University of Florence

International Doctorate in Structural Biology
Cycle XXI (2006-2008)

Metal Cofactor insertion in Cytochrome *c* Oxidase

PhD thesis of
Hadjiloi Theodoros

Tutor
Dr. Simone Ciofi-Baffoni

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

Contents

Acknowledgments

I would like to thank Prof.s Bertini, Banci and Luchinat for their support all the three years of my PhD. My tutor, Dr. Simone Ciofi for his support, guidance and help all through my work.

I would like to thank all my colleagues, friends and family for their support and help.

Theodoros Hadjiloi

Florence, 21 Dec. 08

1. Introduction

Copper in mitochondria

Copper (Cu) is an important catalytic element that drives a wide range of biochemical processes that are essential for life: energy generation, iron acquisition, oxygen transport, cellular metabolism, peptide hormone maturation, blood clotting and signal transduction [1]. The living organisms have developed a trafficking system of protein receptors and chaperons for the acquisition and disruption of copper to the target biomolecules in an efficient and safe way minimizing the negative effects of the element reactivity. Although the copper trafficking system in cytoplasm of both prokaryotic [2,3] and eukaryotic cells [4] has been extensively studied, the transfer of copper to mitochondria remains unclear.

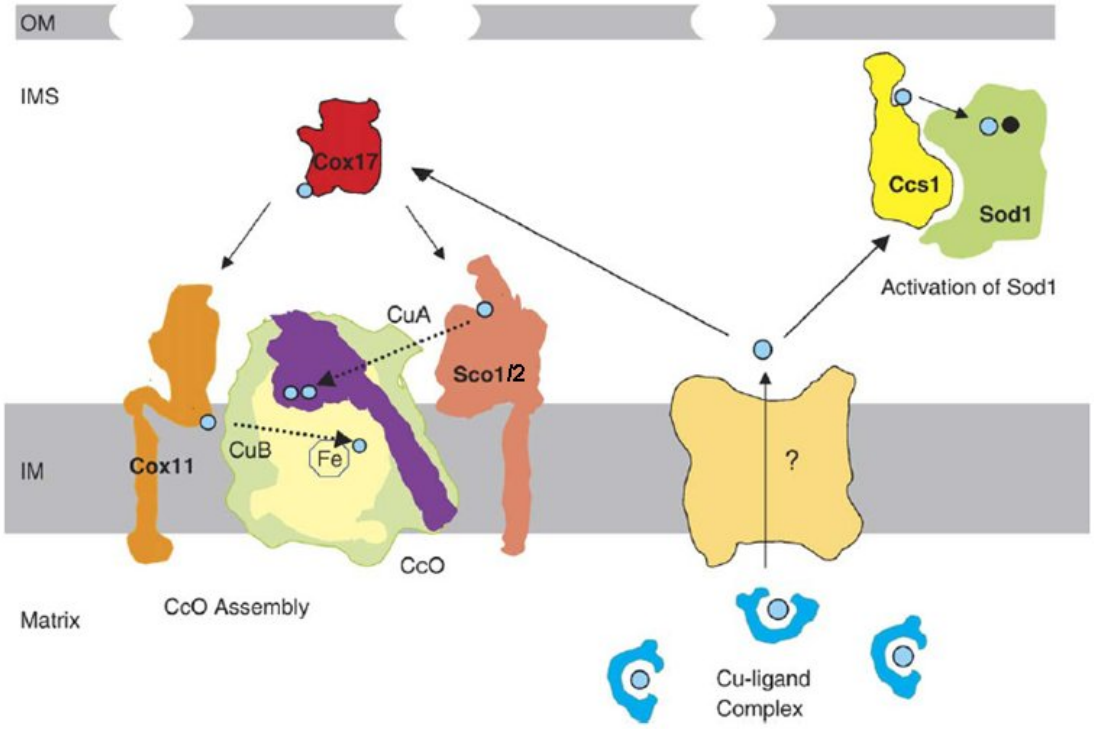


Figure 1.1: Copper delivery and metallation of CcO and Sod1 within the IMS of eukaryotic cells. CcO is metallated by a pathway that consists of Cox17 and the co-chaperones, Sco1 and Cox11. Cox17 delivers copper to Cox11 for subsequent insertion into the Cu_B site in Cox1 and to Sco1/2 proteins (see §1.2.2) for insertion into the Cu_A site of Cox2. Copper is provided for SOD1 by Ccs1. Cox17 and Cox11 assess the matrix copper pool via an unknown transporter. Reprinted from Cobine A. P. et al., BBA, 2006 [5].

Copper is required within the mitochondrion for the assembly of cytochrome *c* oxidase (CcO) and mitochondrial superoxide dismutase (Sod1). The copper-binding subunits of CcO are encoded by the mitochondrial genome, so copper insertion into CcO occurs within the organelle. Also Sod1 is imported as a metal-free polypeptide, so its activation occurs within the mitochondrial inter-membrane space (IMS) by Ccs1 (Cu chaperon of Sod). The presence of Ccs1 and two other Cu(I) chaperons (Cox17, Cox19 see §1.2.2) in both

cytoplasm and IMS make them candidates for shuttling copper ions to the mitochondrion but the unchanged levels of copper in yeast lacking the above proteins or tethering studies of the proteins in the IMS do not support this hypothesis [6-8]. Recent studies demonstrate that mitochondria serve as copper storage organelles where copper is present as soluble, low molecular weight non-proteinaceous Cu(I)-complex in the matrix. This form of copper is present also in the cytoplasm. It is suggested that this Cu(I)-complex can translocate (through an unknown transporter) from the mitochondrial matrix for subsequent use in CcO and Sod1 metallation [9] (**Fig. 1.1**).

1.1 Cytochrome *c* Oxidase

Cytochrome *c* Oxidase (CcO) is a member of a superfamily of heme-copper-containing terminal oxidases that are present in all aerobic organisms [10]. In eukaryotic cells, CcO is the last enzyme of the energy transducing respiratory chain in mitochondria that catalyses the reduction of molecular oxygen and couples this reduction with proton translocation across the inner mitochondrial membrane (IM). It consists of twelve (yeast enzyme) to thirteen (mammalian enzyme) subunits, with the three subunits (Cox1–Cox3) forming the catalytic core encoded by the mitochondrial genome (Fig. 1.2). The remaining ten subunits

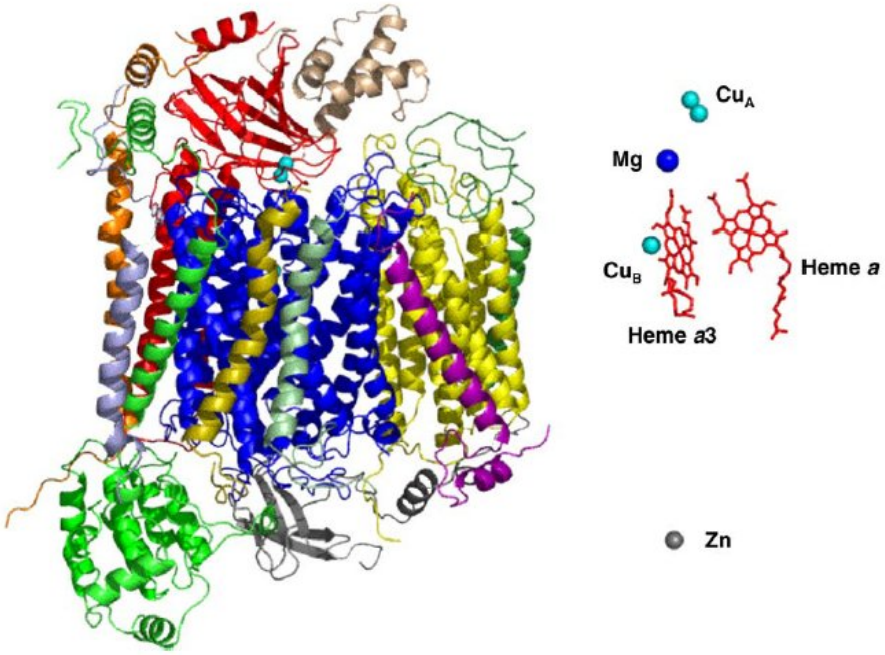


Figure 1.2: Crystal structure of bovine CcO. The mitochondrial-encoded subunits that make up the catalytic core of the enzyme are shown in blue (Cox1), red (Cox2) and yellow (Cox3). The nuclear encoded peripheral subunits are also shown. The cofactors: CuA, CuB, heme *a*, heme *a*3 and Zn cofactor are shown in space in the same orientation. Reprinted from Cobine A. P. et al., BBA, 2006 [5].

are encoded by the nuclear genome and consist of small polypeptides that surround the catalytic core forming a dimeric enzyme buried in the lipid bilayer of the IM.

CcO requires many cofactors to be a functional enzyme; two heme groups, three copper ions, as well as one zinc, one magnesium, and one sodium ion (Fig. 1.2). Both hemes are localized at the Cox1, the first heme (heme *a*) is a six coordinate heme lacking the ability to bind ligands and is functional in electron transfer, the second (heme a_3) is five coordinated heme with an open coordination where the oxygen binds. The heme a_3 interacts with a mononuclear copper site (the Cu_B site) during the catalytic reduction of the oxygen. Finally, in the Cox2 subunit is localized the Cu_A site, a copper bimetallic center which is exposed to the inter-membrane space (IMS). During the catalytic cycle, cytochrome *c* interacts with the Cu_A site, resulting a transfer of electron and oxidation of the cytochrome *c*. The formed electron gradient moves from the Cu_A site to the heme *a* and finally to the oxygen reduction center, heme a_3-Cu_B. In total, four electrons are consumed in the reduction of one oxygen molecule and four protons are used from the matrix; coupled to this reaction is the translocation of four additional protons across the membrane to the IMS. The free redox energy is converted into a proton gradient which drives the synthesis of ATP [11-13].

1.2.1 Cytochrome *c* Oxidase assembly

The CcO assembly is a complex process coordinated by two genomes (mitochondrial and nuclear) that requires the translocation and assembly of the subunits translated by the nuclear and mitochondrial ribosomes in the IM, the formation and insertion of the hemes, the delivery and insertion of the copper ions and finally the maturation of the enzyme. In this process about thirty accessory proteins have been identified necessary for the complex assembly [14].

Subunits translated in cytoplasm are imported as precursor polypeptides through the TOM complex (translocase of the outer membrane, OM) and a N-terminal mitochondrial target sequences direct these polypeptides to the TOM import pathway [15]. Protein import can also occur co-translationally from ribosomes attached to the outer mitochondrial surface at locations where the IM and OM form a contact. Translation of mitochondrially-encoded subunits (Cox1,Cox2 and Cox3) occurs within the matrix on ribosomes associated with the inner surface of the IM. Insertion of Cox1,Cox2 and Cox3 in the IM requires translocation of polypeptide segments across the IM by two conserved translocases, Oxa1 and Cox18 [16].

In mammalian cells the assembly appears to occur in an ordered sequence of subcomplexes interaction, resulting the full maturation of the enzyme (Fig. 1.3). The assembly can be divided into four steps. At the first step, the Cox1 is inserted in the IM, where the heme cofactors are inserted by the interaction with Cox10/Cox15 (heme *a* synthesis proteins [18,19]). The inner-membrane protein Surf1 (see §1.2.3) is possibly involved in heme *a* insertion or in the folding of Cox1 [20]. The first subcomplex is formed between Cox1 and Cox4. In the next step, the Cu_B-heme a_3 site of Cox1 is formed and, simultaneously, Cox2 associates with Cox1 to stabilize the active site. At this stage is believed that the Cu_A site of Cox2 is also created. A number of proteins are involved in the copper insertion at two copper sites of the CcO, including Cox11, Cox17, Sco1 and Sco2 (see. §1.2.2). In the final steps of the assembly, Cox3, Cox5a, Cox5b, Cox6b, Cox6c, Cox7c and Cox8 subunits become part of the subcomplex Cox1/Cox4/Cox2. The monomeric enzyme is formed when Cox6a and Cox7a subunits associate to Cox3 subunit. The assembly is completed with the holoenzyme dimerization [17].

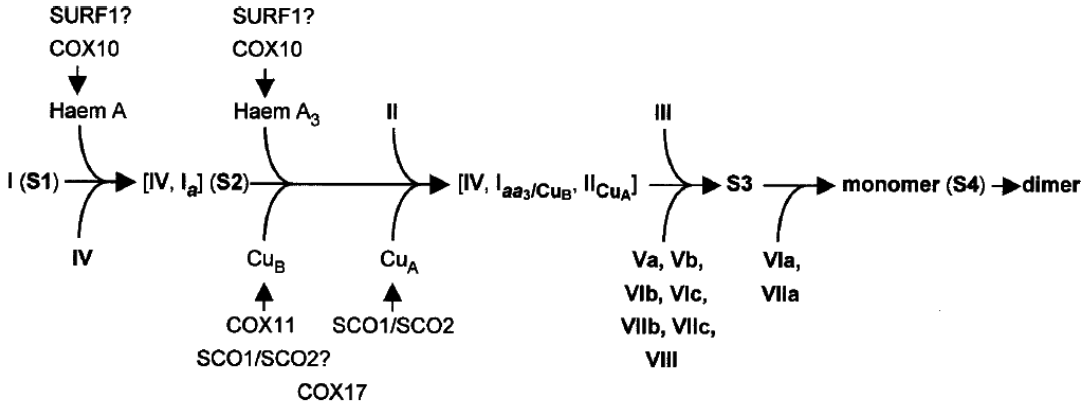


Figure 1.3: Model of the assembly pathway of human CcO. I: Cox1, II: Cox2, III: Cox3, IV: Cox4, V: Cox5, VI: Cox6, VII: Cox7, VIII: Cox8, SURF1: Surf1, COX10: Cox10, Heam A: heme *a*, Heme A_3: heme a_3, SCO:Sco. Reprinted from J.-Taanman, W and Williams, S. L., BST, 2001 [17].

1.2.2 Copper insertion to Cytochrome *c* Oxidase

Copper, zinc, and magnesium ions are necessary for CcO function. Apart from the structural role of the latter two metals, little is known about their delivery and insertion into CcO [21]. On the other hand, copper insertion in the two copper sites, Cu_A and Cu_B has been investigated in both prokaryotic and eukaryotic systems [22].

Cox11, Cox17, Sco1 and Sco2 have been recognized as the four proteins responsible for the copper insertion in an early stage of CcO assembly (Fig. 1.4). Cox17 has been

characterized as the mitochondrial Cu(I) chaperon, whereas Cox11, Sco1 and Sco2 function downstream, for the copper insertion in Cox1 and Cox2 subunits.

COX17 was first identified in yeast as gene involved in copper metabolism and assembly [23]. Yeast lacking *COX17* (*COX17*Δ yeast) are respiratory-deficient due to complete lack of CcO activity [24]. This phenotype is reversed by the addition of copper to the growth medium. This result is consistent with Cox17 functioning in copper delivery to CcO. This was strengthened by the observation that Cox17 is a copper binding protein [25]. Moreover, when as fusion protein Cox17 was directed and tethered to the IM by mitochondrial import sequence and transmembrane segment of Sco2 was able to reverse the respiratory defect of *COX17*Δ cells and restored normal CcO activity [26]. This result indicates that the copper chaperone function of Cox17 is confined into IMS [27].

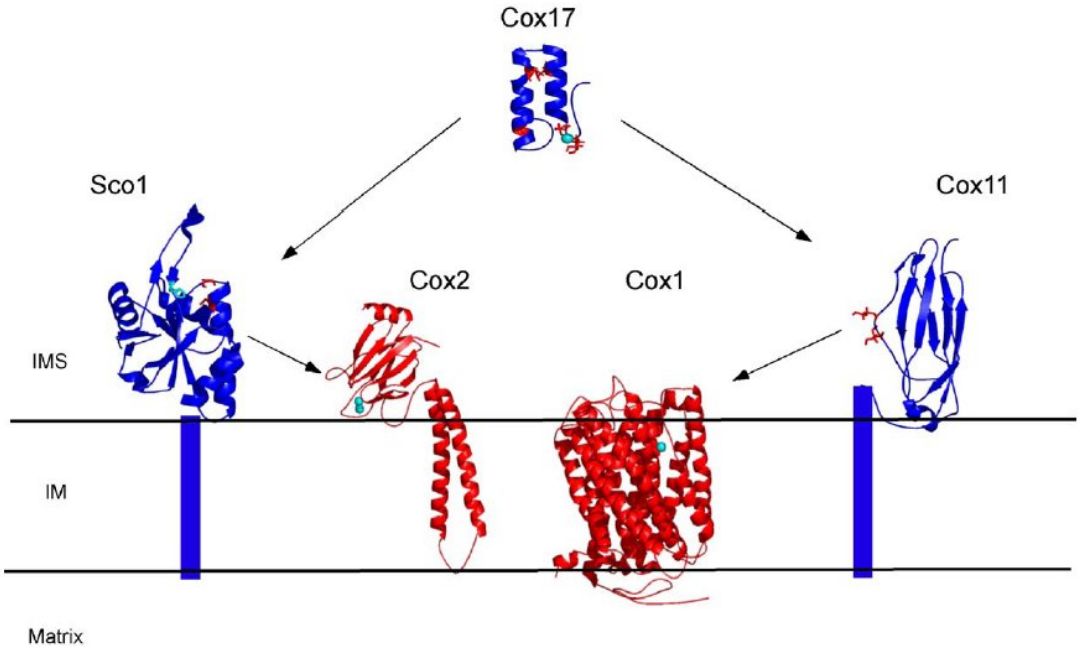


Figure 1.4: Structural representation of the copper-delivery pathway. The crystal structure of Cox1, Cox2 and Sco1 are shown with the solution structures of Cox17 and Cox11. Arrows depict the copper transfer from Cox17 to Cox11 and Sco1 and onto Cox1 and Cox2, respectively. Reprinted from Cobine A. P. et al., BBA, 2006 [5].

Cox17 contains six conserved cysteines, and has a coil-helix-coil-helix (*CHCH*) motif [28]. The four cysteines that have a CX_9C motif form two structural disulfide bonds. Cox17 was found to form two distinct Cu(I) functionally important conformers. One Cu(I) conformer consists of a single Cu(I) coordinated to a monomeric protein stabilized by two disulfide bonds and a second Cu(I) conformer is an oligomeric protein complex coordinating a polycopper–thiolate cluster. Interestingly, *in vitro* experiment showed that only the monomeric conformer can transfer Cu(I) to Sco1, suggesting that this form is the active state in the copper transfer within the IMS [29].

The translated Cox17 is detected both in cytoplasm and IMS [6]. This dual localization of Cox17 led first to the suggestion that Cox17 can transfer Cu(I) ions across the OM. However, *COX17*Δ cells are not deficient in mitochondrial Cu, suggesting that Cox17 is not the primary mitochondrial copper shuttle. The metallation pathway of this chaperon is still unclear. The recent discovery of the matrix copper pool (as soluble, low molecular weight complex) suggests a model where copper is delivered from matrix to IMS [8]. Cox17 is suggested to be translated in the cytoplasm with a molten globule structure of a fully reduced state [30]. The protein translocation in the IMS follows the folding traps pathway [31] similarly to other small proteins that are localized in the IMS. Following its translocation across the OM, the precursor form is bound by the IM protein Mia40 catalyzing the formation of the structural disulfide bonds [32,33].

Sco1 was first implicated in copper delivery to CcO by the observation that the respiratory deficient phenotype of a Cox17 yeast mutant (Cox17 harboring a Cys → Tyr substitution) was suppressed by overexpression of *SCO1* or a related gene designated *SCO2* [34].Yeast strains lacking *SCO1* are also respiratory deficient, and an excess of copper and/ or overexpression of either Cox17 or Sco2 can not compensate for the Sco1-associated CcO deficiency [35]. The requirement of Sco1 in the activation of CcO indicates that Sco1 functions downstream from Cox17 in the delivery of copper to CcO. Indeed, *in vitro* showed that Sco1 can receive copper from the copper chaperone Cox17 [36]. A transient complex with Cox2 with yeast Sco1 confirmed the importance of Sco1 in formation of the Cu_A site [37].

Sco1 and Sco2 are highly similar proteins but their distinct role in eukaryotic cells is still unclear. Both proteins contain a single transmembrane α-helix at the N-terminal segment of the proteins, anchoring these two proteins to the inner mitochondrion while the C-terminal segment of Sco proteins is soluble and extended into IMS. Copper ion is bound by the two cysteins of the highly conserved CXXXC motif and His residue [38]. Mutation of any one of these conserved residues in Sco1 abrogates function, resulting in a nonfunctional CcO complex, which means that the function of Sco1 correlates with Cu(I) binding. Moreover, studies with immortalized fibroblasts from Sco1 and Sco2 patients suggest that human Sco1 and human Sco2 have non-overlapping but cooperative functions in CcO assembly [39]. Recently, it was shown that human Sco2 is also the downstream mediator of the balance between the utilization of respiratory and glycolytic pathways and has a regulatory role in the maintenance of cellular copper homeostasis [40]. In human patients with mutations in the Sco2 gene have a clinical phenotype distinct from that of Sco1 patients (Sco2 patients present with neonatal encephalocardiomyopathy, whereas

Sco1 patients exhibit neonatal hepatic failure), even if Sco1 and Sco2 are ubiquitously expressed and exhibit a similar expression pattern in different human tissues [41].

Structurally both proteins are characterized by a thioredoxin fold and with a single Cu(I) ion bound to be solvent exposed and poised for a ligand exchange transfer reaction. The higher structural flexibility of Sco2 compared to Sco1, especially in its apo form, is suggested to be important for different protein-protein interaction pathways [42]. In addition to Cu(I) binding, Sco proteins bind also Cu(II) [42-44]. The Cu(II) site has a higher coordination number than the three-coordinate Cu(I) site [45]. Replacement of Asp238 in yeast Sco1 abrogated Cu(II) coordination and led to a non-functional protein [46]. These data suggest that copper binding can be critical for normal Sco1 function.

They are mainly three scenarios concerning the function of the Sco based on different observations: 1) Human Sco function as a complex to transfer two coppers to the Cu_A site. This is supported by the dominant-negative effect on CcO activity of the overexpression of either gene in the reciprocal patient cell line and that the full-length proteins are homo-oligomers in vivo [37,47]. It is possible that a Sco1/Sco2 complex may induce one protein to catalyze the oxidation of a Cu(I) to Cu(II) enabling the complex to simultaneously transfer both Cu(I) and Cu(II). Alternatively, an additional assembly factor may be required for Cu(I) oxidation. Otherwise, two successive Cu(I) transfers yield a reduced Cu_A site that can be converted to the oxidized, mixed valent site through electron transfer within the assembly complex [46]. 2) Sco were suggested to function as a redox switch, in which oxidation of Cu(I) to Cu(II) induces release of the Cu(II) ion thereby permitting the two thiolates to participate in a peroxidase reaction [45]. Since Cu(II) is stably bound by Sco, a role for Sco proteins as peroxidases is not compelling. 3) The structural resemblance of Sco1 to peroxiredoxins and thioredoxins raised the possibility that Sco proteins may function as a thiol:disulfide oxidoreductase to maintain the Cu_A site cysteines in the reduced state ready for metallation [45]. The disulfide exchange with Cu(I) insertion between oxidized Cu_A site and Cu(I)Sco, was suggested by the observation of Ni(II) located at the binding site of the human Sco crystal structure with oxidized cysteines [48]. Recently, has been shown that indeed in prokaryotic organisms, Sco1 is able to reduce oxidized Cu_A, while does not transfer Cu(I) to it. Thus its role is, in prokaryotic organisms, that of assisting the copper transfer process maintaining the copper bridging Cys of Cu_A in a reduced state [49].

Cox11 is the co-metallochaperone for the Cu_B site (Fig. 1.4). Recently, was demonstrated that the Cu_B site is absent in CcO purified from *Rhodobacter sphaeroides* lacking *Cox11* gene [50]. Moreover, it is shown that some Cox11 mutants lack CcO activity;

however, RNA and protein synthesis of the core subunits Cox1 and Cox2 are normal, suggesting that Cox11 functions post-translationally to generate active CcO [51]. Yeast Cox11 is a 34 kDa protein that resembles Sco1, with a single transmembrane α-helix downstream of the N-terminal mitochondrial targeting sequence. The C-terminal domain exposed to IMS, similarly to the soluble C-terminal domain of Sco1, binds one Cu(I) ion. There is only one structure available of soluble domain of Cox11 (from *Sinorhizobium meliloti*) showing that the domain is organized into a immunoglobulin-like fold and Cu(I) ion is bound to the putative copper binding motif CFCF (Fig. 1.3) [52]. Similarly for Sco, the mechanism of the insertion of copper to Cu_B by Cox11 is an unresolved question especially how the Cu_B site is formed when is buried 13 Å below the membrane surface.

1.2.3 The role Surf1 protein in CcO assembly

The human *SURF1*, the first gene of the surfeit gene locus (a highly conserved cluster of six housekeeping genes) codes for a 30 kDa protein that is involved in CcO assembly. Located in the IM, Surf1 contains two transmembrane α-helices and a large connecting domain facing the IMS [53]. Mutations in the *SURF1* are responsible for the Leigh syndrome, a fatal neurological disorder associated with severe CcO deficiency, where all tissues show a loss of CcO activity between 80–90% [54]. Sequence alignments revealed many Surf1 homologues in eukaryotes and prokaryotes. They show a striking degree of homology, and all proteins exhibit the same characteristic transmembrane α-helix topology and two highly conserved regions in the domain facing the IMS [55].

The yeast Surf1 homologue, Shy1 has been discovered and characterized in connection with yeast mutants and deletions of the gene leading to reduced oxidase activity, likely as the amount of CcO is highly decreased in mutants. The remaining enzyme behaves enzymatically and structurally like wild type, suggesting Shy contributes to the assembly or the stabilization of CcO [56]. Further experiments on *SHY1* mutants revealed that specifically steady-state levels of Cox1 were decreased, suggesting that Shy1 promotes the formation of an assembly intermediate consisting of Cox1/Cox2 [57]. Moreover, it has been shown that Shy1p interacts with Mss51 (a transcriptional activator of Cox1), Cox14 (CcO assembly factor) and associates with CcO subunits in the assembly subcomplexes [58]. In bacteria, the effect of Surf1 homologue gene deletion was studied in *Paracoccus denitrificans* and *Rhodobacter sphaeroides*. The former bacterium contains two Surf1 homologues (Surf1c and Surf1q) that act independently for the assembly of the aa_3-type CcO and the ba_3-type quinol oxidase, respectively. The latter bacterium synthesizes an aa_3-type CcO. The biochemical features of the individual assembled oxidase

suggest that Surf1 plays a role in facilitating the insertion of heme a_3 (direct or indirect) into the Cox1 where the formation of the heme a_3-Cu_B center is a limiting step for the interaction of Cox1/Cox2 [59,60].

1.3 Aims and topics of the research

In my three-year PhD course, research was focused on the investigation of the function of proteins involved in the assembly and insertion of metal cofactor in an early state of CcO maturation. In particular, I focused my attention on metalloproteins involved in the formation of the Cu_A site and on Surf1 protein that is suggested to play a role in the folding and insertion of heme a_3 in Cox1 subunit.

The mitochondrial metallochaperon Cox17, transfers copper to both Sco1 and Sco2 proteins for subsequent insertion of Cu(I) ions into the Cu_A. In order to clarify the copper exchange mechanism between the Cox17 and the Sco proteins, we investigated their interactions as far as copper and electron exchange. The scenario where the two Sco proteins oligomerize and interact forming a complex able to facilitate two copper ions into the Cu_A site was also investigated. Moreover, the full length Sco1 and Sco2 α-helical trasmembrane proteins were subjected to expression and purification trail.

Surf1 is a protein showing sequence conservation from human to prokaryotes, involved in the assembly of CcO. Apart from observations concerning the relationship of the Surf1 with the Leigh syndrome, interaction with Cox1 or other subcomplexes, the role of the protein is still being speculative and there is no structural information. In order to structurally characterize Surf1, the soluble domains from human and bacteria as well as the full length protein containing the two transmembrane α-helices, were produced and purified.

1.4 Reference list

1. Kim BE, Nevitt T, Thiele DJ. Nat Chem Biol. (**2008**) 4: 176-85.
2. Cavet JS, Borrelly GP, Robinson NJ. FEMS Microbiol Rev. (**2003**) 27:165-81.
3. Gaballa A, Helmann JD.Biometals. (**2003**):497-505.
4. Culotta VC, Lin SJ, Schmidt P, Klomp LW, Casareno RL, Gitlin J. Adv Exp Med Biol. (**1999**)448:247-54.
5. Cobine PA, Pierrel F, Winge DR.Biochim Biophys Acta. (**2006**) 1763:759-72
6. Pierrel F, Cobine PA, Winge DR. Biometals. (**2007**) 20:675-82
7. Rigby K, Zhang L, Cobine PA, George GN, Winge DR. J Biol Chem. (**2007**) 282:10233-42
8. Maxfield AB, Heaton DN, Winge DR.J Biol Chem. (**2004**)13:5072-80.
9. Cobine PA, Pierrel F, Bestwick ML, Winge DR. J Biol Chem. (**2006**) 281:36552-9
10. Castresana J, Lübben M, Saraste M, Higgins DG. EMBO J. (**1994**)13:2516-25
11. Adachi K, Oiwa K, Nishizaka T, Furuike S, Noji H, Itoh H, Yoshida M, Kinosita K Jr. Cell. (**2007**)130:309-21.
12. Khalimonchuk O, Rödel G. Mitochondrion. (**2005**) 5:363-88
13. Richter OM, Ludwig B. Rev Physiol Biochem Pharmacol. (**2003**)147:47-74
14. Carr HS, Winge DR. Acc Chem Res. (**2003**) 36:309-16.
15. Neupert W, Herrmann JM. Annu Rev Biochem. (**2007**) 76:723-49.
16. Szyrach G, Ott M, Bonnefoy N, Neupert W, Herrmann JM.. EMBO J. (**2003**) 22:6448-57.
17. Taanman JW, Williams SL. Biochem Soc Trans. (**2001**) 29:446-51.
18. Antonicka H, Leary SC, Guercin GH, Agar JN, Horvath R, Kennaway NG, Harding CO, Jaksch M, Shoubridge EA..Hum Mol Genet. (**2003**) 12:2693-702.
19. Antonicka H, Mattman A, Carlson CG, Glerum DM, Hoffbuhr KC, Leary SC, Kennaway NG, Shoubridge EA..Am J Hum Genet. (**2003**) 72:101-14
20. Smith D, Gray J, Mitchell L, Antholine WE, Hosler JP.. J Biol Chem. (**2005**)280:17652-6
21. Pan LP, He QZ, Chan SI.. J Biol Chem. (**1991**) 266:19109-12.
22. MacPherson IS, Murphy ME.. Cell Mol Life Sci. (**2007**) 64:2887-99.
23. Glerum DM, Shtanko A, Tzagoloff A. J Biol Chem. (**1996)** 271:14504-9.
24. Glerum, D. M.; Shtanko, A.; Tzagoloff, A.; (**1996**) J. Biol. Chem. 271 20531–20535
25. Heaton, D.N.; George, G.N.; Garrison, G.; Winge, D.R. (**2001**) *Biochem* 40, 743–751
26. Maxfield, A.B.. Heaton, D.N.; Winge, D.R. (**2004**) *J. Biol. Chem.* 279, 5072–5080.
27. Horng, O.C.; Cobine, P.A.; Maxfield, A.B.; Carr H.S.; Winge, D.R. (**2004**) *J. Biol. Chem.* 279, 35334–35340.
28. Banci L, Bertini I, Ciofi-Baffoni S, Janicka A, Martinelli M, Kozlowski H, Palumaa P..J Biol Chem. **(2008**) 283:7912-20.
29. Banci, L., Bertini, I., Ciofi-Baffoni, S., Leontari, I., Martinelli, M., Palumaa, P., Sillard, R., and Wang, S. Proc. Natl. Acad. Sci. U. S. A. (**2007**) 104: 15–20
30. Arnesano F, Balatri E, Banci L, Bertini I, Winge DR. Structure. (**2005**)13:713-22.
31. Neupert, W and Herrmann, JM. Annual Review of Biochemistry, (**2007**) 76:723-749
32. Herrmann JM, Köhl R. J Cell Biol. (**2007**) 176:559-63.
33. Hell K. Acta. (**2008**)1783:601-9

34. Schulze M, Rödel G..Mol Gen Genet. (**1988**) 211:492-8.
35. Tzagoloff, A.; Dieckmann, C. L; (**1990**) *Microbiol. Rev.* 54 211–225
36. Glerum, D. M.; Shtanko, A.; Tzagoloff, A.; (1996) *J. Biol. Chem.* 271 20531–20535
37. Lode A, Paret C, Rödel G. Yeast. (**2002**) 19:909-22.
38. Nittis T, George GN, Winge DR..J Biol Chem. (**2001**)276:42520-6
39. Leary, S.C.; Kaufman, B.A.; Pellecchia, G.; Guercin, G.H.; Mattman, A.; Jaksch, M.; and Shoubridge, E.A.; (**2004**) *Hum. Mol. Genet.* 13 1839–184
40. Matoba, S.; Kang, J.G.; Patino, W.D.; Wragg, A.; Boehm, M.; Gavrilova, O.; Hurley, P.J.; Bunz, F.; Hwang, P.M. (**2006**). *Science* 312 1650–1653.
41. Shoubridge EA..Am J Med Genet. (**2001**) 106:46-52.
42. Banci L, Bertini I, Ciofi-Baffoni S, Gerothanassis IP, Leontari I, Martinelli M, Wang S.Structure. (**2007**)15:1132-40.
43. Beers J, Glerum DM, Tzagoloff A. J Biol Chem. (**2002**) 277:22185-90
44. Balatri E, Banci L, Bertini I, Cantini F, Ciofi-Baffoni S..Structure. (**2003**) 11:1313-4.
45. Williams JC, Sue C, Banting GS, Yang H, Glerum DM, Hendrickson WA, Schon EA. J Biol Chem. (**2005**) 280:15202-11
46. Horng YC, Leary SC, Cobine PA, Young FB, George GN, Shoubridge EA, Winge DR.J Biol Chem. (**2005**) 280:34113-22
47. Leary SC, Kaufman BA, Pellecchia G, Guercin GH, Mattman A, Jaksch M, Shoubridge EA., Hum Mol Genet. (**2004**) 17:1839-48
48. Banci L, Bertini I, Calderone V, Ciofi-Baffoni S, Mangani S, Martinelli M, Palumaa P, Wang S. Proc Natl Acad Sci U S A. (**2006**)103:11814.
49. Abriata LA, Banci L, Bertini I, Ciofi-Baffoni S, Gkazonis P, Spyroulias GA, Vila AJ, Wang S..Nat Chem Biol. (**2008**) 4:599-601.
50. Hiser, L.; Di Valentin, M.; Hamer, A.G.; Hosler, J.P. (**2000**) *J.Biol.Chem.* 275 619-623
51. Tzagoloff, A.; Capitanio, N.; Nobrega, M.P.; Gatti, D. (**1990**) *EMBO J.* 9 2759-2764
52. Banci L, Bertini I, Cantini F, Ciofi-Baffoni S, Gonnelli L, Mangani S.. J Biol Chem. (**2004)** 13:34833-9
53. Péquignot MO, Dey R, Zeviani M, Tiranti V, Godinot C, Poyau A, Sue C, Di Mauro S, Abitbol M, Marsac C..Hum Mutat. (**2001**)17:374-81.
54. Zhu Z, Yao J, Johns T, Fu K, De Bie I, Macmillan C, Cuthbert AP, Newbold RF, Wang J, Chevrette M, Brown GK, Brown RM, Shoubridge EA.. Nat Genet. (**1998**) 20:337-43
55. Poyau A, Buchet K, Godinot C. FEBS Lett. (**1999**) 462:416-20.
56. Nijtmans, L.G.. Artal Sanz, M. Bucko, M M.H. Farhoud, M. Feenstra, G.A. Hakkaart, M. Zeviani, L.A. Grivell, FEBS Lett. (**2001**) 498:46–51
57. Barrientos, A. D. Korr, A. Tzagoloff, EMBO J. (**2002**) 21:43–52.
58. Barrientos, A. Zambrano, A. A. Tzagoloff, , EMBO J. (**2004**) 23:3472–3482.
59. Bundschuh FA, Hoffmeier K, Ludwig B.. Biochim Biophys Acta. (**2008**) 1777:1336–1343
60. Smith D, Gray J, Mitchell L, Antholine WE, Hosler JP..J Biol Chem. (**2005**)280:17652-6

2. Methodological aspects

2.1 Bioinformatics for protein expression

Bioinformatics has an essential role in understanding genomic and proteomic data, and in organizing experimental data. It is an interdisciplinary research area at the interface between the biological and computational science.

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

a) **NCBI**: is a web site which integrates information from several databases (Swissprot, EMBL and all geneBank). [www.ncbi.nlm.nih.gov/Entrez/],

b) **PDB:** A 3-D biological macromolecular structure database. [www.rcsb.org/pdb], c) Pfam: is a collection of different protein families organized in terms of different domains as obtained from multiple alignment. [http://pfam.wustl.edu],

d) **PROSITE**: is a database of protein families and domains. It is based on the observation that, while there is a huge number of different proteins, most of them can be grouped, on the basis of similarities in their sequences, into a limited number of families. Proteins or protein domains belonging to a particular family generally share functional attributes and are derived from a common ancestor. [www.expasy.org/prosite/],

e) **STRING**: is a database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations [www.dag.embl-heidelberg.de].

All bioinformatics data is stored in the above databases and can be browsed by bioinformatics tools for extracting information of a given target or targets. The most used programs for gene or protein browsing are:

a) **BLAST** (The Basic Local Alignment Search Tool): finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families. [www.ncbi.nlm.nih.gov/BLAST/].

b**) CLUSTALW**: is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. [www.ebi.ac.uk/clustalw/]

Once selected a protein target of particular interest is subjected to further bioinformatics investigations. In case where the target will be used for recombinant protein expression the following characteristic of the protein can predicted and be used for defining constructs:

a) **Post-translational modification prediction**: signal peptides, glycosylation, phosphorylation and cleavage sites: TargetP [1], SIGNALP [1].

b) **Topology prediction of transmembrane (TM) regions**: prediction of membrane-spanning regions and their orientation: TMHMM [2]; TMpred [3], HMMTOP [4].

c) **Secondary structure prediction**: of the regions with tendency to form helices, beta-sheet, coil or turns. Combining these predictions with homologue proteins alignment, can be used to select the exact residues that define a construct minimizing the possibility of destroying a region with certain secondary structure. NPS@ [5].

d) **Prediction of intrinsically unstructured /disordered regions**: recognition of such regions from the amino acid sequence based on the estimated pairwise energy content. The underlying assumption is that globular proteins are composed of amino acids which have the potential to form a large number of favorable interactions, whereas intrinsically unstructured proteins adopt no stable structure because their amino acid composition does not allow sufficient favorable interactions to form. IUPred [6].

e) **Rare codon calculation**: Over-expression of recombinant protein in *E. coli* may be severely diminished (to the point of being undetectable) if the ORF that codes for the protein uses "rare" codons that are infrequently used by *E. coli*. In particular, codons for arginine (AGG, AGA, CGA), leucine (CTA), isoleucine (ATA), and proline (CCC) may be a problem. RaCC [7].

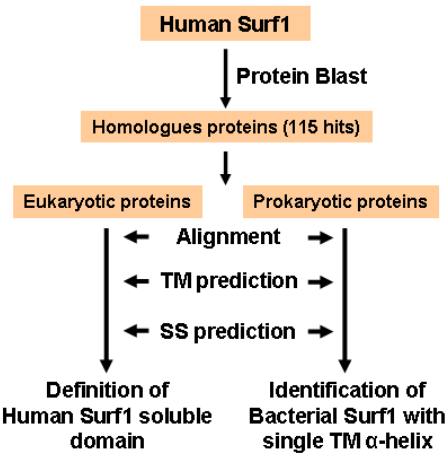


Figure 2.1: Flow chart for definition of human Surf1 protein soluble domain and identification of bacterial Surf1 homologues with single TM α-helix. Abbreviations: TM, transmembrane, SS: secondary structure

In the present thesis, Surf1 protein was chosen as a target (see §1.3). Before cloning and production of the protein was necessary a bioinformatics investigation. In the Fig. 2.1 is shown a flow chat of the general idea followed in order to define the soluble domain of the protein. Using a multi alighment of the eukaryotic proteins and their secondary structure (SS) prediction in

combination with the predicted TM regions, the soluble domain of human protein between the two TM regions (one at each terminus) was predicted from 74Q-294Y (see §3.4). For the expression of the mature human Surf1 (see §3.4) lacking the mitochondrial signal peptide, the cleavage site was predicted at the 45C residue. In order to select a bacterial Surf1 for expression, we browsed the homologues proteins looking for proteins with a single TM region. Under this criterion three bacterial proteins where identified: 1) Caulobacter crescentus: single TM at C-terminus, 28.18% identity to the human Surf1 2) Methylobacterium chloromethanicum (CC): single TM at C-terminus , 25.88% identity to the human Surf1 Sinorhizobium meliloti (SM): single TM at N-terminus , 22.92% identity to the human Surf1. The two of the three proteins (CC, SM) were further studied for the definition of the soluble domain similarly to the human Surf1. Moreover, the human Sco proteins where also investigated as far as the definition of the full length constructs (see §3.4) but also for the sequence conservation of the transmembrane α-helix and charged region at the C-terminal of the transmembrane α-helix (see §3.3).

2.2 Gene cloning

Gene cloning involves molecular biology techniques for the amplification of a DNA construct, purification of DNA construct or vectors, insertion of gene to vectors and finally sequencing of the gene for verifying the correct insertion and absence of mutations.

The first step of the cloning process consists in the amplification of target gene from cDNA (synthetic or extracted from the organism of interest) or plasmids via PCR (Polymerase Chain Reaction) using primers. After purification, the amplified product is inserted into a specific expression vector using one of the methods described below.

The classic method to insert a gene in a vector is using restriction enzymes that cleave DNA at specific recognition sites. To be able to clone a DNA insert into a cloning or expression vector, both have to be treated with two restriction enzymes that create compatible ends. The next step is the ligation of the insert into the open vector. This involves the formation of phosphodiester bonds between adjacent 5'-phosphate and 3'-hydroxyl residues, a reaction catalyzed by the bacteriophage T4 DNA ligase. Finally an aliquot of this reaction is transformed in DH5α *E. coli* competent cells and positives clones are screened both by PCR and DNA sequencing. The yield of this procedure depends on the efficiency of the ligation step which is affected by the DNA concentration, the vector/DNA insert ratio and the transformation efficiency of the cells. Moreover the

parallelization of this approach is not so easy because not all the host vectors and the target genes can be cleaved by the same restriction enzymes.

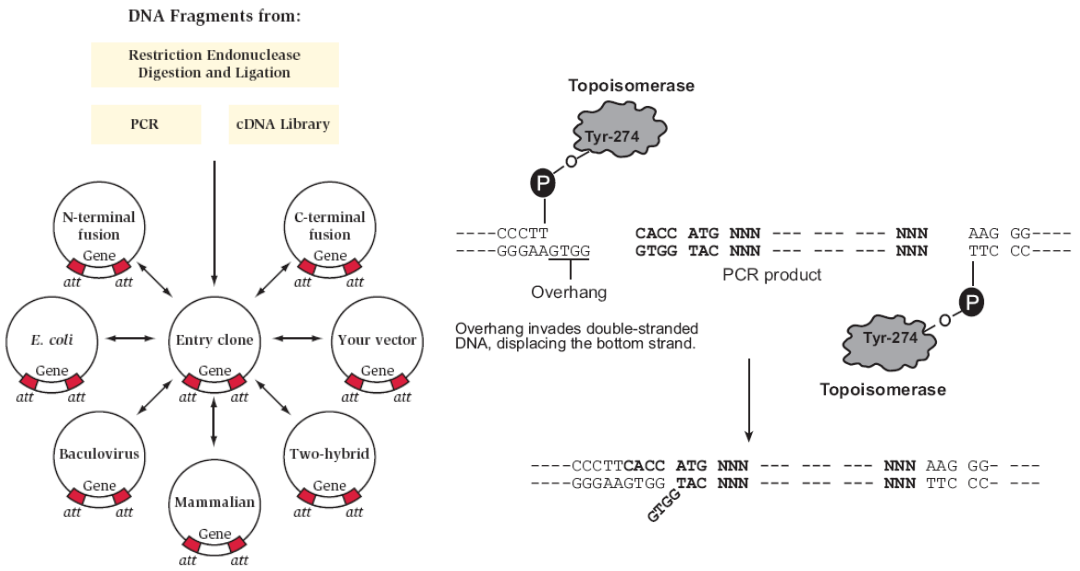


Figure 2.2: **(left)** The flexibility of the Gateway Technology. After cloning the gene of interest into the Entry vector, you can transfer the gene of interest simultaneously into multiple destination vectors. (**right**) The mechanism of the gene ligation by Topoisomerase I. Reprinted from Gateway Technology handbook.

The other cloning strategy exploits the Gateway Cloning System (Fig. 2.2) from Invitrogen, a universal cloning method that takes advantage of the site-specific recombination properties of bacteriophage lambda [8]. Mainly it consists in the creation of a common entry vector that can be recombined in several compatible vectors without the use of any restriction enzyme. The entry vector can be obtained via a ligation reaction of the entry vector with DNA construct amplified; the reaction is catalyzed by the topoisomerase I (Fig. 2.2). After isolation of the entry vector, the second step is to generate an expression vector by performing a recombination reaction between the entry clone and a Gateway destination vector of choice.

In the present thesis as cloning system was used mainly the Gateway Cloning System, by insertion of the DNA constructs in the pENTR/TEV/D-TOPO vector and recombination to expression vectors: pETG-30A (His-GST), pDEST-17/O (His), pDEST-20 (His-Trx), pDEST-His-MBP, pDEST-His-MBPperi, pETG-60A (His-NusA) and pTH34 (His-GB1). The HSco1L (see §3.3) construct was amplified by PCR using the cDNA of human *SCO1* (purchased from RZDB, Germany). The construct was inserted in Gateway Entry vector pENTR/TEV/D-TOPO (Invitrogen) and sequence was verified by sequencing at PRIMM (PRIMM, Italy). The pENTR clone was recombinated with pETG-30A (6His-GST). A

similar procedure was followed for the cloning of the full length proteins Sco1, Sco2 and Surf1 (see §3.4).

2.3 Recombinant protein production

Several host systems are available for protein production, including bacteria, yeast, plants, fungi, insect and mammalian cells. The choice strongly depends upon the specific requirements of the target protein.

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

2.3.1 Recombinant proteins production in *E.coli*.

Similarly to gene cloning, also for protein production the best strategy is the use of a parallel expression strategy; depending on the number of targets the possible choices are: (i) to express more genes in the same bacterial strain, (ii) to express a single gene in different strains or (iii) to express more genes in different strains.

A preliminary expression test is performed in a small-volume scale using three strains (e.g.: Bl21(DE3)pLysS a protease deficient strain, Rosetta(DE3) for rare codons containing genes and Origami(DE3) for disulphide containing proteins), three expression temperatures (37-25-17°C), three inducer (arabinose or isopropyl-beta-D-thiogalactopyranoside) concentrations, different induction times (4h and 16h) and three culture media of different composition. Expression results are checked on SDS polyacrylamide gel (SDS-PAGE). This kind of approach allows to explore a large set of expression conditions and to evaluate which one gives the best yield of soluble protein.

In case of negative or partially positive results some or all variables can be modified, in particular the more influencing ones like the choice of bacterial strain, the induction times, the kind of vectors and expression promoters used. If the main fraction of the protein is produced in the insoluble fraction, refolding trials can be performed (see §2.5). The last choices are to redesign the expressed domains or to use another expression system.

In dependence of the spectroscopic technique of choice, protein expression is performed in differently composed media. In fact, when large amounts of proteins must be isolated for techniques that do not require isotopic labelling, the culture is usually performed in a so-called rich or complex medium. Rich media contain water soluble extracts of plant or animal tissue (e.g., enzymatically digested animal proteins such as peptone and tryptone), and for this reason are rich in nutrients and minerals, assuring a fast bacterial growth and a high expression level. Their exact composition is unknown and this can impair the reproducibility of cultures. Chemically defined (or minimal) media are composed of pure ingredients in measured concentrations, dissolved in purified water; this way the exact chemical composition of the medium is known, allowing high reproducibility of protein yields. Typically, this class of media is composed of a buffering agent to maintain culture pH around physiological values, a carbon source like sugar (glucose) or glycerol, and an inorganic nitrogen source, usually an ammonium inorganic salt. Depending on the bacterial strain and the expressed proteins various mineral salts can be added and, if necessary, growth factors such as purified amino acids, vitamins, purines and pyrimidines. Chemically defined media are easier to isotopically enrich, simply by using ^{15}N and ^{13}C enriched nitrogen and carbon sources.

In the current thesis, the soluble domain of the HSco1 (two constructs, see §3.3) and HSco2 (two construct, see §3.3) were expressed in BL21-Gold(DE3) and BL21-(DE3) Codonplus, respectively. Growth conditions were the same to the ones already reported [9].

2.3.2 Integral membrane proteins production in *E.coli*

Membrane protein production in *E. coli* can take place in at least two states: a functional state involving protein insertion into the bacterial membrane or in an inactive state where the protein usually exists in inclusion bodies as an insoluble aggregate. In cases where fusion tag is used, the protein can be also soluble.

Targeting expression of hydrophobic membrane proteins into inclusion bodies would be preferable to membrane insertion as protein purification from membranes can be labor-intensive. Furthermore, targeting membrane proteins to inclusion bodies also has many advantages over membrane insertion. Inclusion bodies may serve to shield cells from

toxic effects of the expressed protein, resulting in higher protein yields. Also, as the expressed protein is the primary component of the inclusion body, protein purification is simplified to the isolation of the inclusion bodies. However, the major downside of this expression method is that following purification, the protein must be refolded to its native state. In addition, lipids or detergents may be required during purification to maintain protein solubility due to the hydrophobic nature of these proteins, which can impose complications to purification procedures [10].

A bacterial expression procedure used with considerable success is expression of membrane proteins as fusion constructs. The *E. coli* maltose-binding protein (MBP) and thioredoxin (Trx) along with glutathione *S*-transferase (GST) from *Schistosoma japonicum* are the most widely used fusion systems and can facilitate solubility, limit toxicity, decrease proteolysis, and aid in purification of proteins. While GST and MBP are advantageous as fusion partners because of their high bacterial expression levels and affinity purification capacity, Trx is considered a useful fusion protein for hydrophobic proteins because of its high aqueous solubility [11].

Altering the expression medium, temperature and concentration of inducer can also aid membrane protein expression.

The commercially available BL21 *E. coli* is the most commonly used bacterial host for heterologous expression as it is protease deficient and is known to promote plasmid stability. BL21 derivatives with considerable success at overexpressing membrane proteins normally toxic to the cells are the Avidis C41 (DE3) and C43 (DE3) strains (Saint-Beauzire, France). While the genetic mutations resulting in improved protein expression is unknown, current hypothesis suggest the mutations performed may affect the amount of T7 RNA polymerase production [12].

In the present thesis, production of transmembrane proteins (HSurf1, HSco) was detected in the soluble fraction, in the cell membranes and in inclusion bodies. The production in the cell membranes and the conditions used are described in the §3.4.

2.4 Protein purification

The strategy of purification depends mainly upon the localization of the produced protein within the host; for example the protein can be secreted into the growth media, transported in the periplasmic space or produced like a soluble or insoluble (inclusion bodies, IBs) protein within the cytoplasm and finally directed to the cell membranes. In each of the cases the isolation has to be performed in different ways.

Cell lysis is the first step in protein purification and the technique chosen for the disruption of cells must be compatible with the amount of material to be processed and the intended downstream applications. Many techniques are available, spanning from physical to detergent-based methods. Among the first class are mechanical disruption**,** liquid-based homogenization, such as the French press, where cells are lysed by forcing the suspension through a narrow space, and sonication, which uses pulsed, high frequency sound waves to agitate and lyse cells. Gentler ways of disrupting cells are freeze-thaw lysis, detergent lysis, enzymatic lysis and osmotic lysis. The latter method is well suited when the protein is expressed in the periplasm of the cell. The big advantage of this technique is that all the interferents coming from the cytoplasmic region are eliminated, since the inner membrane is not broken, preventing also genomic DNA and proteases release.

The following step of purification procedure is composed of normally two or more rounds of chromatography, playing on the different physical chemical and biological characteristics of the protein. A powerful chromatographic technique is Ion exchange chromatography (IEX) that separates the proteins on the basis of a reversible interaction between the polypeptide chain and a specific, charged ligand attached to a chromatographic matrix. The sample is applied in conditions that favour specific binding, e.g. low ionic strength of the solution and carefully calibrated pH, in order to enhance interaction with the target of interest. The unbound material is washed away, then the bound protein is recovered by changing conditions to those favouring desorption. This can be achieved by changing pH or ionic strength of the eluant. The most used resin is the anionic one, since the majority of proteins carry an overall net negative charge at physiological conditions.

Size-exclusion chromatography (SEC) is the simplest and mildest of all the chromatographic techniques; the support for gel filtration chromatography is composed of beads which contain holes, called "pores," of given sizes. Larger molecules, which can't penetrate the pores, move around the beads and migrate through the spaces which separate the beads faster than the smaller molecules, which may penetrate the pores; in this technique the buffer composition does not play any role in the resolution. SEC can be performed on a “rough” level, separating the components of a sample in major groups to remove for example high or low molecular weight contaminants or to exchange buffers, while high resolution fractionation of biomolecules allows to isolate one or more components of a protein mixture, to separate monomers from aggregates, to determine molecular weights or to perform a molecular weight distribution analysis, provided that suitable standards are available.

In case of peptide fusion tags, a variety of affinity purification methods are available. The basic principle is that the protein tag specifically interacts with an affinity resin and binds the tagged protein whereas other polypeptides (impurities) are not bound. This allows a simple purification by releasing the tagged protein with a suitable competing buffer. In this case, usually an intermediate purification step includes the cleavage of the fusion tag with proteases such as enterokinase, Factor Xa, thrombin or Tobacco Etch Virus (TEV) depending on the cleavage site. The protease recognition site is selected and cloned in the vector codifying for the protein sequence at the cloning step; pilot experiments should be done to find out suitable cleavage conditions. The fusion protein is then separated, either with a second round of affinity purification, or by SEC when favoured by the size difference.

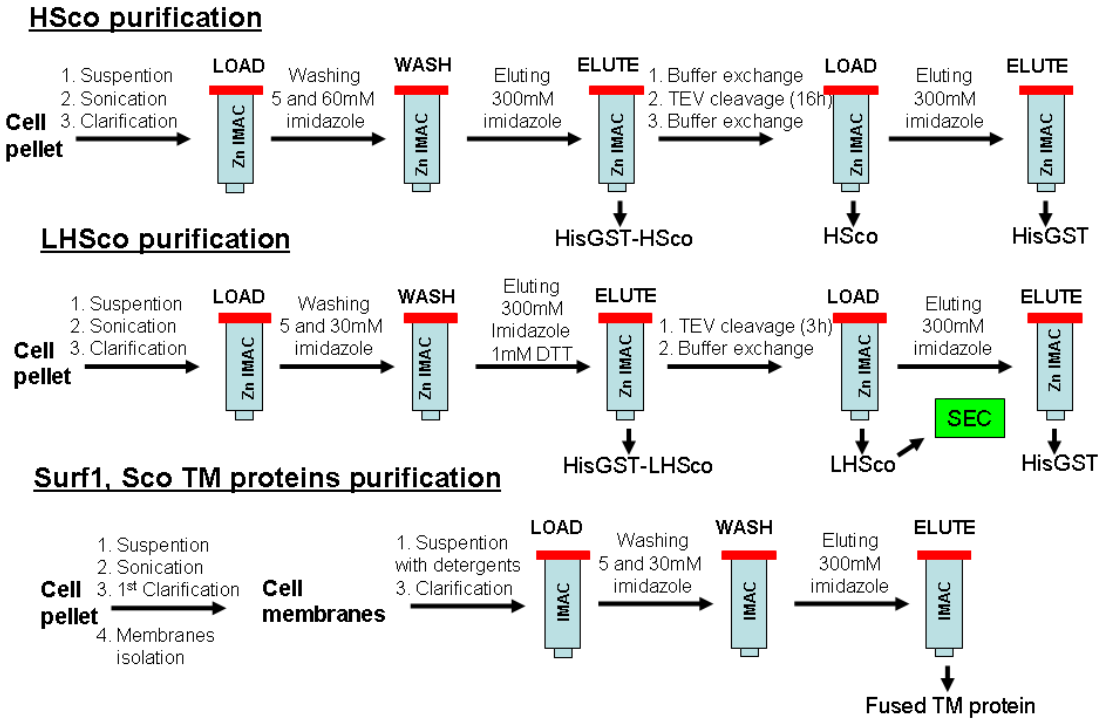


Figure 2.3: Purification of HSco, LHSco and Surf1/Sco TM proteins. Abbreviations: HSco: short constructs of the soluble domain of human Sco1 (G132- 301S) and Sco2 (100G-266S), LHSco: longer constructs of the soluble domain of human Sco1 (113K-301S) and Sco2 (79R-266S), Surf1/Sco TM proteins: see §3.4, IMAC: Immobilized Metal Ion Affinity Chromatography, SEC: Size-exclusion chromatography.

Currently Immobilized Metal Ion Affinity Chromatography (IMAC) is the most used affinity technique; it exploits the interaction between chelated transition metal ions (generally Zn(II), Ni(II) and Co(II)) and side-chains of specific amino acids (mainly histidine) added on the protein. In IMAC chromatography, the target protein is usually washed from the impurities and eluted using increasing concentration of imidazole, which acts as competitive agent. Histidine-tagged (6His-tag) proteins can also be eluted from the

column lowering the pH to decrease the binding affinity or stripping the metal ions from the column with EDTA.

In the present thesis two soluble constructs for each HSco1 and HSco2 have been used. The purification of the short constructs (human Sco1 (G132- 301S) and Sco2 (100G-266S)) has been already reported [9]. These constructs has been used for the interaction with Cox17 (see §3.2). Two longer constructs of HSco, Sco1 (113K-301S) and Sco2 (79R-266S), have been used for the dimerization studies (see §3.3). The purification of the proteins is shown on Fig. 2.2., which involves mainly purification by IMAC (His-trap column charged with Zn(II)). The longer constructs purification involves an addition step of SEC. Moreover the general purification of the transmembrane proteins (see §3.4) is also shown on Fig. 2.3.

2.5 Protein refolding

Heterologous expression of foreign genes in *E. coli* often leads to production of the expressed proteins in IBs. IBs must then be solubilized and refolded into an active conformation. Refolding of IBs is not a straightforward process, often requiring an extensive trial-and-error approach. There are two important issues in recovering active proteins from IBs, i.e., solubilization and refolding. Solubilization must result in monomolecular dispersion and minimum non-native intra- or inter-chain interactions. Choice of solubilizing agents, e.g., urea, guanidine HCl, or detergents, plays a key role in solubilization efficiency, in the structure of the proteins in denatured state, and in subsequent refolding [13].

Refolding is initiated by reducing concentration of denaturant used to solubilize IBs. Protein refolding is not a single reaction and competes with other reactions, such as misfolding and aggregation, leading to inactive proteins. Rate of refolding and other reactions is determined both by the procedure to reduce denaturant concentration and the solvent condition. The procedures to reduce the denaturant will be shortly described below: a) One-step dialysis: Denatured, unfolded protein samples in concentrated denaturant solution are dialyzed against a refolding buffer, and hence, exposed to descending concentration of the denaturant. b) Step-wise dialysis: This protocol uses descending concentration of denaturant for dialysis and has been successfully used for refolding antibodies. Unfolded protein sample is initially brought to equilibrium with high denaturant concentration, then with middle concentration, and with low concentration. c) Buffer-exchange by gel filtration: Gel filtration column is equilibrated with the final

refolding buffer. Unfolded protein sample in denaturant is applied to the column and run through it with the refolding buffer. Use of desalting column will separate proteins from denaturant, while use of protein-sizing column will fractionate protein species. d) Dilution: Protein samples at high denaturant concentration are delivered into a large volume of refolding buffer. Dilution brings the unfolded sample into a rapid collapse, whereby bypassing the intermediate denaturant concentration. e) Solid phase refolding: Denatured protein is initially non-covalently bound to solid matrix such as Ni-resin or ion-exchange resin in the presence of denaturant. Denaturant concentration is then decreased to initiate refolding. Since protein molecules are bound to resin, this procedure minimizes aggregation of unfolded protein or folding intermediates.

During my PhD studies, the production of the soluble domain of human Surf1, 100A-255G, and the soluble domain bacterial of Surf1 (*Caulobacter crescentus*, 107G-276P, *Sinorhizobium meliloti*, 101A-262E) showing that the proteins are expressed in inclusion bodies. Refolding involved purification under denaturing conditions followed by refolding procedures: dilution, step-dialysis and solid phase refolding (IMAC).

2.5.1 Membrane protein refolding

Membrane proteins expressed as inclusion bodies can be refolded in the presence of detergents following strategy similar to soluble protein refolding. In lack of a detailed theoretical understanding of the various factors affecting refolding yields and tendency to aggregate, the most straightforward strategy to find suitable folding conditions is to vary all the parameters, especially the composition of detergent/lipid micelles, in a systematic way and quantify the refolding yield. A proper functional assay that works in detergent is therefore a crucial success factor [14].

During my PhD thesis, I tested the refolding of the HSco1 protein containing the transmembrane helix (THSco1, see §3.4). Refolding was performed according to the protocol of Opella et al. [15], were the protein was extracted from IBs by a harsh detergent (Empigent BB) followed by column refolding and detergent exchange. The refolding was followed by NMR and CD spectroscopy.

2.6 Sample preparation

Human Sco (HSco) proteins metallation protocol is already reported [9]. Protein reduction and metallation were carried out under nitrogen atmosphere in an anaerobic chamber to prevent oxidation of the cysteine residues. Dithiothreitol (DTT) was added to the protein in a 10mM concentration to reduce the cysteine residues in the CXXXC motif

prior to metal reconstitution. The buffer was then exchanged to 50 mM phosphate buffer at pH 7.2 to remove the excess of DTT using PD-10 desalting column or dialysis. In order to obtain Cu(I)-, Cu(II)- and Ni(II)-metallated protein, progressive titration, monitored by UV-Visible spectroscopy or ^{1}H-^{15}N-HSQC experiment were performed, using the metals respectively in the $[Cu(I)(CH_3CN)_4]PF_6$, Cu_2SO_4 and $NiCl_2$ forms. The titration was performed using a maximum protein concentration of 0.2 mM avoiding aggregation of the protein. The excess of metals was removed by PD-10 desalting column or dialysis. All the NMR samples were prepared in phosphate buffer at pH 7.2. Cu(I) metallation of human Cox17 (HCox17) was similarly performed in anaerobic condition. Compared to HSco protocol the HCox17 reduction was induced by lower amount of DTT (1 mM) and the metallation was performed at higher protein concentration (over 0.3mM). For the full metallation of HCox17 3:1 molar ratio (Cu:protein) where added due to the HCox17 lower affinity for Cu(I). Excess of Cu(I) was removed by PD10 or dialysis. Oxidation of HSco CXXXC motif was performed in diluted samples (0.02mM) by incubating the samples for over sixteen hour at room temperature exposed to air. The oxidation state of the HSco was checked by AMS reaction or ^{1}H-^{15}N-HSQC experiment.

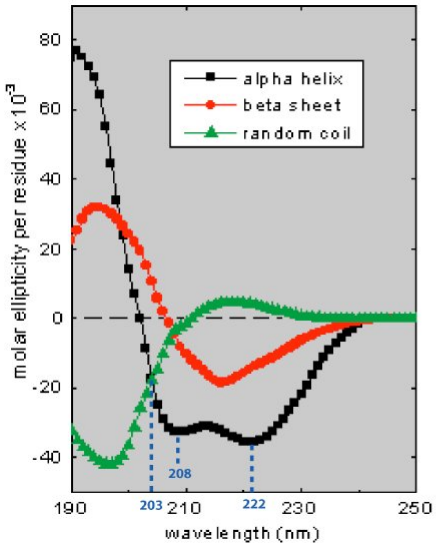
2.7 Biochemical and biophysical characterization

2.7.1 Circular dichroism spectroscopy

Circular dichroism (CD) is a spectroscopy based on the differential absorption of left- and right-handed circularly polarized light. CD is an excellent tool for rapid determination of the secondary structure and folding properties (i.e.arrangement of peptide bonds in secondary protein structure elements like helices and strands) of proteins that have been obtained using recombinant techniques. One of most widely used application of CD in protein studies is to determine whether a purified protein is folded and how a mutation or conditions of sample (denaturant) affects its conformation or stability.

In proteins the major optically active groups are the amide bonds of the peptide backbone (peptide bond: absorption below 240 nm, aromatic aminoacid side chains: absorption in the range 260 to 320 nm and disulphide bonds: weak broad absorption bands centred around 260 nm), typically disposed in highly ordered arrays such as α-helices or b-pleated sheets. A protein consisting of secondary structure elements will therefore display a spectrum that can be deconvoluted into the individual contributions; for this purpose several theoretical methods have been developed, all of them relying on the assumption

that the spectrum of a protein can be represented by a linear combination of the spectra of its secondary structural elements (% α-helix, % β-sheet and % random coil) [16].



Alpha helix has negative bands at 222nm and 208nm and a positive one at 190nm.

Beta sheet shows a negative band at 218 nm and a positive one at 196 nm.

Random coil has a positive band at 212 nm and a negative one around 195 nm.

Figure 2.4: CD spectra of poly-lysine in three conformations. The characteristic bands of secondary structure element are also indicated. Reprinted from Alliance protein laboratories home page [17].

In the current thesis CD spectroscopy was used to monitor the folding of proteins after their recombinant production. Moreover CD was used to monitor the effect of detergent in HSco proteins. CD measurements were performed with protein concentrations of 10-20uM . The CD investigations were carried out in a Jasco J-715 spectropolarimeter using a 0.1-cm pathlength cell with cooling jacket connected to a water thermostatic device at 25^{o}C. Spectra were recorded from 190-250nm at a scan speed of 20 nm/min and resulted from averaging 4 scans. The final spectra were baseline-corrected by subtracting the corresponding buffer obtained under identical conditions. Results were expressed as the mean residue ellipticity [θ] at a given wavelength. The data were fitted with the secondary structure estimation program DicroProt [18] according to Yang et al. [19].

2.7.2 Dynamic light scattering

Dynamic light scattering (DLS), also known as Photon correlation spectroscopy or Quasi-elastic light scattering is a technique which can be used to determine the size distribution profile of small particles in solution like proteins.

When light hits small particles the light scatters in all directions (Rayleigh scattering) so long as the particles are small compared to the wavelength (< 250 nm). If the light source is a laser, and thus is monochromatic and coherent, then one observes a time-dependent fluctuation in the scattering intensity. These fluctuations are due to the fact that the small molecules in solutions are undergoing Brownian motion and so the distance

between the scatterers in the solution is constantly changing with time. This scattered light then undergoes either constructive or destructive interference by the surrounding particles and within this intensity fluctuation, information is contained about the time scale of movement of the scatterers. There are several ways to derive dynamic information about particles' movement in solution by Brownian motion. One such method is dynamic light scattering. The dynamic information of the particles is derived from an autocorrelation of the intensity trace recorded during the experiment. At short time delays, the correlation is high because the particles do not have a chance to move to a great extent from the initial state that they were in. The two signals are thus essentially unchanged when compared after only a very short time interval. As the time delays become longer, the correlation starts to exponentially decay to zero, meaning that after a long time period has elapsed, there is no correlation between the scattered intensity of the initial and final states. This exponential decay is related to the motion of the particles, specifically to the diffusion coefficient. The photon autocorrelation function from a system of monodisperse particles is given by a single exponential decay. The diffusion coefficient D can be determined by fitting the measured function to a single exponential. The hydrodynamic radius of particles R_h can be calculated from the Stokes-Einstein equation:

$$R_h = k_B T / 6\pi\eta D$$

where, k_B is the Boltzmann constant (in J K^{-1}), η is the diluents' viscosity (in kg/ms), D is the diffusion coefficient (in m^2s^{-1}) and T is the temperature in Kelvin.

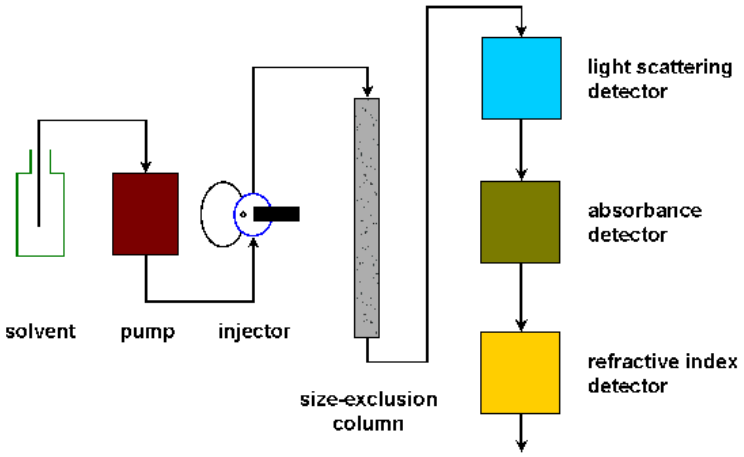


Figure 2.5: Diagram of light scattering in the chromatography mode. The protein sample is injected (injector), separated in the size-exclusion column. After elution the different properties of the protein are measured by the detectors. Reprinted from Alliance protein laboratories home page [17].

In the present thesis DLS was applied in chromatography mode (see Fig. 2.5) for the investigation of the aggregation state of the HSco constructs (SHSco1, LHSco1, SHSco2, LHSco2, see §3.3). 0.1–1 mM protein samples were injected on a Superdex75 HR-10/30 size-exclusion column on an AKTAFPLC system (Amersham Pharmacia Biosciences) connected with a multiangle light scattering (DAWN-EOS, Wyatt

Technologies, Santa Barbara, CA) coupled with quasielastic light scattering detectors. Data analysis and average MW were automatically calculated by the Wyatt's Astra software.

2.7.3 UV-visible spectroscopy

UV-visible spectroscopy involves the spectroscopy of photons in the UV-visible region. It uses light in the visible and adjacent near ultraviolet (UV) and near infrared (NIR) ranges. In this region of the electromagnetic spectrum, molecules undergo electronic transitions. UV-visible spectroscopy is routinely used in the quantitative determination of protein and DNA solution. Protein concentration can be determined by the absorption at 280 nm (mainly due to Tyr and Trp residues) using the protein extinction coefficient (estimated by Protparam [20]) or by any colorimetric assay e.g. Bradford. The molar extinction coefficient (or molar absorptivity) is defined via the Beer-Lambert law:

$$\varepsilon = \mathrm{A} / c \times l$$

where A absorbance, c the sample concentration (M) and l the length of light path through the sample in cm.

DNA quantification is based on the absorption purines and pyrmidines which show absorption maxima around 260nm (alternatively DNA can be quantified on agarose gel by ladder markers). In the present work UV-visible spectroscopy, apart for routine quantification, was used to monitor the metallation of HSco with Cu(II) and Ni(II), see §3.4.

2.8 Reference list

1. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Protoc. (**2007**)2:953-71
2. K. Hofmann & W. Stoffel Biol. Chem. Hoppe-Seyler 374,166
3. Bendtsen JD, Binnewies TT, Hallin PF, Ussery DW, Microbiology, (**2005**)151:2119-2121
4. Tusnády GE, Simon I. J Mol Biol. (**1998**)283:489-506
5. Combet C., Blanchet C., Geourjon C. and Deléage TIBS (**2000**)25:147-150
6. Dosztányi Z, Csizmok V, Tompa P, Simon I Bioinformatics. (**2005**)16:3433-4.
7. Schenk, P.M., Baumann, S., Mattes, R., Steinbib, HBiosci Biotechnol Biochem (**1998**)62:2408-14
8. Ross,W. & Landy,A. *Proc. Natl. Acad. Sci. U.S.A*, (**1982**)7724-7728
9. Banci L, Bertini I, Calderone V, Ciofi-Baffoni S, Mangani S, Martinelli M, Palumaa P, Wang S.A.Proc Natl Acad Sci U S A. (**2006**)23:8595-600.
10. Cunningham F, Deber CM. Methods. (**2007**) 4:370-80
11. Davis, G.D Elisee, C. Newham D.M. and. Harrison, R.G *Biotechnol. Bioeng.* (**1999**)65:382-388
12. Dumon-Seignovert, L. Cariot G. and Vuillard, L. *Protein Expr. Purif.* (**2004**),37:203–206
13. Tsumoto K, Ejima D, Kumagai I, Arakawa T. Protein Expr Purif. (**2003**) 1:1-8
14. Kiefer H. Biochim Biophys Acta. (**2003**)1:57-62.
15. Page RC, Moore JD, Nguyen HB, Sharma M, Chase R, Gao FP, Mobley CK, Sanders CR, Ma L, Sönnichsen FD, Lee S, Howell SC, Opella SJ, Cross TA. J Struct Funct Genomics. (2006)1:51-64
16. Kelly SM, Jess TJ, Price NC. Biochim Biophys Acta. (**2005**) 2:119-39
17. http://www.ap-lab.com
18. Deleage, G. and Geourjon, C. Comput. Appl. Biosci. (**1993**)9:197–199
19. Yang, J. T., Wu, C. S. and Martinez, H. M. (**1986**). Methods Enzymol. 130:208–269
20. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A, Humana Press (**2005**) 571-607

3. Results

3.1 Introduction

In this chapter are presented the results of my PhD thesis which involve a published article, a manuscript (status: submitted in JBC) and a last section with results on a project that is in progress.

In the article under the title '**Mitochondrial copper(I) transfer from Cox17 to Sco1 is coupled to electron transfer**' we show the that the Cox17 can transfer simultaneously copper(I) and two electrons to the human cochaperone Sco1 in the oxidized state. The same reaction of copper–electron-coupled transfer does not occur with the human homolog of Sco1, HSco2. In the article are described a number of NMR titrations with labelled and unlabeled samples of HSco1, HSco2 and Cox17 but also electrospray ionization (ESI) MS experiments. For this article I was responsible for the production of the HSco1 and HSco2 proteins and the preparation of the samples of both Sco and Cox17 in the appropriate metallation and redox state before the NMR studies.

In manuscript present in the §3.3 with the title '**A 19aa segment promotes Sco1 dimerization'** we characterized for the first time the interface of the Sco1 protein dimer using NMR spectra. A number of analytical gel filtration and light scattering data to verify the dimerization of the LHSco1 and the interaction between LHSco1 and LHSco2, were also performed. For this project I defined and cloned the LHSco1 construct, I altered and optimized the purification of the LHSco improving the stability and avoiding degradation of the samples (see §2.4.), purified and prepared the samples needed for the experiments and finally performed the gel filtration and light scattering experiment.

In last section of this chapter '**Production of human Surf1, Sco1 and Sco2 transmembrane proteins'** I describe the production of the human Surf1, Sco1 and Sco2 full length proteins showing that the proteins can be purified in the presence of specific detergents. For Sco proteins, this work primarily aims to investigate the possibility of dimerization of the HSco through their TM regions and for Surf1, the principal aim is to produce the protein for its structural characterization as it is not available. In this project I have cloned, produced and purified all proteins. Moreover, I screened and selected the appropriate detergents, in particular by investigating the folding state of the proteins by CD.

3.1 Mitochondrial copper(I) transfer from Cox17 to Sco1 is coupled to electron transfer.

The article can be found in the CD

3.2 A 19aa segment promotes Sco1 dimerization

Lucia Banci, Ivano Bertini, Theodoros Hadjiloi, Shenlin Wang

Magnetic Resonance Center CERM and Department of Chemistry, University of Florence, Scientific Campus, 50019, Sesto Fiorentino, Florence, Italy.

Abstract

Human Sco1 is a mitochondrial membrane bound protein involved in the assembly of the dinuclear Cu_A center in cytochrome c oxidase. The full length human Sco1 contains in the N-terminus a short matrix segment, a single transmembrane helix buried in the inner membrane and a soluble C-terminal domain extended into IMS. While previous structures determined for the soluble C-terminal domain show that this domain is monomeric, full length human Sco1 has been reported to function as homodimers. The structural information for dimeric human Sco1 is still unavailable. We report here that a 19-aa segment, rich in charged residues, linking the soluble C-terminal domain to the transmembrane helix, promotes human Sco1 dimerization in both the apo and the copper(I) form. The interface of dimeric Sco1, characterized by NMR spectra, contains residues of the N-terminus, metal binding site and their surroundings. A structural model for dimeric Sco1 is calculated accordingly. On the basis of our results, a mechanism of transferring copper cargos from human Sco1 to the binuclear Cu_A center is discussed.

Introduction

Cytochrome *c* oxidase (CcO) is the terminal enzyme of the respiratory chain in eukaryotes and a number of prokaryotes. Eukaryotic CcO is embedded in the inner mitochondrial membrane, where it catalyses reduction of molecular oxygen to water and pumps protons across the membrane[1-3]. Mammalian CcO consists of 13 subunits, three of which (CoxI-CoxIII) are encoded by the mitochondria genome and the remaining by the nuclear genome [4]. The assembly of this enzyme is a complex, multistep process that involves at least 30 proteins [5-8]. Among them, the Sco family proteins, Sco1 and Sco2, are involved in the assembly of the Cu_A center, which is a binuclear copper center localized in the CoxII subunit [9]. Mutations in either Sco1 or Sco2 lead to decreased CcO activity and early death due to the failure of the holoenzyme assembly [10-13]. Studies with immortalized fibroblasts from *SCO1* and *SCO2* patients suggested that Sco1 and Sco2 have non-overlapping but cooperative functions in CcO assembly [14].

Human Sco1 and Sco2 have high sequence identity. Both proteins contain, after a short matrix segment, a single transmembrane helix (TM) in the N-terminal segment of the protein, anchoring them to the inner mitochondrial membrane, while the C-terminal domain is soluble and extended into the IMS. A 19-aa segment, rich in charged residues, connects the TM and the soluble C-terminal domain of both Sco1 and Sco2 (Fig S1). *In vivo* and *in vitro* studies proved that the soluble C-terminal domain of human Sco1 and of Sco2 is able to bind one copper(I) ion in relatively high affinity via the CXXXCP motif and a His residue, both fully conserved [15-17]. Mutations of either these Cys or His residues abrogated the copper(I) binding ability and lead to a non-functional CcO [16], implying that the function of the Sco proteins is related to their metal binding ability.

The Sco proteins show a thioredoxin fold [18-23], i.e. that of proteins responsible of catalyzing the reduction of protein disulfides. This fold similarity suggested the possibility that Sco proteins could be involved also in thiol redox processes [23]. Recently, we have shown that, indeed, in prokaryotic organisms Sco1 is able to reduce oxidized Cu_A, while does not transfer copper(I) to it [24]. Thus in prokaryotic organisms its role is that of assisting the copper transfer process maintaining the copper bridging Cys of Cu_A in a reduced state [24]. At variance of prokaryotic, eukaryotic Sco1s have also the ability of binding copper(I) with high affinity thus suggesting for Sco1 a dual role , being both a copper chaperone and a thioredoxin. Recently, it has been also proposed that human Sco proteins might play a role in regulating the mitochondrial copper content, which is a further phenotype independent from that of the CcO assembly [25].

While the functional role of the soluble C-terminal domain of the Sco proteins has been investigated through structural and functional characterization [15,22], the role of the other parts of the proteins had not been specifically addressed. Size exclusion chromatography suggested that full-length human Sco proteins function as homodimers [25]. However, the soluble C-terminal domain of human Sco1 and Sco2 proteins are both monomeric in solution, as proved by several techniques [15,19,20]. Thus, the N-terminal part of human Sco protein (either the matrix domain, the TM or the segment between the soluble C-terminal domain and the TM) has to be critical for protein association. Furthermore, the segment between the soluble C-terminal domain and the TM is crucial for the Sco1 function. Indeed, in yeast Sco1, this cannot be replaced by its Sco2 counterpart [26], while replacement of the TM with that of Sco2 maintain a fully functional Sco1. The segment linking the TM and the soluble C-terminal domain might therefore have a role for determining the spatial orientation of the soluble C-terminal domain and/or for modulating the aggregation state of the proteins [26].

To test whether this segment has some role in protein self-association and/or mediate the formation of heterodimers between Sco1 and Sco2, we characterized two truncated forms of HSco1 and HSco2, lacking only the TM, the matrix domain and the proceeding mitochondrial target sequence (MTS) (LHSco1 and LHSco2 hereafter). Constructs contains, with respect to the constructs used for structure determination (called HSco1 and HSco2 hereafter), an additional 19-aa segment, which links the soluble C-terminal domain to the TM. From the present investigation, we found that the longer construct of human Sco1, but not that of Sco2, has tendency to dimerize both as apo and copper(I) loaded forms. The interface of dimeric LHSco1 contains long loop containing the fully conserved His residue, the flexible N-terminus, β-7 and loop 10.

Results

The aggregation state of Sco1 depends on the construct length

Two truncated forms of the *human SCO1* gene, both lacking the N-terminal mitochondrial targeting sequence, matrix segment and the single-transmembrane helix, have been engineered and characterized (Fig. S1). These forms differed by the presence (LHSco1) or not (HSco1) of the 19-aa segment, which, particularly in Sco1, is rich in positively and negatively charged residues. Size-exclusion chromatography experiments in conjunction with multiangle light scattering measurements, showed that, while both the apo and the copper(I) bound form of HSco1 eluted in a pure monomeric state in the concentration range have investigated (5 μM to 20 μM), with an average molecular weight of 20.0 kDa, the apo and the copper(I)-bound LHSco1 had a concentration dependent elution profile (Fig 1A and Table S4). Although they eluted as a single fraction, their elution volumes indicated a protein size larger than the monomeric one, with the size increasing with concentration. These data suggested that LHSco1 has a tendency to partially form soluble homodimers, which are in fast equilibrium with the monomer. Consistently, the fitted molecular weight values, derived from light scattering data, also showed a concentration dependent pattern. An average molecular weight of 24.9 kDa, higher than the theoretical value of 22.1 kDa, was obtained for Cu(I)LHSco1 at 5 μM concentration.

Since the additional 19-aa segment of LHSco1 is rich in charged residues, electrostatic interactions may play an important role in protein dimerization. Indeed, consistently with this hypothesis, the dependence of protein dimerization on salt concentration, as followed by multiangle light scattering measurements (Fig 1B), showed

that the average molecular weight of both forms of LHSco1 decreased with increasing of NaCl concentration.

The overall behavior of the two constructs of HSco1, therefore, suggests that the additional 19-aa segment is implicated in the formation of the dimeric form of LHSco1.

Both truncated human Sco2 forms are monomer in solution

Our previous work showed that HSco2 is monomeric in both apo and copper(I) loaded forms [20]. Size exclusion chromatography experiments performed on LHSco2 in its apo, Cu(I), forms, showed that, at variance with LHSco1, the two forms of LHSco2 have neither concentration dependent elution volumes nor experience formation of homodimers. The additional segment therefore does not affect the aggregation state of HSco2.

No evidence for stable complex LHSco1-LHSco2 heterodimer formation

Having shown that LHSco1 has tendency to form homodimers, we wanted then to address the question of whether a stable LHSco1-LHSco2 heterodimer is formed. Size-exclusion chromatography on mixtures containing equimolar amounts of both proteins, either in the apo or in the Cu(I) bound forms, indicated that the proteins retained a molecular weight similar to those of the isolated proteins, suggesting that LHSco1 and LHSco2 do not interact or, if they do so, they form a very transient complex that cannot be detected through this analysis.

Structural and dynamical properties of apo and Cu(I)LHSco1

NMR spectra, chemical shift perturbation mapping, and backbone amide ^{15}N relaxation parameters of apo and Cu(I)LHSco1 at different concentrations were used to characterize the model of LHSco1 dimerization at the atomic level. Overall, 162 out of the expected 177 NH cross peaks were detected for Cu(I)LHSco1. Resonances were not detected for residues, from K-15 to K-11 and from K-3 to R1, located in the 19-aa additional segment, and I135 and M136, which were not observed also in Cu(I)HSco1[22]. Backbone chemical shifts of the soluble C-terminal domain of Cu(I)LHSco1 (0.1mM) were in very good agreements with those of Cu(I)HSco1, reported previously [22], indicating that the additional 19-aa segment does not affect Sco1 conformation.

A lower number of ^{1}H-^{15}N amide signals was detected in the apo form, as also found for apoHSco1[22]. Only 116 out of the expected 177 cross peaks were observed. In addition to those not observed in the Cu(I) bound form, resonances were not detected also for residues, L7-G10, Y35-F38, H40-L49, I73-D76, A110-M136 and F149-N152. The above residues comprise the N-terminal region, the CXXXCP metal-binding motif, loop 8, where the metal binding His residue is localized, loop 5 and loop 10. This different behavior

between the apo and the copper bound forms is similar to our previous observation on the apo forms of HSco1 and HSco2 proteins [20,22]. Backbone chemical shifts of apoLHSco1 match those of apoHSco1, indicating a similar overall conformation for the apo protein as well. The lack of several signals in the apo form originates from conformational disorder and/or exchange processes among multiple conformation states of the apo form. Upon metal binding, these regions take a more rigid conformation, as it occurs in HSco1 [22].

The NMR spectra of Cu(I)LHSco1 are concentration dependent with a number of peaks changing their shift as a function of protein concentration, consistent with monomer-dimer equilibra, fast on the NMR time scale. The residues showing the largest combined chemical shift variations as a function of concentration are localized in the N-terminus of the construct (L7, G9 and G10), loop3 (H40), loop 8 (S118, H132, S133, and T134), β-6 (Y137), loop 9 (L144) and β-7 (T146 and D147) (Fig 2A). With the exception of the residues on the β-6, these residues are all solvent exposed and located on the same side of the protein surface (Fig 2B), suggesting that protein dimerization occurs in a specific manner, which involves the metal binding site, β-7 and the N-terminus of the construct. In apoLHSco1, none of the residues experiencing chemical shift change in the Cu(I) bound form are detected. Therefore the concentration dependence of NMR spectra of apo form is not informative with respect to protein dimerization.

^{15}N relaxation parameters on Cu(I)LHSco1, which were analyzed for 144 backbone NH groups, showed that the residues of the additional 19-aa segment are characterized by lower R_2 and higher R_1 values (Fig 3). These residues experience faster motions than the overall protein tumbling. The average values of R_2 and R_1 for the folded domain, estimated at 0.1mM concentration, are 24 ± 3 s^{-1} and 0.82 ± 0.15 s^{-1}. ^{15}N relaxation parameters are also concentration dependent, with higher R_2 values in more concentrated sample (29 ± 3 s^{-1} and 0.79 ± 0.14 s^{-1} for 0.50 mM sample, Fig 3) These R_2 values are significantly higher than those expected for monomeric HSco1 and lower than the values estimated for dimeric HSco1 (average R_2 of 18 s^{-1} and 34 s^{-1} for monomer and dimer, respectively). Also the experimental R_1 values are lower than those expected for monomeric HSco1 and higher than the values for dimeric HSco1 (average R_1 of 0.96 s^{-1} and 0.50 s^{-1} for monomer and dimer, respectively). These data confirmed that Cu(I)LHSco1 is partially in a dimeric state with a fraction of about 40% dimer at 0.5mM protein concentration.

Concentration dependent relaxation rates were also obtained for apoLHSco1 (Fig S2); the average R_1 and R_2 values of apoLHSco1 are the same as those of Cu(I)LHSco1

measured in the same concentration, suggesting a formation of a similar fraction of dimeric form at the same concentration.

Model of dimeric Cu(I)HSco1

A structural model for dimeric Cu(I)HSco1, calculated on the basis of concentration dependent chemical shift data of Cu(I)LHSco1 (Fig 4) shows that dimerization occurs mainly through electrostatic interactions between the two monomers. At the dimer interface, D131, which is next to the metal binding His residue, forms salt bridges interactions with residues R154 and K153 of the other monomer. R114, at the beginning of loop 8, is involved in salt bridge interaction with E144 of the other monomer. The dimer is further stabilized by a hydrophobic patch, i.e. residue Y116 makes contact with Y148 of the other monomer.

The distance between the N-terminus residues is 10-15 Å, a distance which can be easily bridged by two stretches of 19 residues (Fig 4). The two metal binding sites are on the same side of the dimer unit with a distance between the two copper ions of 20 Å.

Discussion

The N-terminus 19-aa segment of Sco1 which connects the folded C-terminal domain with the TM helix, is responsible for the Sco1 tendency to dimerize. On the other hand, the same construct of Sco2 never forms dimeric species either with itself or with Sco1. It had been previously shown that Sco1, having the 19-aa segment substituted with that of Sco2, is not functioning in yeast [26], thus indicating that this segment has a key role in the *in vivo* Sco1 activity. As it comes out from the present results, its main role is that of modulating Sco1 dimerization, which involves the N-terminus, the solvent exposed loop containing the metal binding His residue and their surrounding. The driving force for protein dimerization resides in the additional 19-aa segment, which is rich in charged residues (10 positively and 4 negatively charged residues). Several salt bridge interactions were also detected in the C-terninal folded domain in the structural model. The electrostatic driven interaction was also confirmed by the ionic strength dependence of protein dimerization. This high content in charged residues is also conserved in all the sequences of eukaryotic Sco1, but not in prokaryotic Sco proteins. Indeed the latter are proposed to be monomer [23].

Protein dimerization in the cell could be further strengthened by interactions between the two TM helices. Indeed, the latter are characterized by a side with several small residues, i.e. Gly and Ala, which are essential for stabilizing helix-helix interactions [27]. The two metal binding sites are on the same face of the dimer and located at the opposite side with respect to the additional 19-aa segment linking the folded domain to the TM.

This location implies that the metal binding sites of dimeric HSco1 extend in the IMS, instead of facing the inner membrane of mitochondria. This conformation makes the copper ions bound to HSco1 for being transferred to the Cu_A site, which is 8 Å outside the membrane surface [28]. We could also speculate that the two copper ions bound by dimeric Sco1 may transfer simultaneously to the Cu_A site.

Finally, LHSco2 is never present in a dimeric state in all the forms have investigated, i.e. apo, Cu(I), bound forms, nor it forms heterodimers with Sco1. The different behavior with respect to Sco1 may arise from the lower number of charged residues (6 positively and 3 negatively charged residues) in the additional 19-aa segment. The dimeric state, which had been observed in cell might be, however, formed through interaction between two TM, which, similarly to Sco1, feature a number of small residues on the same side. The soluble C-terminal domain of Sco2 has an internal dynamic and conformational disorder higher than that of Sco1[20].These two features, i.e. different dynamical and interaction properties, can contribute to make the dimeric state of Sco2 different from that of Sco1, thus contributing to differentiating the behavior and function of these two otherwise similar proteins.

Materials and Methods

Sample preparation

Two truncated forms of HSco1 and of HSco2 were prepared as described previously [22,22]. The Cu(I) metallated form was obtained by addition of stoichiometric amounts of the metal ions $(Cu(I)(CH_3CN)_4)PF_6$) to diluted protein solutions in 50mM phosphate buffer (pH 7.2), followed by protein concentration under nitrogen atmosphere.

Size exclusion chromatography and light scattering experiment procedure

All the size-exclusion chromatography experiments were done on a Superdex 75 HR-10/30 size-exclusion column on an AKTA-FPLC system (Amersham Pharmacia Biosciences) connected to a multiangle light-scattering system (DAWN-EOS; Wyatt Technologies) equipped with quasielastic light-scattering detector. The column was equilibrated with 50mM phosphate buffer (pH 7.2), containing 1mM DTT. A flow rate of 0.5ml/min was used for all the studies. 100µl protein solution was injected into the column. The loading concentrations had a range from 0.10 to 0.50 mM. The final concentrations of the eluted protein, estimated by the absorbance at 280 nm, had a range from 5 to 20 µM. Elution profiles were obtained from the optical density values at 280 nm. The elution volumes were reported in Table S3 and Table S4.

NMR spectroscopy

NMR spectra for backbone assignments were obtained through standard triple resonance experiments. The backbone assignment experiments of apo and Cu(I)LHSco1 were performed on 0.50 mM samples at 500 MHz (298K) using standard pulse sequences.

The relaxation data on apo and Cu(I)LHSco1 were obtained through ^{15}N-R_1 and ^{15}N-R_2 experiments at 700 MHz (298K) for samples containing 0.10 and 0.50 mM protein. Processing of NMR data was accomplished with the XWINNMR program. Relaxation parameters were determined by fitting the integration values obtained from XWINNMR program, measured as a function of the delay within the pulse sequence, to a single-exponential decay.

HADDOCK docking calculations

HADDOCK calculations were performed with standard protocols (as described at www.nmr.chem.uu.nl/haddock/haddock.html [29]) to obtain a model of the dimeric form of the soluble C-terminal domain of Cu(I)HSco1. The best conformer of the solution structures of the Cu(I)SHSco1 (PDB code 2GT6) was used as the monomeric input.

The "active" residues and "passive" residues were selected on the basis of the concentration dependent chemical shift variation (CSV) data of Cu(I)LHSco1. The high solvent accessible residues (>50%) that exhibited a CSV larger than average were designated as active residues and the residues neighboring those active residues with a high solvent accessibility (>50%) were designated as passive residues. The per residue solvent accessibility area was calculated by NACCESS program.

In the first stage of the calculation, an initial set of 1000 rigid-body docking models was generated. The 200 lowest energy complex models were then selected and submitted for a second stage of calculations with semi-flexible simulated annealing. The 100 lowest energy models in the second stage were refined in water solvent. The final 100 calculated structures resulted all converge to a single ensemble of conformations, displaying a pairwise rms deviation <1.0 Å. The 20 lowest-energy structures were analyzed in terms of intermolecular contacts (hydrogen bonds and nonbonded contacts).

HYDRONMR

Theoretical prediction of the relaxation parameters for monomeric and dimeric Cu(I)HSco1 was calculated by using the ‘‘shell modeling’’ strategy implemented in HYDRONMR software (30). The best represent conformer of solution structures of the Cu(I)HSco1 (PDB code 2GT6) was used to obtain expected relaxation values for monomeric Cu(I)HSco1. Those values for dimeric Cu(I)HSco1 were obtained by using the structural model of dimeric Cu(I)HSco1. The calculations were performed by using a

viscosity of water at 298K of 0.89 Ns/m^2. The value of the atomic element radius was set to 3.3 Å (30).

Figures and Tables

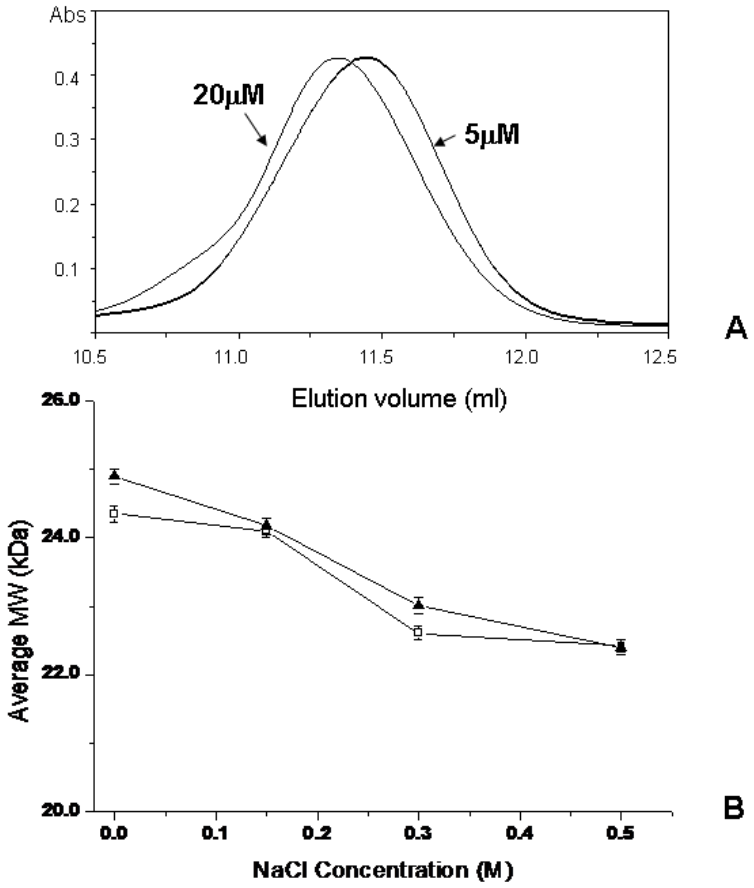


Figure 1: (A) Concentration dependent elution profiles of Cu(I)LHSco1 at two concentrations, 5 and 20 μM, are shown. The elution profile for 5 μM sample is scaled by a factor of four. (B) Salt dependence of the average molecular weight (MW) of LHSco1; Apo and Cu(I) LHSco1 are shown in "□" and "▲", respectively. The sample concentrations are 5 μM.

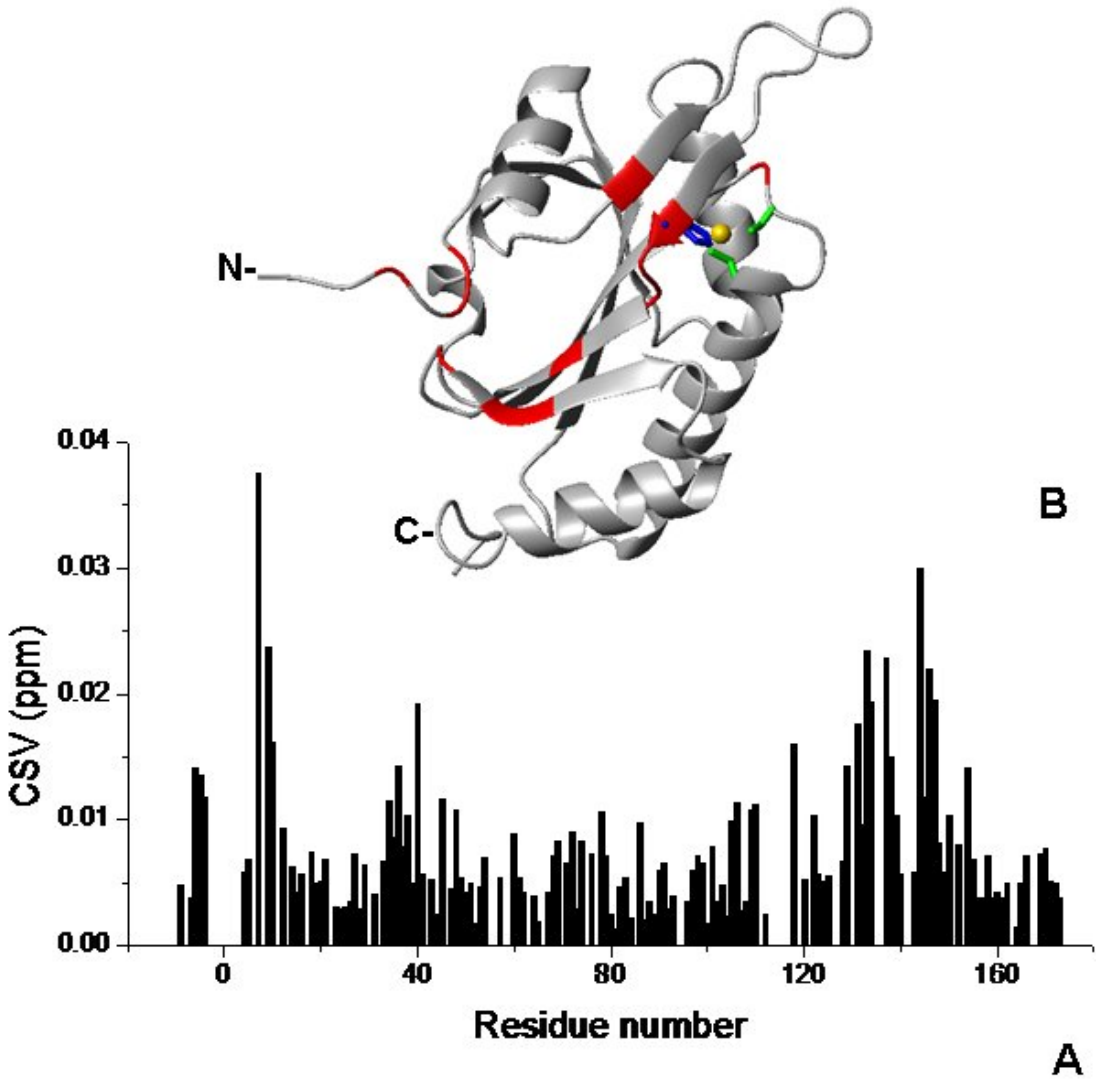


Figure 2: Chemical shifts of Cu(I)LHSco1 are concentration dependent. (A) Sequential plot of combined chemical shift variations (CSV) between 0.10 and 0.50mM Cu(I)LHSco1 is shown. (B) Residues showing CSV values above the threshold (> 0.015 ppm) are mapped in red on the structure of Cu(I)SHSco1. The copper binding Cys and His residues are shown in green and blue, respectively. Cu(I) ion is depicted in gold sphere.

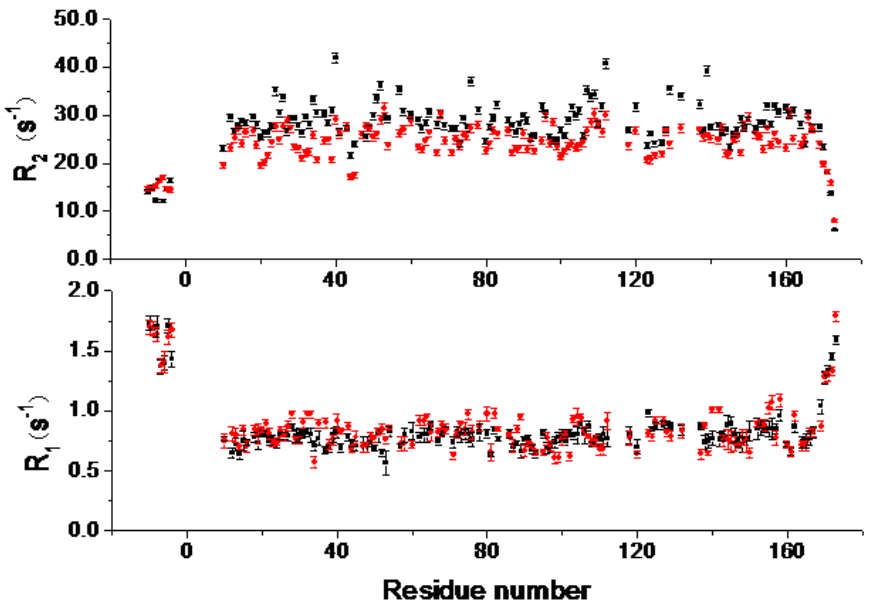


Figure 3: Sequential plot of concentration dependent ^{15}N relaxation parameters for Cu(I)LHSco1. Data for 0.10 and 0.50 mM sample are shown in red and black, respectively.

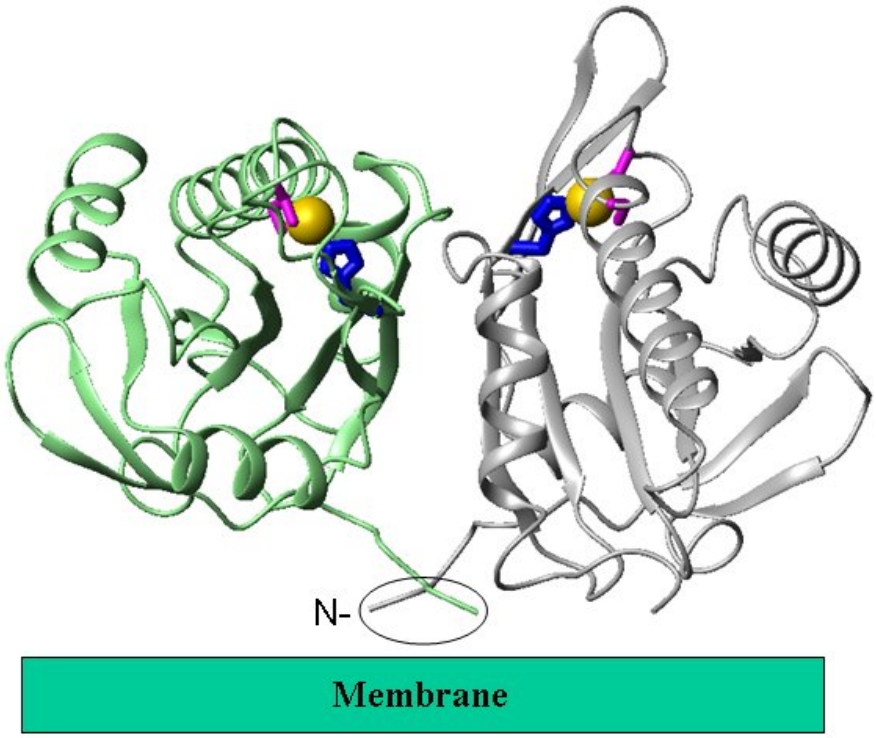


Figure 4: Model of dimeric Cu(I)HSco1. The two monomers are shown in green and grey, respectively. Copper ions are shown in gold. The copper binding Cys and His residues are shown in pink and blue, respectively.

Supplementary materials

Table S3. Elution volumes of Sco1 and Sco2 in different constructs and forms in 50mM phosphate buffer at 298K, pH 7.2.

Constructs	Elution volume (ml) [a]	
	Adducts	
	apo-	Cu(I)-
HSco1	12.32	12.34
HSco2	12.52	12.56
LHSco2	12.25	12.24
LHSco1	11.45[b]	11.35[b]

a: The sample concentrations were 20μM.
b: The value is for 20μM conc. The elution volumes are concentration dependent (see Table S4).

Table S4. Elution volumes of apo and Cu(I)LHSco1 at different concentrations in 50mM phosphate buffer at 298K, pH 7.2

	Elution volume (ml)		
	5 μM	**10 μM**	**20 μM**
apo LHSco1	11.55	11.47	11.45
Cu(I)LHSco1	11.47	11.41	11.35

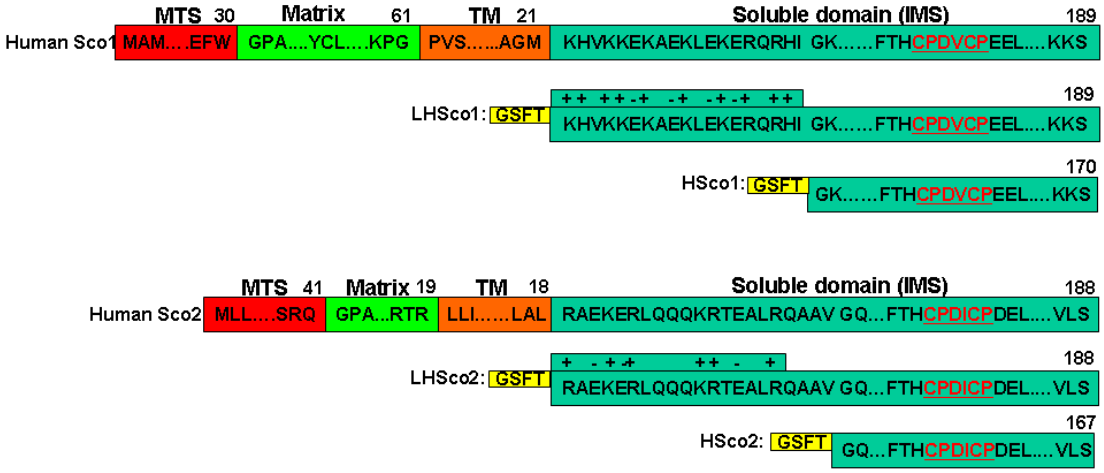


Figure S1: Protein sequence and cloned constructs of HSco1 and HSco2.

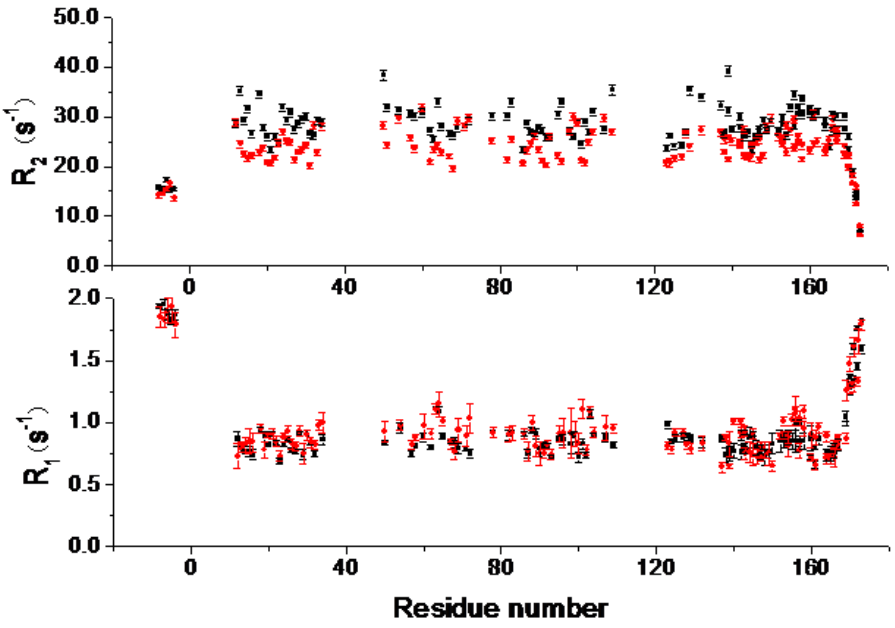


Figure S2: Concentration dependent ^{15}N relaxation parameters for apoLHSco1. Data for 0.10 and 0.50 mM sample were shown in red and black, respectively.

Reference list

1. Wikstrom, M. *Nature* (**1984**) 308**:** 558-560

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

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

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

5. Tzagoloff, A. and Dieckmann, C. L. *Microbiol.Rev.* (**1990**) 54**:** 211-225

6. Carr, H. S. and Winge, D. R. *Acc.Chem.Res.* (**2003**) 36:309-316

7. Khalimonchuk, O. and Rodel, G. *Mitochondrion* (**2005**) 5: 363-388

8. Herrmann, J. M. and Funes, S. *Gene* (**2005**) 354:43-52

9. Yoshikawa, S., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., Yamashita, E., Inoue, N., Yao, M., Fei, M. J., Libeu, C. P., Mizushima, T., Yamaguchi, H., Tomizaki, T., and Tsukihara, T. *Science* (**1998**) 280: 1723-1729

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

11. Jaksch, M., Paret, C., Stucka, R., Horn, N., Muller-Hocker, J., Horvath, R., Trepesch, N., Stecker, G., Freisinger, P., Thirion, C., Muller, J., Lunkwitz, R., Rodel, G., Shoubridge, E. A., and Lochmuller, H. *Hum Mol Genet* (**2001**) 10:3025-3035

12. Sacconi, S., Salviati, L., Sue, C. M., Shanske, S., Davidson, M. M., Bonilla, E., Naini, A. B., De Vivo, D. C., and DiMauro, S. *Pediatr Res.* (**2003**) 53: 224-230

13. Tarnopolsky, M. A., Bourgeois, J. M., Fu, M. H., Kataeva, G., Shah, J., Simon, D. K., Mahoney, D., Johns, D., MacKay, N., and Robinson, B. H. *Am.J.Med.Genet.A* (**2004**) 125: 310-314

14. Leary, S. C., Kaufman, B. A., Pellecchia, G., Guercin, G. H., Mattman, A., Jaksch, M., and Shoubridge, E. A. *Hum.Mol.Genet.* (**2004**) 13:1839-1848

15. Horng, Y. C., Leary, S. C., Cobine, P. A., Young, F. B., George, G. N., Shoubridge, E. A., and Winge, D. R. *J Biol Chem* (**2005**) 280:34113-34122

16. Nittis, T., George, G. N., and Winge, D. R. *J.Biol.Chem.* (**2001**) 276:42520-42526

17. Banci, L., Bertini, I., Cavallaro, G., and Rosato, A. *J.Proteome Res.* (**2007**) 6:1568-1579

18. Balatri, E., Banci, L., Bertini, I., Cantini, F., and Ciofi-Baffoni, S. *Structure* (**2003**) 11:1431-1443

19. Williams, J. C., Sue, C., Banting, G. S., Yang, H., Glerum, D. M., Hendrickson, W. A., and Schon, E. A. *J.Biol.Chem.* (**2005**) 280: 15202-15211

20. Banci, L., Bertini, I., Ciofi-Baffoni, S., Gerothanassis, I. P., Leontari, I., Martinelli, M., and Wang, S. *Structure* (**2007**) 15: 1132-1140

21. Abajian, C. and Rosenzweig, A. C.*J.Biol.Inorg.Chem.* (**2006**) 11: 459-466

22. Banci, L., Bertini, I., Calderone, V., Ciofi-Baffoni, S., Mangani, S., Martinelli, M., Palumaa, P., and Wang, S. *Proc.Natl.Acad.Sci.USA* (**2006**) 103: 8595-8600

23. Ye, Q., Imriskova-Sosova, I., Hill, B. C., and Jia, Z. *Biochemistry* (**2005**) 44: 2934-2942

24. Abriata, L. A., Banci, L., Bertini, I., Ciofi-Baffoni, S., Gkazonis, P., Spyroulias, G. A., Vila, A. J., and Wang, S. *Nature Chemical Biology* (**2008**) 4:599-601

25. Leary, S. C., Cobine, P. A., Kaufman, B. A., Guercin, G. H., Mattman, A., Palaty, J., Lockitch, G., Winge, D. R., Rustin, P., Horvath, R., and Shoubridge, E. A. *Cell Metab.* (**2007**) 5:9-20

26. Lode, A., Paret, C., and Rodel, G. *Yeast* (2002) 19: 909-922

27. Melnyk, R. A., Kim, S., Curran, A. R., Engelman, D. M., Bowie, J. U., and Deber, C. M. *Journal of Biological Chemistry* (**2004**) 279:16591-16597

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

29. Dominguez, C., Boelens, R., and Bonvin, A. M. *J.Am.Chem.Soc.* (**2003**) 125:1731-1737

30. Garcia, d. l. T., Huertas, M. L., and Carrasco, B. *J.Magn.Reson.* (**2000**) 147: 138-146

3.4 Production of human Surf1, Sco1 and Sco2 transmembrane proteins

In the previous section (see Results 3.3), we investigated the HSco oligomerization and the formation of a stable complex between the soluble domains, showing that these domains do not interact and HSco1 but not HSco2 forms dimeric species However, the interaction of the proteins and the dimerization of HSco2 maybe mediated through the TM α-helices. In order to investigate the role this TM domain, human Sco1 (HSco1 hereafter) and Sco2 (HSco2 hereafter) containing the single TM α-helix were overexpressed in *E.coli*. Additionally, the full length human Surf1 (HSurf1 hereafter) containing two predicted TM α-helices was also tested for expression. In this section, we show the results of the overexpression of three human TM proteins, Sco1, Sco2 and Surf1 isolated from *E.coli* membranes. Having in mind of the difficulty of studying membrane proteins that contain soluble domain that tends to unfold in the presence detergent micelles, we have performed the initial purification of the TM heterologous proteins with detergents that do not unfold the soluble domain of the HSco proteins.

Introduction

Cytochrome *c* oxidase (CcO) is the terminal enzyme of the energy-transducing respiratory chain in the mitochondria of eukaryotes and certain prokaryotes [1]. The eukaryotic enzyme catalyzes the reduction of molecular oxygen and couples this reduction with proton translocation across the inner membrane of mitochondria to generate the membrane potential necessary for ATP synthesis [2-4]. Eukaryotic CcO consists of 12 or 13 subunits, with the three core enzyme subunits (Cox1–Cox3) being encoded by the mitochondrial genome. The catalytic cofactors in CcO include two modified heme moieties and three copper ions. One metal site, designated heme a_3-Cu_B, is deeply buried within the Cox1 subunit. The other two copper ions exist in the binuclear Cu_A center within Cox2 [5].

The assembly of the CcO complex requires various nuclear-encoded assembly factors. In yeast, more than 30 different protein factors for CcO assembly have been indeed identified [6].

Assembly of CcO is coordinated by two genomes (nuclear and mitochondrial) and consist of different steps, including the assembly of subunits translated on cytoplasmic and mitochondrial ribosomes, modification of protoheme to heme *a*, and the delivery and subsequent insertion of copper ions into the enzyme during the assembly process [7]. A number of copper chaperons, including Cox11, Cox17, Cox19, Cox23, Sco1, Sco2 are known to be important or have been implicated in the copper metallation of CcO [8,9].

Copper insertion into Cox1 and Cox2 subunits occurs in the IMS, since the proteins are localized within this compartment. Cox11 and Sco proteins mediate the copper metallation of the Cu_B and Cu_A sites, respectively [10].

Sco1 and Sco2 are highly similar proteins but their distinct role in eukaryotic cells is still unclear [11]. Both proteins contain a single transmembrane (TM) α-helix at the N-terminal segment, anchoring them to the IM while the C-terminal segment of Sco proteins is soluble and extended to the IMS. A highly conserved CXXXC motif and a conserved His residue form the copper binding site [12,13]. Structurally both proteins are characterized by a thioredoxin fold and with a single Cu(I) ion bound to be solvent exposed and poised for a ligand exchange transfer reaction. *In vitro* binding assays or gel filtration of mitochodrial extract showed that the eukaryotic Sco can form oligomeric and heteromeric complexes [11,14]. Interestingly, the TM α-helices of both proteins contain motifs that have a degree of propensity to promote TM helix-helix interaction [15-17]. Moreover, the TM regions of human Sco proteins have lower similarity than the soluble domain; actually, the N-terminal part (matrix and TM α-helix) is the most diverse part of the two proteins. *In vivo* experiments in yeast showed that the TM α-helices and the matrix segment are important for the function of the proteins since only the full length proteins can mediate the formation of a functional CcO [11,18].

Heme *a* biosynthesis and insertion is mediated by another group of assembly factors. Cox10 and Cox15, two intrinsic IM proteins catalyze the modification of protoheme to heme *a* by their successive catalytic action. Cox10 and Cox15 carry out the farnesylation and oxidation reactions, respectively [19,20]. Another assembly protein that that is necessary in this process is the Surf1 protein. The *SURF1* gene was identified in a highly conserved cluster of housekeeping genes known as the *surfeit* locus, but its function remained unknown until pathogenic mutations in the gene were identified in CcO-deficient Leigh Syndrome (LS) patients [21]. The CcO deficiency in LS patients results from the failure in assembling physiological amount of active enzyme, suggesting a role for the Surf1 protein in the maturation of the CcO [22,23]. Moreover, in bacteria (*R. sphaeroides* and *P. denitrificans*) Surf1 homologue gene deletion leads to misassembled CcO lacking the heme a_3, suggesting that Surf1 (direct or indirect) mediate the insertion of heme a_3 to the CcO [24,25]. The Surf1 protein is anchored to the IM by two predicted TM α-helices, one each at the N- and C-terminus of the protein. The extramembrane domain of the protein faces the IMS. Sequence alignments revealed many Surf1 homologues in eukaryotes and prokaryotes. They show a striking degree of homology, and all proteins exhibit the same

characteristic TM α-helical topology and two highly conserved regions in the soluble domain [26].

Results and discussion

Transmembrane HSco1, HSco2 and HSurf1 proteins as targets

The transmembrane proteins selected for hetegologous expression, HSco1, HSco2 and HSurf1 are involved in the assembly of the CcO. HSco1 and HSco2 are involved in the formation of the Cu_A site and HSurf1 is suggested to function during the folding of the Cox1 subunit [21]. Compared to Sco proteins whose soluble domains have been greatly investigated [12,13, 27-30], the Surf1 protein structure and biochemical characteristics remain unknown. On the other hand, the functional role of the TM α-helices of the Sco1 and Sco2 is not known at the molecular level.

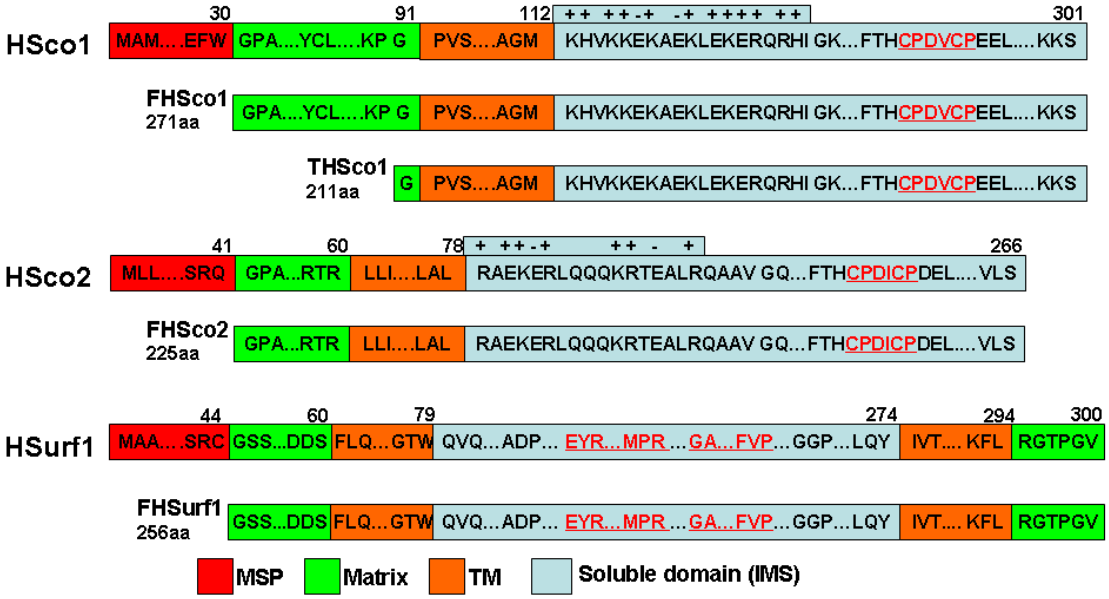


Figure 1: Domain definition of human Sco1, Sco2 and Surf1 proteins. Two constructs for the human Sco1 (FHSco1 and THSco1), one construct for the human Sco2 (FHSco2) and one construct for the human Surf1 (HSurf1) were generated. In Sco proteins, the charged residue region at the C-terminus of the predicted TM α-helix (TM) is also shown. The CPXXCP motif of HSco and the two conserved regions at the soluble domain of HSurf1 are indicated with red underlined letters. MSP: mitochondrial signal peptide; Matrix: domain exposed to the matrix of mitochondrion; Soluble domain (IMS): soluble domain of the proteins exposed to the intermembrane space of mitochondrion.

We defined two constructs for the HSco1 (FHSco1 and THSco1) and one for HSco2 (FHSco2) as shown in Fig.1. The two full length protein constructs of the HSco proteins lack the predicted the mitochondrial signal peptide. The different size of the matrix domain of the HSco1 compared to HSco2 (61 vs 19 aa) and a predicted α-helix at

the position 34A-52A of HSco1 rationalize the selection of the extra construct for HSco1 lacking the matrix domain of the protein (THSco1, Fig.1).

The Surf1 protein is anchored to the IM by two predicted TM α-helices, one each at the N- and C-terminus of the protein. We defined one construct for the HSurf1 protein which lacks the mitochondrial signal peptide (FHSurf1) shown in Fig.1.

Production and localization of HSco1, HSco2 and HSurf1 transmembrane proteins

The production of recombinant membrane proteins in *E.coli* has been recently tested (using 6His, Flag-peptide, MBP [31,32]), proving the possibility of producing high yields of membrane proteins for structural determination.

Table 1: Production and localization of the HSco1, HSco2 and HSurf1 heterologous proteins

	FHSco1			**THSco1**			**FHSco2**			**FHSurf1**		
	E	*M*	*S*	*E*	*M*	*S*	*E*	*M*	*S*	*E*	*M*	*S*
6His	+	++	--	--	--	--	--	--	--	--	--	--
GST	+	++	+	+	++	+	--	--	--	--	--	--
Trx	--	--	--	--	--	--	+	--	--	--	--	--
MBP	--	--	--	+	++	+	+	++	+	--	--	--
Msc	+	--	--	+	--	--	+	--	--	+	--	--
GB1	--	--	--	--	--	--	+	++	+	+	++	+
MBPperi	--	--	--	--	--	--	--	--	--	--	--	--

***** Abbreviations: E, production detected in the whole cell lysate; M, production detected in the cell membrane; S, production detected in the soluble fraction, +, positive production; ++; positive production detected in the cell membranes; --, negative production.

We have tested the production of the HSco1, HSco2 and HSurf1 TM proteins fused with 6His, GB1, Trx, GST, MBP, MBPperi and Msc, considering as positive result only the detection of the proteins in the cell membrane. The three strains tested (BL21-Gold(DE3), C41(DE3), C43(DE3)) showed different profile of production and localization. The C41(DE3) strain was found to produce high amount of protein in inclusion bodies and was no further used. The BL21-Gold and C43(DE3) strains had similar production profile but different protein production levels. C43(DE3) strain produced higher amount the target proteins and was chosen for further analysis. In Table 1 are shown the production and localization results of heterologous TM proteins expressed in the C43(DE3) strain. From the 28 fused proteins tested only the 7 have been detected in the cell membrane, without any preference in fusion protein or protein target. Moreover, when

the proteins are detected in the cell membranes they are also detected in the soluble fraction (5-10 fold higher amounts), except the 6His-FHSco1, which is only detected in the cell membrane. Finally, the Trx, Msc, and MBPperi fused TM proteins were not detected in the cell membranes.

Detergent selection: Effect of detergent micelles on HSco soluble domain

In NMR studies of membrane proteins, a number of detergents have been used with success largely limited to small-micelle detergents with small or charged headgroups such as dodecylphosphocholine (DPC), Lyso-palmitoyl phosphatidylcholine (LPPC), Lyso-myristoyl phosphatidylglycerol (LMPG) sodium dodecylsulfate (SDS), N-lauroylsarcosine, lauryldimethylamine oxide (LDAO) and octylglucoside [33,34]. The selection of the more suitable detergent greatly depended on monitoring the stability and folding of the membrane protein [35].

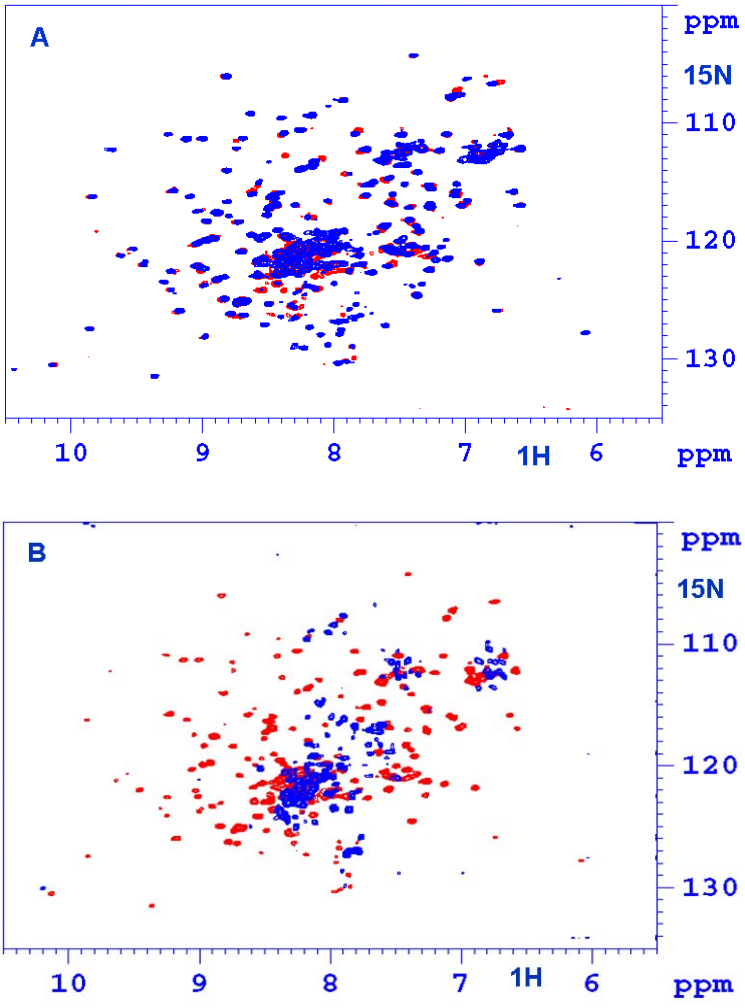


Figure 2: Overlay of ^{1}H-^{15}N HSQC spectra of ^{15}N-LHSco1Cu(I) (red) with: A) ^{15}N-LHSco1Cu(I)/OG micelles, B) ^{15}N-LHSco1Cu(I)/DPC micelles. Concentration of detergent: OG 1.0% and DPC 0.26%

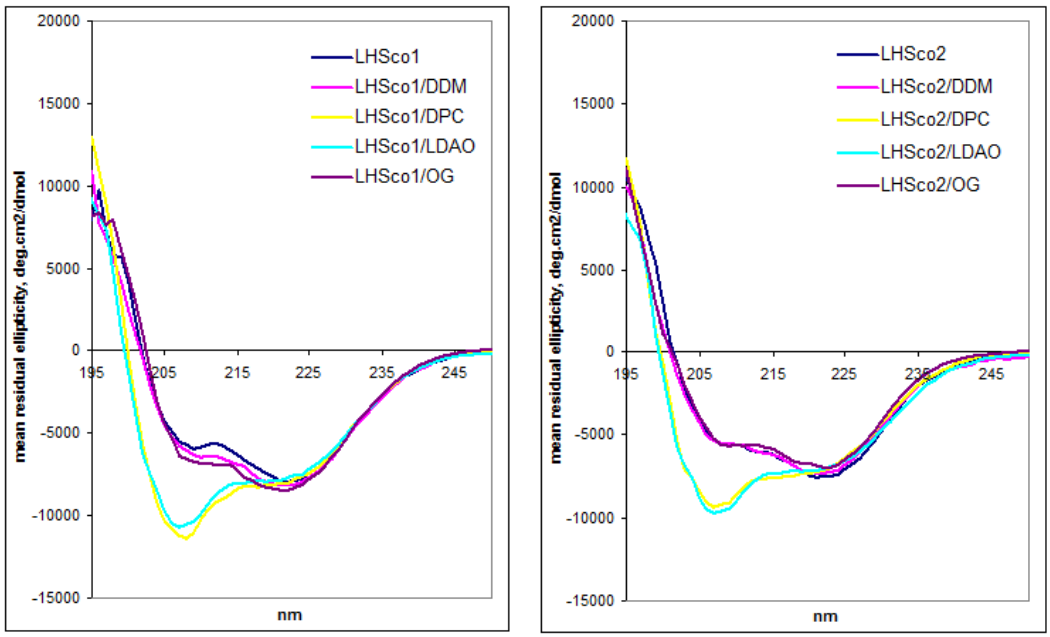


Figure 3: CD spectra of LHSco1 and LHSco2 in detergent micelles. DDM 0.02%, DPC 0.1%, LDAO 0.06%, OG 1.0%.

The TM proteins have a soluble domain that should behave differently in detergent micelles compared to the TM domain. An ideal detergent for the HSco and HSurf1 TM proteins would be a detergent that do not unfold the soluble domain and stabilize the TM helices avoiding aggregation.

In order to choose the detergent for the purification of the membrane proteins we tested the effect of the detergent micelles on the soluble domain of HSco and the ability of HSco to be metallated in the presence of detergent micelles.

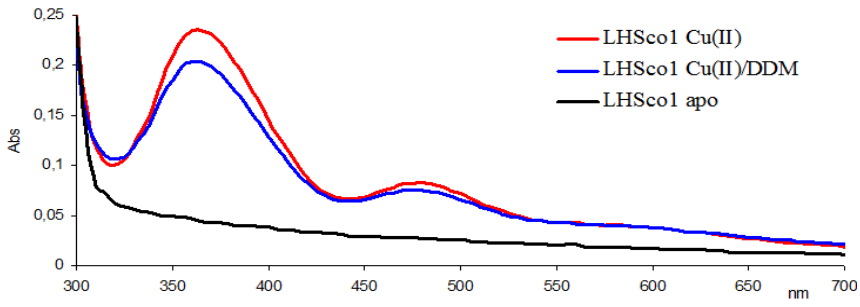


Figure 4: UV-Visible spectra of Cu(II)HSco soluble domain (LHSco1) in DDM micelles. The electronic spectra of the LHSco1 after metallation (1:1 Cu(II): protein ratio) in the presence and absence of detergent micelles showing the characteristic absorbance at 360, 470 and 610nm.

The NMR (Fig. 2) and CD (Fig. 3) spectra collected show that zwitterionic detergents (like DPC and LDAO) unfolds the soluble domain of HSco. Although the CD

spectra of both HSco indicate that the secondary structure elements are different but not lost with respect to that present in the absence of detergents, the ^{1}H-^{15}N-HSQC spectra show the unfolding of the HSco soluble domain. On the other hand, the two non-ionic detergent (OG, DMM) micelles do not unfold the soluble domain and are proven to be the best choice for the purification of the TM proteins. Additionally, the metallation of HSco in non-ionic detergents indicates that the proteins are well folded (Fig. 4).

Purification and cleavage of the fused transmembrane proteins

Three non-ionic detergents OG, DM, and DDM were chosen for the purification of the seven positive fusion proteins detected in the *E.coli* cell membrane. These detergents have different characteristics mainly due to the size alkyl chain, with OG having the shortest (C8) and the higher CMC value (25mM) compared to DM (C10) and DDM (C12) that have 10- and 100-fold lower CMC values, respectively. The length of the alkyl chain affects also the aggregation number and the size of the micelles formed. OG forms the smallest micelles with an average size of 23kDa.

Upon purification of the target proteins, the detergents were evaluated for their effect on the targets stability and TEV activity. OG was found to destabilize the purified proteins and cause precipitation of all the targets. Moreover, there was no indication of cleavage when the samples were incubated with TEV protease in the presence of OG. In the presence of DM and DDM the purified proteins were stable and the cleavage of certain fused proteins was also detected. The results are shown in Table 2.

Table 2: Purification and digestion of target proteins in the presence of DM and DDM micelles

Target	Fusion protein (MW, kDa)	Purification (yield, mg)	TEV digection %	
			DM	DMM
His-FHSco1	34.1	n/a	n/a	n/a
GST-FHSco1	57.6	n/a	0	0
GST-THSco1	51.2	3	0	0
MBP-THSco1	66.4	3	70	70
MBP-FHSco2	67.6	4	70	70
GB1-FHSco2	35.1	3	50	50
GB1-FHSurf1	38.9	5	50	50

The amount of protein was determined by absorbance at 280 nm after eluting the sample from the metal affinity column. Abbreviations are as follows: MW, molecular weight; n/a, not applicable.

Typically, high yields (>5mg/L of the target protein) are desirable for structural studies. The yields of the purified protein obtained from 1L LB cultures are shown in Table 2. The yield can be further improved by longer incubation of the targets to the resin during the purification since, due to inefficient binding in presence of detergents, unbound protein accounts to 90% .

Conclusions

The three TM proteins, HSco1, HSco2 and HSurf1 were produced as fused proteins and detected in the cell membrane of the *E.coli*. The purification trails with selected detergent showed that some of the targets (like MBP-THSco1, MBP-FHSco2, GB1-FHSco2 and GB1-FHSurf1) can be purified. Biophysical tools (NMR, CD and light scattering) are now used to characterize the produced proteins.

Materials and methods

Detergents: Octylglucoside(OG), Decylmaltoside (DM), Dodecylmaltoside (DDM), Dodecylphospho- choline (DPC), N-Lauroyldimethyl amineoxide(LDAO), where purchased from Glycon Biochemical. Vectors were purchased form European Molecular Biology Laboratory Protein Expression and Purification Facility apart from the homemade pDEST-Mistic.

Definition of constructs

The transmembrane (TM) regions of the HSco1 (NP_004580), HSco2 (NP_005129) and HSurf1 (NP_003163) proteins were predicted using TMbase [37], PRED-TMR [38] and TMHMM2.0 programs [39]. The mitochondrial signal peptide was predicted using MitoProt II 1.0a4 program [40]. Rare condons determination by *Rare Codon* calculator (*RaCC*). Disordered regions were predicted using IUPred [41]. The designed constructs are: the full length human Sco1 lacking the predicted mitochondrial signal peptide, 31G-301S, (FHSco1 hereafter), the human Sco1 lacking the predicted mitochondrial signal peptide and the matrix segment, 90G-301S, (THSco1 hereafter), the full length human Sco2 lacking the predicted mitochondrial signal peptide, 45G-266S (FHSco2 hereafter) and full length human Surf1 lacking the predicted mitochondrial signal peptide: 45G-300V (FHSurf1 hereafter). Constructs are shown in Fig.1.

Cloning

The cDNA of human *SCO1*, *SCO2* and *SURF1* (purchased from RZDB, Germany) were amplified by PCR. The constructs were inserted in Gateway Entry vector pENTR/TEV/D-TOPO (Invitrogen) and sequences were verified by sequencing at PRIMM (PRIMM, Italy). The pENTR clones were recombinated with pETG-30A (6His-GST, GST hereafter), pDEST-17 (6His), pDEST-HisMBP (6HisMBP, MBP hereafter), pDEST-periHisMBP, (6HisMBP, MBPperi hereafter), pTH34 (6His-GB1, GB1 hereafter), pETG20A (6HisTrx, Trx hereafter) and pDEST-Mistic (6HisMistic, Msc hereafter). Insertion of the DNA constructs was verified by PCR screening and the plasmid of the positive clones was purified using the QIAquick PCR Purification Kit (Qiagen) to

transform BL21-Gold(DE3) (Stratagene), BL21(DE3) C43 (Avidis) and BL21(DE3)C41 (Avidis) cells.

Protein production screening

The cells were grown in LB medium at 30 °C until the cell cultures reached the OD_{600} of ~0.6. The cultures were then cooled down to 18°C and induced with 0.2 mM isopropyl-b-D-1-thiogalactoside (IPTG). Cell cultures were harvest after overnight incubation at 18°C. Protein production was first tested in 5 mL cultures. Clones that showed heterologous protein production were grown in 100 mL cultures. The cell pellet obtain was analysed for protein expression in the soluble and membrane fraction of the cell lysate. Clones that the heterologous protein was detected in cell membranes were grown in 1L culture for protein purification. Heterologous protein production of the target proteins was detected via SDS-PAGE.

Cell lysis, membranes isolation and protein purification

Frozen cell pellets were resuspended in 20 mM Tris-HCl, 100 mM NaCl, pH 7.5 (B-buffer) 1 mg/mL lysozyme, complete protease inhibitor cocktail EDTA-free (Roche). After sonication (5 times 30 sec/5 min pause) the suspension was centrifuged at 15,000 *g* for 15min to remove inclusion bodies and unbroken cells. The cell membranes were isolated by 1 h centrifugation at 150,000*g*. The membranes were resuspended (using a glass homogenizer) in B-buffer and solubilized by addition of detergent and incubated for 2h, followed by centrifugation for 45 min at 200,000 *g*. The supernatants containing solubilized membrane proteins were diluted (with B-buffer to reach the working concentration of detergent) and incubated 1h with metal chelating agarose resin (Ni^{2+} was used for Surf1 and Zn^{2+} for Sco proteins) pre-equilibrated with the B-buffer, including 5 mM of imidazole and the corresponding detergent. The resin was washed with the B-buffer containing 10 mM of imidazole and the corresponding detergent. The recombinant proteins were eluted with 300 mM of imidazole in the same buffer supplemented with 1mM dithiothreitol (DTT). Tobacco etch virus (TEV) protease cleavage of the fusion proteins was perfomed in B-buffer supplemented with 1 mM DTT, 0.5 mM EDTA and the corresponding detergent. The yields were calculated by measuring the absorbance at 280 nm.

HSco folding characterization in detergent micelles

Human Sco1 and Sco2 soluble domain (113K-301S: LHSco1, and 79R-266S: LHSco2) were expressed, purified and metallated with Cu(I) or Cu(II) according to already reported protocol [12]. 15N-LHSco1-Cu(I) sample in phosphate 50 mM, 1 mM DTT, pH 7.2

was mixed with the corresponding detergent. The final concentration of protein was 0.1 mM. The final concentration of detergent (C_D) used was based on the formula [34]:

$$C_D = CMC + \left(A_{num} \times C_P\right)$$

where *CMC*, critical micelle concentration of detergent; A_{num}, aggregation number of detergent in a micelle and C_p, the concentration of the protein. ^{1}H-^{15}N-HSQC spectra were collected at 298 K on Avance 800 spectrometer. CD measurements of apo LHSco1 and LHSco2 were performed with protein concentrations of 10 uM in 10 mM phosphate buffer, pH 7.2 and the corresponding detergent ($C_D = 1.5 \times CMC$). The CD investigations were carried out in a Jasco J-715 spectropolarimeter using a 0.1-cm pathlength cell with cooling jacket connected to a water thermostatic device at 25^{o}C. Spectra were recorded from 190-250nm at a scan speed of 20 nm/min and resulted from averaging 4 scans. The final spectra were baseline-corrected by subtracting the corresponding buffer obtained under identical conditions. Results were expressed as the mean residue ellipticity [θ] at a given wavelength. Electronic spectra of the Cu(II) form of the HSco after metallation with $CuSO_4$ (1:1.2 protein:Cu ratio) in detergent micelles were recorded on a Cary 50 spectrophotometer (Varian) using 50 uM protein concentration and detergent concentration according to the formula above.

Reference list

1. Castresana J, Lübben M, Saraste M, Higgins DG. EMBO J. (**1994**)13:2516-25
2. Adachi K, Oiwa K, Nishizaka T, Furuike S, Noji H, Itoh H, Yoshida M, Kinosita K Jr. Cell. (**2007**)130:309-21.
3. Khalimonchuk O, Rödel G.. (**2005**) 5:363-88
4. Richter OM, Ludwig B.. Rev Physiol Biochem Pharmacol. (**2003**)147:47-74
5. Cobine PA, Pierrel F, Winge DR..Biochim Biophys Acta. (**2006**) 1763:759-72
6. Tzagoloff A, Dieckmann CL..Microbiol Rev. (**1990**) 54:211-25
7. Carr HS, Winge DR.. Acc Chem Res. (**2003**) 36:309-16
8. Rigby K, Zhang L, Cobine PA, George GN, Winge DR. e. J Biol Chem. (**2007**) 282:10233-42
9. Maxfield AB, Heaton DN, Winge DR. e.J Biol Chem. (**2004**)13:5072-80
10. Hiser, L.; Di Valentin, M.; Hamer, A.G.; Hosler, J.P. (**2000**) *J.Biol.Chem.* 275 619-623
11. Leary, S.C.; Kaufman, B.A.; Pellecchia, G.; Guercin, G.H.; Mattman, A.; Jaksch, M.; and Shoubridge, E.A.; (**2004**) *Hum. Mol. Genet.* 13 1839–184
12. Banci, L., Bertini, I., Calderone, V., Ciofi-Baffoni, S., Mangani, S., Martinelli, M., Palumaa, P., Wang, S., (**2006**)103: 8595
13. Banci L, Bertini I, Ciofi-Baffoni S, Gerothanassis IP, Leontari I, Martinelli M, Wang S.Structure. (**2007**)15:1132-40
14. Lode A, Paret C, Rödel G. Yeast. (**2002**) 19:909-22.
15. Melnyk RA, Kim S, Curran AR, Engelman DM, Bowie JU, Deber CM., , J Biol Chem. (**2004**) 16:16591-7.
16. Kleiger G, Grothe R, Mallick P, Eisenberg D., Biochemistry. (**2002**)41:5990-7.
17. Schneider D, Engelman DM., J Mol Biol. (**2004**) 343:799-804
18. Beers J, Glerum DM, Tzagoloff A. J Biol Chem. (**2002**) 21:22185-90
19. Antonicka H, Leary SC, Guercin GH, Agar JN, Horvath R, Kennaway NG, Harding CO, Jaksch M, Shoubridge EA..Hum Mol Genet. (**2003**) 12:2693-702.
20. Antonicka H, Mattman A, Carlson CG, Glerum DM, Hoffbuhr KC, Leary SC, Kennaway NG, Shoubridge EA..Am J Hum Genet. (**200**3) 72:101-14
21. Nijtmans, L.G.. Artal Sanz, M. Bucko, M M.H. Farhoud, M. Feenstra, G.A. Hakkaart, M. Zeviani, L.A. Grivell, , FEBS Lett. (**2001**) 498:46–51
22. Zhu Z, Yao J, Johns T, Fu K, De Bie I, Macmillan C, Cuthbert AP, Newbold RF, Wang J, Chevrette M, Brown GK, Brown RM, Shoubridge EA.. Nat Genet. (**1998**) 20:337-43
23. Barrientos, A. Zambrano, A. A. Tzagoloff, , EMBO J. (**2004**) 23:3472–3482.
24. Bundschuh FA, Hoffmeier K, Ludwig B.. Biochim Biophys Acta. (**2008**) 1777:1336–1343
25. Smith D, Gray J, Mitchell L, Antholine WE, Hosler JP..J Biol Chem. (**2005**)280:17652-6
26. Poyau A, Buchet K, Godinot C.FEBS Lett. (**1999**) 462:416-20.
27. Horng YC, Leary SC, Cobine PA, Young FB, George GN, Shoubridge EA, Winge DR.J Biol Chem. (**2005**) 280:34113-22
28. Balatri E, Banci L, Bertini I, Cantini F, Ciofi-Baffoni S..Structure. (**2003**) 11:1313-4
29. Williams JC, Sue C, Banting GS, Yang H, Glerum DM, Hendrickson WA, Schon EA. J Biol Chem. (**2005**) 280:15202-11
30. Nittis T, George GN, Winge DR..J Biol Chem. (**2001**)276:42520-6

31. Eshaghi S, Hedrén M, Nasser MI, Hammarberg T, Thornell A, Nordlund P.Protein Sci. (**2005**) 14:676-83
32. Korepanova A, Moore JD, Nguyen HB, Hua Y, Cross TA, Gao F.Protein Expr Purif. (**2007**):24-30.
33. Vinogradova O, Sönnichsen F, Sanders CR .J Biomol NMR. (**1998**)11:381-6
34. Sanders CR, Sönnichsen F.Magn Reson Chem. (**2006**) 44:24-40
35. Prive,GG Methods. (**2007**)4:388
36. Lundbäck AK, van den Berg S, Hebert H, Berglund H, Eshaghi S. Anal Biochem. (**2008**)382:69-71
37. Hofmann, K., Stoffel, W., Biol. Chem. Hoppe-Seyler (**1993**), 374,166
38. Pasquier,C.,Promponas, V.J., Palaios,G.A.,Hamodrakas, J.S.,Hamodrakas,S.J,Protein Eng (**1999**): 12, 381-5
39. Krogh,A.,Larsson,B.,Von HeijneG.,Sonnhammer,E.L.L.,Journal of Molecular Biology, (**2001**), 305:567-580
40. Claros, M.G., Comput Appl Biosci., (**1995**), 11:441-7
41. Dosztányi, Z., Csizmók, V., Tompa, P., Simon, I Bioinformatics, (**2005**) 21:3433-3434
42. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A, Humana Press (**2005**) 571-607

4. Conclusions and perspectives

Structural biology seeks to understand the structure and function of macromolecules through structure determination by X-ray crystallography, NMR, electron microscopy, theoretical analyses, and a wide range of correlative biochemical and genetic studies of function. In recent years, the analysis of metalloenzymes and metalloproteins has accelerated[1]. Many developments and initiatives have taken place, particularly due to progress in X-ray and NMR studies, in order to clarify the role and trafficking of the metals in the cell.

Cytochrome *c* oxidase (CcO), the last enzyme of the respiratory chain, is a member of a superfamily of heme-copper-containing terminal oxidases that are present in all aerobic organisms. Defects involving genetic mutations altering CcO functionality or structure can result in severe, often fatal metabolic disorders. Among the many classified mitochondrial diseases, those involving dysfunctional CcO assembly are thought to be the most severe. Currently, mutations have been identified in six CcO assembly factors: *SURF1*[2], *SCO1*[3], *SCO2*[4], *COX10*[5], *COX15*[6], and *LRPPRC*[7]. Mutations in these proteins can result in altered functionality of sub-complex assembly, copper transport, or translational regulation.

During my PhD course, research was focused on clarifying the functional role of human Sco1, Sco2 and Cox17 in Cu_A site of CcO assembly mechanism, and the production of the Surf1 protein.

The formation of the Cu_A site of CcO proceeds through protein-protein interaction processes involving Cu(I) transfer from Cox17 to the Sco proteins and finally to the apo Cu_A site. In this process, Sco1 and Sco2 are working cooperatively with non-overlapping functional role [8]. We investigated the metal transfer between the human Sco proteins and the human Cox17, showing that mitochondrial Cu(I) transfer from Cox17 to Sco1 can be coupled to electron transfer and the same reaction does not occur in Sco2. The different redox behaviour of Sco1 compared to Sco2 can strongly contribute to define the molecular basis of the functional differentiation of the two proteins. The electron transfer-coupled metallation mechanism could occur within the IMS for an efficient specific transfer of the metal to the proteins, when metal-binding thiols of Cu_A site are oxidized.

The human Sco proteins are detected as dimeric proteins. The dimeric aggregation state and the formation of a Sco1/Sco2 heterodimeric complex were suggested to be the key complex for the formation of the bimetallic Cu_A site [8,9]. We investigated the aggregation state of the soluble domain of human Sco, showing that a 19aa segment promotes the dimerization of human Sco1, but not human Sco2. The investigation of the apo and Cu(I) forms showed that the Sco1 dimerization is driven by the charged residue of

the 19aa segment at the N-terminus and the dimer is orientated in such way that the copper sites are close and solvent exposed. The dimeric Sco1 strongly supports the model of simultaneous transfer of two copper ions to the apo Cu_A site. Moreover, the formation of a stable heterocomplex between human Sco1 and Sco2 soluble domain was also investigated, showing that these soluble domains do not interact. However, the interaction between the two proteins and the dimerization of human Sco2 maybe reside in the TM segments. To investigate this matter, the human Sco1 and Sco2 containing their single transmembrane α-helix were expressed and extracted from the cell membranes of *E.coli*. Following steps will be to characterize them by NMR and investigate their interaction at a molecular level.

From our data, the distinct functional role of human Sco1 and Sco2 can be related to the observed different behaviour related to their redox properties when interacting with Cox17 as well as to their aggregation state. Human Sco1 and Sco2 function should be more deeply investigated at the molecular level. The interaction of the human Sco proteins with the Cu_A site can definitively unravel the "mystery" of the mechanism of the insertion of copper and the different function of each Sco in this process. The role of the human Cu(II) Sco form is also unknown. Does this form exist in the mitochondria? And if it does, how is it formed since Cox17 transfer exclusively Cu(I) ions? Further studies on the Cu(II) form of the Sco proteins are necessary to clarify the role of this controversial state.

Surf1 is a transmembrane protein anchored to the mitochondrial inner membrane by two transmembrane α-helices. The protein is involved in the folding of Cox1 subunit and suggested to mediate the heme a_3 insertion. Apart from the implication of the protein in the mitochondrial diseases, the biochemical and structural characteristics of the protein are unknown. We expressed the Surf1 transmembrane protein in *E.coli* and detected production in the cell membrane. The biochemical and structural characterization of the protein will be the next step.

Reference list

1. Bertini I, Cavallaro G.J Biol Inorg Chem. (**2008**) 13:3-14.
2. Péquignot MO, Dey R, Zeviani M, Tiranti V, Godinot C, Poyau A, Sue C, Di Mauro S, Abitbol M, Marsac C., Hum Mutat. (**2001**)17:374-81.
3. Petruzzella V, Tiranti V, Fernandez P, Ianna P, Carrozzo R, Zeviani M. Genomics. (**1998**) 54:494-504
4. Papadopoulou LC, Sue CM, Davidson MM, Tanji K, Nishino I, Sadlock JE, Krishna S, Walker W, Selby J, Glerum DM, Coster RV, Lyon G, Scalais E, Lebel R, Kaplan P, Shanske S, De Vivo DC, Bonilla E, Hirano M, DiMauro S, Schon EA.Nat Genet. (**1999**) 23:333-7
5. Kennaway NG, Harding CO, Jaksch M, Shoubridge EA.Hum Mol Genet. (**2003**) 12:2693-702
6. Antonicka H, Leary SC, Guercin GH, Agar JN, Horvath R, Antonicka H, Mattman A, Carlson CG, Glerum DM, Hoffbuhr KC, Leary SC, Kennaway NG, Shoubridge EA.Am J Hum Genet. (**2003**) 72:101-14
7. Mootha VK, Lepage P, Miller K, Bunkenborg J, Reich M, Hjerrild M, Delmonte T, Villeneuve A, Sladek R, Xu F, Mitchell GA, Morin C, Mann M, Hudson TJ, Robinson B, Rioux JD, Lander ES.oc Natl Acad Sci U S A. (**2003**) 100:605-10
8. Leary, S.C.; Kaufman, B.A.; Pellecchia, G.; Guercin, G.H.; Mattman, A.; Jaksch, M.; and Shoubridge, E.A.; (**2004**) *Hum. Mol. Genet.* 13 1839–184
9. Horng YC, Leary SC, Cobine PA, Young FB, George GN, Shoubridge EA, Winge DR.J Biol Chem. (**2005**) 280:34113-22