.[12] Moreover, the fact that the full-length, IMS-localized

and well-known Mia40 protein partner Erv1 is specifically required for the maturation of Fe/S proteins outside mitochondria suggests that anamorsin, since being a Mia40-substrate, can be the target protein that is linking the Mia40/Erv1-dependent mitochondrial protein trapping pathway to the cytosolic iron/sulfur cluster maturation process.

All the above considerations suggest that anamorsin can play a functional role linked to cytosolic Fe/S cluster biogenesis not only in the cytoplasm as recently reported,[12] but also in the IMS once trapped there through the Mia40-Erv1 pathway. A similar double functionality in different cellular compartments has been found for other IMS proteins such as the copper chaperone of superoxide dismutase (CCS) which, similarly to anamorsin, has both cytosolic and IMS localization playing the same function in both compartments and being trapped into the IMS through the Mia40/Erv1 machinery.[41,64] Other functions can however be associated with anamorsin molecules in the IMS, in particular related to oxidative mitochondrial stress response. The yeast Dre2 cytosolic protein partner Tah18 is indeed relocalized to mitochondria upon oxidative stress to induce cell apoptosis and anamorsin has been found to be linked to apoptosis-related processes.[13,16] Therefore, the IMS-localized anamorsin could be a component of the cell death program.

References

1. Horowitz, M.P. & Greenamyre, J.T. Mitochondrial iron metabolism and its role in neurodegeneration. *J. Alzheimers Dis* **20 Suppl 2**, S551-568 (2010).
2. Andreini, C., Bertini, I., Cavallaro, G., Najmanovich, R.J. & Thornton, J.M. Structural Analysis of Metal Sites in Proteins: Non-heme Iron Sites as a Case Study. *Journal of Molecular Biology* **388**, 356-380 (2009).
3. Bertini, I., Sigel, A. & Sigel, H. *Handbook on metalloproteins*. (Marcel Dekker: 2001).
4. Lill, R. Function and biogenesis of iron-sulphur proteins. *Nature* **460**, 831-838 (2009).
5. Rouault, T.A. & Tong, W.H. Iron-sulfur cluster biogenesis and human disease. *Trends Genet* **24**, 398-407 (2008).
6. Sheftel, A., Stehling, O. & Lill, R. Iron-sulfur proteins in health and disease. *Trends Endocrinol. Metab* **21**, 302-314 (2010).
7. Lill, R. & Mühlenhoff, U. Maturation of iron-sulfur proteins in eukaryotes: mechanisms, connected processes, and diseases. *Annu. Rev. Biochem* **77**, 669-700 (2008).
8. Mühlenhoff, U. & Lill, R. Biogenesis of iron-sulfur proteins in eukaryotes: a novel task of mitochondria that is inherited from bacteria. *Biochim. Biophys. Acta* **1459**, 370-382 (2000).
9. Lange, H. et al. An essential function of the mitochondrial sulfhydryl oxidase Erv1p/ALR in the maturation of cytosolic Fe/S proteins. *EMBO Rep* **2**, 715-720 (2001).
10. Lill, R. & Mühlenhoff, U. Iron-Sulfur Protein Biogenesis in Eukaryotes: Components and Mechanisms. *Annu. Rev. Cell Dev. Biol.* **22**, 457-486 (2006).
11. Zhang, Y. et al. Dre2, a conserved eukaryotic Fe/S cluster protein, functions in cytosolic Fe/S protein biogenesis. *Mol. Cell. Biol* **28**, 5569-5582 (2008).
12. Netz, D.J.A. et al. Tah18 transfers electrons to Dre2 in cytosolic iron-sulfur protein biogenesis. *Nat Chem Biol* **6**, 758-765 (2010).
13. Vernis, L. et al. A Newly Identified Essential Complex, Dre2-Tah18, Controls Mitochondria Integrity and Cell Death after Oxidative Stress in Yeast. *PLoS ONE* **4**,
14. Banci, L. et al. MIA40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria. *Nat. Struct. Mol. Biol* **16**, 198-206 (2009).
15. Riemer, J., Fischer, M. & Herrmann, J.M. Oxidation-driven protein import into mitochondria: Insights and blind spots. *Biochim Biophys Acta* (2010).doi:10.1016/j.bbamem.2010.06.003
16. Shibayama, H. et al. Identification of a cytokine-induced antiapoptotic molecule anamorsin essential for definitive hematopoiesis. *J. Exp. Med* **199**, 581-592 (2004).
17. Hao, Z. et al. Distribution of CIAPIN1 in Normal Fetal and Adult Human Tissues. *J. Histochem. Cytochem.* **54**, 417-426 (2006).
18. Li, X., Wu, K. & Fan, D. CIAPIN1 as a therapeutic target in cancer. *Expert Opin. Ther. Targets* **14**, 603-610 (2010).
19. Finn, R.D. et al. The Pfam protein families database. *Nucleic Acids Research* **38**, D211-D222 (2009).
20. Larkin, M. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947 -2948 (2007).
21. Prilusky, J. et al. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435-3438 (2005).
22. Holm, L., Kääriäinen, S., Rosenström, P. & Schenkel, A. Searching protein structure databases with DaliLite v.3. *Bioinformatics* **24**, 2780 -2781 (2008).
23. Kennedy, M.C. & Beinert, H. The state of cluster SH and S2- of aconitase during cluster interconversions and removal. A convenient preparation of apoenzyme. *J. Biol.*

Chem **263**, 8194-8198 (1988).
24. Siegel, L.M. A direct microdetermination for sulfide. *Analytical Biochemistry* **11**, 126-132 (1965).
25. Inubushi, T. Efficient detection of paramagnetically shifted NMR resonances by optimizing the WEFT pulse sequence. *J Magn Reson* **51**, 128-133 (1983).
26. Farrow, N.A. et al. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984-6003 (1994).
27. Grzesiek, S. & Bax, A. The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. *Journal of the American Chemical Society* **115**, 12593-12594 (1993).
28. Lee, L.K., Rance, M., Chazin, W.J. & Palmer, A.G. Rotational diffusion anisotropy of proteins from simultaneous analysis of 15N and 13C alpha nuclear spin relaxation. *J. Biomol. NMR* **9**, 287-298 (1997).
29. Keller, R. The Computer Aided Resonance Assignement Tutorial. *Cantina Verlag: Goldbau* **ISBN**,
30. Güntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol* **278**, 353-378 (2004).
31. Herrmann, T., Güntert, P. & Wüthrich, K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR* **24**, 171-189 (2002).
32. Herrmann, T., Güntert, P. & Wüthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol* **319**, 209-227 (2002).
33. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213-223 (2009).
34. Grishaev, A. & Bax, A. An empirical backbone-backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation. *J. Am. Chem. Soc* **126**, 7281-7292 (2004).
35. Case, D. et al. AMBER 10. *Universityof California* **San Francisco**,
36. Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R. & Thornton, J.M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477-486 (1996).
37. Vriend, G. WHAT IF: a molecular modeling and drug design program. *J Mol Graph* **8**, 52-56, 29 (1990).
38. Bhattacharya, A., Tejero, R. & Montelione, G.T. Evaluating protein structures determined by structural genomics consortia. *Proteins* **66**, 778-795 (2007).
39. Schuster-Bockler, B., Schultz, J. & Rahmann, S. HMM Logos for visualization of protein families. *BMC Bioinformatics* **5**, 7 (2004).
40. Hao, Z. et al. Subcellular localization of CIAPIN1. *J. Histochem. Cytochem* **54**, 1437-1444 (2006).
41. Kawamata, H. & Manfredi, G. Import, maturation, and function of SOD1 and its copper chaperone CCS in the mitochondrial intermembrane space. *Antioxid. Redox Signal* **13**, 1375-1384 (2010).
42. Sideris, D.P. & Tokatlidis, K. Trapping oxidative folding intermediates during translocation to the intermembrane space of mitochondria: in vivo and in vitro studies. *Methods Mol. Biol* **619**, 411-423 (2010).
43. Terziyska, N. et al. The sulfhydryl oxidase Erv1 is a substrate of the Mia40-dependent protein translocation pathway. *FEBS Lett* **581**, 1098-1102 (2007).
44. Milenkovic, D. et al. Identification of the Signal Directing Tim9 and Tim10 into the

Intermembrane Space of Mitochondria. *Mol. Biol. Cell* **20**, 2530-2539 (2009).
45. Sideris, D.P. et al. A novel intermembrane space-targeting signal docks cysteines onto Mia40 during mitochondrial oxidative folding. *J. Cell Biol* **187**, 1007-1022 (2009).
46. Andreini, C., Banci, L., Bertini, I., Elmi, S. & Rosato, A. Non-heme iron through the three domains of life. *Proteins* **67**, 317-324 (2007).
47. Bertini, I., Luchinat, C., Provenzani, A., Rosato, A. & Vasos, P.R. Browsing gene banks for Fe2S2 ferredoxins and structural modeling of 88 plant-type sequences: an analysis of fold and function. *Proteins* **46**, 110-127 (2002).
48. Hosein, B., Friesner, R. & Holzwarth, G. Ferredoxin circular dichroism at 3600 c-1. *Biochim. Biophys. Acta* **368**, 18-21 (1974).
49. Morimoto, K., Nishio, K. & Nakai, M. Identification of a novel prokaryotic HEAT-repeats-containing protein which interacts with a cyanobacterial IscA homolog. *FEBS Letters* **519**, 123-127 (2002).
50. Palmer, G. & Sands, R.H. On the Magnetic Resonance of Spinach Ferredoxin. *Journal of Biological Chemistry* **241**, 253 (1966).
51. Banci, L. The 1H NMR parameters of magnetically coupled dimers—The Fe2S2 proteins as an example. *Structure and Bonding* **72**, 113-1135
52. Bertini, I., Turano, P. & Vila, A.J. Nuclear magnetic resonance of paramagnetic metalloproteins. *Chemical Reviews* **93**, 2833-2932 (1993).
53. Nagayama, K., Ozaki, Y., Kyogoku, Y., Hase, T. & Matsubara, H. Classification of iron-sulfur cores in ferredoxins by 1H nuclear magnetic resonance spectroscopy. *J. Biochem* **94**, 893-902 (1983).
54. Kagan, R.M. & Clarke, S. Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. *Arch. Biochem. Biophys* **310**, 417-427 (1994).
55. Martin, J.L. & McMillan, F.M. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol* **12**, 783-793 (2002).
56. Schluckebier, G., O'Gara, M., Saenger, W. & Cheng, X. Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol* **247**, 16-20 (1995).
57. Herrmann, J.M. & Riemer, J. The intermembrane space of mitochondria. *Antioxid. Redox Signal* **13**, 1341-1358 (2010).
58. Cavadini, P. et al. RNA silencing of the mitochondrial ABCB7 transporter in HeLa cells causes an iron-deficient phenotype with mitochondrial iron overload. *Blood* **109**, 3552-3559 (2007).
59. Kispal, G., Csere, P., Prohl, C. & Lill, R. The mitochondrial proteins Atm1p and Nfs1p are essential for biogenesis of cytosolic Fe/S proteins. *EMBO J* **18**, 3981-3989 (1999).
60. Kuhnke, G., Neumann, K., Mühlenhoff, U. & Lill, R. Stimulation of the ATPase activity of the yeast mitochondrial ABC transporter Atm1p by thiol compounds. *Mol. Membr. Biol* **23**, 173-184 (2006).
61. Chloupková, M., LeBard, L.S. & Koeller, D.M. MDL1 is a high copy suppressor of ATM1: evidence for a role in resistance to oxidative stress. *J. Mol. Biol* **331**, 155-165 (2003).
62. Nigel J. Robinson & Dennis R. Winge Copper Metallochaperones. (2010).a <http://www.annualreviews.org/doi/abs/10.1146/annurev-biochem-030409-143539>
63. Banci, L. et al. A Structural-Dynamical Characterization of Human Cox17. *Journal of Biological Chemistry* **283**, 7912 -7920 (2008).
64. Culotta, V.C., Yang, M. & O'Halloran, T.V. Activation of superoxide dismutases: Putting the metal to the pedal. *Biochim Biophys Acta* **1763**, 747-758 (2006).

Suplementary materials

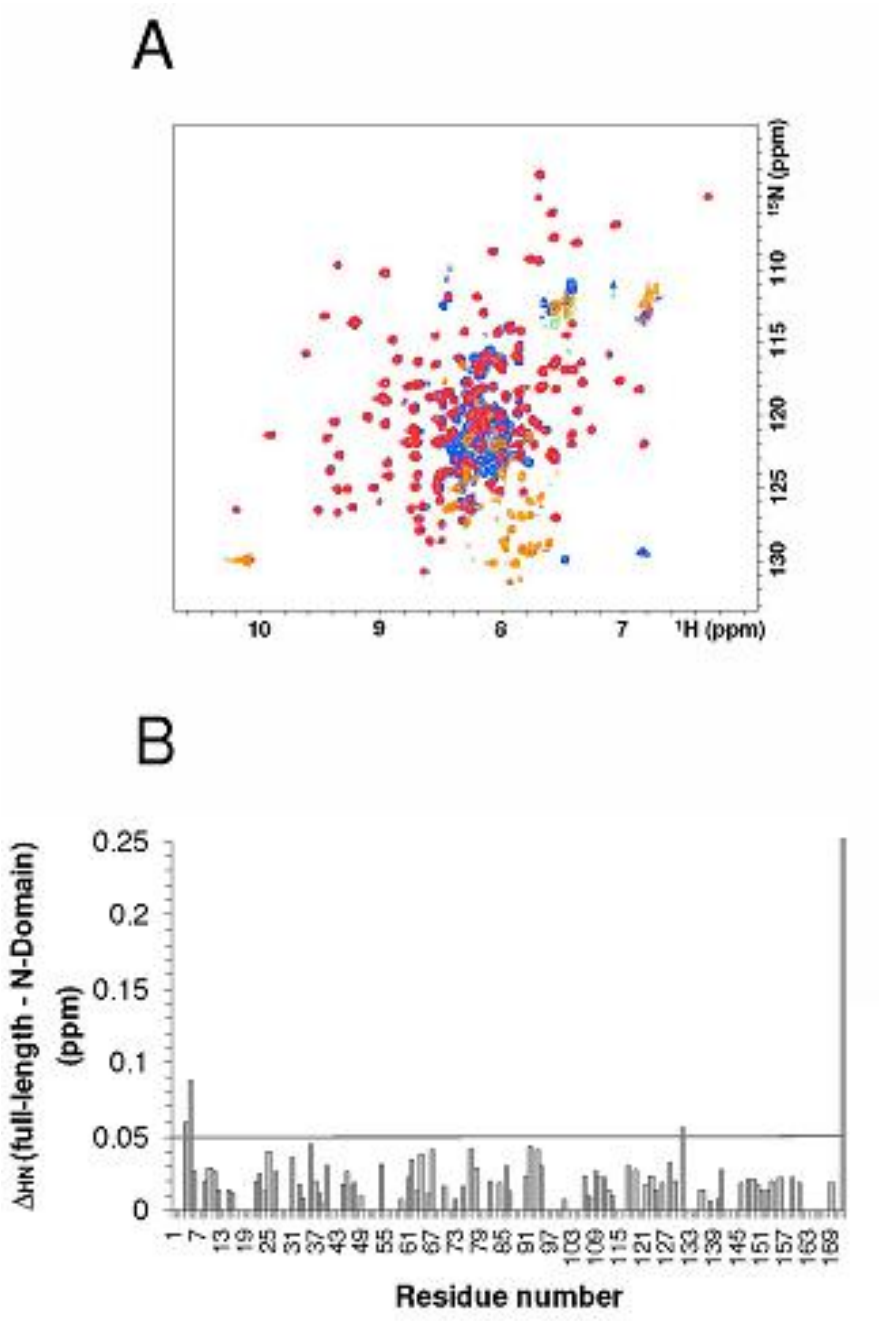


Fig. S1. Related to Figure 2. The C-terminal part of anamorsin is flexible and does not interact with the N-terminal domain. (A) Overlay of $^{15}N\{^{1}H\}$ NOE experiments recorded with (positive peaks in red, negative in orange) and without saturation (peaks are in blue). (B) The weighted-average chemical shift differences Δ_{HN} between full-length anamorsin and the N-terminal domain plotted against the amino acid residue number. Δ_{HN} is $([(\Delta H)^2 + (\Delta N/5)^2]/2)^{1/2}$, where ΔH and ΔN are chemical shift differences for ^{1}H and ^{15}N, respectively.

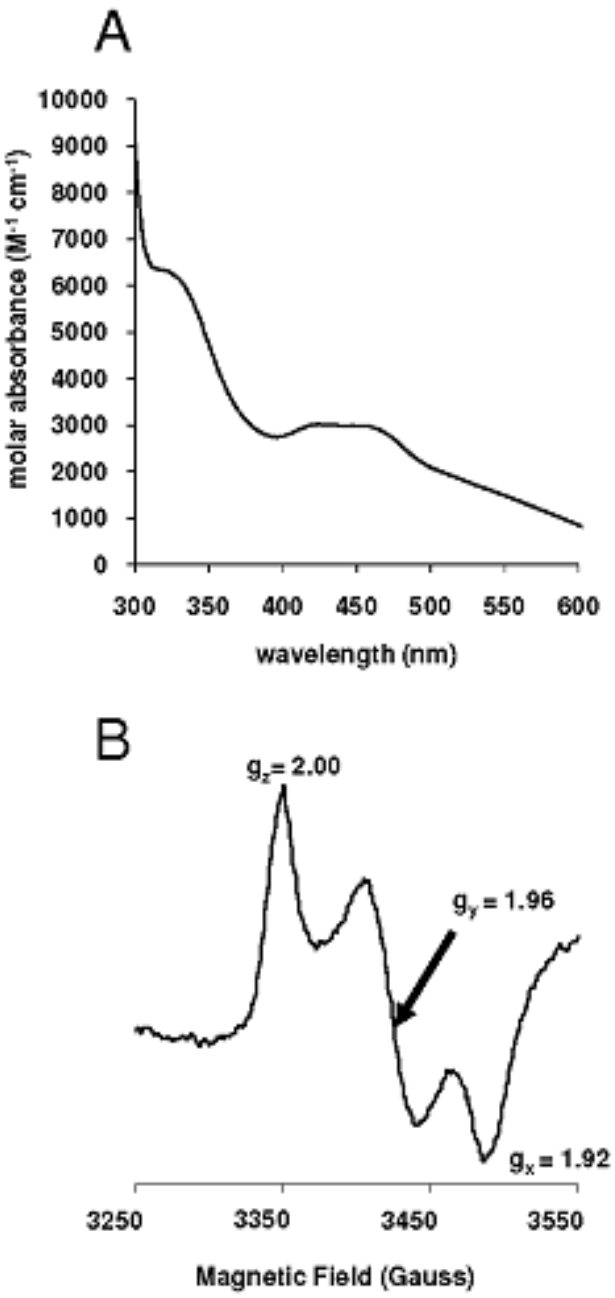


Fig. S2. Related to Figure 3. Spectral features of anamorsin after reaction with Mia40

UV/vis (A) and EPR (B) spectra of 2Fe2S-anamorsin after reaction with Mia40. EPR conditions were the same like in **Fig. 3** except that 10 mM dithionite was used for reduction of the cluster.

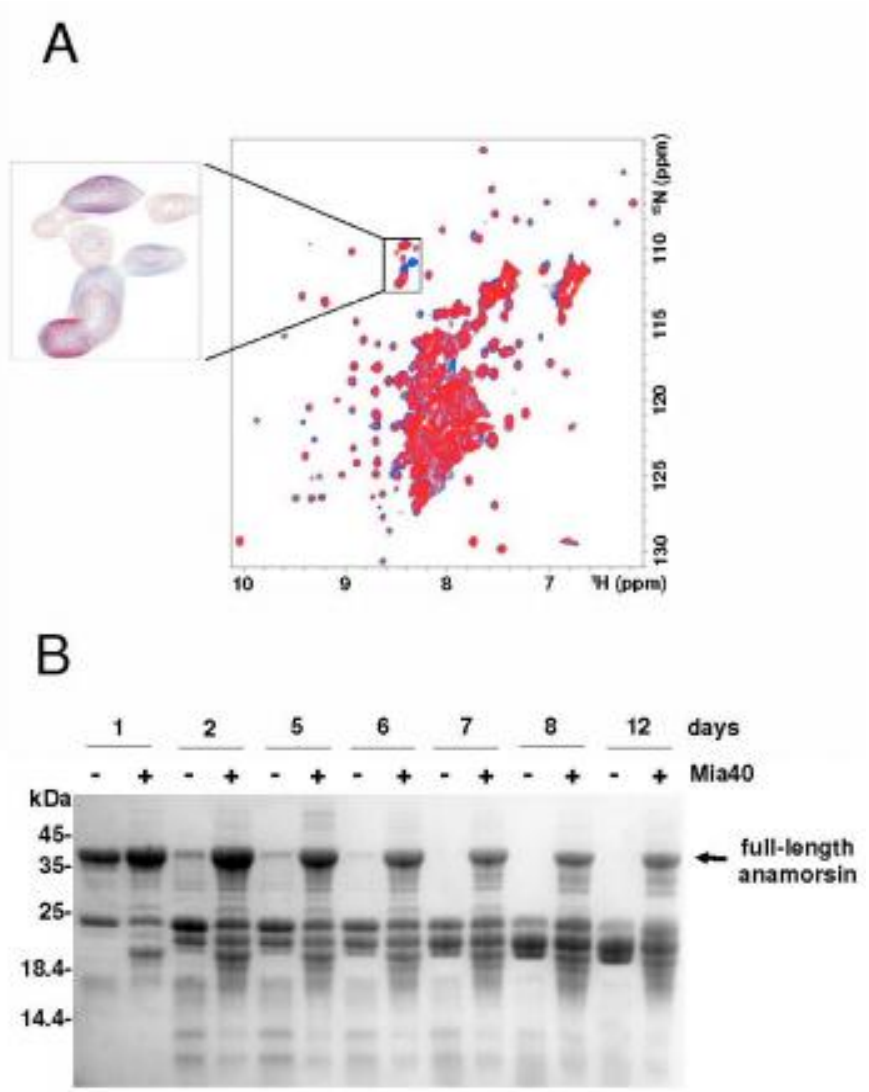


Fig. S3. Related to Figure 4. Backbone chemical shift variations and stability of 2Fe2S-anamorsin upon Mia40 addition

(A) Overlay of 1H-15N HSQC spectra (800MHz, 298K) of 15N-labeled 2Fe2S-anamorsin in the presence of 0 (blue) and 1.2 equivalents (red) of unlabeled Mia403S-S. A selected region of the spectra between 8 and 8.5 ppm reporting chemical shift changes of Gly residues is shown in detail. (B) 17% SDS PAGE showing the stability of anamorsin before (-) or after (+) reaction with Mia40 over 12 days.

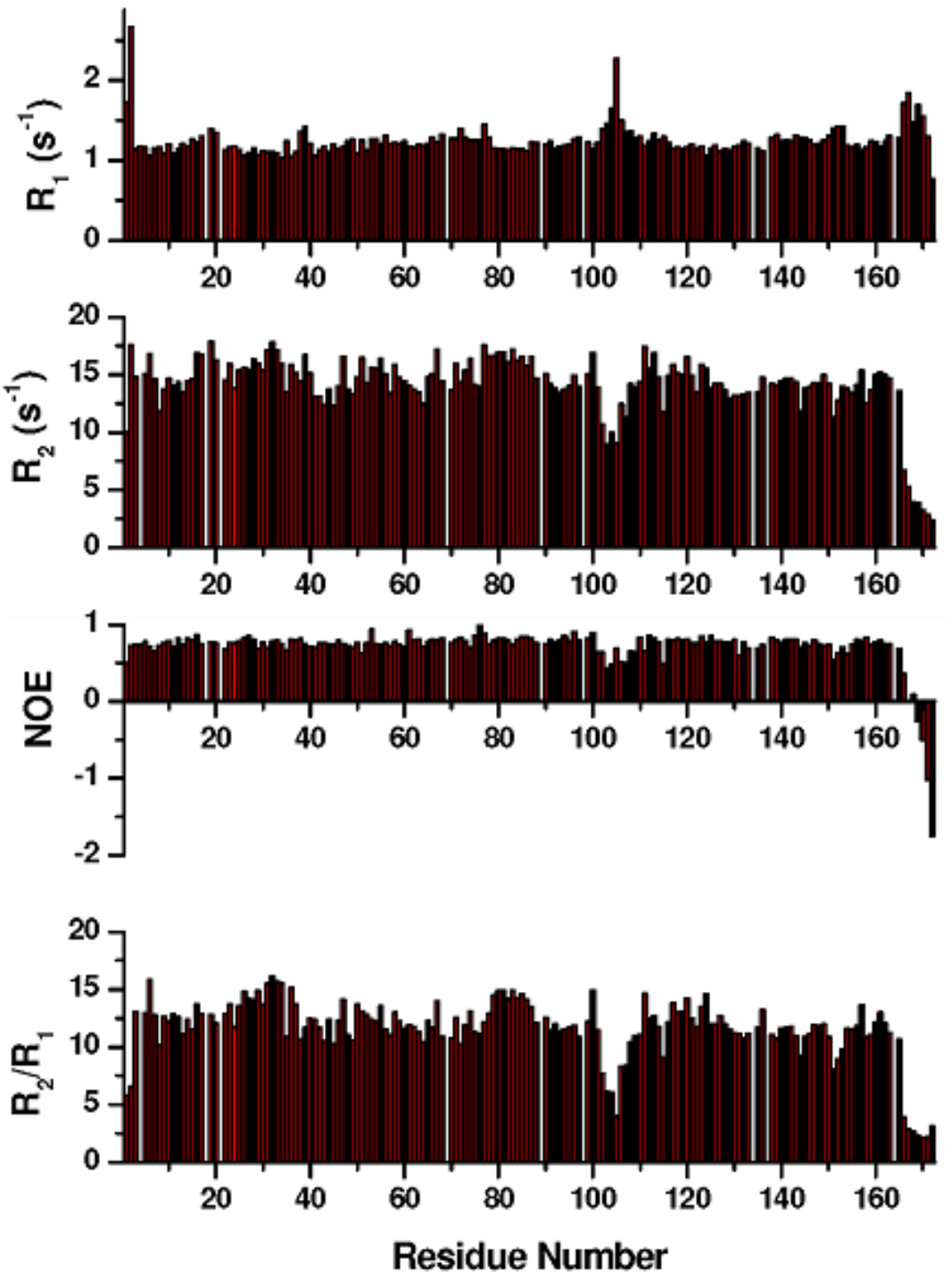


Fig. S4. Related to Figure 7. ^{15}N relaxation parameters R_1, R_2, ^{15}N{^{1}H} NOE and ^{15}N R_2/R_1 ratio versus residue number of anamorsin at 600 MHz and 298K.

The protein was in 50 mM phosphate buffer pH 7.0, 2 mM DTT containing 10% (v/v) D_2O.

Table S1. Related to Figure 5. Correlation between the masses and the corresponding peptides of anamorsin before and after reaction with Mia40

Mia40 reaction	m/z	corresponding amino acid sequence	Carboamido-methylations	Digesting enzyme
-	2557.446	90-112 ILRPGGCLFLKEPVETAVDNNS	1X	Trypsin
+	2557.336	90-112 ILRPGGCLFLKEPVETAVDNNS	1X	Trypsin
-	2016.191	119-136 LCSALTLSGLVEVKELQR	1X	Trypsin
+	2016.081	119-136 LCSALTLSGLVEVKELQR	1X	Trypsin
-	1063.395	238-247 AASCGEGKKRK	1X	Trypsin
+	1063.450	238-247 AASCGEGKKRK	1X	Trypsin
-	1910.075	248-263 RKACKNCTCGLAEELE	3X	Trypsin
+	1909.982	248-263 RKACKNCTCGLAEELE	3X	Trypsin
-	2756.05	271-294 SSQPKSACGNCYLGDAFRCASCPY	4 X	Chymotrypsin
+	2524.073	271-294 SSQPKSACGNCYLGDAFRCASCPY	-	Chymotrypsin
-	1416.558	283-294 LGDAFRCASCPY	2X	Chymotrypsin
+	1300.595	283-294 LGDAFRCASCPY	-	Chymotrypsin
-	1490.683 1912.946	276-288 SACGNCYLGDAFR 289-305 CASCPYLGMPAFKPGEK	2X 2X	Trypsin
+	3170.425	276-288 SACGNCYLGDAFR 289-305 CASCPYLGMPAFKPGEK Fragments linked by 2 disulphide bridges	- -	Trypsin

Table S2. Related to Figure 7. NMR and refinement statistics for the N-terminal domain of anamorsin

NMR distance and dihedral constraints	
Distance constraints	
Total NOE	3543
Intra-residue	771
Inter-residue	2772
Sequential ($\|i-j\|=1$)	945
Medium-range ($\|i-j\|<4$)	678
Long-range ($\|i-j\|>5$)	1149
Hydrogen bonds	63
Total dihedral angle restraints	284
ϕ	139
ψ	145
Structure statistics	
Violations (mean and s.d.)	
Distance constraints (Å)	0.0091 ± 0.0006
Dihedral angle constraints (°)	0.3949 ± 0.0404
Max. dihedral angle violation (°)	2.82 ± 0.36
Max. distance constraint violation (Å)	0.27 ± 0.12
Deviations from idealized geometry	
Bond lengths (Å)	0.0564 ± 0.0005
Bond angles (°)	1.962 ± 0.019
Impropers (°)	7.850 ±0.434
Average pairwise r.m.s. deviation* (Å)	
Heavy	0.69 ± 0.06
Backbone	0.38 ± 0.05

*Pairwise r.m.s. deviation was calculated among 30 refined structures.

3.2. Characterization of glutaredoxin5, a protein involved in Fe-S cluster biogenesis in mitochondria

Lucia Banci, Ivano Bertini, Simone Ciofi-Baffoni, Rebecca Del Conte, Maciej Mikolajczyk

I.B. and L.B. planned the research. M.M. coordinated and performed protein production and characterization; S.C.-B. and R.D.C. guided the flow of experiments, planned and recorded NMR spectra.

Introduction

Glutaredoxins (GRX) are redox proteins widely present in both prokaryotes and eukaryotes which share the thioredoxin fold and use glutathione as a cofactor. These proteins are divided into two classes; monothiol (C-G-F-S), and dithiol (C-P-Y-C) on the basis of the presence of one or two cysteines in the highly conserved active sites. Whereas dithiol GRX have been characterized as reducers of disulfides (protein-S-S) or glutathione mixed disulfides (protein-S-SG) thereby serving important roles in redox homeostasis and signal transduction, the mechanism and function of monothiol GRX remain puzzling.

In each class, several crystal structures of GRX have been shown to form [2Fe-2S] cluster coordinating homodimers or tetramers (dimmer of dimers). The dithiol GRX homodimer is proposed to serve as a sequestration form and its iron-sulfur cluster as an oxidative stress sensor. In contrast, the monothiol GRX homodimer has been suggested to act as a scaffold for [2Fe-2S] cluster delivery.[1] The discovery that a wide variety of recombinant monothiol GRXs can be purified containing [2Fe-2S] cluster under anaerobic conditions and that the cluster can be reconstituted on apo-forms in the presence of glutathione, has raised the possibility that they may function in the assembly, storage, or transfer of [2Fe-2S] clusters, or as sensors of the cellular Fe-S cluster status.[2] Monothiol glutaredoxins coordinating the Fe-S cluster, bind non covalently the GSH molecule, which is required together with the active site cysteine in the CGFS motif for cluster assembly. The cluster is coordinated at the subunit interface by one cysteine from each GRX and two glutathiones.

A series of recent papers indicate that the recombinant versions of plant, yeast and human GRX5 orthologues (belonging to the monothiol subfamily) overexpressed in *E. coli* are able to incorporate labile Fe–S clusters.[1,3] Yeast monothiol GRX5 has been shown to be a member of mitochondrial iron sulfur cluster assembly machinery (ISC). The protein plays a role in cluster delivery to the apo-recipients.[4,5] Furthermore, complementation experiments of the yeast GRX5 deletion strain demonstrated that most CGFS monothiol GRXs from prokaryotic or eukaryotic sources other than yeast, when targeted to the yeast mitochondrial matrix, can functionally substitute for GRX5, suggesting that this role might be conserved throughout evolution.[6,7] It has been demonstrated that GRX5 interacts *in vivo* with Isa1 and Isa2 proteins (both mitochondrial proteins involved in Fe-S cluster delivery to the apo-proteins) by observing bimolecular fluorescence complementation. These results suggest that GRX5 plays a central role in

Fe-S assembly process through the interaction with previously mentioned A-type Fe-S scaffold proteins (Isa1 and Isa2), each of which is necessary for the proper function of mitochondrial ISC assembly machinery.[8]

Human glutaredoxin5 (GLRX5) is a monothiol glutaredoxin harbouring the conserved CGFS active site motif. Besides the cysteine in the CGFS motif, human GLRX5 possesses another cysteine residue at the C-terminal part of the protein. GLRX5 like its yeast homologue is localised in the mitochondria where it is involved in the synthesis and/or transfer of iron-sulfur clusters to apo-proteins.[9] It was shown that this member of the mitochondrial ISC assembly machinery functions downstream of ISCU. After a transient Fe-S cluster has been assembled on the ISCU scaffold, it must be transferred and incorporated into apo-proteins by coordinating with specific amino acid ligands. This process in human mitochondria is specifically assisted by the group of chaperon proteins including e.g. GLRX5, ISCA1/2 (homologues of yeast Isa1/2) and recently reported Ind1 (responsible for cluster delivery to mitochondrial respiratory complex1).[10,11] However, the exact molecular function of GLRX5 is still unclear. Deletion experiments in model organism *S. cerevisiae* showed that, mutant lacking GRX5 is highly sensitive towards oxidative stress, accumulates iron in the cells and causes inactivation of Fe-S cluster containing enzymes. A knock out of the zebrafish *Shiraz* GRX5 homologue causes hypochromic anemia due to impaired Fe-S assembly and as a result defects the haem biosynthesis. In humans a mutation that causes a splicing defect in GLRX5 is implicated as a cause of sideroblastic anemia.[12] This example shows that Fe-S cluster biogenesis is directly connected with the overall cellular iron homeostasis.

Recombinantly expressed and purified human GLRX5 coordinates a [2Fe-2S] cluster similar to other monothiol glutaredoxin proteins.[13,14,15,16] The crystal structure of a tetrameric form of human GLRX5 has been solved showing that iron-sulfur clusters is coordinated by two glutathione molecules and one cysteine of the active site from each subunit of the tetramer (2WEM). In this work we have structurally characterized human Glutaredoxin 5 in solution and its iron-sulfur binding properties were investigated through UV-vis and paramagnetic NMR spectroscopies. The data indicate that human GLRX5 binds a [2Fe-2S] cluster in GSH dependent mode in a monomeric state, in contrast with what was found in the crystal structure. This suggests that the C-terminally located cysteine which is conserved in GLRX5 homologues is an additional Fe-S cluster ligand.

Materials and methods

Bioinformatic tools

The mitochondrial targeting signal located in the N-terminal part of GLRX5 sequence was identified by the TargetP and SignalP online tools.[17] The folding probability was predicted by the algorithm FoldIndex.[18] Homologue sequences to GLRX5 were searched via Blastp (www.ncbi.nlm.nih.gov/BLAST/), and aligned using ClustalW (www.ebi.ac.uk/clustalw/).

Molecular cloning expression and purification of glutaredoxin5

The DNA encoding mitochondrial version of GLRX5 was amplified by PCR reaction using specific primers: Sense: CAC CAT GGG CGC GGG CGG CGG CGG C; Anti sense: CTA CTA TTC ATC TAA AAG GGC GGA GTG GAT CCC C, and subsequently cloned into pETG20A vector using the Gateway technology (Invitrogen), in order to obtain a plasmid producing the protein fused with N-terminal 6xHis-Trx tag. The vector was transformed in BL21 GOLD *E.coli* strain (Novagen) and cells were grown in LB media at 37° C under vigorous shaking overnight, diluted 1:1000 into LB or minimal medium (with $^{15}(NH4)_2(SO)_4$ as the nitrogen source) containing 100ug/ml ampicillin until OD at 600nm reached 0.6. Then, protein expression was induced by adding 1 mM IPTG.

In order to produce a recombinant protein containing Fe-S cluster cells were grown in presence of 250 μM $FeCl_3$. Expression was performed at 25° C for 14-16 hours. Cells were harvested by centrifugation at 6000 x g and resuspended in degassed lysis buffer (50 mM Tris-HCl pH 8.0 containing 500 mM NaCl, 10 mM Imidazole and 5 mM GSH). Cells disruption was performed on ice by sonication alternating 30 seconds of sonication and 5 minutes of resting for 10 times. All the following purification steps were performed under anaerobic conditions. The soluble extract, obtained by ultracentrifugation at 40000 x g, was loaded on a HiTrap chelating HP column (GE Healthcare) previously charged with Ni^{2+} (for selective His-tag binding) and equilibrated with lysis buffer. After washing the column with different step corresponding to different, intermediate imidazole concentrations (W1 10 mM and W2 40 mM, **Fig. 3**), recombinant fused protein was eluted with 50 mM Tris-HCl pH 8.0 containing 500 mM NaCl, 5mM GSH and 500 mM imidazole (E, **Fig. 3**). The protein was then concentrated with an Amicon Ultra-15 Centrifugal Filter Units with a MwC of 3kDa (Millipore). Cleavage of the tag was

performed by TEV protease (3ul TEV/1 mg GLRX5) in 1X TEV Buffer (50 mM Tris-HCl, pH 8.0, 0.5 mM EDTA, 5 mM GSH and 1 mM DTT) at room temperature with overnight incubation. The tag and the TEV itself (that present a His-tag at the N-terminus), were then removed from the cut protein by reverse affinity chromatography. Protein expression level and purity was monitored by SDS-PAGE electrophoresis. GLRX5 was used directly for analysis or flash-frozen in liquid nitrogen and stored at -80°C.

Analytical gel filtration

The aggregation state of apo- and holo-glutaredoxin5 was analyzed using analytical gel filtration approach. Purified samples were run on ÄKTAFPLC™ (GE Healthcare) using SuperdexTM 75 HR 10/30 analytical column (Amerscham Bioscience). In both cases degassed 50 mM Tris-HCl buffer pH 8.0 (containing 5 mM GSH, 500 mM NaCl) was used as an eluent with the flow rate of 0.5 ml/min.

Iron and Sulfide quantification

Iron concentration of purified samples of GLRX5 was estimated by using iron chelator bathophenanthrolinedisulfonic acid disodium salt (BPS) colorimetric assay[19] and by ICP measurements. In the case of the colorimetric assay 350 µl of each sample dilution were treated with 50 µl of 5 % SDS for 5 minutes and later reduced with 50 µl of 10 mM sodium dithionite followed by addition of 50µl of 10mM BPS. Samples were incubated in dark for 20 minutes, mixed and centrifuged for 5 minutes at 12000 x g. Later absorbance at 515 nm was measured and concentration was estimated by plotting the obtained values on the standard curve prepared with different concentration (0 - 100 µM) of ferric ammonium sulfate $NH_4Fe(SO_4)_2$. Sulfide content of purified protein was determined by the colorimetric assay.[20] Samples (600 µl of several sample dilutions in water) were treated with 50 µl of 6 % NaOH and incubated for 5 minutes. Samples were mixed with 125 µl of 0.1 % DPD (N,N-dimethyl-p-phenylenediamine dihydrochloride, Sigma-Aldrich) (1 mg DPD dissolved in 1 ml 5 N HCl) followed by addition of 50 µl of 11.5 mM $FeCl_3$ (dissolved in 1.2 M HCl) and incubated for 30 minutes at room temperature. Sulfide concentration was estimated by plotting absorbance values (recorded at 670 nm) on the standard curve generated from different concentrations of (0 - 100 µM) sodium sulfide Na_2S.

UV-visible spectroscopy

In order to identify the Fe-S cluster type coordinated by GLRX5 freshly purified protein samples were analyzed by UV-vis spectroscopy technique. Spectra were recorded using Cary 50 Eclipse Spectrophotometer, for samples in degassed 20 mM Ammonium-acetate buffer pH 7.5. Spectra were recorded in the range between 250 and 750 nm. The stability of the cluster after exposure to air was checked by the same technique.

Mass spectroscopy

MALDI-MS experiments were performed on Bruker Daltonics Ultraflex III MALDI TOF/TOF instrument in order to confirm the molecular mass of the expressed and subsequently purified protein. 1 µl of protein solution was mixed with a 1 µl of matrix solution (SA 10 mg/ml in 70 % acetonitrile/30 % water, 0.1 % TFA) and analysed. Flex Control 3.0 was used as data acquisition software in positive linear mode. The instrument was externally calibrated prior to analysis using the Bruker Protein I calibrant standard kit (5000-17 000 Da).

ESI/LTQ-Orbitrap high resolution mass spectrometer (Thermo Electron Corporation) was used for mass analysis. Mass spectra of GLRX5 were recorded using ammonium-acetate buffer pH 7.5 as a solvent. ESI-MS data were acquired in positive ion mode, setting the spray voltage at 5 kV, the capillary voltage and temperature respectively at 45V and 37° C, and the tube lens at 230V. For acquisition, Xcalibur 2.0 software (Thermo) was used and monoisotopic and average deconvoluted masses were obtained by using the integrated Xtract tool. For spectra acquisition a nominal resolution (at m/z 400) of 100.000 was used.

Paramagnetic ^{1}H-NMR

Cluster characterization was performed using ^{1}H paramagnetic NMR. The spectra of 2 mM oxidized [2Fe-2S]-GLRX5 (50 mM phosphate buffer pH 7.0 with 10% D_2O, 5 mM GSH and 1 mM DTT) were recorded on a Bruker Avance 600 MHz spectrometer equipped with a selective ^{1}H probe head (5 mm SEL ^{1}H) at 298 K. Spectra of the paramagnetic signals were acquired with a recycle delay of 50 ms and an acquisition of 82 ms. The 1D spectra, each acquired with 2048 scans, were summed to obtain the final spectrum.

GLRX5 – relaxation time measurments

For Apo-GLRX5 nine experiments were collected to measure the ^{15}N R_1 relaxation rates, with recovery delays of 10, 70, 200,380, 550, 740, 1000, 1350, 2500 ms. Ten experiments were collected to measure the ^{15}N R_2 relaxation rates using delays of 16.96, 33.92, 67.84, 101.76, 135.68, 169.6, 203.52, 271.36, 305.28, 356.16 ms. All the data points were collected using 8 scans and with a recycle delay of 3 s.

For Holo-GLRX5 ten experiments were collected to measure the ^{15}N R_1 relaxation rates, with recovery delays of 10, 70, 150, 200,380, 550, 740, 1000, 1350, 3000 ms. Eleven experiments were collected to measure the ^{15}N R_2 relaxation rates using delays of 16.96, 33.92, 67.84, 101.76, 135.68, 169.6, 203.52, 237.44, 271.36, 305.28, 356.16 ms. All the data points were collected using 8 scans and with a recycle delay of 3 s. The spectra parameters are reported **i**n the **Table 1**.

For both apo- and holo- protein $^{15}N\{^1H\}$NOEs were measured using 80 scans and acquiring two spectra: one with ^1H saturation and one without. The ^1H saturation was maintained for 5.5 s. Spectra parameters are reported in **Table 1**. All the spectra were processed using the standard Bruker software (XWINNMR).

Table 1. Acquisition parameters for NMR experiments performed at 298 K.

Experiments	Dimension of acquired data (nucleus)		Spectral width (ppm)	
	t_1	t_2	F_1	F_2
^1H-^{15}N-HSQC	200 or 256 (^{15}N)	2048(^1H)	14.0	34.0
^{15}N R1	200 (^{15}N)	2048(^1H)	14.0	34.0
^{15}N R2	200 (^{15}N)	2048(^1H)	14.0	34.0
$^{15}N\{^1H\}$NOEs	200 (^{15}N)	2048(^1H)	14.0	34.0

Relaxation rates R_1 and R_2 were determined by fitting the cross-peak intensities measured as a function of the delay within the pulse sequence to a single-exponential decay. Errors on the rates were estimated through a least-squares method. The heteronuclear NOE values were obtained from the ratio of the peak intensity for [^1H]-saturated and unsaturated spectra. Integration of cross-peaks for all spectra was performed using a standard routine of CARA software package.

The overall rotational correlation time values were estimated from the R_2/R_1 ratio using the program QUADRATIC_DIFFUSION (G.A. Palmer III, Columbia University).[21] In this analysis, care was taken to remove from the input relaxation data those NHs having an exchange contribution to the R_2 value or exhibiting large-amplitude internal motions on a time scale longer than a few hundred picoseconds, identified from low NOE values, as inclusion of these data would bias the calculated tumbling time.[22,23]

Theoretical prediction of the reorientation tumbling rate was also estimated using the available structures with the "shell modeling" approach implemented in HYDRONMR software. The calculation was performed by using a viscosity of water at 298 K or 308 K of 8.9E-04 or 7.1 E-04 Ns/m^2, respectively.[24]

Results

Bioinformatic analysis of human GLRX5

Mitochondrial precursor of human glutaredoxin 5 is a 157 amino acid residues protein, containing a mitochondrial targeting signal in the N-terminal part of the sequence.[17] This peptide signal is cleaved upon protein maturation in the targeted compartment. On the basis of bioinformatic anlalysis, we cloned a GLRX5 construct where the N-terminal 34 residues constituting the targeting signal were removed together with the last seven residues from C-terminal part, which were not present in the crystal structure (PDB 2WEM) and therefore highly unstructured. Obtained construct comprised residues 35-150.

GLRX5 sequence was analysed with folding prediction tools – IndexFold and Phyre server.[18,25] No disordered regions were found (**Fig. 1**) and the protein was predicted to have thioredoxin-like fold. Blast search and alignment also showed that the cysteine (residue 67) from the CGFS motif is conserved among eukaryotic homologues. Moreover, another cysteine (residue 122) is conserved at the C-terminus (**Fig. 2**).

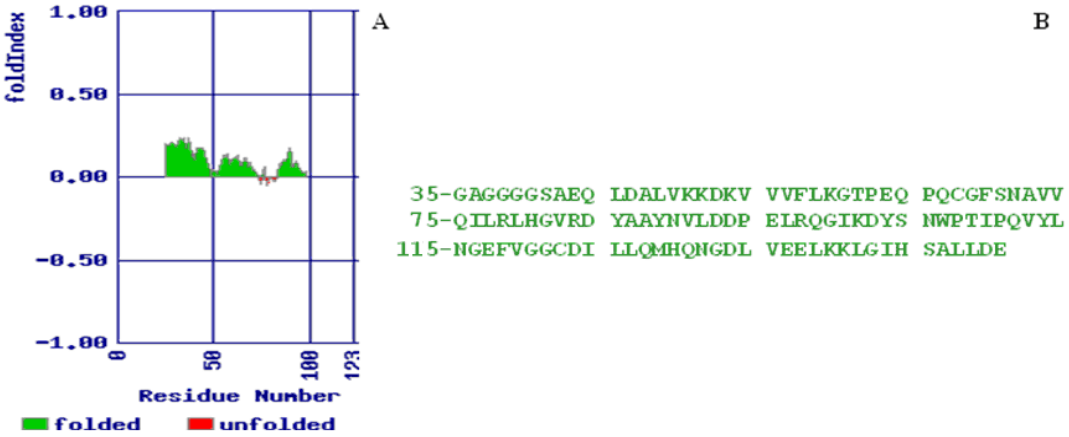


Figure 1. (A) Folding prediction of investigated construct of GLRX5 derived from FoldIndex program. Regions shaded in green are predicted to be in a folded state, while red shaded regions are predicted to be unfolded. (B) Sequence of expressed construct of GLRX5 (residues 35-150 of the mitochondrial precursor).

```
Homo_sapiens                -----MSGSLGRAAAALLRWGRGAGGGGLWGPGVRAAGSGAGGGGSAEQL
Rattus_Norvegicus           -----MSASLSRAAAALLRWGRGAGGGVLPRAGVRAASS----GGQAEQL
Bos_taurus                  -MSGSLGRAAATAASALLRWGRGAGGG-LRGPGVRAAGS--GGSGSSEHL
Gallus_gallus               MSGSLRAALRLGVWAARGRPGRRHLSEAASGSAGGSGGADGGGSGSREAV
Xenopus_laevis              -----MSWCLSKVSAAVLRGSARTPQACGSLSVLRWLSSD---SVSPEHL
Danio_rerio                 -----MNSVFRSTARCLRSATWQYKTAHKNGELSVGRARLMCSSAGQKNL
Saccharomyces_cerevisiae    ------MFLPKFNPIRSFSPILRAKTLLR------YQNRMYLSTEIRKAI
Candida_albicans            --MFRRSLLSTSRFINKSSSSLSPGLLLNNFR-GNFQQSRFISTELKDAL
Neurospora_crassa           --MLTRSLFSRQLFAAASRPAIAPKAVSSAFRPVLFYQNRFLSDATRQAI

Homo_sapiens                DALVKKDKVVVFLKGTPEQPQCGFSNAVVQILRLHGVR--DYAAYNVLDD
GLRattus_Norvegicus         DALVKKDKVVVFLKGTPEQPQCGFSNAVVQILRLHGVR--DYAAYNVLED
Bos_taurus                  DALVKKDKVVVFLKGTPEQPQCGFSNAVVQILRLHGVR--DYAAYNVLDD
Gallus_gallus               ERLVREHPVVVFMKGSPAQPLCGFSNAVVQILRLHGVE--DYRAHDVLQD
Xenopus_laevis              DGLVKKDKVVVFIKGTPAQPMCGFSNAVVQILRMHGVN--DYAAYNVLED
Danio_rerio                 EEMVKKDKVVVFMKGTPAQPMCGFSNAVVQILRMHGVD--NYASYNVLDD
Saccharomyces_cerevisiae    EDAIESAPVVLFMKGTPEFPKCGFSRATIGLLGNQGVDPAKFAAYNVLED
Candida_albicans            DKAVTTSPVVLFMKGTPEFPQCGFSRATIQILGQQGVDPEKFAAYNVLED
Neurospora_crassa           DKAVASAPVVLFMKGTPETPQCGFSRASIQVLGLQGVDPNKFAAFNVLED
                                   ** * ** *  * **** *    *   **        ** *

Homo_sapiens                PELRQGIKDYSNWPTIPQVYLNGEFVGGCDILLQMHQNGDLVEELKKLGI
GLRattus_Norvegicus         PELRQGIKDYSNWPTIPQVYLNGEFVGGCDILLQMHQNGDLVEELKKLGI
Bos_taurus                  PQLRQGIKDYSNWPTIPQVYLNGEFVGGCDILLQMHQNGDLVEELKKLGI
Gallus_gallus               PDLRQGIKNYSNWPTIPQVYLNGEFVGGCDILLQMHQNGDLVEELKKLGI
Xenopus_laevis              QDLRQGIKNYSNWPTIPQVYFNGEFVGGCDILLQMHQNGDLVEELNKLGI
Danio_rerio                 QDVRQGIKTFSNWPTIPQVFFNGEFVGGCDILLQMHQSGDLVEELQKLGI
Saccharomyces_cerevisiae    PELREGIKEFSEWPTIPQLYVNKEFIGGCDVITSMARSGELADLLEEAQA
Candida_albicans            SELREGIKEYSSWPTIPQLYVNGEFIGGCDIITSMAQNGELAELLEESNA
Neurospora_crassa           AELRQGIKEYSDWPTIPQLYIDKEFVGGCDIIVSMHQNGELAKLLEEKDV
                               * ***  * ******     ** ****    *   * *   *

Homo_sapiens                HSALLDEKKDQDSK--------------
GLRattus_Norvegicus         RSALIDEK-DQDSK--------------
Bos_taurus                  RSALLDDTKDQDSK--------------
Gallus_gallus               RSALLDAEKDQEKK--------------
Xenopus_laevis              RSALLDAEPSQEKK--------------
Danio_rerio                 RSALLDQEKESK----------------
Saccharomyces_cerevisiae    LVPEEEEETKDR----------------
Candida_albicans            LIPEENEPTTSEENQETVQSANIKPNRN
Neurospora_crassa           LVKGEEGAAEEQTEKKE-----------
```

Figure 2. Alignment of human glutaredoxin5 with homologue proteins from different organisms. The conserved active site motif (CGFS) is marked in red, second conserved cysteine residue located in the C-terminal part of the proteins is marked in yellow, (*) indicates identity of residues among the species.

Expression and purification of GLRX5

Gene encoding the mature version (residues: 35-150) of GLRX5 was amplified by PCR reaction and inserted into the Gateway® pEntr-TEV-d-Topo vector (Invitrogen™), followed by recombination with pETG20A destination vector and transformation into BL21 (DE3) Gold *E.coli* strain. Protein was recombinantly expressed and purified. Affinity chromatography purification employing the HisTag, performed without addition of GSH in anaerobic conditions lead to a soluble and pure apo-GLRX5. In order to obtain a holo-form of GLRX5, *E.coli* cells were grown in presence of 250 μM $FeCl_3$. Subsequent anaerobic purification (in glove-box) using degassed buffers (in presence of 5 mM GSH) resulted in reddish-brown protein solution, that indicated a presence of bound cofactor. These data show that the binding of Fe-S cluster is strictly dependent from a presence of glutathione. After purification the protein was concentrated in presence of 5 mM DTT and the fusion protein was cleaved using TEV protease at room temperature, and removed by reverse affinity column. For both apo- and holo-GLRX5, the purity of the target protein was higher than 90%, judged from SDS-PAGE electrophoresis (**Fig. 3**).

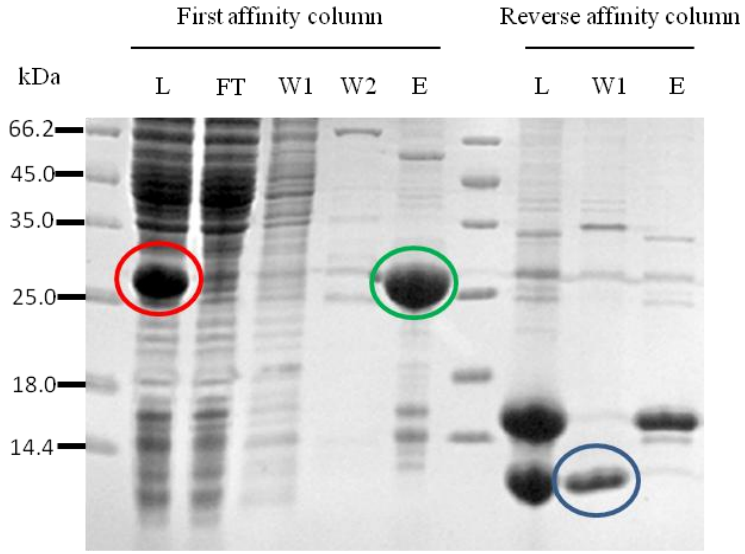


Fig 3. Purification of recombinant GLRX5. Recombinantly expressed in *E.coli* GLRX5 fused with thioredoxin tag (red circle), was applied on the Ni chelating column and subsequently purified (green circle). L – loading, FT – flow through, W1 – 1st washing step, W2 – 2nd washing step, E – elution. Purified fused target protein was digested with TEV protease and applied on the reverse affinity column. GLRX5 is a 13 kDa protein (blue circle).

The expected molecular weight (13157.9 Da) of the overexpressed apo-protein was confirmed by MALDI-MS (obtained value 13158,123 Da) and ESI-MS (obtained value 13157.67 Da) analysis.

The ratio between GRLX5:Fe:S in the holo-protein was estimated by using colorimetric assays (described in Materials and Methods). The iron content in the purified protein sample was also measured through ICP analysis. A GRLX5:Fe:S ratio of 1:2:2 has been achieved in all preparations.

GLRX5 binds a [2Fe-2S] cluster

In order to characterize the holo-GLRX5, UV-visible spectroscopy was performed. Samples were prepared in anaerobic chamber (in 20 mM ammonium-acetate buffer pH 7.5) and transferred to sealed cuvettes. Spectra were recorded in the range from 250-750 nm, showing the characteristic peaks of [2Fe-2S] cluster at 320 nm and 410 nm (**Fig. 4**).[26]

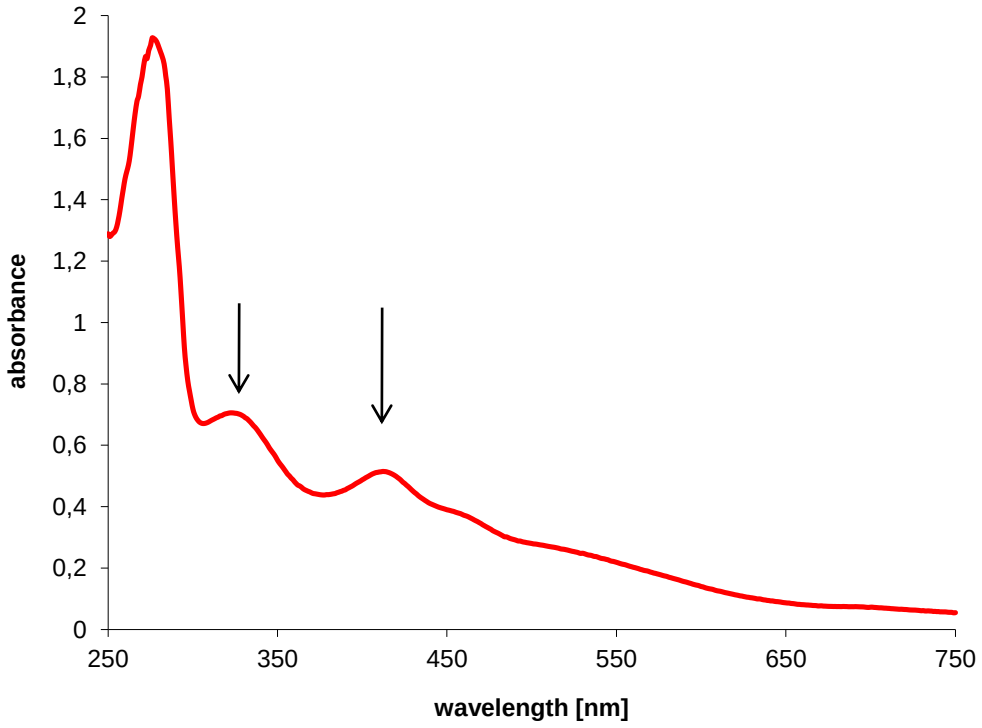


Figure 4. UV-visible spectra of freshly purified GLRX5. Presence of [2Fe-2S] cluster coordinated by GLRX5 was confirmed by UV-vis spectroscopy. Arrows on the spectra indicate peaks at 320 and 410 nm that are characteristic for [2Fe-2S] cluster.

The stability of the bound Fe-S cluster after exposure to air was also monitored by UV-visible spectroscopy (**Fig. 5**). After exposure to air the intensity of spectra was decreased, indicating release of the cluster. After 12h in aerobic conditions these absorbances were quenched, indicating that the Fe-S cluster is coordinated in the labile fashion.

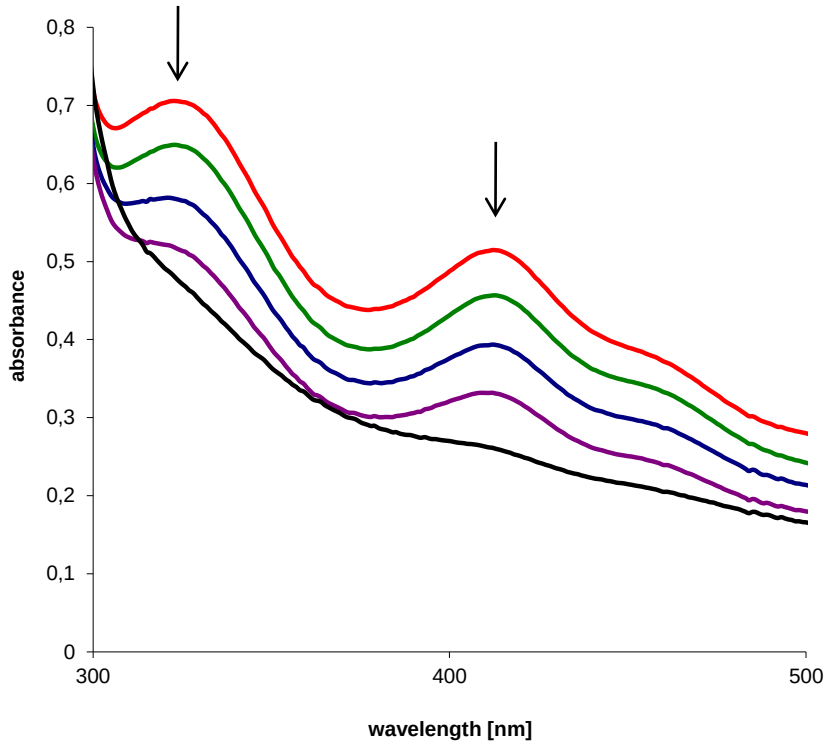


Figure 5. Quenching of the typical peaks upon air exposure. Freshly purified sample (red trace), 1 h air incubation (green trace), 3 h air incubation (blue trace), 6 h air incubation (violet trace), and 12 h air incubation (black trace). Arrows indicate characteristic peaks of 2Fe-2S cluster.

Further cluster characterisation was performed through 1D paramagnetic NMR approach. Data obtained from ^{1}H paramagnetic NMR also indicate a presence of [2Fe-2S] cluster bound by GLRX5. The spectrum of the oxidized protein recorded on 600 MHz magnet at 298K exhibited two broad peaks with maxima at 25 and 29 ppm (CH_{β}-Cys protons) and two sharp signals at 12 and 14 ppm (CH_{α}-Cys protons) (**Fig. 6**). These signals were typical of proteins with [2Fe-2S] cluster.[27,28,29] After anaerobic addition of 1 mM dithionite no paramagnetic signals were detected. We have interpreted this result as a consequence of fully release of the cluster, in agreement with the bleaching of protein solution.

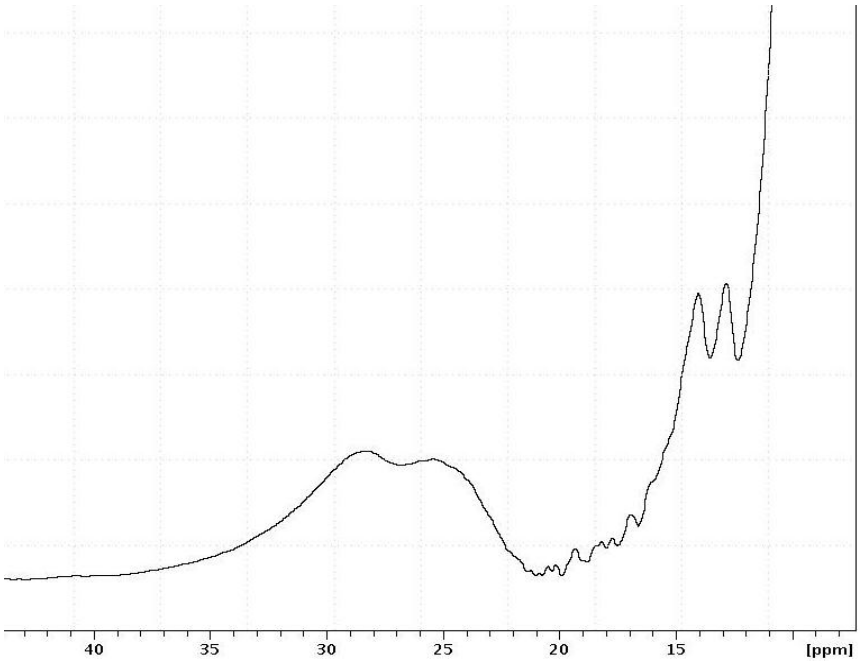


Figure 6. ^{1}H-NMR paramagnetic spectra of oxidized GLRX5

Paramagnetic ^{1}H- NMR spectrum of holo-GLRX5 (2 mM) in 50 mM phosphate buffer pH 7.0 (in presence of 5mM GSH and 5 mM DTT) recorded at 600 MHz magnet and 298K.

Glutaredoxin 5 coordinates [2Fe-2S] cluster as a monomer

It was recently reported that the CGFS-type monothiol Grxs from different organisms (cyanobacterium *Synechocystis* (SyGrx3p), *E.coli* (EcGrx4p)), exist as a dimeric iron-sulfur containing holo-proteins or as a monomeric apo-proteins in solution.[6,13] Following this information we decided to check if the human GLRX5 behave in the same manner. In order to answer this question we performed analytical gel filtration analysis of both apo- and holo-GLRX5. Samples were applied on the Superdex^{TM} 75 HR 10/30 column, equilibrated with degassed elution buffer in presence of 5 mM GSH and 5mM DTT. As described for cyanobacterium *Synechocystis* monothiol glutaredoxin (SyGrx3p) we expected to see a shift in the elution peaks corresponding to the monomeric (apo) and dimeric (holo) protein forms, but our experimental data were different. Apo- and holo-GLRX5 were indeed eluted at the same volume (13.5 ml) and consistently with the cluster being coordinated by a monomeric species (**Fig. 7**).

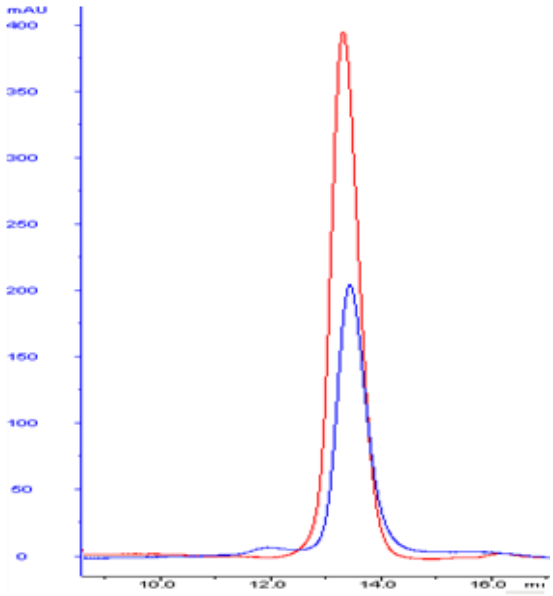


Figure 7. Analytical gel filtration of holo- and apo-GLRX5 (red and blue trace respectively).

^{1}H-^{15}N HSQC NMR experiments were then performed for both apo- and holo-protein. We can observe chemical shifts changes between two GLRX5 forms, indicating that several residues are affected by GSH-dependent Fe-S cluster binding (**Fig. 8**).

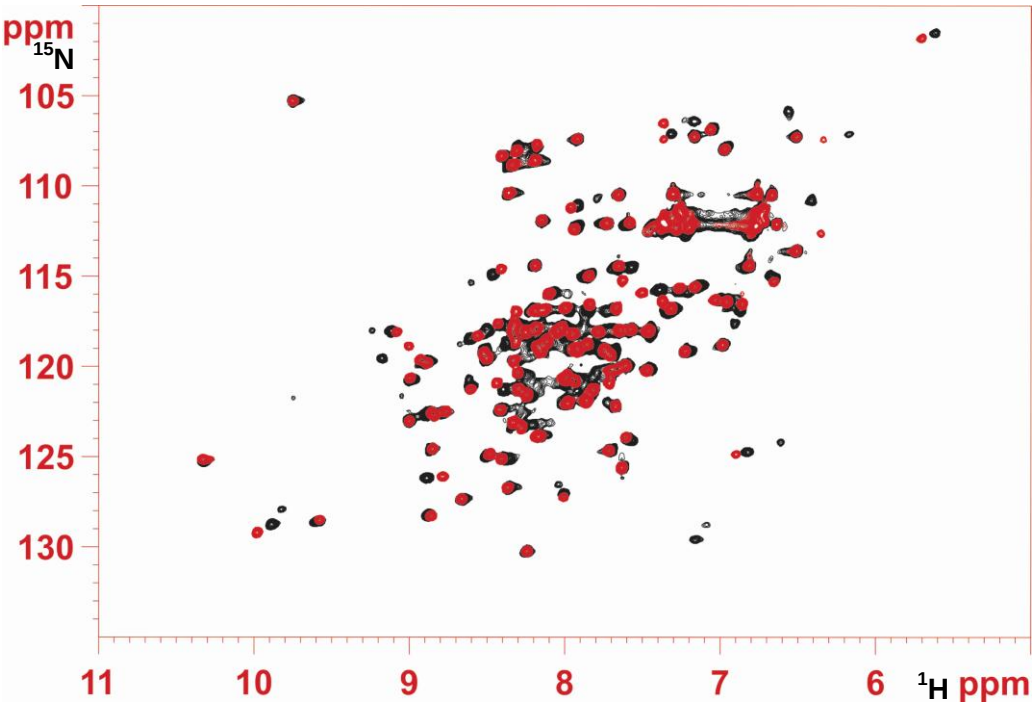


Figure 8. ^{1}H-^{15}N HSQC of the recombinant Glutaredoxin5 apo (red) and holo (black), Samples were used for the relaxation time calculations.

Backbone heteronuclear relaxation data can provide information on the overall protein tumbling rate which is strictly linked to the aggregation state of the molecule in solution. From the analysis of the well-resolved backbone HN signals on both apo and holo forms, average [^{15}N] R_1, R_2, and [^{1}H–^{15}N] NOE values of R_{1apo} 1.68 ± 0.25, R_{2apo} 10.18 ± 2.32, R_{1holo} 1.53 ± 0.18, R_{2holo} 12.54 ± 3.46 s^{-1}; [^{1}H–^{15}N] NOE_{apo} 0.70 ± 0.27, [^{1}H–^{15}N] NOE_{holo} 0.65 ± 0.38 are obtained, respectively. Both apo and holo forms feature essentially homogeneous heteronuclear relaxation properties. The correlation times for overall protein tumbling, as estimated from the R_2/R_1 ratios, are 7.53 ± 0.677 ns and 8.906 ± 0.856 ns for apo- and holo- protein, respectively. The tumbling times of both apo- and holo-GLRX5, experimentally estimated by heteronuclear relaxation data, are in agreement with that predicted through the HYDRONMR program at the same magnetic field and temperature. The latter have been calculated using the crystal structure of yeast and human apo-glutaredoxin 5 forms (9 ns and 8 ns). Therefore, these values are in agreement with a protein of this size (13 kDa) indicating that holo form is in a monomeric state as found for the apo form.

Discussion

GRX5 proteins are mitochondrial monothiol glutaredoxins present in eukaryotes from yeast, plants to humans. Bioinformatic sequence analysis of human GLRX5 shows that the protein contain a conserved, for monothiol subfamily, CGFS motif in the N-terminal part of the sequence, and another conserved, for GRX5 homologues only (not for all monothiol glutaredoxins) cysteine residue in the C-terminus. Several crystal structures of dimeric and tetrameric (dimmer of a dimers) forms of monothiol glutaredoxins (including human GLRX5) coordinating [2Fe-2S] cluster in the GSH dependent manner were reported.[13,30,] These data showed that the cluster is coordinated by two glutathione molecules and one cysteine of the active site from each subunit of the homo-dimer. We recombinantly expressed in *E.coli* and characterized two forms (apo- and holo-) of human GLRX5 protein.

UV-vis and 1D NMR spectroscopy data indicated that holo-GLRX5 binds [2Fe-2S] cluster in the GSH dependent, labile fashion similar to other monothiols previously described in the literature. In contrast to the reported crystal structure of human GLRX5 our aggregation state analysis, performed using analytical gel filtration and NMR relaxation studies, clearly showed that the GLRX5 in solution binds the cofactor as a

monomer. Therefore, we postulate that two conserved cysteines (residues 67 and 122) present in the sequence of the protein and GSH molecules can be necessary for the cofactor coordination. On this basis, two possible ways of cluster ligation are presented in the **Fig. 10**.

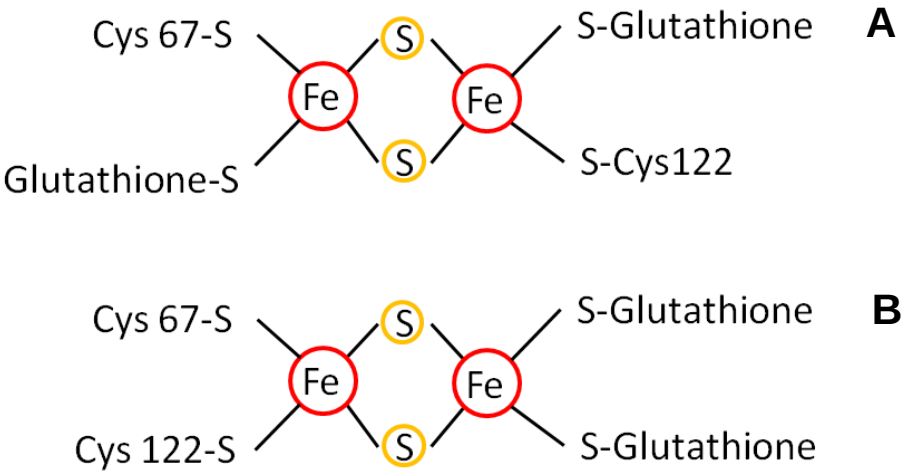


Figure 10. Two proposed models for [2Fe-2S] cluster coordination performed by human GLRX5.

The distance between the conserved cysteine residues measured from the crystal structure of human GLRX5 indicate that model A (**Fig. 10**), in which cysteine residues coordinate two different iron ions, does not require large structural rearrangements to ligate the cofactor, in contrast with model B (**Fig. 10**) in which large conformational changes would be necessary for the cysteine residues to coordinate the same iron ion.

In conclusion, our results suggest that the physiogical mitochondrial form of human GRLX5 contains a [2Fe-2S] cluster in a monomeric state. The latter form is best suited with respect to the tetrameric state to rapidly donate the Fe-S cluster to the protein partner, the recently identified ISCA1, ISCA2 proteins which are involved in the cluster insertion into mitochondrial Fe-S dependent enzymes. Protein-protein interaction studies between GRLX5 and ISCA1 or ISCA2 to follow Fe-S cluster transfer efficiency as well as to characterize the molecular recognition process are now in progress.

References

1. Iwema, T. et al. Structural basis for delivery of the intact [Fe2S2] cluster by monothiol glutaredoxin. *Biochemistry* **48**, 6041-6043 (2009).
2. Bandyopadhyay, S., Chandramouli, K. & Johnson, M.K. Iron-sulfur cluster biosynthesis. *Biochem. Soc. Trans* **36**, 1112-1119 (2008).
3. Rouhier, N., Couturier, J., Johnson, M.K. & Jacquot, J. Glutaredoxins: roles in iron homeostasis. *Trends Biochem. Sci* **35**, 43-52 (2010).
4. Rodríguez-Manzaneque, M.T., Ros, J., Cabiscol, E., Sorribas, A. & Herrero, E. Grx5 glutaredoxin plays a central role in protection against protein oxidative damage in Saccharomyces cerevisiae. *Mol. Cell. Biol* **19**, 8180-8190 (1999).
5. Rodriguez-Manzaneque, M.T., Tamarit, J., Belli, G., Ros, J. & Herrero, E. Grx5 Is a Mitochondrial Glutaredoxin Required for the Activity of Iron/Sulfur Enzymes. *Mol. Biol. Cell* **13**, 1109-1121 (2002).
6. Bandyopadhyay, S. et al. Chloroplast monothiol glutaredoxins as scaffold proteins for the assembly and delivery of [lsqb]2Fe-2S[rsqb] clusters. *EMBO J* **27**, 1122-1133 (2008).
7. Molinanavarro, M., Casas, C., Piedrafita, L., Belli, G. & Herrero, E. Prokaryotic and eukaryotic monothiol glutaredoxins are able to perform the functions of Grx5 in the biogenesis of Fe/S clusters in yeast mitochondria. *FEBS Letters* **580**, 2273-2280 (2006).
8. Kim, K., Chung, W., Kim, H., Lee, K. & Roe, J. Monothiol glutaredoxin Grx5 interacts with Fe-S scaffold proteins Isa1 and Isa2 and supports Fe-S assembly and DNA integrity in mitochondria of fission yeast. *Biochem. Biophys. Res. Commun* **392**, 467-472 (2010).
9. Camaschella, C. et al. The human counterpart of zebrafish shiraz shows sideroblastic-like microcytic anemia and iron overload. *Blood* **110**, 1353-1358 (2007).
10. Sheftel, A., Stehling, O. & Lill, R. Iron-sulfur proteins in health and disease. *Trends Endocrinol. Metab* **21**, 302-314 (2010).
11. Sheftel, A.D. et al. Human ind1, an iron-sulfur cluster assembly factor for respiratory complex I. *Mol. Cell. Biol* **29**, 6059-6073 (2009).
12. Ye, H. et al. Glutaredoxin 5 deficiency causes sideroblastic anemia by specifically impairing heme biosynthesis and depleting cytosolic iron in human erythroblasts. *J. Clin. Invest* **120**, 1749-1761 (2010).
13. Iwema, T. et al. Structural Basis for Delivery of the Intact [Fe2S2] Cluster by Monothiol Glutaredoxin. *Biochemistry* **48**, 6041-6043 (2009).
14. Ye, H. & Rouault, T.A. Erythropoiesis and iron sulfur cluster biogenesis. *Adv Hematol* **2010**, (2010).
15. Rouhier, N. et al. Functional, structural, and spectroscopic characterization of a glutathione-ligated [2Fe-2S] cluster in poplar glutaredoxin C1. *Proc. Natl. Acad. Sci. U.S.A* **104**, 7379-7384 (2007).
16. Feng, Y. et al. Structural insight into poplar glutaredoxin C1 with a bridging iron-sulfur cluster at the active site. *Biochemistry* **45**, 7998-8008 (2006).
17. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protocols* **2**, 953-971 (2007).
18. Prilusky, J. et al. FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435 -3438
19. Zhang, Y. et al. Dre2, a conserved eukaryotic Fe/S cluster protein, functions in cytosolic Fe/S protein biogenesis. *Mol. Cell. Biol* **28**, 5569-5582 (2008).

20. Siegel, L.M. A direct microdetermination for sulfide. *Analytical Biochemistry* **11**, 126-132 (1965).
21. Lee, L.K., Rance, M., Chazin, W.J. & Palmer, A.G. Rotational diffusion anisotropy of proteins from simultaneous analysis of 15N and 13C alpha nuclear spin relaxation. *J. Biomol. NMR* **9**, 287-298 (1997).
22. Kay, L.E., Torchia, D.A. & Bax, A. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972-8979 (1989).
23. Tjandra, N., Feller, S.E., Pastor, R.W. & Bax, A. Rotational diffusion anisotropy of human ubiquitin from 15N NMR relaxation. *Journal of the American Chemical Society* **117**, 12562-12566 (1995).
24. García de la Torre, J., Huertas, M.L. & Carrasco, B. HYDRONMR: Prediction of NMR Relaxation of Globular Proteins from Atomic-Level Structures and Hydrodynamic Calculations. *Journal of Magnetic Resonance* **147**, 138-146 (2000).
25. Kelley, L.A. & Sternberg, M.J.E. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protocols* **4**, 363-371 (2009).
26. Picciocchi, A., Saguez, C., Boussac, A., Cassier-Chauvat, C. & Chauvat, F. CGFS-Type Monothiol Glutaredoxins from the Cyanobacterium Synechocystis PCC6803 and Other Evolutionary Distant Model Organisms Possess a Glutathione-Ligated [2Fe-2S] Cluster†. *Biochemistry* **46**, 15018-15026 (2007).
27. Banci, L. The 1H NMR parameters of magnetically coupled dimers—The Fe2S2 proteins as an example. *Structure and Bonding* **72**, 113-1135
28. Bertini, I., Turano, P. & Vila, A.J. Nuclear magnetic resonance of paramagnetic metalloproteins. *Chemical Reviews* **93**, 2833-2932 (1993).
29. Nagayama, K., Ozaki, Y., Kyogoku, Y., Hase, T. & Matsubara, H. Classification of iron-sulfur cores in ferredoxins by 1H nuclear magnetic resonance spectroscopy. *J. Biochem* **94**, 893-902 (1983).
30. Luo, M. et al. Structural and Biochemical Characterization of Yeast Monothiol Glutaredoxin Grx6. *Journal of Molecular Biology* **398**, 614-622 (2010).

3.3. Wilson and Menkes proteins: cloning, expression and purification of ATP-binding and P-domain

Lucia Banci, Ivano Bertini, Manuele Migliardi, Maciej Mikolajczyk

I.B. and L.B. planned the research. M.M. and M.M. planned, coordinated and performed cloning, protein production and characterization.

Introduction

P-type ATPases are a superfamily of membrane proteins, which are responsible for the active transport of a variety of cations across cell membranes. More than 200 P-type ATPases have been identified so far and they play critical roles in ion homeostasis in species as diverse as bacteria and humans. Translocation of cations against their electrochemical potential gradient is achieved by utilizing the energy of the hydrolysis of ATP.[1,2] The human Cu-transporting P_{1B}-type ATPases ATP7A and ATP7B (Menkes MNK and Wilson WLN proteins, respectively) are two key proteins that regulate the Cu status of the body. They transport Cu across cellular membranes for biosynthetic and protective functions, enabling Cu to fulfill its role as a catalytic and structural cofactor for many essential enzymes, and to prevent a toxic build-up of Cu inside cells.[3,4] Mutations in genes encoding MNK and WLN proteins lead to Menkes and Wilson diseases, respectively (see also chapter 1).

The core structure of Menkes and Wilson proteins (that share ~60% amino acid identity) is composed of eight transmembrane helices and two soluble domains, the Actuator-domain (A) and the ATP-binding domain. The latter one is composed by two subdomains: the Phosphorylation-domain (P) and the Nucleotide-domain (N) which is an insert in the P-domain sequence. In addition, their long N-terminal tail contains six independently folded copper(I)-binding domains (MBD1-6) (**Fig. 1**).[5]

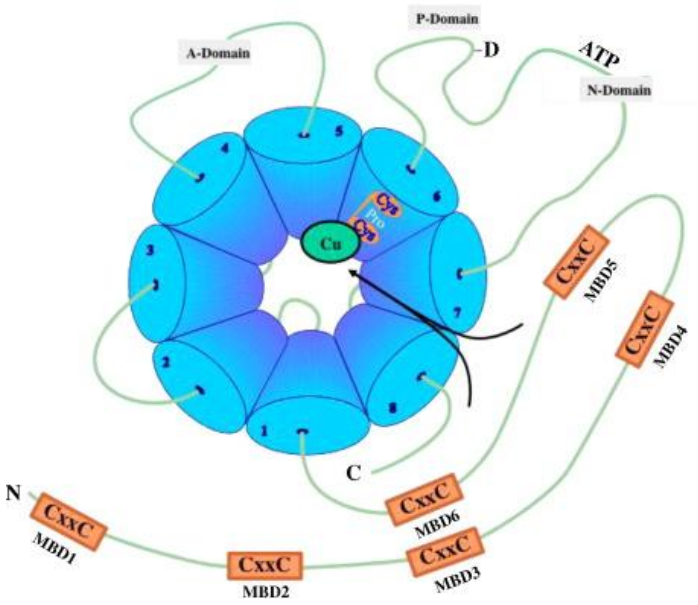


Figure 1. Schematic diagram of the structure of the Cu-ATPases, ATP7A and ATP7B. Reprinted from La Fontaine et al., *Journal of Biochemistry & Cell Biology* (2010)

These Cu-ATPases transport copper from the cytosol across cellular membranes. The central step in the catalytic cycle is the binding of an ATP molecule by the N-domain. γ-phosphate is transferred from ATP to the invariant Asp residue in the DKTG motif, located within the P-domain, resulting in the formation of a transient phosphorylated intermediate. The prerequisite for this reaction is the binding of copper to the sites within the membrane portion of the enzyme containing the conserved CPC motif, while the release of copper from these sites stimulates the dephosphorylation of Asp performed by the conserved TGE motif located in the A-domain.[6,7,8] During copper transfer, MNK and WLN P_{1B}-type ATPases undergo large conformational changes. Originally, two distinct conformations were called E1 and E2 (enzyme-1 and enzyme-2) (**Fig. 2**), with each having a different affinity for the nucleotide and the transported ions.[4,7,9,]

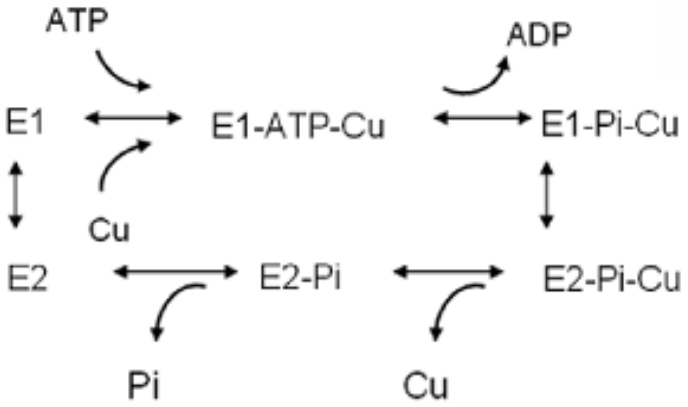


Figure 2. The simplified catalytic cycle of human Cu-ATPases. Two major conformational states associated with high affinity for ATP and Cu (E1) and lower affinity for these ligands (E2) as well as phosphorylated intermediates (E1–Pi–Cu and E2–Pi–Cu) are shown.

Several soluble domains of MNK and WLN proteins had been structurally characterized in our laboratory, followed by intermolecular and partner (human copper(I) chaperone) interaction investigations.[7,10,11,12,13] Moreover, mutational studies were performed to get insights on the mechanism of protein copper-dependent translocation.[14] With the aim of structurally characterizing and understanding at the molecular level the interplay between the ATP-binding domain and N-terminal copper binding domain, I attempted to produce ATP-binding domains from both MNK and WLN proteins as well as the P-domain of MNK.

Materials and methods

Bioinformatic tools

Sequences of ATP-binding domain from Menkes and Wilson proteins were aligned using the ClustalW program and default parameters. Folding prediction of MNK ATP-binding domain and P-domain constructs was performed using FoldIndex© and PSI-PRED online bioinformatic tools.[15,16] A 3D model of MNK ATP-BD was built based on multiple-threading alignments by LOMETS and iterative TASSER simulations. The PDB entry 3A1C was used as a template for modeling.[17,18]

Cloning strategies

Investigated constructs:

1. WLN ATP-binding domain (K1010-R1324),
2. MNK ATP-binding domain (K1027-R1354),
3. MNK ATP-binding domain (without helices)(K1037-D1337),
4. MNK P-domain (K1027-T1050/fused with/V1232-R1354),
5. MNK P-domain Ala (K1027-T1050/AlaAla/V1232-R1354),
6. MNK P-domain without helices (K1037-T1050/fused with/V1232-D1337),
7. MNK P-domain Ala without helices(K1037-T1050/AlaAla/V1232-D1337)

Genes encoding WLN (construct 1) and MNK (construct 2) ATP-binding domains were amplified from human cDNA using PCR and inserted into the Gateway® pEntr-TEV-d-Topo vector (Invitrogen™). The entry clone containing MNK ATP-binding domain (construct 2) was used as a template for the amplification of constructs 3,4,5,6 and 7. Primers for all the constructs were designed for Gateway Technology®. Constructs of MNK P-domain (4 and 5) were build-up by several rounds of PCR reaction in order to eliminate the sequence of the N-domain and fuse the C- and N-terminal part of the P-domain. Two predicted alpha-helices from the C-terminal and N-terminal part of the MNK-ATP and P-domain constructs (marked - without helices) were eliminated using PCR reaction.

Expression and purification

Positive clones of constructs 1 and 2 were subcloned into pDEST-17(6xHis), PTH34 (6xHis-Gb1), PETG-20A (TRX-6xHis), PETG-30A (6xHis-GST) and pDEST-MBP (6xHis-MBP) destination vectors in order to perform expression test. Positive clones of constructs 3, 4, 5, 6, 7 were subcloned into destination vectors: pDEST-17(6xHis), PTH34 (6xHis-Gb1), PETG-20A(TRX-6xHis), PETG-30A(6xHis-GST). All the plasmids were transformed into Rosetta-2 (DE3), BL21 (DE3) pLysS and BL21 (DE3) gold *E.coli* strains. A series of expression and solubility tests were performed using: standard expression test (100 ml culture, IPTG concentrations: 1.0 - 0.25 mM and temperature: 37, 25 and 17^{o}C) for constructs 1 and 2. High-throughput expression and solubility test (5 ml cultures grew on 24-well plates induced with IPTG concentration of 0.5 mM at 37, 25 and 17^{o}C,) was performed for constructs 3, 5, 6 and 7. Samples were collected after 5-16 hours after induction and analysed by SDS-PAGE electrophoresis.

Soluble proteins were purified on nickel-charged HiTrap chelating HP column (Amersham Pharmacia Biosciences) followed by over-night TEV digestion and reverse Ni(II) affinity chromatography in order to remove the cleaved tag. Additionally, the amylose resin affinity chromatography column was employed in case of MBP fused proteins. Size exclusion chromatography on Superdex 75 Hi Load-16/60 column (GE Healthcare) was used as the final purification step.

In order to obtain sufficient amount of WLN ATP-binding domain fused with 6xHisTag (expressed as inclusion bodies), bacterial cells were harvested after 5 h of induction (1 mM IPTG) at 37° C. Cells were resuspended in 50 ml of lysis buffer (50 mM Tris at pH 8, 150 mM NaCl, 1 mM EDTA, 0.1% Triton X100, 1 mM PMSF, 0.25 mg/ml lysozyme) and frozen overnight at -80°C. After thawing the cell suspension, DNAse and $MgSO_4$ were added at final concentrations of 10 μg/ml and 20 mM, respectively. The lysate was incubated for 30 min at 37°C (or until it was no longer viscous), and then spun for 30 min at 17,000*g*. After discarding the supernatant, the pellet was thoroughly resuspended in 50 ml of Tris buffer (50 mM Tris at pH 8, 150 mM NaCl), disrupted by sonicating it four times with a 15-sec pulse, and spun again. This washing procedure was repeated three times. The final pellet was solubilized in 20 ml of 50 mM Tris, 150 mM NaCl, 10 mM imidazole, 8 M guanidinium chloride. After a 30-min run at 17,000*g*, the supernatant was loaded onto a HiTrap chelating HP column (Amersham Bioscience) preequilibrated with 50 mM Tris, 150 mM NaCl, 10 mM imidazole, and 8 M urea (buffer

A). The column was washed with buffer A supplemented with 40 mM imidazole, and the recombinant protein was eluted with buffer A containing 250 mM imidazole. Protein purity was checked by SDS-PAGE. Elution fractions containing the protein of interest were pooled, and the imidazole was removed using a desalting column.

Refolding

In order to find the best conditions for refolding of 6xHisTag WLN ATP-binding domain, 27 different buffer compositions (Tris-HCll pH 8.0, phosphate pH 7.0 and ammonium-acetate pH 5.0, buffers with different NaCl and glycerol concentration) were tested using UV-visible spectroscopy to follow the precipitation by measuring solution turbidity at wavelength 340 nm.[19] The best buffer conditions (the lowest absorbance), were scaled-up.

Light scattering

Aggregation states of purified proteins were investigated by using 0.3 mM samples (100 µl) on a G2000SW_{XL} analytical column (Tosoh Bioscience) with a void volume of 150 kDa. The column was connected with a multiangle light scattering device (DAWN-EOS Wyatt Technologies, Santa Barbara, CA) coupled with quasi-elastic light scattering detectors. Data were analysed with Astra software (Wyatt Technology) to estimate the molar mass.

Results and discussion

Wilson and Menkes ATP-binding domains

Wilson and Menkes ATP-binding domains (constructs 1 and 2, see materials and methods) were expressed as inclusion bodies in three different *E.coli* strains (Rosetta-2(DE3), BL21(DE3) pLysS and BL21(DE3) gold) using several destination vectors (pDEST-17(6xHis), PTH34 (6xHis-Gb1), PETG-20A (TRX-6xHis), and PETG-30A (6xHis-GST)) during the standard expression test.

The only soluble expression was obtained using pDEST-MBP (6xHis-MBP) vector. The highest yield for both proteins was obtained by expressing the recombinant proteins in the BL21 (DE3) pLysS strain at 25° C using 0.5 mM final IPTG concentration. These conditions were scaled-up and extracted proteins after affinity chromatography purification were digested with TEV protease. Efficiency of cleavage was lower than 50%. Proteins were subsequently applied on reverse affinity column followed by amylose resin purification. Elution fractions from the amylose resin column contained cleaved and fused species. Further purification using the Gel filtration showed two peaks; one in the void volume of the column containing both cleaved and un-cleaved proteins and the second containing only MBP fusion. These results suggest that the WLN and MNK ATP-binding domains have high tendency to aggregate.

Refolding of Wilson ATP-binding domain

Wilson-ATP binding domain fused with 6xHisTag (pDEST17 vector) was the subject of refolding condition screening. The protein was overexpressed in the BL21(DE3) pLysS *E.coli* strain at 37° C for 5 h in the presence of 1 mM IPTG and purified from inclusion bodies (IB). Refolding buffer conditions were screened by quick dilution method (with protein : buffer ratio equal 1:100) followed by precipitation analysis using UV-visible absorbance. The tested buffer conditions are presented in (**Tab. 1**).

Table 1. High-throughput buffer condition test. Three different 50 mM buffers (ammonium-acetate (pH 5.0), phosphate (pH 7.0) and Tris-HCl (pH 8.0)) were tested. The buffer with the lowest absorbance at 340 nm was chosen for the refolding scale-up (highlighted in yellow)

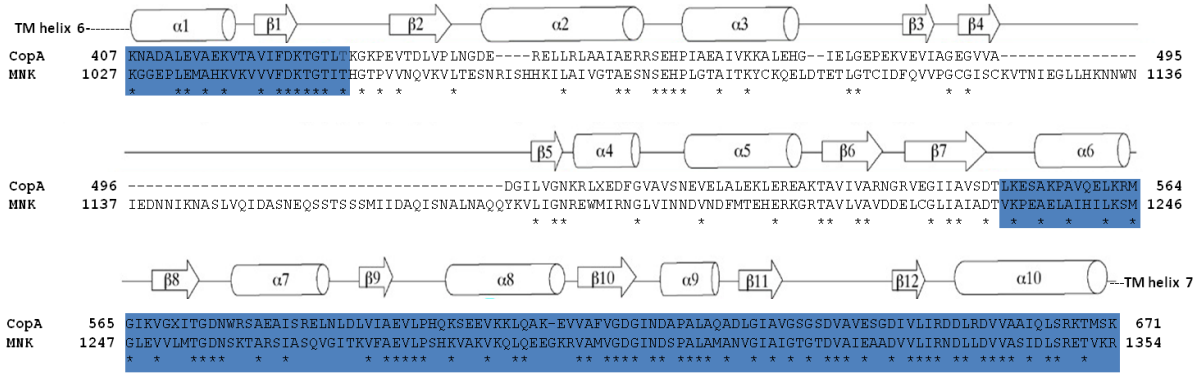
	BUFFER	NaCl [M]	Glycerol		BUFFER	NaCl [M]	Glycerol		BUFFER	NaCl [M]	Glycerol
1	pH 7.0	0	0%	**10**	pH 5.0	0,2	10%	**19**	pH 5.0	0,1	0%
2	pH 8.0	0,2	0%	**11**	pH 5.0	0,2	0%	**20**	pH 8.0	0	10%
3	pH 8.0	0,1	10%	**12**	pH 8.0	0	5%	**21**	pH 8.0	0,1	0%
4	pH 5.0	0	5%	**13**	pH 5.0	0	0%	**22**	pH 5.0	0,1	10%
5	pH 8.0	0,2	10%	**14**	pH 8.0	0,2	5%	**23**	pH 7.0	0,2	5%
6	pH 7.0	0	10%	**15**	pH 8.0	0,1	5%	**24**	pH 7.0	0,1	5%
7	pH 5.0	0	10%	**16**	pH 7.0	0,2	0%	**25**	pH 5.0	0,1	5%
8	pH 5.0	0,2	5%	**17**	pH 7.0	0,2	10%	**26**	pH 7.0	0	5%
9	pH 7.0	0,1	10%	**18**	pH 7.0	0,1	0%	**27**	pH 8.0	0	0%

Scaled-up refolding of the purified WLN ATP-binding domain was performed in anaerobic conditions (inert atmosphere chamber – Coy Lab) in an ice bath using multi-step dialysis against degassed solutions containing 50 mM Tris-HCl (pH 8.0), 0.2 M NaCl, decreasing concentrations of urea (8 – 0 M) and 5 mM DTT. Aggregation state analysis of the refolded sample (using analytical gel filtration approach - column G2000 SW_{XL}) showed that the protein forms aggregates bigger than 150 kDa (only one peak in the void volume of the column). After 3-4 hours of incubation at room temperature all the protein precipitated. The addition of detergents (TritonX-100, PEG 3000), glycerol or Arginine during or/and after the refolding process did not inhibit the aggregate formation.

Menkes ATP-binding and P -domain

P_{1B}-ATPases are found in all kingdoms of life, from extremophilic archaea to man and despite of sequence differences they share a common basic core architecture. One of the common protein domains - the ATP-binding domain - is the large cytoplasmic loop located between the transmembrane helix 6 and helix 7. The domain encompasses two subdomains; the nucleotide binding (N) and the phosphorylation (P) domains. The P-domain is the most conserved among all $P1_B$-ATPases. Based on the homology (**Fig. 3A**) to the structurally characterized ATP-binding domain from the hyperthermophile *Archaeoglobus fulgidus, namely* CopA, which is a Cu^+-ATPase that contains all of the key elements present in eukaryotic Cu^+-ATPases, a 3D model of the MNK ATP-binding domain was created using the I-TASSER server (**Fig. 3B**) and new constructs of the Menkes ATP-binding domain were cloned. [17,18,20]

A



B

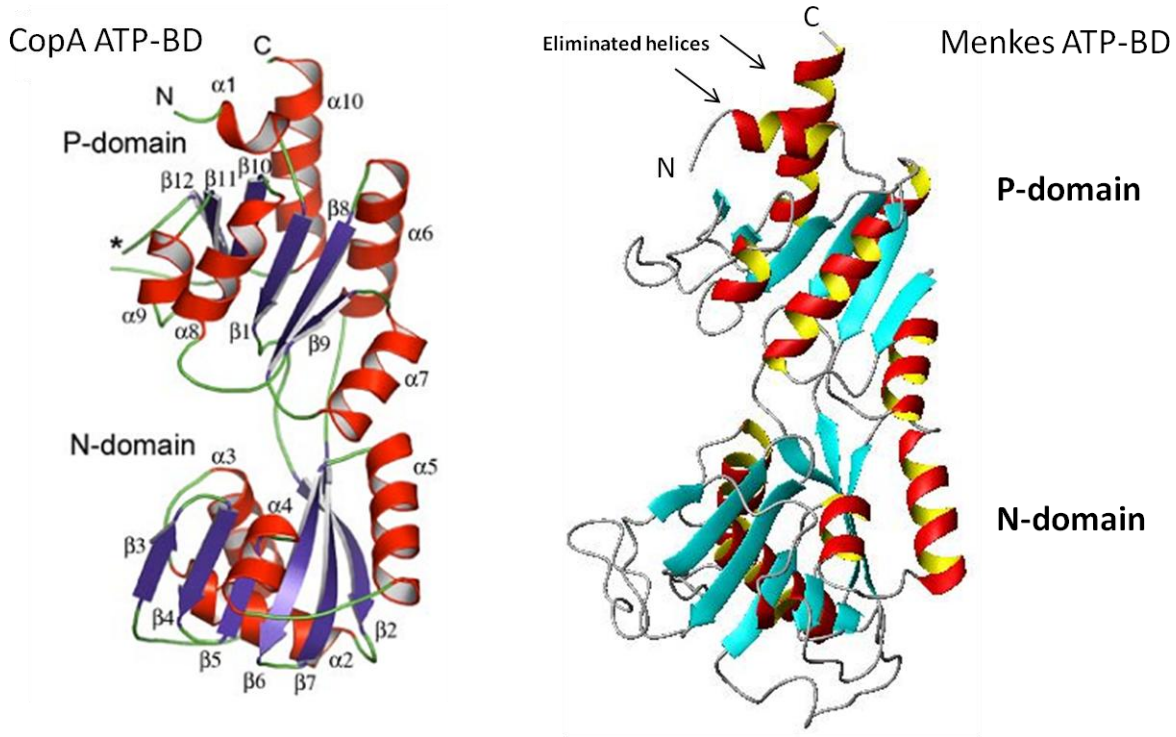


Figure 3. (A) Sequence alignment between ATP-binding domains of CopA and MNK proteins. Barrels and arrows represent secondary structure components of the CopA ATP-binding domain. Residues comprising the P-domain are highlighted in blue. Asterisks (*) indicate conserved residues (homologues share 36% of identity). (B) On the left, crystal structure of ATP-binding domain of CopA from *Archaeoglobus fulgidus* (Reprinted from Sazinsky et al. *JBC* 2006). On the right, 3D model of ATP-binding domain of Menkes. N- and C-terminal α-helices eliminated during cloning of the new constructs of ATP binding and P-domain are marked with arrows.

In order to increase the solubility of the recombinantly expressed MNK ATP-binding domain, two α-helices (one from the C- and the other from the N-terminus, helices indicated by arrows **Fig. 3B**), which are directly connected with the hydrophobic transmembrane part (TM helix 6 and 7, **Fig. 3A)** were removed from the coding sequence using PCR reaction (construct 3). Additionally two constructs of the MNK P-domain were created; first, by elongation of the C-terminus and thus "fusion" of two separated (by N-domain) parts of the P-domain sequence (construct 4), and second using the same approach but separating the sequence of the P-domain with two alanine residues (construct 5). Finally two other P-domain constructs (6 and 7) were created by eliminating the α-helices from the C- and N-terminal part of construct 4 and 5 sequences (**Fig. 4**).

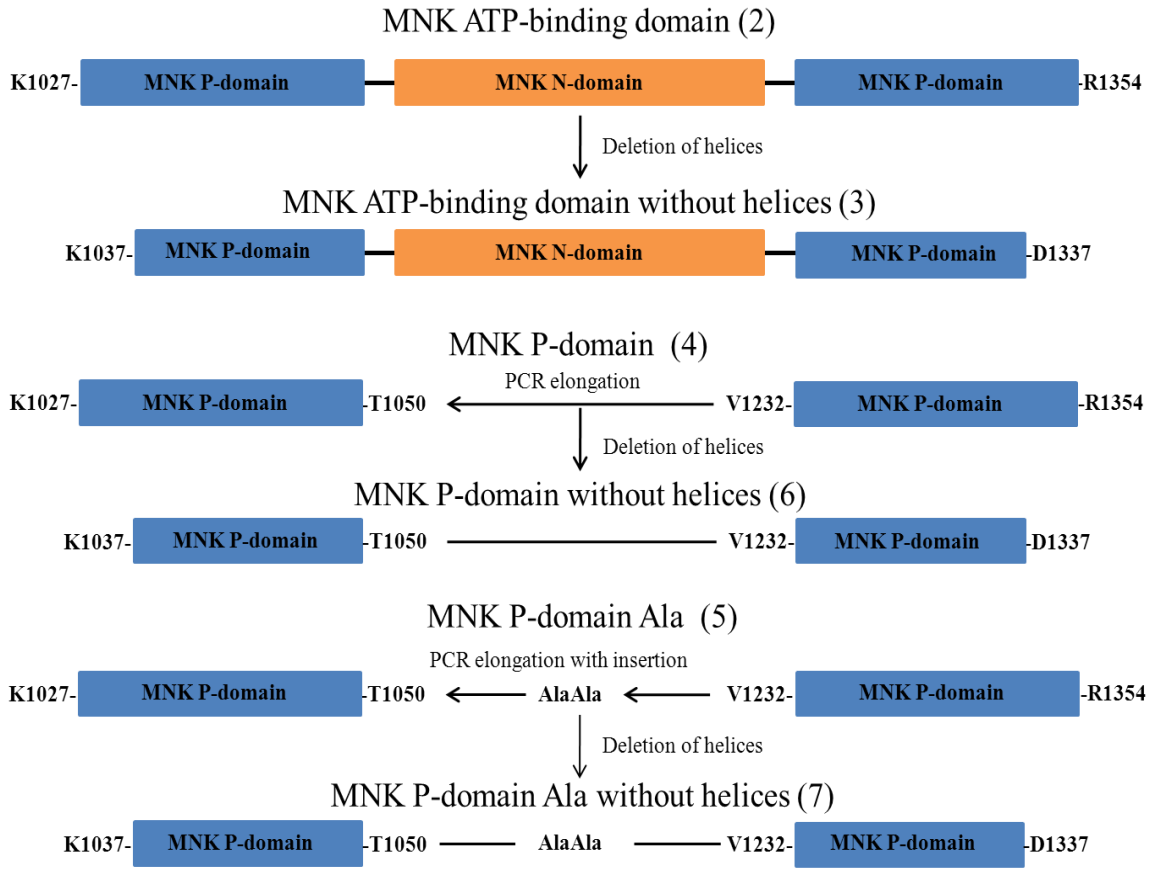


Figure 4. Schematic representation of the cloning process performed in order to obtain new constructs of Menkes ATP- and P-domains. Numbers in brackets correspond to the constructs described in material and methods and text.

Constructs 3, 5, 6 and 7 were amplified using PCR and inserted into the Gateway® pEntr-TEV-d-Topo vector (Invitrogen™). After recombination with several destination vectors and subsequent transformation into *E.coli* host strains the expression and solubility test was performed using the high throughput approach. Cells were grown in the 24-well plates containing LB medium in presence of appropriate antibiotics at 37° C. After induction with 0.5 mM IPTG, plates were further incubated at different temperatures (17, 25 or 37°C). After 5 (37°C) to 16 h (17 and 25°C) cells were pelleted by centrifugation and resuspended in lysis buffer (A) (50 mM Tris-Hcl, 500 mM NaCl, 5 mM imidazole) containing lysozyme and DNA-se I. After subsequent sonication the insoluble fraction was separated from the soluble fraction by centrifugation. The applied HisTag (present in all the constructs) allowed purifying the expressed proteins by affinity to nickel charged agarose beads. The obtained supernatant was incubated with affinity agarose beads for 1 h at 25° C. In order to separate the beads from the cell extract, the whole samples were transferred into a large pore 96-well filter plate that retains the beads with bound proteins. The beads were washed (buffer A) and subsequently the target proteins were eluted from the beads, and analyzed by SDS-PAGE electrophoresis (**Fig. 5**). Tested conditions are listed in **table 2.**

Table 2. Conditions tested during the expression test. Cells: G - BL21 (DE3) gold; P - (DE3) pLysS; R - Rosetta-2 (DE3).

Construct	Plasmid	Cells	Temperature	Expected MW (kDa)
(3) ATP - H	pDEST 17	G, P, R	17, 25, 37°C	35
	pETG 20	G, P, R	17, 25, 37°C	47
	pETG 30	G, P, R	17, 25, 37°C	61
(5) P-dom Ala	pDEST 17	G, P, R	17, 25, 37°C	20
	pETG 20	G, P, R	17, 25, 37°C	32
	pETG 30	G, P, R	17, 25, 37°C	46
(6) P-dom -H	pDEST 17	G, P, R	17, 25, 37°C	17
	pETG 20	G, P, R	17, 25, 37°C	29
	pETG 30	G, P, R	17, 25, 37°C	43
(7) P-dom Ala -H	pDEST 17	G, P, R	17, 25, 37°C	17
	pETG 20	G, P, R	17, 25, 37°C	29
	pETG 30	G, P, R	17, 25, 37°C	43

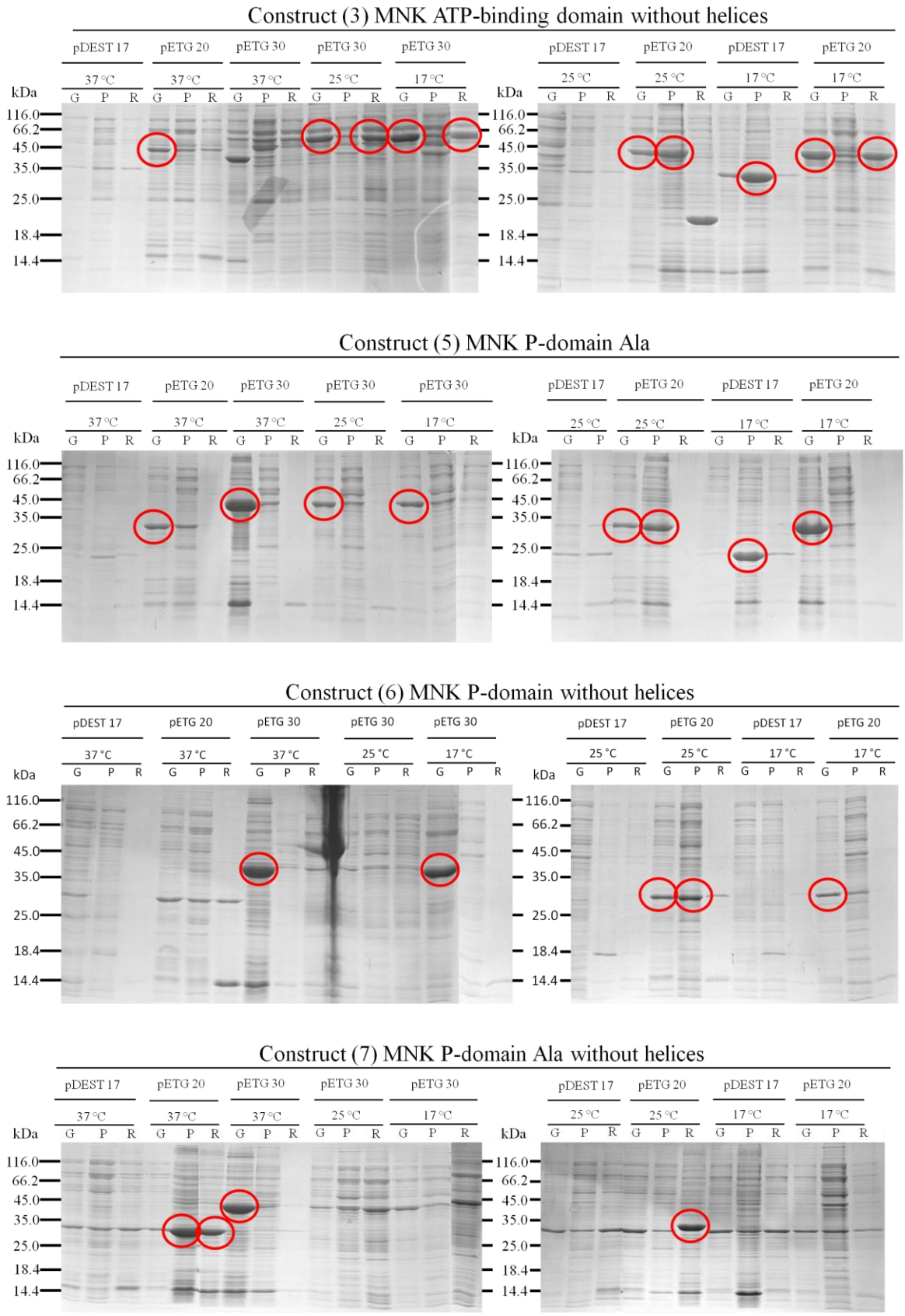


Figure 5. SDS-PAGE electrophoresis performed after expression/solubility test. Positive results for each construct are marked with red circle.

Several positive results were obtained for each construct (**Fig. 5**). Conditions in which proteins were overexpressed and soluble had been scaled up. Recombinant proteins were purified by affinity chromatography employing the HisTag present in all the fusions. Subsequent concentration was followed by digestion with TEV protease. In all the cases the cleavage efficiency was not higher than 5 – 10% (judged from SDS-PAGE). Addition of either L-Arginine, 10% glycerol or 0.1% Tween 80 to the purified proteins as well as purification in presence of detergents (Tween 80, TritonX 100) did not increase the cleavage efficiency. The same behaviour was observed for the P-domain (construct 4). Samples run on the analytical gel filtration column (G2000 SW_{XL}) were eluted in the void volume indicating the presence of high molecular weight aggregates (> 150kDa) (**Fig. 6**).

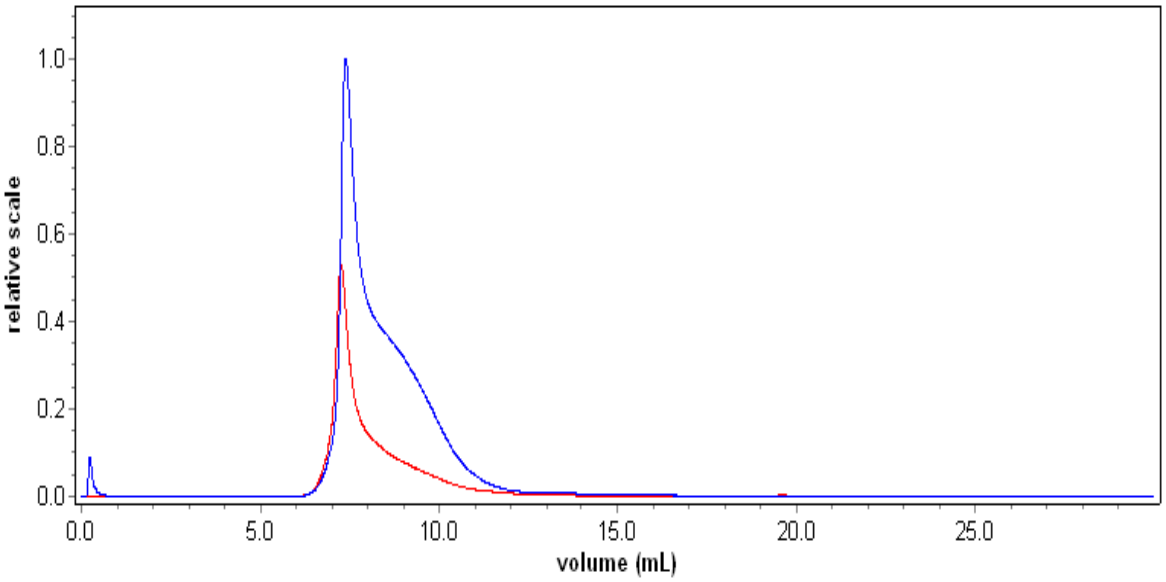


Figure 6. Analytical gel filtration elution profile for MNK ATP-binding domain without helices (construct 3) fused with GST, purified in absence (red trace) and presence (blue trace) of 1% Tween 80. Similar results were obtained for other constructs.

Light scattering experiments performed in the same conditions confirmed that the overexpressed proteins form soluble aggregates (**Fig. 7**)

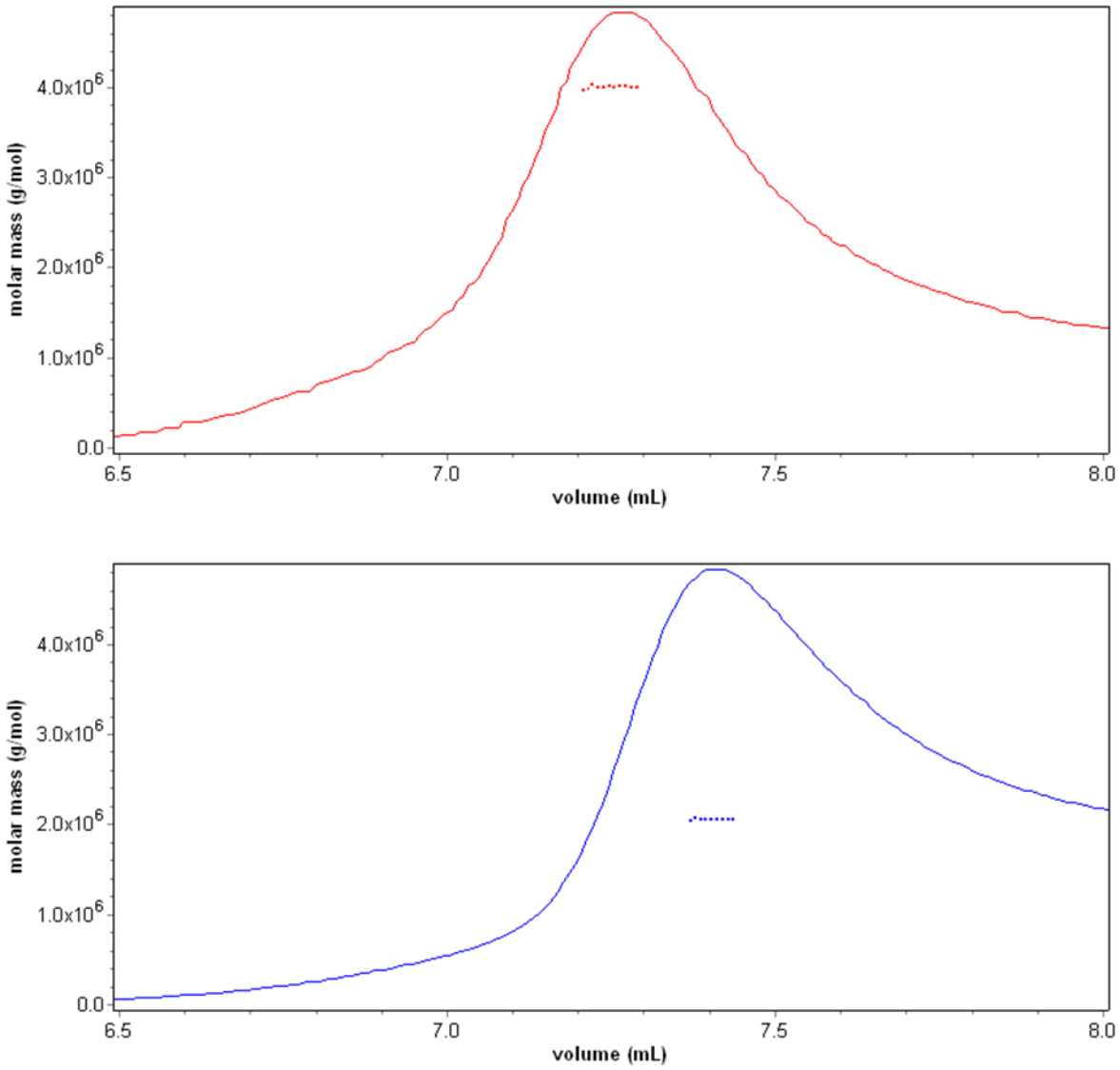


Figure 7. Light scattering experiments performed on the MNK ATP-binding domain without helices (construct 3) fused with GST, purified in absence (red trace) and presence (blue trace) of 0.1% Tween 80. High molecular weight soluble aggregates are formed; MW of the samples prepared in absence (red trace) and in presence (blue trace) of detergent was estimated to 4.013 MDa and 2.05 MDa, respectively. The expected MW of the monomeric protein fused with GST is 61 kDa.

The copper ion translocation cycle performed by WLN and MNK protein involves several stages in which ATP hydrolysis drives translocation of the target ion across cellular membranes. The first step is the binding of copper to the conserved CPC motif in the transmembrane part of the protein, and binding of ATP to the nucleotide-binding N-domain. ATP hydrolysis and phosphorylation of the P-domain, lead to conformational changes and translocation of the target ion. Finally dephosphorylation of the P-domain by the actuator A-domain allows restarting the cycle. Several mutations causing Menkes syndrome and Wilson disease are found in ATP-binding domain and are correlated with malfunctions of ATP binding (N-domain) or acylphosphate intermediate formation (P-domain). At present, not all the molecular events of each domain of ATP7A and ATP7B in the catalytic cycle are clear. Structural and functional characterization (of wild type and mutated) domains will be required for understanding pathogenesis of Wilson and Menkes diseases.

Numerous variations on several different levels of protein production (e.g. cloning, expression and purification) were performed in order to recombinantly express the pure, soluble monomeric ATP-binding domain of MNK and WLN and P-domain of MNK proteins. However, none of these variations was sufficient to avoid protein aggregation either during protein synthesis or purification.

The obtained results for MNK and WND ATP-binding and P-domain clearly suggest that a different approach for construct building and protein expression is necessary to obtain not aggregated protein. The expression of a protein as inclusion bodies or presence of high molecular aggregates indicates that the protein is not folded properly. Wilson and Menkes proteins are composed of several soluble domains and transmembrane channel, it is probable that the correct folding of the ATP-binding and P-domain requires a presence of other subdomain/s (e.g. A-domain). The specific interactions between subdomains could lead to their correct folding.

References

1. Axelsen, K.B. & Palmgren, M.G. Evolution of substrate specificities in the P-type ATPase superfamily. *J. Mol. Evol* **46**, 84-101 (1998).
2. Lutsenko, S. & Kaplan, J.H. Organization of P-type ATPases: significance of structural diversity. *Biochemistry* **34**, 15607-15613 (1995).
3. La Fontaine, S. & Mercer, J.F.B. Trafficking of the copper-ATPases, ATP7A and ATP7B: role in copper homeostasis. *Arch. Biochem. Biophys* **463**, 149-167 (2007).
4. Lutsenko, S., Barnes, N.L., Bartee, M.Y. & Dmitriev, O.Y. Function and Regulation of Human Copper-Transporting ATPases. *Physiol. Rev.* **87**, 1011-1046 (2007).
5. La Fontaine, S., Ackland, M.L. & Mercer, J.F. Mammalian copper-transporting P-type ATPases, ATP7A and ATP7B: Emerging roles. *The International Journal of Biochemistry & Cell Biology* **42**, 206-209 (2010).
6. Voskoboinik, I., Camakaris, J. & Mercer, J.F.B. Understanding the mechanism and function of copper P-type ATPases. *Adv. Protein Chem* **60**, 123-150 (2002).
7. Banci, L. et al. Solution structures of the actuator domain of ATP7A and ATP7B, the Menkes and Wilson disease proteins. *Biochemistry* **48**, 7849-7855 (2009).
8. Barry, A.N., Shinde, U. & Lutsenko, S. Structural organization of human Cu-transporting ATPases: learning from building blocks. *J. Biol. Inorg. Chem* **15**, 47-59 (2010).
9. Voskoboinik, I. & Camakaris, J. Menkes copper-translocating P-type ATPase (ATP7A): biochemical and cell biology properties, and role in Menkes disease. *J. Bioenerg. Biomembr* **34**, 363-371 (2002).
10. Banci, L. et al. An NMR study of the interaction between the human copper(I) chaperone and the second and fifth metal-binding domains of the Menkes protein. *FEBS J* **272**, 865-871 (2005).
11. Banci, L. et al. Solution structure and intermolecular interactions of the third metal-binding domain of ATP7A, the Menkes disease protein. *J. Biol. Chem* **281**, 29141-29147 (2006).
12. Banci, L., Bertini, I., Cantini, F., Rosenzweig, A.C. & Yatsunyk, L.A. Metal binding domains 3 and 4 of the Wilson disease protein: solution structure and interaction with the copper(I) chaperone HAH1. *Biochemistry* **47**, 7423-7429 (2008).
13. Banci, L. et al. The binding mode of ATP revealed by the solution structure of the N-domain of human ATP7A. *J. Biol. Chem* **285**, 2537-2544 (2010).
14. Banci, L. et al. An atomic-level investigation of the disease-causing A629P mutant of the Menkes protein, ATP7A. *J. Mol. Biol* **352**, 409-417 (2005).
15. Prilusky, J. et al. FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435 -3438
16. Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol* **292**, 195-202 (1999).
17. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**, 725-738 (2010).
18. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).
19. Vincentelli, R. et al. High-throughput automated refolding screening of inclusion bodies. *Protein Sci* **13**, 2782-2792 (2004).
20. Sazinsky, M.H., Mandal, A.K., Argüello, J.M. & Rosenzweig, A.C. Structure of the ATP Binding Domain from the Archaeoglobus fulgidus Cu+-ATPase. *Journal of Biological Chemistry* **281**, 11161 -11166 (2006).

Chapter 4- CONCLUSIONS AND PERSPECTIVES

Genome sequencing projects have provided unprecedented number of molecular information that promises to revolutionize our understanding of life and lead to new treatments of its disorders. However, genome sequences alone offer only limited insights into the biochemical pathways that determine protein, cell and tissue function. The vast number of uncharacterized proteins suggests that our knowledge of cellular biochemistry is far from complete. Therefore, it is clear that an integrated approach in the study is essential, in order to obtain information about the protein function. Indeed a large number of different scientific fields like bioinformatics, biochemistry, molecular and cell biology are necessary to fully understand the cellular pathways of life.

My research was focused on cloning, expression and characterization of human proteins involved in i) Fe-S cluster biosynthesis and ii) copper homeostasis.

We demonstrated that CIAPIN1 (anamorsin), a protein connected with apoptosis processes and ISC assembly in cytosol coordinates a [2Fe-2S] cluster itself, and that could be trapped in IMS through an interaction with Mia40, a component of disulfide relay system of the mitochondrial IMS. Specific incorporation of two disulfide bridges performed by Mia40 is independent of the cluster binding ability of anamorsin. Therefore, IMS localized anamorsin could be the connecting component of mitochondrial and cytosolic Fe-S cluster assembly machineries. Since another essential component of the disulfide relay system is the sulfhydryl oxidase ALR, which is as well necessary for cytosolic Fe-S cluster biogenesis, CIAPIN1 could be a link between ISC biogenesis and the disulfide relay system of the IMS. The next step in this project will be focused on the characterization of interaction, at the molecular level, between anamorsin and ALR and recently identified protein Ndor1 involved in cytosolic ISC assembly process

Glutaredoxin 5 is a member of mitochondrial Fe-S cluster assembly machinery that functions in the cluster delivery to the apo-proteins. Our data showed that the protein coordinates a [2Fe-2S] cluster in the glutathione dependent fashion as a monomer. Basing on experimental results we proposed a model in which the cluster is coordinated by two conserved cysteines and two GSH molecules. The protein is now being structurally characterized and the characterization of the interaction with a putative protein partners ISCA1 and ISCA2 is in progress.

In the last part of this PhD thesis I described the trials that I have performed in order to obtain ATP-binding and P- domain of Wilson and Menkes proteins. Numerous variations on several different levels of protein production (e.g. cloning, expression and purification) were performed. However, none of these variations was sufficient to avoid protein

aggregation either during protein synthesis or purification. The aggregation of proteins strongly suggests that they are not properly folded. The most plausible solution could be the expression of full length proteins, this would indeed allow the intramolecular interactions between the sub-domains which are probably essential, on the basis of our results, for the correct folding of the ATP-binding and P- domains.

Acknowledgments

I deeply thank my tutor, Prof. Lucia Banci for help, advice and supervision.

I thank Prof. Ivano Bertini for the opportunity of studying at C.E.R.M.

I also thank Simone Ciofi and Julia Winkelmann, for their kindness and substantial input in coordination of my work and all the support they gave me during last years.

I am very grateful to all former and present members of my laboratory group, especially Tania, Chiara, Mirella, Miguela, Nikolaki, Doros and Joao for all the patients, friendliness and help.

I want to thank Manuele and Valentina for help and answering all my questions, Fabrizzio and Alfonso for the "balcony" group meetings.

Agostino and Pippo thank you for everything.

Chciałbym podziękować mojej ukochanej rodzinie, a w szczególności rodzicom, babci, mojej siostrze i szwagrowi za wsparcie, na które zawsze mogę liczyć. Dziękuję za zrozumienie, cierpliwość i wiarę we mnie. Bez Was nigdy by mi się nie udało.....

Specjalne podziękowania dla wszystkich Gajdów. Dziękuję Seniorom, Juniorom i Karolowi za to, że traktują mnie jak rodzinę.

Misiakina, dziękuję za to, że jesteś.....

Dziękuję wszystkim moim ciociom, wujkom, kuzynkom i kuzynom. Mieć taką rodzinę to samo szczęście.

Serdeczne dzięki dla wszystkich przyjaciół w Polsce za to, że jak wracam chociaż na moment, to czuję się tak, jakbym nigdy nie wyjechał.

Anioł... szkoda, że nie możemy się cieszyć razem.