University of Florence

International Doctorate in Structural Biology

Cycle XX (2005-2008)

The effect of single point mutations on protein structure and stability of SOD1 pathogenic mutants involved in familial ALS

Ph.D. thesis of

Miguela Vieru

Tutor

Prof. Lucia Banci

Coordinator

Prof. Claudio Luchinat

S.S.D CHIM/03

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

…in memory of **George-Valentin Rãgãlie**.

Contents

1

INTRODUCTION

1.1 The role of copper and zinc in biological systems

Metal ions are essential for the biological function of many proteins. Thus, at least one-third of all proteins encoded in the human genome contains metal ions or need metal ions to perform their function [(1)]. Metal ions are used either as a structural component or as a catalytic co-factor [(2)]. The metals are implicated in various biological processes such as electron transfer reactions, oxygen transport, and in a large variety of catalytic processes. Copper and zinc are defined as "micronutrients", these metal ions are essential for the normal growth and reproduction of all living organisms.

In biological systems, Cu ions can exist in two oxidation states: Cu^{1+} (reduced) and Cu^{2+} (oxidized). This redox activity has been harnessed for catalysis by a plethora of enzymes whose activities are critical to a broad range of cellular biochemical and regulatory functions in organisms from microbes to plants and mammals. In humans, copper is necessary for the healthy development of connective tissue, nerve coverings, and bones. It is also involved in both iron and energy metabolism. The coordination chemistry of ions is often distinct, with Cu^{1+} preferring sulfur donor ligands such as cysteine or methionine, whereas Cu^{2+} prefers nitrogen donors such as histidine or oxygen donors such as glutamate or aspartate[(3,4)]. Given that Cu ligands can exist in distinct combinations and locations in proteins, it has been challenging to identify the constellation of natural Cu-binding proteins based on the coding capacity of many sequenced genomes. Recent bioinformatic studies estimate the Cu proteome to be less than 1% of the total proteome encoded by bacterial, archaeal and eukaryotic genomes[(5)] allowing the identification of many putative Cu-binding proteins with unknown function.

In living organisms, zinc is the second most abundant transition metal ion after iron. In contrast to other transition metal ions, such as copper and iron, Zn^{2+} does not undergo redox reactions thanks to its filled *d* shell. Zinc is a cofactor of many proteins and is indispensable for their catalytic activity and/or structural stability. It is also a ubiquitous component of enzymes involved

in transcription and of the Zn finger proteins, which are involved in regulation of gene [6]. Besides its involvement in protein structure and function [7, 8], the interaction of zinc with lipids contributes to the regulation of membrane fluidity [9] and its interaction with nucleic acids helps to prevent deleterious radical reactions [10]. Interacting mainly with amino-acid side chains and occasionally with non-protein ligands, zinc is generally in a tetrahedral coordination with all the four ligands from the protein in structural Zn sites. The zinc ion is usually coordinated by histidines and occasionally by cysteine thiols or by carboxylate bearing residues. By contrast, in the catalytic Zn sites, water or inhibitor usually occupies the fourth ligation site.

Human copper and/or zinc enzymes include:

i) Ceruloplasmin (CP)

Ceruloplasmin is the primary copper carrier in serum. It is also a potent ferroxidase, capable of oxidizing Fe(II) to Fe(III) in the presence of molecular oxygen and accelerating the binding of iron by apotransferrin [11]. Oxidation of iron is important for its binding to transferrin. The latter oxidation step is critical for iron release from cells. In the case of copper deficiency, iron remains in the liver and upon addition of copper containing ceruloplasmin, iron is immediately released from liver into the blood [12, 13].

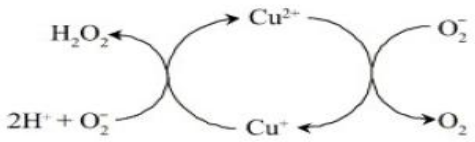
ii) Cytochrome *c* Oxidase (CcO)

It is the terminal enzyme in the respiratory chain, whose function is the conversion of O_2 to water. This enzyme anchors on the inner membrane of mitochondria. A number of metal cofactors are needed for its function. The metal binding sites include two copper specific sites, Cu_A in Cox2 subunit and Cu_B in Cox1 subunit [14]. Cu_A is a binuclear, mixed valent center (one Cu^{2+} state and one Cu^{1+} state), located in the inner membrane space (IMS) of mitochondria [15]. Cu_A site receives one electron from reduced cytochrome *c*, thus switching from the Cu^{2+}form to Cu^{1+}form. The electron is further transferred to the Cu_B center via two heme centers. Cu_B is located in the inner membrane of mitochondria, where O_2 is reduced. Since the formation of two water molecule from O_2 consumes four protons, a proton gradient is generated between IMS and the matrix. The energy stored by the

proton gradient is subsequently utilized for ATP synthesis. The copper co-factors in this enzyme are critical for electron transfer and the overall catalytic reactions.

iii) Cu/Zn superoxide dismutase

Cu/ZnSOD (SOD1) is a member of the SOD enzyme family, which catalyzes the disproportionation reaction turning two molecules of superoxide into one molecule of O_2 and one of H_2O_2 [16]. The mechanism can be described as a cycling 'ping-pong':



SOD1 is a homodimer which is largely localized in intracellular cytoplasmic spaces, but it is present also in the IMS [16].

iv) Metallothioneins (MT)

The primary purpose of this group of proteins is to sequester metal ions when they are present in excess [17]. Metallothioneins can bind Cu, Zn, Cd, Hg, Ag, or Ni [17]. However, Cu has the highest apparent affinity and can displace other metal ions [18]. It has been reported that metallothioneins with bound copper possess some SOD activity [18].

v) Lysyl oxidase

Lysyl oxidase is a copper-dependent amine oxidase that plays a critical role in the biogenesis of connective tissue matrices by cross linking the extra cellular matrix proteins, collagen and elastin [16, 19]. In Menkes disease, connective tissue development is altered due to failure of lysyl oxidase [20, 21]. If the function of lysyl oxidase is impaired, patients show hyperelastic skin, hernias, and aortic aneurysms among other connective tissue related abnormalities [19, 22].

vi) Carboxypeptidases

Carboxypeptidase A (CPA) is a digestive enzyme that favours C-terminal cleavage of residues with large aromatic side chain while carboxypeptidase B (CPB) favours N-terminal

residues. The zinc cofactor in CPA has a tetrahedral geometry and is located at the bottom of a cavity. It is coordinated by two histidine ligands, a bidentate glutamate and one water molecule which is hydrogen bonded to a serine and a glutamate residues. In the presence of substrate, the bound carboxylate shift to an unidentate coordination. The enzyme can also hydrolyse esters.

vii) Liver alcohol dehydrogenase

Its function is the oxidative conversion of an alcohol to an aldehyde. The zinc ion is coordinated in a tetrahedral geometry by two cysteine and one histidine residues with a bound hydroxide. The Zn^{2+} must expand its coordination number to five after coordinating the alcohol which was first deprotonated by the bound hydroxide.

Disruption of copper or zinc homeostasis can have effects on all of these enzymes and their activities as well as the development of toxic levels of the metal and reactive oxygen species.

1.2 Copper trafficking in human cell

In the most aerobic organisms, copper serves as a catalytic and structural cofactor for enzymes that function in energy generation, iron acquisition, oxygen transport, cellular metabolism, peptide hormone maturation, blood clotting, signal transduction and a host of other processes. It is also toxic for both eukaryotes and prokaryotes. Free copper ions contribute to produce reactive oxygen species (ROS). Furthermore, oxidative stress carried out by ROS could be amplified by copper reactivity, leading to the impairment of essential molecules such as lipids, proteins and DNA. Thus, biological systems have developed various copper homeostasis systems, ranging from copper storage to translocation, controlled by specific proteins. Free copper concentration in cell is indeed lower than 10^{-18}M [(23)].

A model summarizing how Cu^{1+} is taken up by intestinal epithelial cells (IECs), routed for incorporation into Cu-dependent proteins and mobilized across the basolateral membrane into the peripheral circulation is shown in Fig. 1.

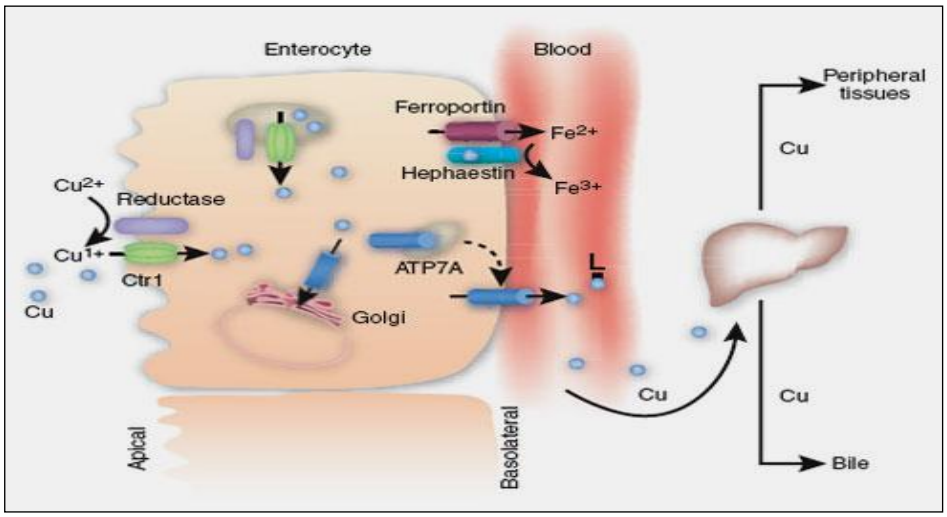


Fig.1: *Model for intestinal Cu absorption and peripheral distribution.* In this model of Cu absorption and distribution, the Ctr1 high-affinity Cu^{1+} transporter is shown both at the apical membrane of IECs and in intracellular vesicles. A putative metalloreductase reduces Cu^{2+} to Cu^{1+} for import by Ctr1. Ctr1 and a metalloreductase may also function in the mobilization of Cu^{1+} from endosomal compartments. Cu^{1+} is pumped into the secretory compartment for loading onto Cu-dependent enzymes, or out from the basolateral membrane by the Cu^{1+}-transporting P-type ATPase ATP7A. In the bloodstream Cu is transported via the portal vein to the liver, a central organ of Cu homeostasis. Cu is transported to the peripheral tissues via the systemic circulation in a complex with one or more ligands (L) that have not been identified. (Reprinted from Byung-Eun Kim, Tracy Nevitt & Dennis J Thiele, Nature Chemical Biology 4, 176-185 (2008)).

Once Cu crosses the intestinal lumen, it is transported into the portal circulation where it is delivered to the liver, a central organ for Cu homeostasis, mobilized into the peripheral circulation or secreted into the bile for excretion. In the liver, Cu is incorporated into, among other proteins, the secreted multi-Cu ferroxidase ceruloplasmin, which catalyzes the oxidation of ferrous iron Fe^{2+} to ferric iron Fe^{3+} in serum, thereby facilitating its loading onto transferrin for systemic iron distribution [(24)]. Although more than 90% of the serum Cu is bound by ceruloplasmin, mice lacking ceruloplasmin have no apparent defect in Cu absorption or distribution [(25)]. Therefore, given the ability of free Cu to generate reactive oxygen species (ROS), other as-yet poorly defined physiological Cu-binding proteins or ligands must facilitate peripheral Cu distribution. Copper is transported inside the eukaryotic cell by Ctr1, an integral membrane protein that is structurally and

functionally conserved from yeast to humans [26]. A single Ctr1 polypeptide harbors three transmembrane domains, a methionine-rich N terminus, a cysteine-histidine cluster in the C terminus and an MX_3M motif in the second transmembrane domain. Mutagenesis and functional studies demonstrate that, while the MX_3M motif in the second transmembrane domain is essential for function in Cu uptake, the methionine-rich ectodomain is dispensable except under Cu-limiting conditions [27]. Different studies indicate that Ctr1 is present in the membrane as a homotrimer that is likely to form a pore between the subunit interfaces [28, 29].

1.2 a Copper deliveries to mitochondria

Recent studies demonstrate that steady-state Cu levels in the mitochondrial matrix are nearly an order of magnitude above that predicted to be required for the activation of the abundant mitochondrial Cu-dependent enzyme cytochrome oxidase [30] (Fig.2).

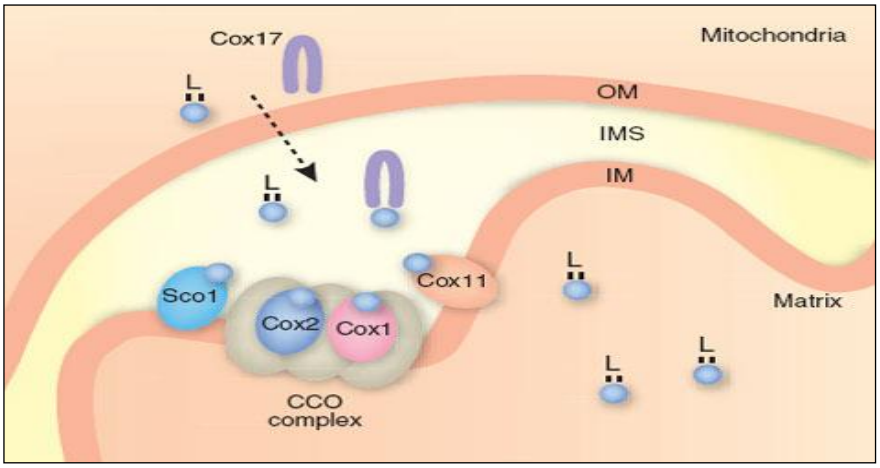


Fig.2: *Cu loading of mitochondrial cytochrome oxidase*. Mitochondrial Cu is thought to be shuttled from the cytosol via as-yet uncharacterized ligands (denoted L). In the intermembrane space (IMS), Cu^{1+} is bound by Cox17 and delivered either to Sco1, which transfers the Cu to the Cox2 subunit, or to Cox11, which delivers Cu to the Cox1 subunit of cytochrome oxidase (CCO). (Reprinted from Byung-Eun Kim, Tracy Nevitt & Dennis J Thiele, Nature Chemical Biology 4, 176-185 (2008)).

Several proteins are involved in the correct copper binding of CcO, including Cox17, Sco1, Sco2, Cox11, Cox19 and Cox23. Cox17, which is only a 69-residue protein, acts as the copper chaperone in mitochondria. The overall fold of Cox17 is composed of two helixes and a flexible N-terminus [31]. *In vitro* studies showed that Cox17 is able to bind up to 4 copper ions by six Cys residues [32]. Cox17 is encoded by nucleic DNA and leads to be imported into IMS by the TOM

protein, located in the outer membrane of mitochondria. In cytosol, all six Cys residues of Cox17 are in the reduced state, since a large amount of reduced glutathione is present. In mitochondria, four Cys residues of Cox17 are specifically oxidized by Mia40 to form two disulphide bonds. The remaining two reduced Cys residues are able to bind one copper ion. Because Cox17 is imported to mitochondria in the apo form, the mechanism by which copper is transported into mitochondria is still unknown.

1.3 Copper and diseases

Although copper is an essential trace metal, copper in excess of cellular needs is toxic. The alteration of copper homeostasis leads to errors of metabolism and damage to cells, resulting in pathological conditions. Such conditions are well exemplified by two inherited diseases of copper metabolism in humans, Menkes and Wilson syndromes: in both cases, the impairment by mutation of two homologous copper transport ATPases, ATP7A and ATP7B, which are selectively expressed in different tissues, leads either to the decrease or to the overload of copper in the cells.

Menkes disease (MD) is a fatal X-linked copper deficiency disorder. About one in 200,000 males are born with this devastating genetic disorder. Its symptoms include profound mental retardation, and connective tissue (collagen) abnormalities that result in soft bones and cartilage and weakened artery walls [33]. Mutations on the ATP7A gene (MNK) cause the patho-physiology of this disease. The latter defect creates a systemic copper deficiency since MNK is primarily responsible for absorbing copper through diet. The copper content is particularly low in the brain of the patients with mutated ATP7A, because this protein is also expressed at the blood-brain barrier [34, 35]. The current treatment of Menkes disease is administration of copper-histidine to the patient. Individuals affected by MD usually die in early childhood.

Wilson's disease (WD) is an autosomal recessive hereditary disease, with an incidence of about 1 in 30,000 in most parts of the world and a male preponderance. Symptoms usually appear

around the ages of 10 to 21 years. The mutant form of ATP7B expressed in people with Wilson's disease inhibits the release of copper into bile, leading to liver failure as a result of the very high concentrations of copper accumulation in this organ. The disease is treated with lifelong use of chelating agents such as D-penicillamine or trienine hydrochloride, drugs that help remove copper from tissue.

Disorders due to respiratory chain deficiency also determine fatal diseases, leading to death at an early age as well. Isolated CcO deficiency represents one of the most commonly recognized causes of respiratory chain defects in humans associated with a wide spectrum of clinical phenotypes, primarily affecting those organs with high-energy demand such as brain, skeletal, muscle and heart [(36)]. A number of mutations in genes encoding CcO assembly factors (Sco1, Sco2, Cox15, Cox10, Surf) have been described as a frequent cause of CcO deficiency and have been assigned with specific clinical symptoms.

Moreover, the imbalance of copper homeostasis have been linked to some neurodegenerative diseases as Alzheimer's disease (AD) [(37)], Creutzfeldt-Jakob disease, Parkinson's disease (PD) [(38)] and Amyotrophic Lateral Sclerosis disease (ALS) [(39)]. Aberrant reactivity of copper is a major source of reactive oxygen species production, which in turn is responsible for the higher levels of oxidative stress observed in these diseases. These disorders may all be classified as "conformational" in that they all show, as hallmarks, proteins with altered conformations, which precipitate, form aggregates, and may be responsible of cell damage. Direct interaction between copper and the proteins that are the pathological hallmark of these diseases has been reported, and thus, redox activity of the metal may also be involved in the process of protein misfolding. The molecular mechanisms leading to neurodegeneration in all these diseases are still unknown. However, metal-mediated oxidative stress could induce a cascade of events, including mitochondrial dysfunction that may be responsible for cell death.

1.4 Zinc and diseases

The biological functions of zinc in numerous cellular processes are based on its occurrence in over one thousand enzymes as a catalytic metal and in at least the same number of proteins as a structural metal. These pleiotropic actions require tight regulation in allocating zinc at the right time to the correct protein(s). Metalloregulatory proteins perform this function. Thus, both the availability of zinc ions from the diet and the proper functioning of the proteins that handle zinc are critical for maintaining good health. Mutations in the genes coding for these proteins are the basis for inborn errors of zinc metabolism, as has now been recognized in several diseases. The diseases include acrodermatitis enteropathica, a rare disorder with a mutation in the Zip4 (SLC39A4) importer gene and resulting in severe zinc deficiency [40, 41], and transient neonatal zinc deficiency with a mutation in the ZnT-2 (SLC30A2) exporter gene, resulting in low zinc concentrations in the mother's milk [42]. Furthermore, the risk of type II diabetes in several human populations has been associated with a single nucleotide polymorphism (SNP) resulting in a Trp325Arg substitution in the pancreatic β-cell-specific zinc transporters protein ZnT-8 (SLC30A8), which provides zinc for insulin maturation and/or storage [43-45]. This gene variation (W genotype) also confers resistance against posttransplantational diabetes mellitus, a major complication that may develop after renal transplants [46]. Last but not least, ZnT-8 is a major autoantigen in type 1 diabetes [47].

Alzheimer's disease (AD) is a neurodegenerative disease characterized by deposits of extracellular amyloid plaques and intracellular neurofibrillary tangles in the brain. The principal components of the plaques are fibrils formed of 40–42 residues long amyloid β peptides (Aβ) derived from cleavage of amyloid precursor protein (APP) by β- and γ-secretases [48-50]. Aβ aggregation is an established pathogenic mechanism of AD and therefore factors influencing the Aβ aggregation are of high interest. It has been demonstrated that the interaction of Aβ peptides with zinc [51-56] and copper [57, 58] ions may induce the aggregation of Aβ peptides and these ligands may act as seeding factors in the formation of amyloid plaques. Indeed, the levels of Zn^{2+} and Cu^{2+} are

elevated in the amyloid deposits [59] and the homeostasis of these metals is mis-regulated in the AD brain [58].

1.5 Superoxide dismutases

Reactive oxygen species are frequently proposed to exert a toxic effect on cells via oxidation of a wide range of biomolecules including DNA, proteins, and lipids. Oxidative modifications of these biomolecules often lower enzymatic activities and/or induce mutations in DNA, and can lead to cell death through necrosis or apoptosis [60]. Hydrogen peroxide (H_2O_2) and hydroxyl radical ($HO^{\cdot}$) exhibit high redox potential and have been shown to damage biomolecules *in vitro* and *in vivo*. One of the cellular antioxidant systems are the superoxide dismutases (SODs). These proteins protect redox-sensitive cellular machinery from damage by catalyzing the dismutation of superoxide anion radical ($O_2^{\cdot-}$) into oxygen (O_2) and hydrogen peroxide (H_2O_2). The activity of superoxide dismutase relies upon a specific redox active metal ion, and depending on the SOD molecules, this could either be a manganese, iron, copper or nickel ion. A generic mechanism for the metalloenzyme-dependent dismutation steps is described below, where M is Cu [(n=1)], Mn [(n=2)]; Fe [(n=2)]; Ni [(n=2)], the oxidation state of the metal ion oscillates between n and n+1:

$$M^{(n+1)+}\text{–SOD} + O_2^{-} \leftrightarrow M^{n+} - \text{SOD} + O_2$$

$$\underline{M^{n+} - \text{SOD} + O_2^{-} + 2H^{+} \leftrightarrow M^{(n+1)+} - \text{SOD} + H_2O_2}$$

$$2\ O_2^{-} + 2H^{+} \leftrightarrow O_2 + H_2O_2$$

The metal cofactors catalyze both a one-electron oxidation (see first step) and a one-electron reduction (second step) separate superoxide anions to give the overall disproportionation reaction. These reactions typically require no external source of redox equivalents and are thus self-contained components of the antioxidant machinery.

Superoxide dismutases are typically soluble secreted or cytosolic proteins but are also found in a number of sub-cellular compartments such as the cell envelope of gram-negative bacteria or the

mitochondria of eukaryotic cells, as well as the extra-cellular milieu. Several common forms of SOD exist: they are proteins co-factored with copper and zinc (Cu,ZnSOD) (SOD1) or manganese, iron, (Mn,FeSOD)(SOD2) or nickel. While there is no similarity in sequence or structure between SOD1 and SOD2 families, these metalloproteins do share some intriguing properties. For instance, both are quite stable relative to most proteins found in mesophilic organisms. Members of these families of enzymes are found in prokaryotes and eukaryotes. The SOD1 and SOD2 are the only forms found in yeast and mammals. Nickel and iron SODs are found in different prokaryotes.

SOD1, in mammals and *S. cerevisiae*, is mainly localized in cytosol and in the intermembrane space of mitochondria [61-63]; it has also been identified in the nucleus, peroxisomes, and lysosomes [64]. Mammalian SOD1 is highly expressed in the liver and kidney [65] and is also abundant in motor neurons [66]. SOD2 is located in the mitochondrial matrix [67] in close proximity to a primary endogenous source of superoxide, the mitochondrial respiratory chain. In humans, three forms of superoxide dismutase are present: SOD1 is located in the cytoplasm and intermembrane space of mitochondria, SOD2 in the mitochondria and SOD3 is typically made in vascular smooth muscle cells and secreted into the extracellular environment where it binds to extracellular matrix and endothelial surface components[68].

1.6 Copper, Zinc Superoxide dismutases (SOD1)

Copper, zinc superodixe dismutase (SOD1) was earlier known as erythrocuprein, the major copper containing protein in erythrocytes [69], its enzymatic function was discovered in 1969 by McCord and Fridovich who published that ‘erythrocuprein’ dismutes superoxide[70]. SOD1 is found in almost all eukaryotic cells and very few prokaryotes. Structural studies reveal that all eukaryotic SOD1s are homodimeric proteins that contain one copper, one zinc, and one intrasubunit disulfide bond per monomer. A considerable sequence homology is present among SOD1 proteins from various species so far examined, within the mammalian species, human, rat, pig, and horse the

homology is around 80%. In 1982 the X-ray structure of the oxidized form of SOD1 for the bovine enzyme was published [71, 72]. Nowadays, several other structures are available [73, 74]. Solution structures of monomeric and dimeric SOD1 were also solved in our laboratory [75, 76] (Fig.3).

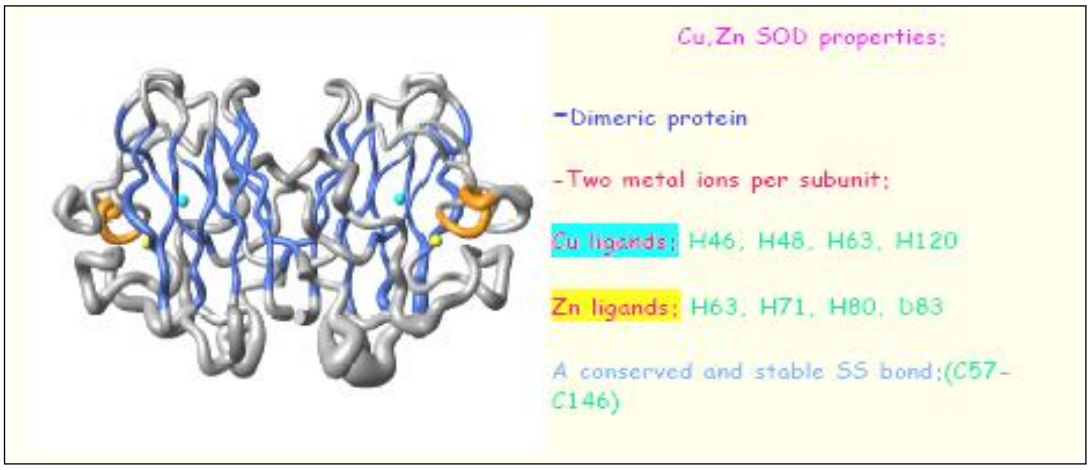


Fig.3: *Tube representation of the family of 30 conformers of human reduced native SOD1.* Elements of secondary structure are highlighted (blue: β structure, orange: α structure). The Cu ion is shown in cyan, the Zn ion in yellow. *(*Reprinted from Lucia Banci, Ivano Bertini, Fiorenza Cramaro, Rebecca Del Conte and Maria Silvia Viezzoli, *Eur J Biochem*. 2002*).*

The human SOD1 is constituted by two identical subunits bearing the same amino acidic sequence of about 150 residues. Within each subunit the polypeptide chain is folded in a characteristic eight stranded β-barrel. The latter is formed by two four-stranded antiparallel β-sheets which are connected by turns and loops in a Greek key β-barrel motif (Fig. 4). The eight β-strands forming the walls of the β-barrel display an overall right handed twist. The first sheet is formed by strands 1, 2, 3 and 6 and the second by strands 4, 5, 7 and 8. The loops connecting the secondary structure elements can be divided in two groups: the odd loops are located on the opposite side of the barrel with respect to regions involved in the subunit-subunit interface, while the even loops are in part located at the subunit-subunit interface.

Inspection of the SOD1 structure reveals that loop IV, containing, Cys57, which forms the disulfide bridge with residue Cys146 in strand β8, can influence the conformation of the catalytically important residue, Arg143, through a hydrogen-bonding network [77]. Portions of this loop also contribute to the dimer interface [78], leading to the possibility that the disulfide bond

influences the protein dimerization and thereby the SOD1 quaternary structure. This intra-subunit disulfide bond is stable and conserved among all SOD1 structures published to date. Crystallographic studies of SOD1 suggest a structural role for the disulfide in guiding the substrate into the active site [79]. Furthermore, the effect of the disulfide on dismutase activity and quaternary structure is not clear, even though altered or aggregated forms of the protein are thought to be at the center of a familial form of amyotrophic lateral sclerosis (fALS) [80]. The two monomers of SOD1 are held together by hydrophobic interactions. The contact surface between the monomers comprises the N terminus, strand β1, the C terminus, strand β8, the two loop regions: loop IV and loop VI. The extension of the contact region explains the high stability of the dimer to thermal and chemical denaturation. SOD1 is one of the most thermally stable enzymes known in mesophilic organisms. Dismutase activity declines at 80 °C with a corresponding melting temperature, Tm, above 90 °C [81].

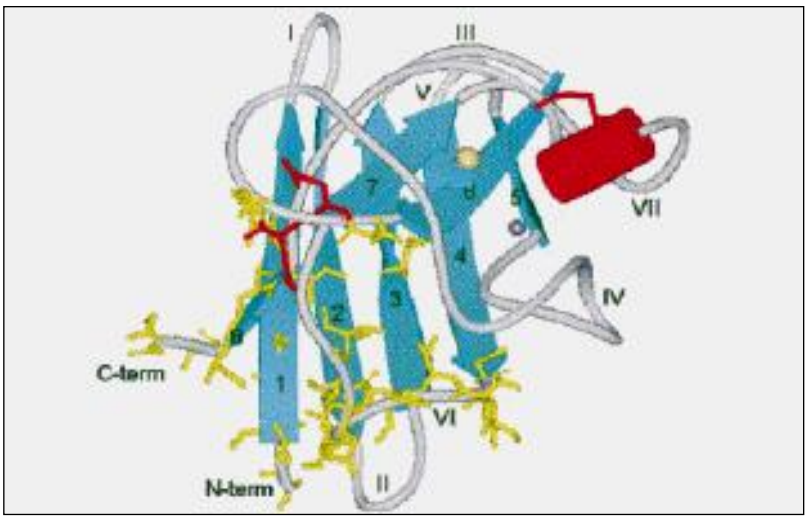


Fig.4: *Schematic view of the Q133M2SOD structure displaying the secondary structure* elements numbered by Getzoff et al., Proteins. 1989;5(4):322-36. The side - chains of the residues homologous to those involved in the subunit- - subunit interface in the WT dimer are shown as yellow sticks. The mutated residues E50, E51, Q133 are shown as red sticks. Orange, yellow and grey spheres of arbitrary radius represent copper and zinc ions. (Reprinted from Ferraroni et.al Journal of Molecular Biology, 1999).

The protein is stable in the presence of strong denaturants, and enzymatic activity is observed even in 4% SDS or 8 M urea [82]. SOD1, in stable monomer form, has been obtained by substituting specific hydrophobic residues at the subunit–subunit interface with hydrophilic ones. In

particular, Phe50 and Gly51 have been substituted by two Glu residues, yielding a soluble single subunit. This mutant displays an activity of about 10% respects to the wild–type human SOD1 [83]. Therefore, to partially restore the activity, Glu133 has been neutralized by substituting it with Gln [84]. This particular monomeric Cu,ZnSOD is referred as F50E/G51E/E133Q monomeric triple mutant or CuZnQ133M2SOD mutant. Each SOD1 monomer contains one Cu(II) and one Zn(II) ion. The Zn(II) ion is tetrahedrally coordinated to three histidyl imidazoles (His63, His71, His80) and to an aspartyl carboxylate group (Asp83). The Cu(II) ion, in the oxidized enzyme, is coordinated by four histidines (His46, His48, His63, His120) and, weakly, by a water molecule in a distorted square planar pyramid. The water molecule occupies the apical position toward the opening of the cavity. The two metal ions are bridged, in the oxidized enzyme, by His63 that keeps them apart of about 6 Å. His63 binds Zn(II) ion and Cu(II) ion through the $N\delta_1$ and $N\varepsilon_2$ atoms, respectively. In the reduced Cu(I) form, a trigonal planar arrangement is obtained through the migration of the copper ion of about 1 Å which is accompanied by protonation of the histidine ligand. As a consequence, histidine does not coordinate the Cu ion (Fig. 5).

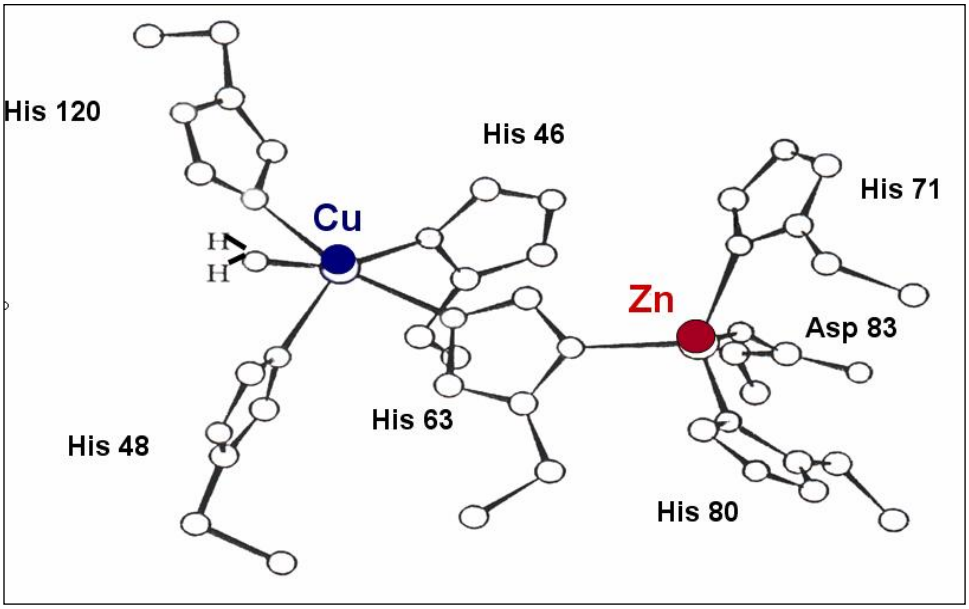


Fig.5: *Overall view of the metal coordination in the active site of SOD1.*

In SOD1, the zinc is completely buried in the protein and its role appears to be primarily structural [71]. Removal of the Zn(II) ion, results in a diminished thermal stability, i.e., the zinc-

depleted protein denaturise at a lower temperature [85]. However, in conditions where the copper ion remains bound to the copper site, the removal of Zn(II) ion does not significantly reduce SOD1 activity. The copper ion has a solvent-exposed surface of 5.2 Å and lies at the bottom of a narrow channel which is large enough to admit only water, small anions and similarly small ligands. Its role in the enzyme appears to be mainly catalytic (Fig. 6). The redox properties of copper ion allow the disproportionation of superoxide anion to hydrogen peroxide and dioxygen. The insertion of copper(II) into SOD1 requires a copper metallochaperone protein, named CCS (the copper chaperone for superoxide dismutese) which was first identified in the yeast strain *Saccharomyces cerevisiae* [86] and has emerged as an important posttranslational regulator of SOD1 structure, function, and physiology (Fig. 7).

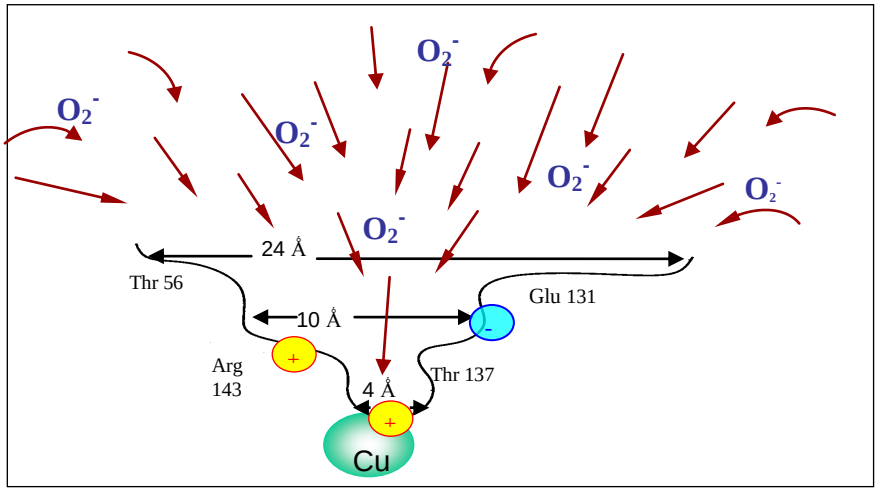


Fig.6: *Schematic diagram of a cross section of the active site channel in SOD1.*

Yeast cells lacking the CCS gene, known as LYS7, express a form of SOD1 protein that is essentially copper depleted [87] but contains a single atom of zinc per subunit [88, 89]. CCS proteins have been also identified in various species including humans [90], rodents [91], insects [92], and plants [93]. Moreover, cell biology studies have shown that CCS largely co-localize with SOD1, it is less abundant than SOD1 in mammalian [94] and yeast cells [95]. Mice with disruptions in CCS exhibit

marked reductions in SOD1 activity, stressing the conserved requirement for CCS in activation of eukaryotic SOD1 [96].

CCS is highly specific for SOD1 and cannot deliver a copper ion to other proteins, such as cytochrome *c* oxidase or copper transport ATPase, that have different metallochaperones. CCS proteins are the largest copper chaperon known to date, with molecular masses of 30-32 kDa versus 7–8 kDa of ATX1 and COX 17 (other copper chaperons). While the ATX1 and COX17 are single-domain proteins, CCS is represented by three functionally distinct protein domains [97]. The N-terminus domain I has a structure similar to ATX1, including the "CXXC" copper-binding motif. Based on this homology, it was initially assumed that this domain is critic for the copper-transfer reaction. However, a CCS molecule lacking this domain can still insert copper into SOD1 *in vivo* but only when copper is plentiful [96]. It therefore appears that this N terminal-domain work to maximize CCS activity specifically under copper starvation conditions.

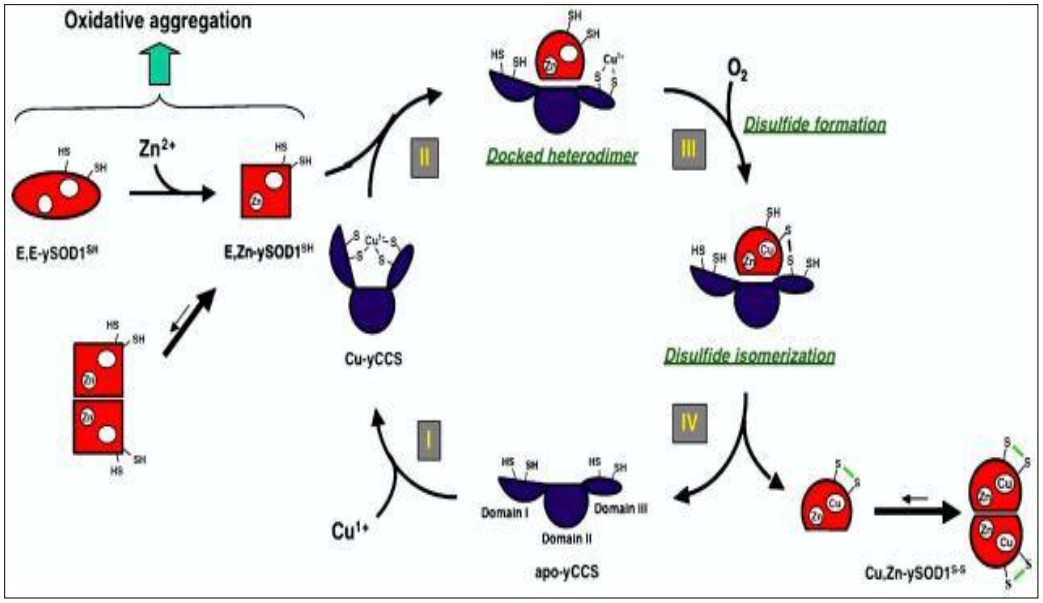


Fig.7*: Proposed mechanism of copper insertion into SOD1 by CCS*. CCS acquires copper through unknown routes and then docks with a disulfide reduced form of SOD1 (steps I and II). This complex is inert to further reaction unless exposed to oxygen or superoxide (step III), at which point a disulfide-linked heterodimeric intermediate forms. In the case of the WT protein, this complex undergoes disulfide isomerization to an intramolecular disulfide in SOD1 (step IV). *(Reprint from Culotta, VC; Yang, M; and O'Halloran, TV; Biochim Biophys Acta 2006).*

Domain II in the center of CCS is homologous to SOD1 [96]. In the case of human CCS, this noted homology is so strong that a single mutation in the copper site of Domain II activates a superoxide scavenging activity of CCS. This domain does not appear to directly participate in copper transfer but physically interacts with SOD1 to secure the enzyme during copper insertion [98]. The SOD1 usually exist as homodimer of two identical subunits and it has been proposed that CCS docks with SOD1 to form either a transient heterodimer or heterotetramer as a prerequisite to copper transfer. In the case of yeast proteins, a CCS-SOD1 heterodimer has been resolved by X- ray crystallographic analyses [99]. As predicted, the complex is stabilised by interactions at the dimer interface of SOD1 and the corresponding homologous region of CCS domain II. The C terminal domain III of CCS is quite small (30 amino acids), yet it is extremely crucial for activating SOD *in vivo*. This domain is highly conserved among CCS molecules from diverse species and includes an invariant CXC motif that can bind copper [96]. Domain III is disordered in the crystal structure; however, it is predicted to lie in the vicinity of the N terminal domain I. It has been proposed that the C-terminal Domain III in concert with the N-terminal domain I of CCS directly inserts copper into the active site of SOD1 [97].

The precise mechanism of copper transfer is not completely understood. Questions as how does copper move from a sulphur-coordination environment (CCS) at nitrogen-copper site (SOD1) remain to be elucidate. Additionally, little is known on how SOD1 acquires zinc. It has been proposed that CCS may also function as a "Zn chaperone" protein given that human CCS can bind a zinc ion in Domain II and possibly Atx1-like Domain I. However, these are the challenges of SOD1 metalation that remain to be elucidated.

1.7 SOD1 and amyotrophic lateral sclerosis disease

Amyotrophic lateral sclerosis (ALS) is a late-onset progressive neurodegenerative disease affecting motor neurons. It has been described for the first time by the neurobiologist and clinician

Jean-Martin Charcot [100] with premature degeneration and death of upper and lower motor neurons provoking fatal paralysis as its salient clinical features. The name of the disease derives from Charcot's observation, in the lateral portions of the spinal cord, of a distinct ''myelin pallor'' representing degeneration and loss of the axons of upper motor neurons as they descend from the brain to connect directly or indirectly onto the lower motor neurons within the spinal cord. Initially known as Charcot's sclerosis, it is now more familiarly known, in the United States, as Lou Gehrig's disease. Onset of disease is in midlife (usually between age 45 and 60), with a typical disease course of one to five years. Consideration of incidence (frequency of new cases per year) and prevalence (the proportion of affected individuals in the population; 1–2 and 4–6 per 100.000, respectively) understates the impact of ALS, with the lifetime risk at about 1 in 1000 [101]. Most incidences (90%) of ALS are sporadic, that is, without an obvious genetic component. Approximately 10% are inherited in a dominant manner and referred to as familial ALS (fALS) (see below).

The pathogenesis of ALS was considered quite obscure for more than a century after its first description. Anatomic and morphologic examination of post-mortem samples conveyed the notion of selective vulnerability of motor neurons mainly in the anterior of the spinal cord, associated with the activation of astrocytes and microglia and muscle atrophy. Biochemical studies showed alterations in the cerebrospinal fluid of patients and in samples from spinal cord and brain. Nowadays, ALS is considered both a multi-factorial and multi-systemic disease involving a combination of mechanisms that affects several cell types, similarly to other disorders such as spinal cord injury, Alzheimer's disease, Huntington's disease and Parkinson disease. Several loci have been identified by genetic analyses as linked to ALS but only few of them have been assigned to specific genes. These includes genes coding for alsin [102], senataxin [103], dynactin [104], VAMP (vesicle-associated membrane protein)-associated protein B (VAPB) [105] and angiogenin [106]. These genes encodes proteins involved in a wide range of cellular processes, from oxidation to axonal transport, RNA processing, DNA repair, vesicular transport and angiogenesis.

A major step forward in understanding of molecular mechanism underlying ALS was provided in 1993 by the observation that mutations in the gene coding for the antioxidant enzyme Cu,Zn SOD1 are carried by one fifth of fALS patients (20% of all the fALS forms, corresponding to 2% of all ALS cases) [107]. Since the first time a connection was established between SOD1 mutations and fALS the number of known mutations has increased to more than 100 (http://www.alsod.org). Clinical evidences showed that SOD1 is the only gene among the ones connected to ALS, that, when mutated in almost any way, leads specifically to classic ALS and not to any other motor neuron or neurodegenerative disorder.

Sporadic and familial ALS produce similar pathological hallmarks, including progressive muscle weakness, atrophy, and spasticity, each of which reflects the degeneration and death of upper and lower motor neurons. Swelling of the perikarya and proximal axons is observed. It is also observed the accumulation of phosphorylated neurofilaments, Bunina bodies, Lewy body-like inclusions, and the deposition of inclusions (spheroids) and strands of ubiquitinated materials in these axons.

Two hypotheses have been proposed to explain the gain of toxic function of ALS mutant SOD1 proteins: the oxidative damage hypothesis and the oligomerization hypothesis [108]. The first one maintains that fALS mutants acquire one or more toxic properties, catalysing reactions with oxidants such as peroxynitrite [87] and possibly hydrogen peroxide [109]. These reactive nitrogen and oxygen species cause toxicity spoiling proteins, nucleic acids and lipids. The oxidative damage hypothesis seems to require copper (or some other redox active metal ion) bound to the fALS mutant SOD1 protein to promote the oxidative reaction, therefore the role of copper mediated oxidative damage in neuronal SOD1-mediated toxicity has been questioned. Indeed, studies on transgenic mice expressing a mutant SOD1 characterized by the disruption of the active site through artificial mutation of four histidines showed a motor neuron disease similar in clinical and pathological appearance to other mouse models of SOD1-linked fALS [110]. On the other hand, the oligomerization hypothesis maintains that mutant SOD1 proteins are, or became, misfolded and

consequently oligomerize into increasingly high-molecular-mass species that ultimately lead to the death of motor neurons [111] (Fig.8). This hypothesis is supported by the finding of proteinaceous inclusions rich mutant SOD1 in tissues from ALS patients, ALS-SOD1 transgenic mice, and in cell culture model systems [112]. These finding suggest that SOD1-associated fALS is a protein conformational disorder, similar to Alzheimer's disease, Parkinson's disease, Huntington's disease, transmittable spongiform encephalopathies, and other neurodegenerative diseases in which aggregation and deposition of abnormal proteins are found [113]. A common trait of many neurodegenerative diseases including ALS is also damage to mithocondria that contributes to the degenerative phenotype [114]. Indications that mithocondrial dysfunction may participate in the pathogenesis of ALS came from studies conducted essentially in post-mortem tissues from ALS patients [115].

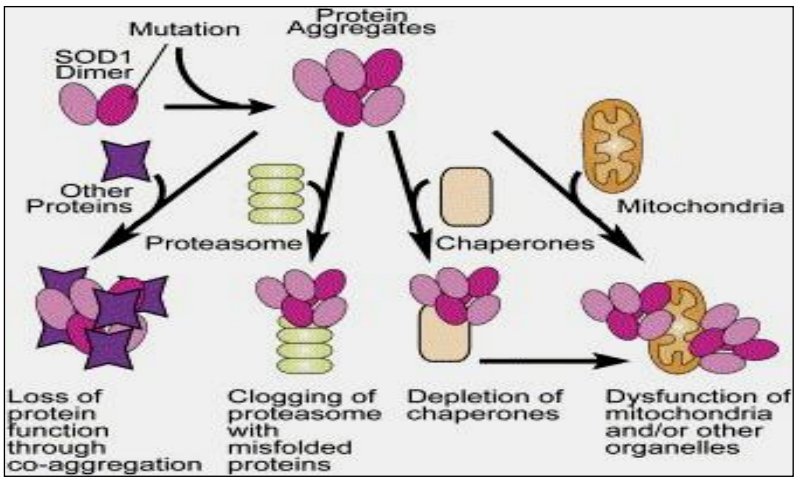


Fig.8: *Proposed Toxicities of ALS-Causing SOD1 Protein Aggregates*. Cell machinery that might be affected by misfolded, mutant SOD1 includes coaggregation of essential cytoplasmic components, poisoning of the proteasome thereby inhibiting timely degradation of many cellular proteins, saturation of cytoplasmic chaperones that catalyze essential protein folding and refolding, and damaging mitochondria by aggregation onto the cytoplasmic surface and/or transport into the mitochondrial intermembrane space. (Reprinted from Séverine Boillée, Christine Vande Velde, and Don W. Cleveland, Neuron, 2006).

New competing proposals such as vacuolated, dilated, and disorganized mitochondria, which were confirmed in mutant SOD1 mouse models expressing dismutase active [116], but not - inactive mutants [117] are now tested. It is interesting to note that these alterations in mitochondrial structure occur before any other pathological and/or clinical disease feature.

The visible inclusions in SOD1-linked fALS contain neurofilament proteins, ubiquitin, and a variety of other components in addition to SOD1, but it is not known whether copper, zinc, or any other metal ions are present in the inclusions or are involved in their formation. Nowadays, different observations regarding the protein composition of aggregated SOD1 polypeptide are published. Studies from transgenic mouse models of fALS have shown the presence of intermolecular disulfide-linked SOD1 species in the spinal cord aggregates [118, 119]. One study used immunoprecipitation to conclude that an apparent SOD1 species of 35 kDa was a complex between SOD1 and Bcl-2 [120]. A recent study on humans patients identify a covalently crosslinked, biotinylation-dependent, SOD1-IR, partially misfolded species of 32 kDa [121]. On the contrary, another recent study show the presence of the unmodified, full-length SOD1 polypeptide in the mouse spinal cord aggregates [122]. A common agreement is that those inclusions, and other visible protein aggregates, common to all of these neurodegenerative diseases, represent an end stage of a molecular cascade of several steps, and that earlier steps in the cascade may be more directly tied to pathogenesis than the inclusions themselves [123]. The relatively large fibrils or insoluble inclusions observed in ALS may be the result of a protective mechanism that forms inclusions when the burden of misfolded or damaged proteins exceeds the capacity of the protein degradation machinery to eliminate them [124].

1.8 Familial ALS mutations in SOD1

As reported before, there are more than 100 single point mutations of SOD1 protein which are related to the familial form of ALS (fALS). The majority of these mutations are amino acid substitutions at one of at least 64 different locations, but some cause frame shifts, truncations, deletions, or insertions.

The vast majority of the mutations are genetically dominant. The only one known exception, D90A, is an oddity since in certain families it is recessive [125], whereas in others it is dominant [126].

These mutations are spread over the entire sequence of the protein (Fig. 9). The age of onset of SOD1-linked fALS varies widely, even within the same families, but relative survival, i.e., the length of time between onset and death, clearly depends upon the identity of the specific mutation. The survival data for the more common SOD1 mutations are as follows: >17 years for G37R, G41D, H46R, and E100K; ~10 years for G93C; ~ 6 years for G85R ~5 years for E100G; ~1–3 years for L38V, H43R, and G93A; and ~1 year for A4V. The I113T and D90A SOD1 mutations gave highly variable survival times [127] whereas, H48Q [128] and S134N [129] seem to have very short survival times.

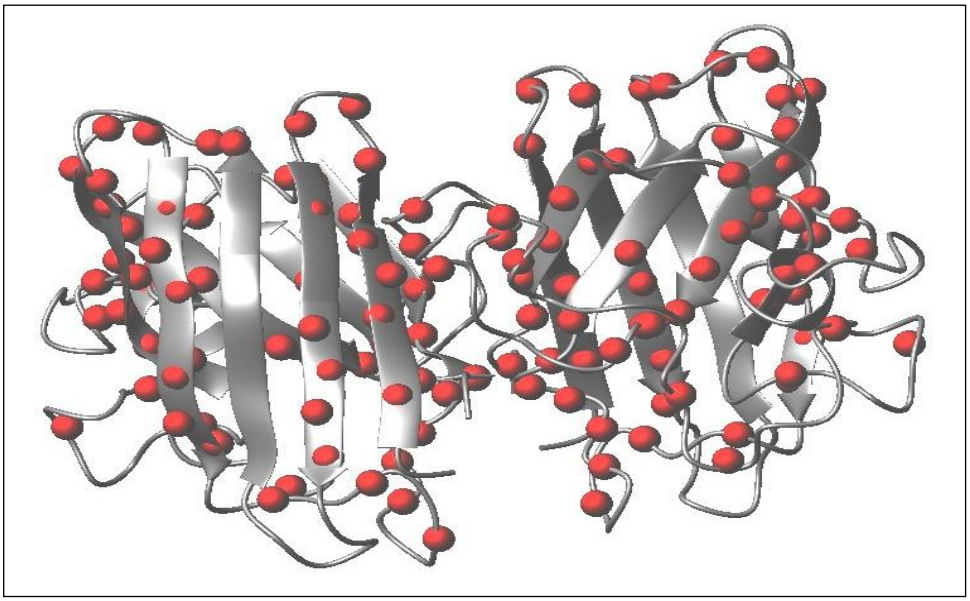


Fig.9: *Locations of point mutation (red sphere) fALS-associated mapped on SOD1 structure.*

Currently, a large number of fALS mutant SOD1 proteins were isolated and expressed *in vitro* by different groups in an effort to identify the source of their toxicity. Biophysical studies of isolated fALS mutants suggest that these proteins can be divided in two groups with distinctly different biophysical characteristics with respect to metal content, SOD1 activity, and spectroscopy [130]. These two groups have been termed metal-binding region (MBR) and wild-type-like (WTL) fALS mutant SOD1 proteins [131]. The MBR subset of SOD1 proteins has mutations that are localized in and around the metal-binding sites (H46R, H48Q, H48R, and H80R), including the electrostatic and zinc loops (e.g., S134N and D125H), and they were found to have significantly

altered biophysical properties relative to wild-type SOD1. By contrast, the WTL subset of SOD1 proteins (G10V, V14G, L38V, G93A) was found to be remarkably similar to wild-type SOD1 in most of their properties.

The activity and the stability of wild-type SOD1 are strongly dependent on the level of metallation of the polypeptide. However, in most cases, the in vivo metallation status is unknown and could vary in different tissues or in different compartments of the cell. Studies on the biologically metallated WTL fALS-mutants showed that these proteins have similar metallation levels suggesting that the metal-binding affinities and the mechanisms for the *in vivo* metallation are little affected by this class of mutations. On the contrary, the metal-binding properties of the MBR SOD1 mutants are very different. Each MBR SOD1 mutation is unique in its effect on the metal-binding properties of the protein because of either an alteration to the liganding amino acid that normally binds either copper (H46R, H48Q) or zinc (H80R), or a major alteration of the metal-binding affinity (G85R), or a major structural perturbation nearby the metal-binding region. H46R SOD1, for example, *in vitro*, is capable of binding either copper or zinc at the native zinc site but does not bind metal ions to the native copper site [(132)].

WT SOD1 is an incredibly stable protein in its fully metallated, disulfide-oxidized form. The enzyme does not denature in the presence of 8 M urea or 1% sodium dodecyl sulfate (SDS) [(133)]. The thermostabilities of a large subset of WTL and MBR fALS SOD1 mutations have been measured for the metal-loaded or as-isolated fALS mutants using differential scanning calorimetry (DSC) [(134)]. The WTL mutants have similar endotherm profiles and their metallated forms are almost as stable as wild-type SOD1. MBR mutants, which usually contain very low amounts of metal, have melting profiles significantly different than wild-type or the WTL mutants, and such difference is likely due to the absence of the stabilizing metals.

A number of X-ray structural studies of fALS mutant SOD1 proteins have been published in the last years (G37R, A4V, H46R crystallized both in the presence and absence of metal, I113T, D125H, S134N, and G85R) [(135, 136)]. Mutants SOD1 crystallized in three different forms that

resembled amyloid-like filaments and water filled nanotubes [135]. All of the described structures depend on aligned beta-sheet interactions that may be promoted by non native conformational rearrangement in the metal-depleted mutant enzyme.

One of the most striking features of ALS is the cell specificity: motor neurons are selectively affected in the patients, whereas the mutant SOD1 is expressed ubiquitously. The neurotoxic effect of mutant SOD1 may be not just a direct consequence of its expression inside the neuron, but may require functional alternation of non-neural cells, particularly microglia and astrocytes. Expression of mutated SOD1 selectively in either motor neurons [137] or astrocytes [138] has failed to cause ALS-like disease in mice. A recent study show that ALS–linked SOD1 mutants with different biochemical characteristics disrupted the blood–spinal cord barrier in mice by reducing the levels of the tight junction proteins ZO-1, occludin and claudin-5 between endothelial cells. This resulted in microhemorrhages with release of neurotoxic hemoglobin-derived products, reductions in microcirculation and hypoperfusion. SOD1 mutant–mediated endothelial damage accumulated before motor neuron degeneration and the neurovascular inflammatory response occurred, indicating that it was a central contributor to disease initiation [139].

Nowadays, ALS is recognized as multi-systemic disease, and the contribution of signals originating in cells other than motor neurons to the progression of the disease has been assessed. Even if the cellular mechanism is not elucidated yet major advantages in understanding ALS have been found out. This has changed the perspective on therapeutic strategies worthy of consideration in the near future. Therefore, nowadays drug screening *in vitro* and *in vivo* research are cooperating in order to further understand the mechanism and defeat the disease.

1.10 References

(1) Finney, L.A. & O'Halloran, T.V. *Science* 300, 931-936, (2003).
(2) Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C. *Nature* 409 860-921, (2001).
(3) Bertini, I.; Gray, H.B.; Stiefel, E. & Valentine, J.S. *Bioloical Inorganic Chemistry* 1–712 (Sausalito, California, USA, 2007).
(4) Lippard, S.J. & Berg, J.M. *Principles of Bioinorganic Chemistry* 3–388 (University Science Books, Mill Valley, California, USA, 1994).
(5) Andreini, C.; Banci, L.; Bertini, I. & Rosato, A. *J. Proteome Res.* 7, 209–216 (2008).
(6) Böhm, S.; Frishman, D., and Mewes, H. W. *Nucleic Acids Res*. 25:2464-2469 (1997).
(7) Magonet, E.; Hayen, P; Delforge, D.; Delaive, E. and Remacle, J. *Biochem. J.* 287:361-365, (1992).
(8) Vallee, B.L. and Auld, D. S. *Biochemistry* 29:5647-5659, (1990).
(9) Binder, H.; Arnold, K.; Ulrich, A. S. and Zschornig, O. *Biophys. Chem. 9*0:57-74, (2001).
(10) Berg, J. M. and Y. Shi., *Science* 271:1081-1085 (1996).
(11) Sarkar, J.; V. Seshadri, N.A.; Tripoulas, M.E.; *J. Biol. Chem.* 278 45 44018-24, (2003).
(12) Osaki, S.; Johnson, D.A., *J. Biol. Chem*. 244 20 5757-8, (1969).
(13) Harris, Z.L.; Durley A.P.; Man T.K.; Gitlin J.D. *Proc. Natl. Acad. Sci*. USA 96 19 10812-7, (1999).
(14) Ferguson-Miller, S.; Babcock, G.T.; *Chem. Rev.* 96 7 2889-2908, (1996).
(15) Yoshikawa, S. et al., *Science* 280(5370) 1723-9, (1998).
(16) Zelko, I.N.; Mariani, T.J.; Folz, R.J.; *Free Radic Biol Med* 33 3 337-49, (2002).
(17) Cousins, R.J., *Physiol Rev* 65 2 238-309, (1985).
(18) Kang, Y.J., *Proc Soc Exp Biol Med* 222 3 263-73, (1999).
(19) Mercer, J.F., *Am J Clin Nutr* 67 5 Suppl 1022S-1028S, (1998).
(20) Rucker, R.B.; Kosonen, T.;et al. *Am J Clin Nutr* 67, 5 Suppl, 996S-1002S, (1998).
(21) Tang, C.; Klinman, J. P., *J. Biol. Chem*. 276 33 30575-8, (2001).
(22) Kuivaniemi, H.; Peltonen, L.; Palotie, A.; Kaitila, I.; Kivirikko, K.I. *J Clin Invest* 69 3 730-3, (1982).
(23) Rae, T.; Schmidt, P.J.; Pufahl, R.A.; Culotta, V.C. & O'Halloran, T.V. *Science* 284, 805-808, (1999).
(24) Hellman, N.E. & Gitlin, J.D., *Annu. Rev. Nutr*. 22, 439–458, (2002).
(25) Harris, Z.L., Durley, A.P., Man, T.K. & Gitlin, J.D., *Proc. Natl. Acad. Sci*. USA 96, 10812–10817, (1999).
(26) Maryon, E.B., Molloy, S.A., Zimnicka, A.M. & Kaplan, J.H., *Biometals* 20, 355–364, (2007).
(27) Puig, S., Lee, J., Lau, M. & Thiele, D.J., *J. Biol. Chem.* 277, 26021–26030, (2002).
(28) Pena, M.M., Puig, S. & Thiele, D.J, *J. Biol. Chem.* 275, 33244–33251, (2000).
(29) Aller, S.G. & Unger, V.M., *Proc. Natl. Acad. Sci*. USA 103, 3627–3632, (2006).
(30) Cobine, P.A., Ojeda, L.D., Rigby, K.M. & Winge, D.R., *J. Biol. Chem.* 279, 14447–14455, (2004).
(31) Arnesano, F.; Balatri, E.; Banci, L.; Bertini, I.; D.R. Winge, *Structure* 13 713-22, (2005).
(32) Palumaa, P.; Kangur, L.; Voronova, A.; Sillard, R., *Biochem J*. 382 (Pt 1) 307-14, (2004).
(33) Tang, C.; Klinman, J.P.; *J. Biol. Chem.* 276 33 30575-8, (2001).
(34) Pase, L.; Voskoboinik, I.; Greenough, M.; Camakaris *J. Biochem J* Pt; 39, (2003).
(35) Sarkar, B., *Chem Rev* 99 9 2535-44, (1999).
(36) Shoubridge, E.A.; *Am J Med Genet*. 106 46-52, (2001).
(37) Waggoner, D.J.; Bartnikas, T.B.; Gitlin, J.D. *Neurobiol.Dis*. 6 221-230, (1999).
(38) Rotilio, G.; Ciriolo, M. R.; Carri, M. T.; Rossi, L. *Handbook of Copper Pharmacology and Toxicology*, Humana Press, Totowa, NJ. 277–96, (2002).
(39) Valentine, J.S.; Hart, P.J. *Proc. Natl. Acad. Sci.* USA 100, 3617–22, (2003).
(40) Wang, K., Zhou, B., Y.M. Kuo, J. Zemanski, et al., *Am. J. Hum. Genet.* 71 66–73, (2002).
(41) Kury, S., Dreno, B., et al *Nat. Genet*. 31 239–240, (2002).
(42) Chowanadisai, W., Lonnerdal, B., Kelleher, S.L., *J. Biol. Chem.* 281 39699–39707, (2006).
(43) Chimienti, F., Favier, A., Seve, M., *BioMetals* 18 313–317, (2005).
(44) Sladek, R., et al., *Nature* 445 881–885, (2007).
(45) Zeggini, E., Weedon, M.N., et al., *Science* 316 1336– 1341, (2007).
(46) Wenzlau, J.M., et al., *Proc. Natl. Acad. Sci.* USA 104 17040– 17045, (2007).
(47) Kang, S.E., et al., Diabetes, doi:10.2337/db07-0761, (2008).
(48) Airola K, et al., *J Invest Dermatol* 109:225–231, (1997).
(49) Uría JA, et al., *Am J Pathol* 153:91–101, (1998).

(50) Tanzi R. E. and Bertram L., *Cell* 120, 545–555, (2005).
(51) Blennow K., de Leon M. J. and Zetterberg H., *Lancet* 368, 387–403, (2006).
(52) Haass C. and Selkoe D. J., *Nat. Rev. Mol. Cell Biol*. 8, 101–112, (2007).
(53) Mantyh P. W., Ghilardi J. R., Rogers S., et al., J. Neurochem. 61, 1171–1174, (1993).
(54) Bush A. I., Pettingell W. H., Multhaup G., d Paradis M., Vonsattel J. P., Gusella J. F., Beyreuther K., Masters C. L. and Tanzi R. E., *Science* 265, 1464–1467, (1994b).
(55) Esler W. P., Stimson E. R., Jennings J. M., et al., *J. Neurochem*. 66, 723–732, (1996).
(56) Cuajungco M. P. and Faget K. Y., *Brain Res. Brain Res. Rev*. 41, 44–56, (2003).
(57) Bush A. I., et al., *Proc. Natl Acad. Sci.* U. S. A. 100, 11193–11194, (2003).
(58) Adlard P. A. and Bush A. I., *J. Alzheimers Dis*. 10, 145–163, (2006).
(59) Miller L. M., Wang Q., et al., *J. Struct. Biol.* 155, 30–37, (2006).
(60) Davies, K.J., *IUBMB Life*, 50:279–289, (2000).
(61) Field, L.S.; Furukawa, Y.; et al.; *J. Biol. Chem.*; 278, 28052–28059, (2003).
(62) Sturtz, L.A.; Diekert, K.; et al.*J. Biol. Chem.*; 276, 38084–38089, (2001).
(63) Weisiger, R.A.; Fridovich, I.; *J. Biol. Chem.*, 248:4793–4796, (1973).
(64) Chang L.Y.; Slot, J.W.; Geuze, H.J.; Crapo, J.D.; *J. Cell Biol.*, 107:2169–2179, (1988).
(65) Asayama, K.; Burr, I.M.; *J. Biol. Chem.* 260:2212–2217, (1985).
(66) Pardo, C.A.; Xu, Z.; Borchelt, D.R.; et al, *PNAS USA*; 92, 954–958, (1995).
(67) Weisiger, R. A.; and Fridovich, I.; *J. Biol. Chem.* 248, 4793–4796, (1973).
(68) Stralin, P., Karlsson, K., et al., *Arterioscler. Thromb. Vasc. Biol.* 15:2032–2036, (1995).
(69) Markowitz, H.; Cartwright G.E.; Wintrobe, M.M., *J. Biol. Chem.* 234, 40–45, (1959).
(70) McCord, J. M.; Fridovich, I., *J. Biol. Chem.* 244, 6049-6055, (1969).
(71) Tainer, J.A.; Getzoff, E.D.; Richardson, J.S.; et al., *Nature* 306,284-287, (1983).
(72) Tainer, J.A.; Getzoff, E.D.; Beem, K.M.; et al., *J. Mol. Biol.* 160, 181-217, (1982).
(73) Parge, H.E.; et al. *Proc. Natl. Acad. Sci*. USA, 89, 6109-6114, (1992).
(74) Ferraroni, M.; Rypniewski, W.; et al., *J.Mol.Biol.* 288, 413-426, (1999).
(75) Banci, L.; Benedetto, M.; Bertini, I.; et al., *Biochemistry* 37, 11780-11791, (1998).
(76) Banci, L.; Bertini, I.; Cramaro, F.; Del Conte, R.; Viezzoli, M.S. *Eur J.Biochem.* 269, 1905-1915, (2002).
(77) Fisher, C.L.; Cabelli, D.E.; Tainer, J.A.; Hallewell, R.A.; and Getzoff, E.D. *Proteins* 19, 24–34 (1994).
(78) Bertini, I., et al., *Advanced Inorganic Chemistry* (Sykes, A. G., ed.) 127–250, Academic Press, SanDiego (1998).
(79) Fisher, C.L.; Cabelli, D.E.; Tainer, J.A.; Hallewell, R.A.; et al., *Proteins* 19, 24–34, (1994).
(80) Cleveland, D.W.; Rothstein, J.D. *Nat.Rev.Neurosci.* 2, 806-819, (2001).
(81) Roe, J.A.; Butler, A.; Scholler, D.M.; et al., *Biochemistry* 27, 950–958, (1988).
(82) Forman, H.J.; and Fridovich, I., *J. Biol. Chem.* 248, 2645–2649, (1973).
(83) Bertini, I.; Piccioli, M.; Viezzoli, M.S, et al., *Eur.J.Biophys.* 23, 167-176, (1994).
(84) Banci, L.; Bertini, I.; Chiu, C.Y.; et al., *Eur.J.Biochem.* 234, 855-860, (1995).
(85) Estévez, A.G.; Crow, J.P.; et al., *Science* 286, 5449, (1999).
(86) Culotta, V.C., Klomp, L.W.J ; Strain J, et al., *J Biol Chem.*; 272:23469–23472, (1997).
(87) Gamonet, F.; Lauquin, G.J.M., *Eur.J.Biochem.* 251, 716-723, (1998).
(88) Culotta, V.C.; Klomp, L.W.; Strain, J.; et al., *J.Biol.Chem.* 272, 23469 -23472, (1997).
(89) Wong, P.C.; Waggoner, D.; et al, *Proc. Natl. Acad. Sci*. USA, 8997, 2886-2891, (2000).
(90) Casareno RLB, Waggoner D, Gitlin JD., *J Biol Chem*.273:23625–23628, (1998).
(91) Wong, P.C; Waggoner, D.; et al., *Proc. Natl. Acad. Sci*. USA, 89, 97:2886–2891 (2000).
(92) Southon,A.; Burke, R.; Norgate, M.; et al., *Biochem J*., 383:303–309 (2004).
(93) Wintz H, Vulpe C., *Biochem Soc Trans.*; 30:732–735, (2002).
(94) Rothstein JD, Dykes–Hoberg M, Corson LB, et al., *J Neurochem.*, 72:422–429 (1999).
(95) Rae,T.D.; Schmidt, P.J.; Pufahl, R.A.; et al., *Science*; 284:805–808 (1999).
(96) Lamb, A.L.; Wernimont, A.K., Pufahl, R.A.; et al., *Nature Struct.Biol.* 6, 724-729, (1999).
(97) Zhu, H.; Shipp, E.; Sanchez, R.J.; Liba, A.; Stine, J.E.; Hart, P.J.; et al., *Biochemistry*,(2000).
(98) Casareno, R.L.; Waggoner, D.J.; Gitlin, J.D., *J.Biol.Chem.* 273, 23625-23628, (1998).
(99) Lamb, A.L.; Torres, A.S.; O'Halloran, T.V., *Nat. Struct. Biology*, 8,751-755, (2001).
(100) Charcot, J.M., and Joffroy, A., *Arch. Physiol. Neurol. Path*. 2, 744–754, (1869).

(101) Mitsumoto, H., Chad, D.A., and Pioro, E.P., *ALS* (Philadelphia: Davis), (1998).
(102) Hadano, S.; Yanagisawa, Y.; Skaug, J. et al, *Genomics* 71, 200–213, (2001).
(103) Chen, Y.Z., Bennett, C.L. et al., *Am J Hum Genet* 74: 1128–1135, (2004).
(104) Puls, I., Jonnakuty, C., et al., *Nat Genet* 33: 455–456, (2003).
(105) Nishimura, A.L.; Mitne-Neto, M.; Silva, H.C., et al., *Am J Hum Genet* 75, 822–831, (2004).
(106) Greenway, M.J.; Andersen, P.M.; et al. , *Nat Genet* 38: 411–413, (2006).
(107) Rosen, D.R., et al., *Nature* 362, 59–62, (1993).
(108) Valentine, J.S. *Free Radic.Biol.Med.* 33, 1314-1320, (2002).
(109) Yim, M.B.; Kang, J.H.; et al., *Proc. Natl. Acad. Sci*. USA, 93, 5709-5714, (1996).
(110) Wang, J.; Slunt, H.; Gonzales, V.; et al. *Hum Mol Genet* 12, 2753–2764, (2003).
(111) Valentine, J.S. and Hart, P.J., *Proc Natl Acad Sci U S A* 100, 3617–3622, (2003).
(112) Bruijn, L.I.; Miller, T.M.; Cleveland D.W *Annu. Rev. Neurosci.* 27, 723–49, (2004).
(113) Taylor, J.P.; Hardy, J.; Fischbeck, K.H.; *Science*, (2002).
(114) Beal, M.F.; *Ann Neurol* 58, 495–505, (2005).
(115) Manfredi, G. and Xu, Z.; *Mitochondrion* 5: 77–87, (2005).
(116) Higgins, C.M.; Jung, C., Xu, Z.; *BMC Neurosci* 4:16, (2003).
(117) Bruijn, L.I. ; et al. *Neuron* 18:327–338, (1997).
(118) Furukawa, Y.; Fu, R.; Deng, H.X.; et al., *PNAS USA* 103:7148–7153, (2006).
(119) Deng, H.X.; Shi, Y.; Furukawa, Y.; et al., *PNAS USA* 103:7142–7147, (2006).
(120) Pasinelli, P. ; Belford, M.E.; Lennon, N. ; et al. *Neuron* 43:19–30, (2004).
(121) Gruzman, A.; Wood, W.L.; et al., *Proc. Natl. Acad. Sci*. USA, 104 (30):12524-9, (2007).
(122) Shaw, B.F., et al., *J. Biol. Chem*., Vol. 283, 8340-8350, (2008).
(123) Ross, C.A.; Poirier, M.A., *Nat Med.* 10, Suppl: S10-7, (2004).
(124) Sherman, M.Y.; Goldberg, A.L., *Neuron* 29, 15–32, (2001).
(125) Jonsson PA, Backstrand A, et al., *Neurobiol. Dis.* 10:327–33, (2002).
(126) Parton MJ, Broom W, Andersen PM, Al- Chalabi A, et al., *Hum. Mutat.* 20:473, (2002).
(127) Orrell RW, Habgood JJ, Gardiner I, King AW, et al., *Neurology* 48:746–51, (1997).
(128) Aoki M, Abe K, Itoyama Y., *Cell. Mol. Neurobiol.* 18:639–47, (1998).
(129) Tiwari A, Hayward LJ, *J. Biol. Chem.* 278:5984–92, (2003).
(130) Potter SZ, Valentine JS., *J. Biol. Inorg. Chem.* 8:373–80, (2003).
(131) Higgins CM, Jung C, Xu Z., *BMC Neurosci*. 4:16, (2003).
(132) Liu H, Zhu H, Eggers DK, et al., *Biochemistry* 39:8125–32, (2000).
(133) Lyons TJ, Gralla EB, Valentine JS, *Met. Ions Biol. Syst.* 36:125–77, (1999).
(134) Rodriguez JA, Valentine JS, et al., *J. Biol. Chem.* 277:15932–37, (2002).
(135) Hough MA, Grossmann JG, et al., *Proc. Natl. Acad. Sci. USA* 101:5976–81, (2004).
(136) Cao, X.; Antoniuk, S.; et al., J. Biol. Chem., 283, 16169-16177, (2008).
(137) Wong PC, Pardo CA, Borchelt DR, Lee MK, et al., *Neuron* 14:1105–16, (1995).
(138) Nagai M, Aoki M, Miyoshi I, Kato M, Pasinelli P, et al., *J. Neurosci.* 21: 9246–54, (2001).
(139) Bruijn LI, Becher MW, Lee MK, et al., *Neuron* 18:327–38, (1997).
(140) Wang J, Xu G, Gonzales V, et al., *Neurobiol. Dis.* 10:128–38, (2002).
(141) Wang J, Slunt H, Gonzales V, et al., *Hum. Mol. Genet.* 12:2753–64, (2003).
(142) A. Pramatarova, J. Laganiere, J. Roussel, , *J. Neurosci*. 21, 3369, (2001).
(143) M. M. Lino, C. Schneider, P. Caroni, *J. Neurosci*. 22, 4825, (2002).
(144) Y. H. Gong, A. S. Parsadanian, A. Andreeva, W. D. Snider, J. L. Elliott, *J. Neurosci*. 20, 660, (2000).
(145) Boillée, S. et al. *Science* 312, 1389–1392, (2006).
(146) Zhong, Z., et al., *Nature Neuroscience* 11, 420 – 422, (2008).

1.9 Aims and topics of the research

During the three years of my studies, the PhD research was focused on the investigation of the effects induced by single point mutations on the structure and stability of human copper, zinc superoxide dismutase (SOD1), an enzyme linked to familial amyotrophic lateral sclerosis (familial ALS).

SOD1 is a 32-kDa very stable dimeric protein. The mature, correctly folded and enzymatically active form of SOD1 is obtained *in vivo* through several post-translational modifications, such as acquisition of zinc and copper ions, disulfide bond formation and dimerization. The structural disorder and/or an increase in mobility of SOD1 mutants is suggested to play a role in the pathology of familial ALS. This increase in protein flexibility might favour protein aggregation and high-molecular weight protein complexes formation. The propensity for oligomerization of wild type (WT) human SOD1 and a number of its mutants were investigated under relatively mild conditions, likely to be encountered in the protein *"in vivo"*, as well as the roles of metalation in hindering or promoting such oligomerization. The results of this first set of studies showed that WT human SOD1, when lacking both its metal ions, forms large, stable, soluble, amyloid-like protein oligomers in solutions exposed to air, under physiological conditions (37 ºC, pH 7, and 100 μM protein concentration). We found that these soluble oligomers are formed by intermolecular disulfide covalent bonds and by non-covalent interactions between beta strands, forming amyloid-like structures.

Secondly, we investigated how the familial ALS-linked mutations in SOD1 influence this new mechanism of SOD1 oligomerization. With this aim, we selected a number of mutants, mainly on the basis of their location on the structure and nature of the residue changes. The mutant proteins were characterized in the apo and zinc-reconstituted states with respect to their ability to form soluble large molecular weight oligomers, and the kinetics of their formation were compared. Just like WT SOD1, all of the mutant SOD1 proteins that we studied formed soluble oligomers when

incubated at pH 7, 37 °C, 100 μM concentration, but the rates of oligomerization were different. This result points out that the presence of a mutation and its location and type on the protein sequence would only modulate the rate of oligomerization of the protein. The finding that WT and all the mutants, independently of the nature and location of the mutation, undergo the same type of oligometization suggest a general, unifying picture of SOD1 aggregation that could operate when either wild type or mutant SOD1 proteins are in the metal-free state.

To further elucidate the mechanism of oligomerization a structural characterization of the metal fee form of WT SOD1 and its mutants is required. Therefore, we have studied the apo form of WTSOD1 through NMR in solution and through x-ray diffraction in crystals the apo forms of T54R and I113T mutant proteins. In order to evaluate the effects of temperature on protein structure, we have carried out crystallization trials at two different temperatures (16 °C and 37 °C) on the apo and metalated WTSOD1, T54R and I113T proteins. The metallated proteins gave crystals at both temperatures while the crystals of the apo proteins are obtained at 16°C only. These results indicate that temperature has a much greater influence on the crystallization of the apo proteins rather than on the metallated ones.

Solution state NMR spectroscopy was chosen as the appropriate technique to obtain valuable structural information on the initial steps of the aggregation process: mono-, double- and triple- labelled samples of apo-WTSOD1 protein were prepared. Moreover, knowing that a critical role in the SOD1 oligomer formation is played by the two free cysteines (Cys6 and Cys111) a specific attention was focused on variations of the structural arrangements and oxidation states of cysteine residues during the oligomerisation process: selectively labelled ($^{2}H^{15}N^{13}C$ Cys) samples were prepared for apo WTSOD1. ^{13}C 1D spectra show the presence of two oxidized (Cys146 and Cys57) and two reduced (Cys6 and Cys111) cysteine residues as expected. At low temperature, double forms due to cis-trans isomerism are detectable for the Cβ carbon signals of both cysteines Cys57 and 146, involved in the formation of the intrasubunit disulfide bond. Shifts induced by temperature and local rearrangements are observed for all the four cysteines Cβ carbon signals both

in the apo and metallated forms. The extent of such shifts is quite different for the four cysteines. The smaller perturbation is observed for Cys146 which is consistent with its location in the middle of a strand (β8), preserved also in the metal free form of SOD1. A significant shift is instead observed for Cys111 also consistent with its oxidation state and location on an unstructured loop, confirming that this residue is the most affected one by temperature.

2

METHODOLOGICAL ASPECT

2.1 Gene cloning

Isolation of sufficient amounts of protein from native sources is often impossible owing to low yields, culturing difficulties of the organism and instability during purification. Therefore, protein production from cloned recombinant DNA is normally the method of choice to obtain the desired protein. Proteins have wide variability in their structure and stability, no single production method and characterization scheme will be applicable to every protein. Thus, several methods were simultaneously developed in the past years, including all appropriate variations on cell-based expression systems (*E. coli*, *yeast*, *insect cell lines*) and cell-free expression systems. Solid-state chemical synthesis is a possible approach for important proteins that failed to express in all DNA-based expression systems. Each technology application has its own set of challenges. For the easy, soluble proteins, the challenge is scale up, while the more difficult proteins (i.e. human proteins) require exploration of thus methods to produce and stabilize them.

Large scale cell-based expression systems have been used worldwide in structural genomics centres with the Gram-negative bacterium *E. coli* as the mainstay system. Despite the attractiveness of this route, problems, i.e. solubility, possible absence of necessary cofactors or chaperones, are commonly encountered in this protein expression system. However, especially for NMR purposes, which require the production of high yield of labelled ^{15}N- and ^{13}C- samples, the *E. coli* expression system is nowadays, the most widely used.

Many efforts are currently focused on the optimization of the *E. coli* expression system. Factors such as reduced temperature [(1)], changes in the *E. coli* expression strain [(2)], different promoters or induction conditions [(3)] and co-expression of molecular chaperones and folding modulators [(4)] have all been examined and, in some specific cases, they led to enhancements of soluble protein production. Many years ago it was discovered that some affinity tags are able to enhance the solubility of some of the partner proteins to which they are attached [(5)]; even if the number of fusion partners is increasing progressively during the years none of these tags work

universally with each partner protein. 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 (Fig.10).

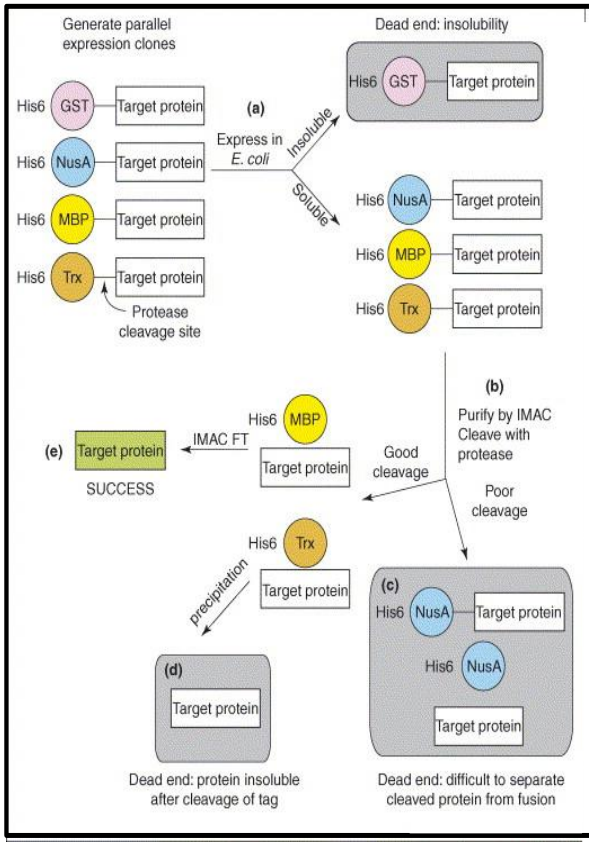


Fig.10: *Schematic representation of the pathway from protein expression to purification using solubility tags.* Four arbitrary tagged versions of the protein of interest are generated in E. coli.

(a) After expression, some fusions will remain in the insoluble fraction and be lost from the pathway.
(b) Soluble fusions are purified by IMAC using the attached His6 tag. A protease is then used to cleave the fusion tag from the partner protein.
(c) Some fusions will not cleave efficiently, and will leave behind a mixture of cleaved and uncleaved proteins that cannot be easily separated.
(d) Other fusions will cleave efficiently, but when separated from the solubility tag the partner protein will become insoluble and precipitate.
(e) However, a well-behaved fusion will remain in solution and can be purified by a second IMAC step to remove the His6-tagged solubility tag and protease, leaving only the target protein in the flow through (IMAC FT).

IMAC=immobilized metal affinity chromatography. (Reprinted from Esposito D et.al (2006) Current Opinion in Biotechnology 17, I 353-358)

Anyway, the experience shows that in some cases the classic approach, that is to express the native protein without any tag results the only successful one. While classical approaches do not require further sub cloning, fusion partners' impact on protein solubility leads to sub cloning the gene of interest in a library of expression vectors that becomes laborious when handling a large number of genes. A recent study [(6)] has described a cloning method (Gateway technology) that enables rapid cloning of one or more genes into virtually any expression vector using site-specific and conservative recombination (LR reaction) eliminating the requirement to work with restriction enzymes and ligase. On the basis of the former considerations, in our laboratory, we selected at least two different domains for each new target. These were cloned with the classic methods to express the protein in the native form and with Gateway system (Invitrogen) in order to express them with different N-Terminal fusions tag (GST, MBP, NusA, Trx, GB1).

2.1 a SOD1 gene cloning

WTSOD1 and its mutants were cloned in the classical approach. The *sod1* gene was cloned by PCR and inserted in pPSOD-Iq plasmid. The plasmid was propagated and purified by MINI KIT (Invitrogen) and the sequence was verified. Mutations were performed using a QuikChange™ site-directed mutagenesis kit (Stratagene).

At the same time, in our lab, WTSOD1 protein was cloned using Gateway Technology. The *sod1* gene was cloned by PCR and inserted in Gateway pENTR/TEV/D-TOPO plasmid (Invitrogen). The plasmid was propagated and purified by MINI KIT (Invitrogen) and the sequence was verified. The LR recombination reaction was performed in order to transfer the *sod1* gene from pENTR/TEV/D-TOPO plasmid into different pDest plasmids (pDEST-HisMBP, pDEST-periHisMBP, pETG20A, pETG30A and pTH34) suitable for protein expression and codifying for various N-terminal fusion tags (MBP, TRX, GST and GB1). The most efficient expression and purification system was conferred from the small-size fusion partner GB1.

2.2 Protein expression

The variables in a protein expression experiment can be divided into two groupes: genetic and environmental. Genetically encoded variables that affect protein expression include the sequence of the open reading frame, the choice of promoter, codon usage, mRNA secondary structure and addition of fused tags. Environmental variables include host strain, growth medium and induction parameters, *e.g.* temperature, IPTG concentration and duration of induction step.

At first, expression and solubility screening on a small scale (1-10 ml LB) was performed generally using different *E. coli* strains (TOPP1, BL21Gold (DE3), Origami (DE3)) and inducing the protein expression at two different temperatures (37 ºC and 25ºC). The result of this first test is analysed and, on the basis of these data, it is decided whether or not proceeding to the scale-up and testing the expression in minimum media. A second screening is sometimes performed in order to

refine expression conditions or, in the case in which all the tests are negative, to redefine the strategy.

1) If no expression is observed we can take into consideration the following choice:

- Redefine the domain.
- Use an expression vector containing an inducible promoter different from T7 (for example ARA or Cold inducible promoters) [3]
- Test other *E.coli* strains. In some cases, using promoters different from T7, as for instance LacIq, better results were obtained with *E.coli* strain lacking the DE3 episome.
- Consider the possibility of moving to an eukaryotic expression system.

2) If all the proteins are expressed in the insoluble fraction as inclusion bodies the choices are:

- Change the expression condition (temperature, induction time, medium ecc)
- Redefine the domain.
- Proceed with an "in vitro" refolding screening
- Clone again the domain in a "classic vector" codifying for the same fusion partners but lacking the His-tag. In some case it was observed that the presence of an N-or C-terminal His-tag [7] and/or the sequences corresponding to the Gateway recombination cassette, has a noticeable negative affect on protein solubility.
- Consider the possibility of moving to an eukaryotic expression system.

If from the expression screening we obtain positive results we select the more promising conditions and move to the scale-up, which is performed according to the following scheme:

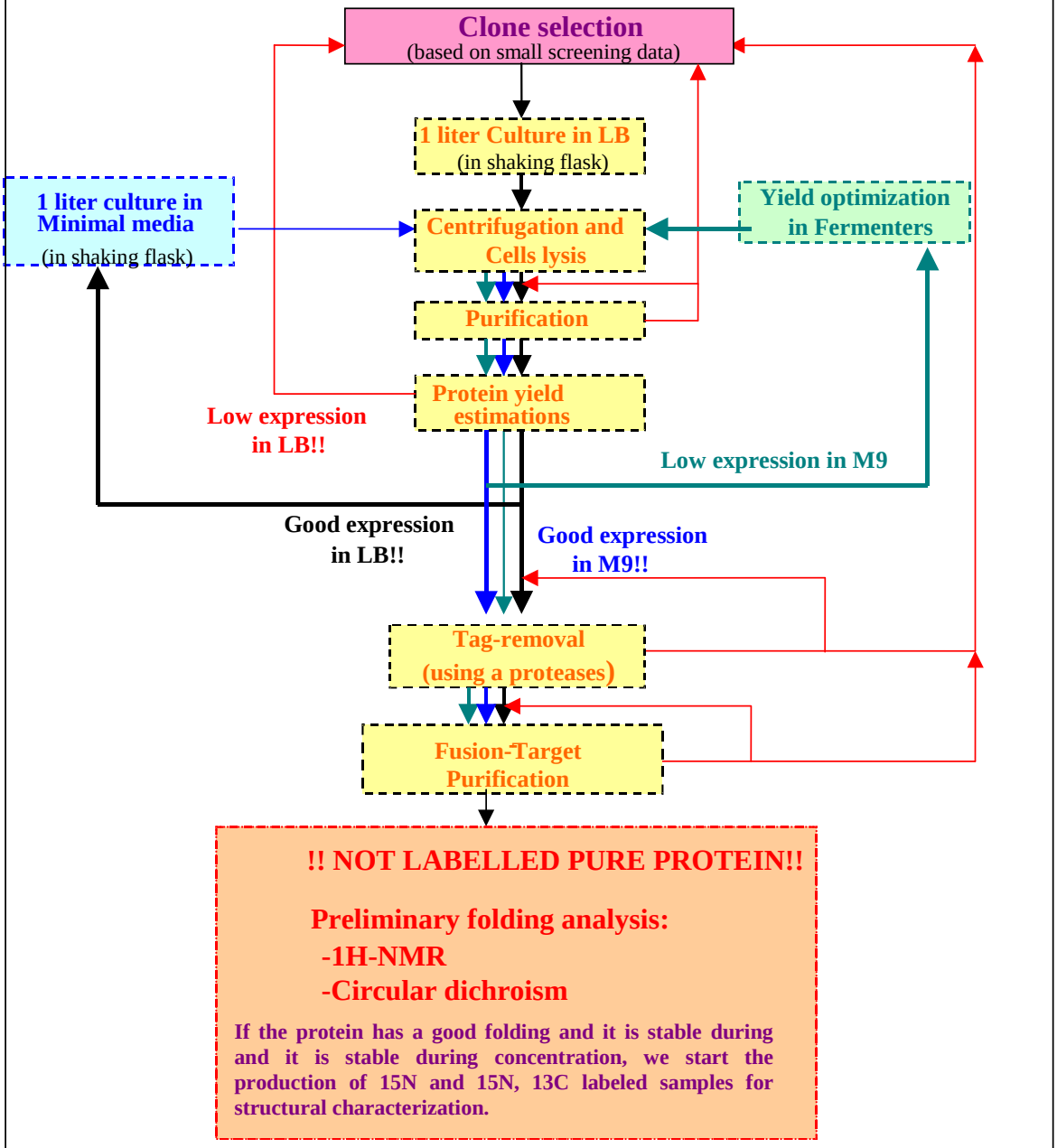


Fig. 11: *Typical scale-up optimization.* Expressions, purification, tag-removal by proteases are optimized on the unlabelled samples from the LB culture. In parallel, simulations of culture in minimal media (M9) are performed to optimize the labelling procedure. Fermentors are used only for M9 when low yields are obtained from shaking-flask.

The set-up of an expression protocol generally requires the optimization of different steps. In Fig. 11**,** the frequent situation in which it is necessary to return to the cloning selection and restart from the beginning if the optimisation of some intermediate step fails, is also shown.

2.2 a SOD1 protein expression

WTSOD1 protein and its mutants were expressed in *E. coli* BL21-Gold (DE3). Twenty five protein mutants, divided in subgroups localized in different secondary structural regions were selected for this study; a test expression in rich medium (big culture) was run for all the mutants and the attention was focused on the ten (T54R, V87M, A89V, G93A, G93D, V97M, I113F, I113T, L144F, L67V) which presented the highest expression results.

The ^{15}N-labeled proteins were obtained by growing cells in minimal medium (M9) containing 1g/l ^{15}N-$(NH_4)_2SO_4$ whereas LB medium was used for the non-labelled protein. For the latter, the cells were grown in shaking flaks until OD_{600} = 0.6 and then induced with 0.7mM IPTG for 12h at 37°C. The same procedure repeated in minimal medium results in loss of protein expression level. A quite good yield (around 10mg/l of the SOD1 proteins) was recovered, optimizing the growth and expression condition in 1.5 liter Minifors fermenter (Infors HT). For GB1-WTSOD1 protein expression, competent *E.coli* Origami (DE3)plyss cells were used. In this case, the shaking flask culture in minimal medium provided a good yield (>40mg/l of the GB1-WTSOD1 fused proteins), and the optimization with fermenters was not necessary.

The ^{15}N-labeled, ^{15}N- ^{13}C-labeled proteins were obtained by growing cells in M9 containing 3g/l ^{13}C-Glucose and 1g/l ^{15}N-$(NH_4)_2SO_4$ and induced at 25°C over night. The deuterium enriched protein (^{15}N- ^{13}C- ^{2}H- labelled WT SOD1) was expressed under similar conditions except for the growth of the *E. coli* strain, which was carried out in minimal medium containing 90 % (volume) of deutarated water (^2H_2O). In order to obtain 80 % of deuterium incorporation into WTSOD1 protein, it was necessary to isolate the *E. coli* cells with high tolerance to deutereted water. Therefore, the cells were subjected to growth conditions of increasing ^2H_2O content. The first step was the inoculation of *E. coli* strain into 3 ml LB medium and successive incubation at 37°C, overnight. The cells culture was diluted 20-folds by inoculation into 3 ml LB medium containing 25 % (volume) ^2H_2O and incubated at 37°C for 10 hours (OD_{600} = 1.2–1.5). The process was repeated with LB medium containing 50 %, 75 % and 90 % (volume) ^2H_2O, step by step. These cells were used

immediately for protein expression in 100ml of minimal medium culture with 90 % (volume) $^{2}H_2O$ and 3g/l ^{13}C-Glucose and 1g/l ^{15}N-$(NH_4)_2SO_4$, in a 250mL flask.

The GB1-WTSOD1 protein was also expressed in a cysteine auxotrophic strain of *E. coli*, Gold BL21 (DE3)cysE, in a medium containing triple labelled cysteines. The protein expression was done according to an already published protocol [8].

2.2 b SOD1 protein extraction

Depending on the location of the expression, the protein has to be brought into solution by breaking the cells containing it. There are several methods to achieve this: repeated freezing and thawing, sonication, homogenization by pressure or permeabilization by organic solvents. The method of choice depends on how fragile the protein is and how sturdy the cells are.

For WTSOD1 and mutant proteins, the extraction from cells periplasma was done by osmotic shock. The protein extraction protocol for thermo-stable mutant AS-SOD1 (where Cys6 and Cys111 have been changed to alanine and serine, respectively), previously developed in our lab, was optimized. The proteins were isolated in a 20mM Tris, 150mM NaCl, 0.1 EDTA, 5mM dithiothreitol (DTT) buffer at pH 8. After incubation for 30 minutes at 37 ºC, the proteins were centrifuged at 40000 rpm for 20 minutes. Supernatants were dialyzed against a 20mM Tris, 1mM DTT buffer in order to be loaded on the anionic exchange column. For GB1-WTSOD1, the protein extraction from cells cytoplasm was done by sonication in a 5mM imidazole buffer, pH 8 followed by centrifugation at 40000 rpm for 20 minutes. The latter protocol was also applied for extraction of the GB1-WTSOD1 protein expressed in the cysteine auxotrophic strain, BL21(DE3)cysE.

2.3 Protein purification

Protein purification after expression presents a number of challenges; particularly in the presence of many other contaminant proteins. Separation of one protein from all others is typically

the most laborious aspect of protein purification. All the purifications involve several chromatographic steps performed exploiting differences in protein size, physical, chemical and biological properties and binding affinity. The strategy of purification depends mainly upon the location of the expressed protein within the host. In fact, the protein can be transported in the periplasmic space or expressed like a soluble or insoluble (inclusion bodies) protein within the cytoplasm. In each case the isolation can be performed with a variety of different techniques. Ion exchange and size exclusion chromatography are commonly used to purify proteins expressed in their native states. The purification of the recombinant proteins with the specific affinity tags is usually done by affinity techniques. Immobilized metal ion affinity chromatography (IMAC) is currently the most commonly used technique which exploits the interaction between transition metal ions (generally, Zn^{2+} or Ni^{2+}) and side-chains of specific amino acids (mainly histidine) on the protein. In some cases, solubility tags have been combined with simple His-tag, allowing the fusion partner to maintain its solubilizing functionality and the His-tag its efficiency as an affinity tag.

For the recombinant proteins, enzymatic digestion with a specific protease is necessary to remove the fusion partner from the target protein and a second IMAC purification is generally performed in order to separate the two proteins. TEV, Fatt.Xa, Thrombin, Prescission Protease, recombinant Enterokinase are some examples of proteases that are normally used for the cleavage of fusion proteins. The protease specific recognition site is selected and cloned in the vector codifying for the protein sequence at the cloning step. For each protease/fusion protein pilot experiments should be done to find out suitable conditions.

2.3 a SOD1 purification

Native WTSOD1 and mutant proteins were purified by FPLC using anionic exchange and/or gel filtration techniques. In the first step, proteins were loaded onto a DEAE FF (Amersham Biosciences) anion exchange column and eluted during increasing salt gradient, using chromatographic buffers with 1 mM DTT each. Proteins were typically eluted in two different

peaks, one at low salt concentration (around 60mM) and the other at higher salt concentration, thus suggesting the presence of two differently charged species. In the case of mutant proteins, the first purification was followed by a size exclusion chromatographic column.

GB1-WTSOD1 fused protein purification was performed using IMAC by His-trap column charged with Ni^{2+}. The protein was eluted in Tris 20mM, NaCl 500 mM and imidazol 300mM pH 8. After concentration the sample was loaded on a PD-10 desalting column in order to exchange the buffer in Tris 50mM, EDTA 0.5mM, DTT 1mM. GB1 tag was cleaved with 2ul of Ac-TEV protease/1mg of fusion protein (Invitrogen, Carlsbad, CA) under overnight incubation at room temperature. The protease digestion is followed by a second purification in order to separate the protein from the fusion parter. At this point, the WTSOD1, lacking the fusion parter containing the His-tag does not bind the column, while, the GB1 tag is eluted at 300mM of imidazole.

Proteins purity was checked on a 17% polyacrilamide gel and protein concentration was determined by optical spectroscopy; the extinction coefficient at 265 nm for SOD1 is 15900 $M^{-1}cm^{-1}$.

2.4 Sample preparation

Copper, zinc superoxide dismutase is one of the most stable globular proteins studied so far. Structural investigations have been established that the enzyme stability is due to a combination of different factors, including the intrinsic stability of the eight-stranded β barrel fold, the close packing of the hydrophobic interfaces between the subunits, the presence of intrachain disulfide bond and the active site stabilization induced by metal ions [9-11].

Apo-form of the WTSOD1 and its mutants were obtained according to the previously published protocol [12] with some small differences. The proteins buffer was exchanged using PD-10 desalting column (protein concentration was usually around 3 mg/ml). The procedure implies two buffer exchanges (50mM acetate, 10 mM EDTA, pH 3.8) followed by 3 of days incubation at 4ºC, then one exchange in the same buffer with 100 mM NaCl overnight, and, finally, two exchanges in 20 mM phosphate buffer, pH 7. WTSOD1 protein was reconstituted with both zinc and copper ions (Cu_2,Zn_2SOD) according to the procedure described above [13]. E_2,Zn_2SOD was obtained by addition of one equivalent of zinc per subunit in a 50 mM acetate buffer, pH 5 followed by 12 hours incubation at room temperature. This form was further reconstituted with copper in a 20mM Tris buffer pH 8 resulting in Cu_2Zn_2SOD. The SOD1 mutant proteins were reconstituted with zinc only by addition to the apo-form of two equivalents of zinc per subunit in 20 mM Tris buffer, pH 8. Metal content of these various forms of SOD1 was checked by inductively coupled plasma mass spectrometry (ICP-MS) using a Thermo Jarrell Ash Atomscan Model 25 Sequential inductively coupled spectrometer.

Polyethylene glycol (PEG)/pH, $(NH)_4SO_4$/pH screening using the vapour diffusion technique vapour were performed in order to obtain SOD1 crystals suitable for X-ray diffraction. Crystals of apo SOD1 (WT, T54R and I113T) were obtained at 18°C from solutions containing 0.1 M MES (pH6.5) or 0.1 M HEPES (pH 7.0), 20% PEG 3350. The final protein concentration was in all cases 0.1 mM.

2.5 Biophysical characterizations

Having a pure protein in hands, several studies can be done in order to determine if it is full-length and in an active form of protein desired. Mass spectroscopy analysis is performed in order to verify the protein identity and understand if the sequence has the N-terminal methionine. Solubility and stability of the proteins, at high concentrations, generally represent an indication of a good folding.

Before proceeding to the preparation of labelled samples the degree of "foldedness" is estimated by ^{1}H-NMR and circular dichroism (CD) spectroscopy. The latter technique could be suitable also to evaluate the thermal stability. Size exclusion chromatography coupled with multiangle light scattering is performed in order to determine the aggregation state of the protein in solution. The metal content is analyzed by atomic absorption measurements. Disulfide status could be checked by SDS−PAGE after modification with AMS, or more accurately, by mass spectroscopy after modification with iodoacetamide.

Once we are sure that validated and stable proteins are produced, a more complete set of biophysical and biochemical characterizations can be made as required by the particular research problem and system.

2.5 a Fluorescence

Fluorescence spectroscopy is a common technique used by physicists, chemists, and biologists to experimentally characterize fluorescent species (proteins, bio-molecules, pharmaceuticals, etc.) and their dynamics.

Proteins, with aromatic amino acids are "intrinsically" fluorescent when excited by UV light. The three amino acid residues that are primarily responsible for the inherent fluorescence of proteins are tryptophan, tyrosine and phenylalanine. These residues have distinct absorption and emission wavelengths and differ in the quantum yields (Table 1).

Table1: Fluorescent characteristic of the aromatic amino acids.

	Absorption		**Fluorescence**	
Amino Acid	Wavelength(nm)	Absorption	Wavelength(nm)	Quantum
Tryptophan	280	5,600	348	0.20
Tyrosine	274	1,400	303	0.14
Phenylalanine	257	200	282	0.04

Protein fluorescence is generally excited at 280 nm or at longer wavelengths, usually at 295 nm. In the first case, we obtain the excitation of both tryptophan and tyrosine residues but, due to tryptophan's large absorptivity, the fluorescence spectrum usually resembles that of tryptophan. In the second case, using an excitation wavelength of 295 nm we can obtain a selective excitation of the tryptophan. The fluorescence of the aromatic residues varies in a somewhat unpredictable manner in various proteins. The quantum yield may be either increased or decreased by the folding. Accordingly, a folded protein can have either greater or less fluorescence than the unfolded form. The intensity of fluorescence is not very informative in itself. The magnitude of intensity, however, can be used as a probe of the perturbation of the folded state.

The fact that protein conformal transitions, corresponding to the transition between different states, like folded and unfolded, oxidized and reduced, are generally characterized by different fluorescence intensities [(14)] was exploited in order to determine the relative stability of this states under different conditions. Progressive protein unfolding in guanidinium chloride [(15)], or a disulfide bond red-ox potential [(16)] are some examples of interesting protein properties that can be monitored in this way. Moreover, proteins can be covalently labelled with various fluorophores, thus producing fluorescent protein conjugates. The emission from these attached tags is called "extrinsic" fluorescence**.** Tagging a protein with fluorescent labels is an important and valuable tool for protein characterization.

- **ThioflavinT (ThT) fluorescence** is a commonly used method to monitor fibril formation. This method is particularly attractive since ThT fluoresces only when bound to fibrils. The reaction between the protein and ThT is completed within 1 minute and ThT does not interfere with aggregation. Free ThT has excitation and emission maxima at 350 nm and 450 nm, respectively. However, upon binding to fibrils the excitation and emission change to 450 nm and 485 nm, respectively [17]. The structure of ThT has a hydrophobic end with a dimethylamino group attached to a phenyl group, linked to a more polar benzothiazole group containing the polar N and S. This combination of polar and hydrophobic regions creates the possibility for thioflavin T molecules to form micelles in aqueouse solution, with hydrophobic interiors and the positively charged N pointing toward the solvent. The benzothiazole moiety is a combination of a hydrophobic phenyl ring linked to a thiazole ring with positively charged nitrogen. ThT micelles binding involves hydrogen bond formation between charged nitrogen in the thiazole group to fibrils [18].

In the present work, ThT fluorescence spectroscopy was used in order to investigate the tendency of the apo, copper depleted and fully metallated, human SOD1 to form fibrillar aggregates under incubation in condition close to the physiological ones (100uM and 37°C) .

- **4-acetamido 4' maleimidylstilbene-2, 2'-disulfonic acid (AMS)** is a reagent that covalently reacts with free thiol groups, adding a group of 490 Da and therefore modifying protein mobility when it runs in a SDS-PAGE gel. This iodoacetamide derivate has high water solubility and is readily conjugated to thiols. AMS is a stilbene derivate, and shows a typical UV absorption at around 328 nm and emission maximum at 408 nm.

Estimation of apo SOD1 free thiols during aggregation was performed by AMS modification. Fluorescence measurements were performed in order to monitor the variation of free thiols during the incubation at 37°C of the protein samples. The methods require a preliminary calibration curve for the free cysteine quantification. In our case, this calibration curve was

determined using freshly prepared apo SOD1 WT at different concentrations as a standard. MALDI analysis proved that there is a maximum of two free cysteines per apo SOD1 monomer.

2.5 b Light scattering

Laser light scattering is a "non-invasive" technique that provides the absolute molecular weight (MW) and size of macromolecules in solution; the amount of light scattered is directly proportional to the product of the weight-average molar mass and the concentration of the macromolecule. Thus monitoring the size of a protein molecule is a way of observing the structure changes which may happen over time, pH, ionic strength, and it also provides information about the oligomeric state of the protein. Since laser light scattering provides the MW average for all molecules in solution it is generally useful to utilize it coupled with the gel filtration technique.

In this work, SOD1 aggregation was monitories by gel filtration and light scattering. 100 µl aliquots of the incubated proteins at 37 ºC were periodically taken and analyzed by gel filtration on Superdex 75 HR 10/30 (Amersham Biosciences) at room temperature. The column was preequilibrated with 20 mM potassium phosphate, pH 7, and the flow rate was 0.6 ml/min. The chromatogram, which monitors the species formed during incubation, was obtained by monitoring the absorbance at 280 nm. 20 µl aliquots of the incubated proteins at 37 ºC were also periodically taken and analyzed by gel filtration on G2000SW$_{XL}$ and G4000SW$_{XL}$ (Tosoh Bioscience) columns at room temperature. The columns were preequilibrated with 20 mM potassium phosphate, pH 7.0, and the flow rates were 0.7 and 1 ml/min respectively. The chromatogram, which monitors the species formed during incubation, was obtained by monitoring the absorbance at 280 nm.

While Superdex 75 HR 10/30 is a semi-preparative gel filtration columns, the G2000/4000SW$_{XL}$ are analytical ones. Their void volumes are 75 kDa, 150 kDa and 7,000 kDa respectively. The Superdex column was used when a separation of the dimer from the rest of the oligomeric species was necessary for further analysis. The G2000SW$_{XL}$ analytical column was used to monitor the very initial steps of the oligomerization process while the successive time points, in

which a larger oligomer, with MW higher than 150kDa, was formed, were better observed with the G4000SW_{XL} column.

The G2000SW_{XL} and G4000SW_{XL} columns were also connected to a light scattering spectrometer. The online multiangle light scattering (MALS) detector (DAWN EOS, Wyatt Technology, Santa Barbara, CA) and differential refractive index (DRI) detector (Optilab DRI, Wyatt Technology) setup was used to measure the light scattered as a function of angle and absolute protein concentration of fractions eluting from the size-exclusion chromatography column. The Zimm/Debye approximations were used in the Astra software (Wyatt Technology) to estimate molar mass. Data were fit using a second-order polynomial. The analysis was performed for each one of the 20 µl aliquots periodically taken from the incubation batches so as to monitor the increase in molecular weight of the soluble species formed during aggregation.

2.6 Structural characterization

Determination of protein structure is a top priority for complete understanding of proteins role and function. There are many techniques suitable to study different structural aspects of cellular components, but two techniques allow a resolution at the level of distinguishing individual atoms: X-ray crystallography and Nuclear Magnetic Resonance (NMR).

X-ray crystallography uses the diffraction pattern of X-rays, which are shot through a crystal. The pattern is determined by the electron density within the crystal. The diffraction is the result of an interaction between the high energy X-rays and the electrons in the atom. The electrons get activated and their relaxation to the initial energy state emits new X-rays. Bundles of such waves can be enhanced if they are in phase, and they get cancelled out if they are out of phase. Therefore, the diffraction of parallel X-rays from an object containing thousands of unit molecules arranged in a regular lattice, results in the enhancement and cancellation of the diffracted waves. A

resulting pattern of this vectorial process can be correlated with the distribution of the electrons in the crystal.

X-ray crystallography requires the growth of protein crystals up to 100/300 micron in size from a highly purified protein source. The most time consuming process on the path to determining molecular structure is protein crystallization. This involves screening a large number of buffer conditions until the ones ideal to induce protein crystallization are found. Once a well-diffracting crystal (< 2.5Å) is obtained the structure determination can proceed quickly especially if the structural model of a protein with good sequence homology to the unknown one has already been determined.

X-ray structures are high resolution structures enabling resolutions of the order of 1Å. Yet they depict a static structure, the result of a technique which requires large, stable protein crystals, within which each protein unit is lined up in a regular lattice. It was soon recognized that these static structures didn't really help explaining function because the structures are mostly the average of millions of identical units. 'Loose' structural parts like surface loops often failed to be resolved leaving some protein structures incomplete. The development of nuclear magnetic resonance techniques, NMR, was also used to overcome such kind of problems.

NMR measurements are carried out in solution under conditions that can be as close as possible to the physiological state. Some times even if crystal structures are available, additional data in solution are needed to determine the potential biological function of the protein. NMR is not only capable of solving protein structures to atomic resolution but it also has the unique ability of accurately measure the dynamic properties of proteins and can also supply information on protein folding and on intra-, as well as, intermolecular interactions. Furthermore, the analysis through NMR spectroscopy easily allows the characterization under several, different experimental conditions, such as different ionic strength and pH. A protein sample characterized by a good circular dichroism spectrum, a good ^{1}H-NMR or better a good ^{1}H-^{15}N HSQC spectrum and if it is stable in time (in the mM range), has high probability to be suitable for NMR structural

characterization. However, the two major bottlenecks limiting NMR in structural biology are represented by the long time requested for the data analysis and the drawback size limit of protein. The current size limit is around 35kDa but recent advances in both hardware and experimental design promise to allow the study of much larger proteins [19].

2.6 a Structure determination of proteins with NMR spectroscopy

Nuclear magnetic resonance technique uses a strong, high frequency magnetic field which stimulates atomic nuclei of the isotopes ^{1}H, ^{2}D, ^{13}C, or ^{15}N (they have a magnetic spin) and measures the frequency of the magnetic field of the atomic nuclei during its oscillation period back to the initial state. The steps towards NMR structure determination can be summarized as follows: preparation of the protein solution, NMR measurements, assignment of NMR signals to individual atoms in the molecule, identification of conformational constraints (e.g. distances between hydrogen atoms), calculation of the 3D structure on the basis of the experimental constraints. NMR spectra of biological macromolecules contain hundreds or even thousands of resonance lines which cannot be resolved in conventional one-dimensional spectra (1D). In fact, the interpretation of NMR data requires correlations between different nuclei, which are implicitly contained in 1D spectrum but often are difficult to extract.

Multidimensional NMR spectra provide both, increased resolution and correlations which are easy to analyse.In two dimensional NMR (2D), the intensities are plotted as a function of two frequencies.

The basic two dimensional experiments use the same basic scheme which consists of the four following time periods as:

Preparatory period	**Evolution period**	**Mixing period**	**Detection period**
	t1		**t2**

In the preparation period the sample is excited by one or more pulses. During the time period *t1* the sample is allowed to evolve. Mixing period consists of one or more pulses depending on the information we are seeking. Finally the magnetization is detected during the detection period. Depending on what is done in the preparation and mixing period, we can design experiments to derive the information of interest. The two dimensional Fourier transform of the data collected during *t2* as a function of *t1* gives a 2D spectrum where the information is spread in two dimensions. There are several two dimensional experiments designed depending on the choice of preparation and mixing periods:

COSY (abbreviation for **CO**rrelated **S**pectroscop**Y**) **–** it is the simplest of the 2D experiments; it consists of only two pulses with an incremental delay (*t1*) between them. It provides information on the spins coupled to each other. The first pulse creates magnetization in the transverse plane. During the evolution period the delay is incremented systematically to sample the spectra in the other dimension also. The second pulse mixes the magnetization between the spins coupled to each other. Finally, the double Fourier transform gives a spectrum, where the diagonal looks alike the normal one dimensional spectrum. The off-diagonal peaks (called cross peaks) give information on the nuclei which were coupled to each other through chemical bond.

TOCSY (abbreviation for **TO**tal **C**orrelated **S**pectroscop**Y**) - in this experiment, during the mixing time the magnetization exchanges through scalar coupling. Depending on the length of the mixing time, the magnetization is transferred between all the coupled nuclei in a spin system, even if they are not directly coupled. This essentially gives the same information as that of COSY experiment, except that COSY gives information only on the directly coupled spins, whereas TOCSY gives the complete spin coupling network.

NOESY (abbreviation for **N**uclear **O**verhauser **E**ffect **S**pectroscop**Y**) – it is one of the most useful experiment as it is used to correlate the nuclei through space. It gives qualitative and quantitative information on the proximity of two nuclei. Also, by measuring the intensity of the peaks, the distance information can be obtained. In this experiment after magnetization evolves

during *t1* period, exchange of the magnetization takes place during the mixing time through dipole-dipole relaxation mediated relaxation.

The extension from a 2D to a n-dimensional (nD) NMR experiment consists in the combination of (n-1) two-dimensional experiments, which contains only one excitation and one detection period but repeats the evolution and mixing times (n-1) times. The NMR multidimensional measurements almost always use protons (^{1}H) and depending on the isotope labelling, ^{13}C and/or ^{15}N nuclei. A 3D spectrum can for example be obtained by correlating the amide groups with the α-carbon nuclei attached to^{15}N. The chemical shifts of these carbon nuclei are used to spread the resonances from the 2D plane into a third dimension. The sensitivity obtainable with these types of nuclei greatly varies even if the sample is fully isotope labelled with ^{13}C or ^{15}N. The proton offers the best sensitivity and for this reason constitutes the preferred nucleus for detection of the NMR signal. The other nuclei are usually measured during evolution periods of multidimensional NMR experiments and their information is transferred to protons for detection.

- **Sequence Resonance Assignment**

NMR spectra contain information to determine bio-molecular structures in solution. However, none of the embedded information can be used without having the resonances of the bio-molecules assigned, which resonance come from which nuclear spins. The process of associating specific spins in the molecule with specific resonances is called *sequence-specific resonance assignment*. The application of multidimensional NMR spectroscopy allowed the development of general strategies for the assignment of signals in proteins. All procedures use the known protein sequence to connect nuclei of amino acid residues which are neighbours in the sequence. For unlabelled proteins smaller than 10 kDa the combination of the [^{1}H, ^{1}H]-COSY or TOCSY, used for the sequential assignment, with the [^{1}H,^{1}H]-NOESY spectrum allows the assignment of most proton NMR signals to individual protons(23). For larger proteins extensive signal overlap prevents complete assignments of all ^{1}H signals in proton spectra. This barrier can be overcome with 3D

NMR technique and uniformly ^{13}C and ^{15}N labelled proteins. The resonance assignment of single (^{15}N or ^{13}C) labelled proteins using 3D experiments is basically an extension of Wüthrich's strategy which exclusively relies on homonuclear ^{1}H NMR experiments. With these methods, systems with molecular weights up to approximately 30 kDa can be studied. In ^{13}C, ^{15}N-labeled proteins a sequential assignment strategy is based on through-bond correlations across the peptide-bond between sequential amino acids. This procedure circumvents the use of NOESY spectra already in the assignment step. Most of these correlation experiments use three types of nuclei ^{1}H, ^{15}N, ^{13}C and are referred to as triple resonance experiments.

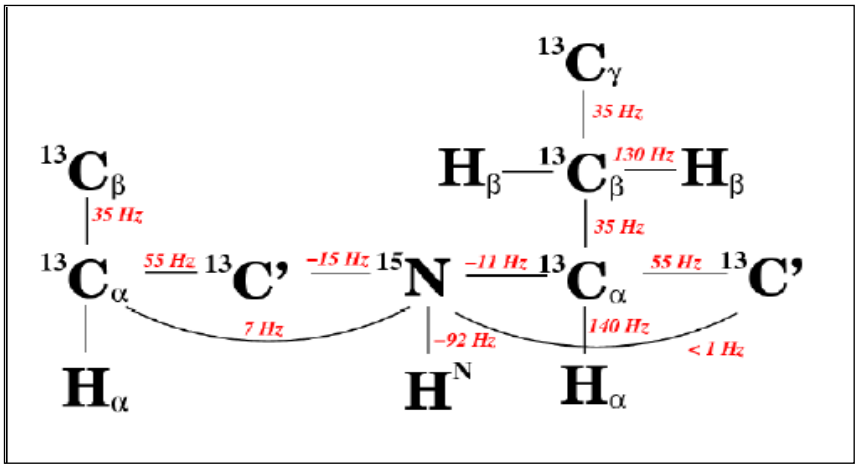


Fig. 13: *Spin **system** of the peptide backbone and the size of the 1J and 2J coupling constants that are used for magnetization transfer in 13C-, 15N- labelled proteins.*

The 3D triple resonance experiments exclusively correlate the resonances of the peptide backbone (HN(i), N(i), Cα(i), Hα(i), Cα(i-1), CO(i) and CO(i-1)).

The 3D experiments used to identify the backbone resonances are, usually, HNCA or HNCACB, HN(CO)CA or HN(CO)CACB, HNCO, HN(CA)CO and HNHA [20]. The HNCACB for example, correlates each H-^{15}N group with both the intra- and the neighbouring inter-residue Cα and Cβ. These four types of connectivity are discriminated using the HN(CO)CACB experiment, in

which only the inter-residue HN-Cα and Cβ couplings are observed. Similar strategy can be used to assign the other resonances in the other triple resonance spectra.

In the case of proteins with a molecular weight larger than 30 kDa the use of TROSY-type experiments [21] is necessarily. TROSY experiment can reduce the signal loss, which is the direct consequence of the slower correlation tumbling of large molecules which results in faster relaxation and consequently broader lines in the NMR spectrum. TROSY uses constructive interference between different relaxation mechanisms and works best at the highest available magnetic field strengths in the range of 700 to 900 MHz proton resonance frequency. With TROSY the molecular size of proteins accessible for detailed NMR investigations has been extended several folds. The TROSY technique benefits a variety of triple resonance NMR experiments as the 3D HNCA and HNCOCA [22] and the TROSY-based NOESY experiments for the collection of structural constraints are also available [23]. Since the Hα and Cα/β chemical shifts have been assigned, 3D H(C)CH-TOCSY and (H)CCH-TOCSY [24] experiments are then used to link the side chain spin systems to the backbone assignments. These two experiments provide information for the assignment of the side chain protons and of the side chain carbons, respectively.

A complete set of backbone chemical shifts for all Hα, Cα, Cβ and CO resonances can be used to predict the secondary structure of the protein [25]. One technique in particular, the Chemical Shift Index (CSI) [26], has been widely used for the quantitative identification and location of secondary structure in proteins. The method relies on the fact that the chemical shifts of the different nuclei in the protein backbone are related both to the type of amino acid and to the nature of the secondary structure they are located in. By comparing the actual chemical shift for a nucleus in a specific amino acid with a reference value, it is possible to predict in what secondary structure element the nucleus resides. The reference value that you compare with is the random coil chemical shift for that same nucleus in the same amino acid.

- **Collection of conformational constraints**

In order to be used in structure calculation, geometric conformational information in the form of distances and /or torsion angles has to be derived from the NMR data. The latter has to be supplemented by information about the covalent structure of the protein - the amino acid sequence, bond lengths, bond angles, chiralities, and planar groups - as well as by steric repulsion between non-bonded atom pairs. Although a variety of NMR parameters contain structural information, the crucial information comes from NOE measurements which provide distance information between pairs of protons. Supplementary constraints can be derived from through bond correlations in the form of dihedral angles. Further, chemical shift data, especially from ^{13}C, provides information on the type of secondary structure and hydrogen bonds can be detected via through-bond interactions. Such information can be included in a structure calculation by restricting the local conformation of a residue to the α−helical or β-sheet region of the Ramachandran plot through torsion angle restraints. Hydrogen bonds can also be experimentally detected via through-bond interactions [27] and they can be useful during structure calculations of larger proteins when not enough NOE data are available yet. Finally one other class of conformational restraints originate from residual dipolar couplings in partially aligned or paramagnetic molecules and gives information on angles between covalent bonds and globally defined axes in the molecule, namely those of the magnetic susceptibility tensor [28, 29].

With sufficient structural constraint, a folded conformation can be determined at atomic resolution. The result of NMR structure determination is not one model, but a set of similar models, all of which fit the experimental structural constraints. The RMSD (root mean square deviation) between these models is used to assess how well the structure calculations have converged. The best structures have backbone RMSD values of less than 1 Å. A final structure is obtained by averaging the models, and then finding the conformation of minimum energy that lies nearest to this average conformation.

2.7 References

(1) Hammarstrom, M.; Hellgren, N.; Van Den Berg, S.; Berglund, H.; Hard, T. *Protein Sci* 11, 313–321, (2002).
(2) Miroux, B.; Walker, J.E. *J Mol Biol* 260, 289–298, (1996).
(3) Qing, G.; Ma, L.C.; Khorchid, A.; Swapna, G.V.; Mal, T.K., Takayama; M.M., Xia; B.; Phadtare, S.; Ke, H. Acton T., *Nat Biotechnol* 22, 877–882, (2004).
(4) De Marco, A.; De Marco, V. *J Biotechnol* 109, 45–52, (2004).
(5) Kapust, R.B.; Waugh, D.S *Protein Sci* 8, 1668–1674, (1999).
(6) Landy, A., *Ann. Rev. Biochem* 58, 913, (1989).
(7) Woestenenk, E.A.; Hammarstrom, M.; Van den Berg, S.; Hard, T.; Berglund, H. J *Struct Funct Genomics* 5, 217-29, (2004).
(8) Strub, M.P.; Hoh, F.; Sanchez, J.F.; Strub J.M.; Böck, A; Aumelas, A; and Dumas, C. *Structure*, 11, 1359-1367 (2003).
(9) Banci, L.; Bertini, I.; Cramaro, F.; Del Conte, R.; Viezzoli, M.S. *Eur J.Biochem.* 269, 1905-1915, (2002).
(10) Fisher, C.L.; Cabelli, D.E.; Tainer, J.A.; Hallewell, R.A.; and Getzoff, E.D. *Proteins* 19, 24–34 (1994).
(11) Bertini, I.; et al., *Advanced Inorganic Chemistry* (Sykes, A. G., ed.) 127–250, Academic Press, SanDiego (1998).
(12) Valentine, J.S. and Pantoliano M.W., *Met. Ions Biol. Syst.* 3, 291 (1981).
(13) Beem, K.M.; Rich, W.E.; Rajagopalan K.V. *J Biol Chem.*, 249:7298–7305, *(*1974).
(14) Holmgren, A. *J Biol Chem.*, 247, 1992-8, (1972).
(15) Tsou, C. L. *Biochim. Biophys. Acta*, *1253*, 151(1995).
(16) Haugstetter, J.; Blicher, T.; Ellgaard, L., *J Biol Chem.* 280, 8371-80, (2005).
(17) Wider, G.; Macura, S.; Kumar, A.; Ernst, R.R.; Wüthrich, K., *J.Magn.Reson.* 56, 207-234, (1984).
(18) Krebs MR, Bromley EH, Donald, AM. *J Struct Biol.* 149:30–37, (2005).
(19) Wuthrich, K. *Nat. Struct. Biol.* 5, 492–495, (1998).
(20) Kumar, A.; Ernst, R.R.; Wüthrich, K. *Biochem.Biophys.Res.Commun.* 95, 1104, (1980).
(21) Kay, L.E.; Ikura, M.; Tschudin, R.; Bax, A. *J.Magn.Reson.* 89, 496-514, (1990).
(22) Pervushin, K., *Q.Rev.Biophys.* 33, 161-197, (2000).
(23) Salzmann, M.; Wider, G.; Pervushin, K.; Senn, H.; Wüthrich, K., *J.Am.Chem.Soc.* 121, 844-848, (1999).
(24) Pervushin, K.V.; Wider, G.; Riek, R.; Wuthrich, K. *PNAS USA* 96, 9607 9612, (1999).
(25) Kay, L.E.; Xu, G.Y.; Singer, A.U.; Muhandiram, D.R.; Forman-Kay, J.D. *J.Magn.Reson.Ser.B* 101, 333-337, (1993).
(26) Wishart, D.S.; Sykes, B.D.; Richards, F.M., *J.Mol.Biol.* 222, 311-333, (1991).
(27) Wishart, D.S.; Sykes, B.D.; Richards, F.M., *Biochemistry* 31, 1647-1651, (1992).
(28) Cordier, F.; Grzesiek, S., *J.Am.Chem.Soc.* 121, 1601-1602, (1999).
(29) Tolman, J.R.; Flanagan, J.M.; Kennedy, M.A.; Prestegard, J.H. *PNAS USA* 92, 9279-9283, (1995).

3

RESULTS

(in this chapter are presented the articles related to the project thesis)

- Banci L., Bertini I., Boca M., Calderone V., Cantini F., Girotto S. and **Vieru M.** (2009) " Apo SOD1 and its mutants: structural and dynamic aspects related to oligomerization", PNAS, 106(17), 6980 5)
- Banci L., Bertini I., Boca M., Girotto S., Martinelli M., Valentine J.S. and **Vieru M.** (2008) "SOD1 and Amyotrophic Lateral Sclerosis: mutations and oligomerization" PloS ONE, 3(2), e1677
- Banci L., Bertini I., Durazo A., Girotto S., Martinelli M., Valentine J.S., **Vieru M.**, and Whitelegge J.P. (2007) "Metal-free superoxide dismutase forms soluble oligomers under physiological conditions: A possible general mechanism for familial ALS", PNAS, 104(27), 11263-7

4

DISCUTION AND PERSPECTIVE

During the three years of PhD my research was focused on the investigation of the effects induced by single point mutations on the structure and stability of human copper, zinc superoxide dismutase (SOD1), an enzyme involved in the inherited form of amyotrophic lateral sclerosis (ALS).

ALS may be classified as a "conformational" disorder. In this class of disorders, protein alteration, through an abnormal folding pathway, generally results in protein aggregation and high-molecular weight complexes formation. Our research is directed towards the understanding of the relationships between structural perturbations of SOD1 enzyme and the neurodegenerative disorder ALS. Over 100 different point mutations of human SOD1 have been identified and linked to the familial form of amyotrophic lateral sclerosis (fALS). A large number of studies on ALS-mutant SOD1 proteins have been conducted and it has been shown that there is great diversity in the biophysical properties of these proteins. Nowadays, no common denominator of SOD1 pathogenic mutant is able to explain their aggregation propensity with respect to WTSOD1. What is known, is that the metallated form of WT SOD1 (Cu,Zn-SOD1) is an exceptionally stable protein to the extent that its melting temperature, Tm, is around 90°C [(1)]; however, removing a copper ion from holo-SOD1 (E,Zn-SOD1) decreases the Tm to ~70°C, and further removal of the Zn ion from E,Zn-SOD1 (E,E-SOD1) results in ~60°C of Tm [(2)].

SOD1 has to undergo several post-translational modifications before reaching its mature form. The protein, requires insertion of zinc and copper atoms, and the formation of a conserved S-S bond between Cys-57 and Cys-146 (human numbering), which makes the protein fully active. Given that these modifications generally increase the protein structural stability, it is well expected that the most immature forms, the apo and disulfide-reduced SOD1 (E,E-SOD1^{SH}), and the other possible intermediate forms (for example: E,Zn-SOD^{SH} or E,E-SOD^{SS}), are more susceptible to destabilization and aggregation. Using biophysical methods, we have investigate the propensity for oligomerization of mature, i.e., disulfide-intact, human SOD1 under the relatively mild conditions,

likely to be encountered in the protein *"in vivo"*; the roles of metalation in hindering or promoting such oligomerization were also investigated. The results of this first set of studies show that WT human SOD1, when lacking both its metal ions, forms large, stable, soluble, amyloid-like protein oligomers in solutions exposed to air, under physiological conditions (37 ºC, pH 7, and 100 μM protein concentration). We found out that these soluble oligomers are formed by intermolecular disulfide covalent bonds and by non-covalent interactions between beta strands, forming amyloid-like structures. This oligomerization occurs *in vitro* as well as *in vivo* in *E. coli*, when the two free cysteine residues (Cys 6 and Cys 111) are present. Lack of aggregate formation for the metallated protein shows that the metallation state is a critical issue in the oligomerization process.

The next question was to understand how the familial ALS-linked mutations, in SOD1 influence this new mechanism of SOD1 oligomerization. A number of mutants, selected mainly on the basis of their location on the structure and nature of the residue changes, characterized in the apo and zinc-reconstituted states with respect to their ability to form soluble large molecular weight oligomers, and the kinetics of their formation were compared. Just like WT SOD1, all of the mutant SOD1 proteins that we studied formed soluble oligomers when incubated at pH 7, 37 °C, 100 μM concentration, but the rates of oligomerization were different. This result points out that the presence of a mutation and its location and type on the protein sequence would only modulate the rate of oligomerization of the protein. The finding that WT and the mutants, independently of the nature and location of the mutation, undergo the same type of oligometization suggest a general, unifying picture of SOD1 aggregation that could operate when either wild type or mutant SOD1 proteins are in the metal-free state. These soluble oligomeric species, formed by the apo form of both WT SOD1 and its mutants through an oxidative process, might represent the precursor toxic species responsible for ALS, pointing at a common behaviour with many other neurodegenerative disorders. In fact, a general pathological feature of many neurological disorders is the presence of inclusions bodies and other visible protein aggregates, which might represent the end stage of a molecular cascade of several steps. Earlier steps in this cascade are more directly tied to

pathogenesis than the aggregates themselves [3], and are the result of the pathogenic species elimination induced by the cells. The formation of the soluble oligomeric species in apo WT SOD1 suggests a common explanation for both, familiar and sporadic ALS. Incorrect and/or incomplete SOD1 metallation starts the onset of pathogenic events, where the mutations are only a time modulation of the disease onset and development.

Another point to address is how SOD1 incorrect metallation can be active also *in vivo*. Neurofilaments-L (NF-L) are one of the most abundant proteins in motor neurons; their presence in aggregates is a pathologic hallmarks of both sporadic and familial ALS [4]; they have a strong affinity for zinc [5] and form aggregates that are positive to the thioflavin T binding assay, in the presence of copper and H_2O_2 [6]. Therefore, under specific conditions, they could compete with immature forms of SOD1 for zinc uptake explaining, in this way, the restricted localization of SOD1 inclusions in motor neurons. This hypothesis is supported by the fact that the absence of neurofilaments slows the disease onset in mice expressing a specific fALS mutant [7].

X-ray crystallography could provide a more defined picture of the protein structural characterisation. In particular, the comparison between WT and ALS-related mutants structures could enlighten the structural basis of some potential different behaviour between WT SOD1 and the mutants. We have selected two mutants with the two extreme behaviours in terms of aggregation rates: T54R aggregates slightly slower than WT SOD1 while I113T has an aggregation rate which is more than twice that of WT SOD1. From the comparative analysis of the three above-mentioned structures it is clear that there are measurable differences among them in terms of RMSD, buried surface areas, residues involved in H-bond interactions or local differences due to mutations; nevertheless it appears hard to fully explain the significant differences in aggregation rates and speed just on the basis of the above-mentioned structural differences.

A generic protein, in certain physical and/or chemical conditions can ‘choose’ whether to form quaternary structures in terms of crystals or in terms of aggregates [8].

In order to evaluate the effect of metallation and mutations on crystal formation, we have carried out crystallization trials at two different temperatures on the apo and metallated WT-SOD1 as well as on the apo and metallated mutants I113T and T54R. For all samples, trials have been performed in parallel at 16 °C and at 37 °C, which is the same temperature at which all aggregation tests have been carried out. Metallated forms of WT SOD1 and its mutants gave crystals at both temperatures. In the case of the three apo proteins, crystals have been obtained at 16 °C only. These results indicate that temperature has a much greater influence on the crystallization of the apo proteins rather than on the metallated ones.

When the protein is demetallated in the crystallization tray at 16°C, although somewhat conformationally unstable, it still finds that the most favoured thermodynamic route is that of selecting the right conformation to form crystals, which allows a lower free energy of the obtained system; conversely, at 37°C, the protein is unlikely to adopt the right conformation to form crystals and 'chooses' the conformation which causes the formation of aggregates. This would imply that the conformation adopted to form crystals is not the same that the protein uses to form aggregates and for this reason we cannot expect to find a crystal structure containing the building block of aggregates.

Although it is not possible to assess the relative weight of metallation and mutations on SOD1 aggregate formation, the crystallization results clearly show that metallation plays a major role when it is time for the protein to 'choose' between aggregation and crystallization. On the basis of the fluorescence experiments assessing that mutations affect the kinetic of the aggregate formation, we can conclude that metallation mostly affects the thermodynamics of the process.

On this regard, solution state NMR experiments were undertaken to investigate the dimeric form of apo SOD1.

The information which we obtained from a partial assignment of the backbone atoms of apoSOD1 suggests that the protein maintains the same secondary structural elements and the same overall fold of the metallated protein with the exception of the regions close to the metal binding

residues. Even though apoSOD1 maintains most of the secondary structural elements and a part of the beta barrel folding, it is characterized by an extensive flexible region containing the metal binding residues and beta strands 4 and 5 (Fig. 4), which could be described as a molten globule-like state. The information obtained from the NMR spectra indicate that the solution structure of the WT SOD1 protein in the demetallated form is quite unfolded if kept for prolonged periods of time even at 288 K. It is possible to argue that higher temperatures would induce, due to kinetic reasons, a more pronounced loss of structure in the protein and therefore a higher mobility. This behaviour perfectly matches the requirement of extensive mobility necessary to move from the rigid structure of the protein in the presence of metal to the partially unfolded apo protein prone to oligomerize at 37 °C. The behaviour here described for the WT SOD1 protein justifies the propensity for oligomerization of WT SOD1 as well as of its mutants and it is mainly due to the absence of metals.

What is not clear at this point is the influence of the mutations on the oligomerization process, i.e. at which stage and how the mutations are acting along this process to induce different oligomerization rate. Detailed analyses of the mechanism of oligomerization for each specific mutant are on the way to clear out this issue.

4.1 References

(1) Forman, H.J.; Fridovich, I. *J Biol Chem* (1973) 248, 2645–2649.
(2) Roe, J.A.; Butler, A.; Scholler, D.M.; Valentine, J.S.; Marky, L.; Breslauer, K.J. (1988) *Biochemistry* **27**, 950–958.
(3) Ross, C. A.; Poirier, M. A. (2006) *Nat. Med.* **10**, S10-17
(4) Chou, S.M.; Wang, H. S.; Komai, K. (1996) *J. Chem. Neuroanat.* **10**, 249-258.
(5) Crow, J. P.; Sampson, J.B.; Zhuang, Y.; Thomson, J.A.; Beckman, J.S. (1997) *J. Neurochem.* **69**, 1936-1944.
(6) Kim, N.H.; Kang, J.H. (2003) *J Biochem Mol Biol.* **36**, 488-92.
(7) Williamson, T.L.; Bruijn, L.I.; Zhu, Q.; Anderson, K.L.; Anderson, S.D.; Julien, J.P.; Cleveland, D.W. (1998) *Proc. Natl. Acad. Sci. U. S. A* **95**, 9631-9636.
(8) Chiti, F., Stefani, M., Taddei, N., Ramponiand, G., Dobson, C.M., *Nature* **424**, 805-808 (2003).

Thanks to:

Prof. Ivano Bertini

Who introduced me in the field of 'Structural Biology' and gave me the opportunity to carry out this Ph.D. research at CERM.

Prof. Lucia Banci

For her guidance throughout my project and for teaching me how to be efficient in scientific research.

Dr. Stefania Girotto

For her guidance throughout this project, for her support, stimulating suggestions and for the kind assistance that she showed during the revisions of this thesis.

Dr. Manuele Martinelli

For his guidance in the lab and for sharing with me his experience in the 'world of protein expression and purification'.

All the other colleagues that worked during the last three years in the molecular biology labs of Chemistry department and CERM and the administrative personnel of CERM.

Special thanks to my family:

my husband and my lovely daughter, *Alessia-Maria,*
my parents, my sister and my brother:

without their support none of this could have been possible.